ISSI 2015

15th INTERNATIONAL CONFERENCE ON SCIENTOMETRICS & INFORMETRICS

29 June - 4 July, 2015

BOĞAZİÇİ UNIVERSITY • ISTANBUL-TURKEY

www.issi2015.org

PROCEEDINGS OF ISSI 2015

BOĞAZİÇİ UNIVERSITY 1863

HACETTEPE UNIVERSITY

TÜBİTAK ULAKBİM

PROCEEDINGS OF ISSI 2015 ISTANBUL

15th International Society of
Scientometrics and Informetrics Conference

Istanbul, Turkey
29th June to 4th July 2015

Editors

Albert Ali Salah, Yaşar Tonta, Alkım Almıla Akdağ Salah,
Cassidy Sugimoto, Umut Al

Partners

Boğaziçi University, Turkey
Hacettepe University, Turkey
TÜBİTAK ULAKBİM, Turkey

Sponsors

Thomson Reuters
Springer
EBSCO Information Services, USA
Emerald
Elsevier B.V.

Boğaziçi University Cataloging –in-Publication Data
Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference, Istanbul, Turkey, 29 June to 3 July, 2015 / editors Albert Ali Salah, Yaşar Tonta, Alkım Almıla Akdağ Salah, Cassidy Sugimoto, Umut Al.

1275 p.; 29 cm.
ISBN 978-975-518-381-7
ISSN 2175-1935
1. Bibliometrics - Congresses. 2. Information science - Congresses. 3. Communication in science - Congresses. 4. Scientific literature - Congresses.
Z669.8|b.I58

Printed at the Boğaziçi University Printhouse
First Printing: June 2015

Boğaziçi Universitesi, ETA B Blok, Zemin Kat, Kuzey Kampüs, İstanbul / TÜRKİYE
Tel ve Fax: (90) 212 359 44 06

ORGANISATION AND COMMITTEES

Conference Chairs

Albert Ali Salah
Yaşar Tonta
M. Mirat Satoğlu

Programme Chairs

Alkım Almıla Akdağ Salah
Cassidy Sugimoto
Umut Al

Doctoral Consortium Chairs

Andrea Scharnhorst
Judit Bar-Ilan

Workshops & Tutorials Chair

Caroline Wagner

Local Organization Chair

Heysem Kaya

Scientific Program Committee

Giovanni Abramo
Isidro Aguillo
Isola Ajiferuke
Alkım Almıla Akdağ Salah
Umut Al
Jens-Peter Andersen
Eric Archambault
Clément Arsenault
Joaquin Azagra-Caro
Judit Bar-Ilan
Aparna Basu
Sada Bihari-Sahu
Maria Bordons
Lutz Bornmann
Hamid Bouabid
Kevin W. Boyack
Guillaume Cabanac
Juan Miguel Campanario
Chaomei Chen
Andrea D’Angelo
Hans-Dieter Daniel
Cinzia Daraio
Hamid Darvish
Koenraad Debackere
Gernot Deinzer
Brad Demarest
Swapan Deoghuria
Fereshteh Didegah
Ying Ding
Güleda Doğan
Tim Engels
Claire François
Jonathan Furner
Antonio Garcia
Aldo Geuna
Wolfgang Glänzel
Alicia Gomez
Isabel Gomez
Juan Gorraiz
Philippe Gorry
Abdullah Gök
Raf Guns
Nabi Hasan
Stefanie Haustein
Sybille Hinze
Michael Hofer
Marianne Hörlesberger
Zhigang Hu
Peter Ingwersen
Siladitya Jana
Evaristo Jiménez-Contreras
Milos Jovanovic
Yuya Kajikawa
Hildrun Kretschmer

ORGANISATION AND COMMITTEES

J P S Kumaravel
Benedetto Lepori
Jacqueline Leta
Jonathan Levitt
Loet Leydesdorff
Liming Liang
Judith Licea
Junwan Liu
Yuxian Liu
Szu-Chia Lo
Carmen Lopez Illesca
Bob Losee
Terttu Luukkonen
Marc Luwel
Domenico Maisano
Wolfgang Mayer
Kate McCain
Eustache Megnigbeto
Lokman Meho
Raul Mendez-Vasquez
Alexis Michel Mugabushaka
Ulle Must
Anton Nederhof
Ed Noyons
Michael Ochsner
Carlos Olmeda-Gómez
José Luis Ortega
Maria Antonia Ovalle-Perandones
Adèle Paul-Hus
Antonio Perianes-Rodríguez
Bluma C. Peritz
Fernanda Peset
Anastassios Pouris
Ismael Rafols
Emanuela Reale
John Rigby
Nicolas Robinson-Garcia
Ivana Roche
Jürgen Roth
Ronald Rousseau
Santanu Roy
Jane Russell
Bibhuti Sahoo
Albert Ali Salah
Ulf Sandström
Elias Sanz
Andrea Scharnhorst
Edgar Schiebel
Christian Schloegl
Jesper Schneider
Torben Schubert
Robert Shelton
Gunnar Sivertsen
Stig Slipersæter
Henry Small
Andreas Strotmann
Cassidy Sugimoto
Yuan Sun
Zehra Taşkın
Mike Thelwall
Bart Thijs
Yaşar Tonta
Andrew Tsou
Saeed Ul Hassan
Peter van den Besselaar
Nees Jan Van Eck
Thed Van Leeuwen
Bart Van Looy
Anthony Van Raan
Benjamin Vargas-Quesada
Liwen Vaughan
Peter Vinkler
Cathelijn Waaijer
Lili Wang
Xianwen Wang
Jos Winnink
Matthias Winterhager
Dietmar Wolfram
Paul Wouters
Qiang Wu
Yishan Wu
Erjia Yan
Dangzhi Zhao
Michel Zitt
Alesia Zuccala

CHAIRS' WELCOME

The 15th International Society of Scientometrics and Informetrics Conference took place at Boğaziçi University in Istanbul, from June 29 to July 4, 2015. The Conference was jointly organised by Boğaziçi University, Hacettepe University, and the TÜBİTAK ULAKBIM (Turkish Academic Network and Information Center – The Scientific and Technological Research Council of Turkey) under the auspices of ISSI – the International Society for Scientometrics and Informetrics.

The ISSI biennial conference is the premier international forum for scientists, research managers, authorities and information professionals to discuss the current status and progress in informetric and scientometric theories, concepts, tools, platforms, and indicators. In addition to theoretical and quantitative focus of the conference, the participants had the opportunity to discuss practical, cross-cultural, and multi-disciplinary aspects of information and library science, R&D-management, and science ethics, among other related topics.

The focus theme of ISSI2015 was "**the future of scientometrics**". Scientometrics and informetrics together represent a broad field with a rich history. Scientometrics has been responsible for creating tools for research assessment and evaluation, as well as for use in charting the flow of scientific ideas and people. Today, with the advancements of computing power, technology, and database management systems, the impact of scientometrics has become ubiquitous for scientists and science policy makers. However, the high diffusion of scientometric and informetric research has also brought a new wave of criticism and concern, as people grapple with issues of goal displacement and inappropriate use of indicators. The question facing the field is how best to move forward given the computational opportunities and the sociological concerns. Therefore, the goal of ISSI2015 was to highlight the best research in this field and to bring together scholars and practitioners in the area to discuss new research directions, methods, and theories, and to reflect upon the history of scientometrics and its implications.

The keynote given by Loet Leydesdorff demonstrated the potential of thinking of science as a complex institution. By building on the Triple Helix Model of University-Industry-Government relations, Dr. Leydesdorff showed that innovation systems can provide institutional mediation between knowledge production, wealth generation, and governance.

The second keynote, by Kevin Boyack, directly answered the challenge of the focus theme of ISSI2015, and proposed several opportunities to expand the field of scientometrics. Dr. Boyack called for increasing attention to funding, workforce, data and instrumentation, research objects, and innovation.

The conference included four special sessions on a range of topics, including performance indicators, algorithms for topic detection, empirical evaluation of education, research and innovation, and how scientometrics can be used to improve and inform university rankings. These special sessions included poster presentations, panel discussions, invited speakers, and public debates.

The increasing number of open-source software for scientometrics presents great opportunities for researchers. Four tutorials, organized on the first day of the conference, aimed to introduce a number of tools in depth: open source data analysis and visualization tools, citation exploration software, measurement of scholarly impact, and on social network analysis with the popular R software.

The Doctoral Forum, organized by Andrea Scharnhorst and Judit Bar-Ilan, is a meeting of senior researchers and selected doctoral students for presenting and discussing research projects and an excellent way for students of getting valuable feedback, along with strong networking opportunities. This is the sixth ISSI Doctoral Forum and we are extremely happy about the interest it continues to receive from the community. Additionally, the prestigious Eugene Garfield Doctoral Dissertation Scholarship is given by the Eugene Garfield Foundation.

During the Conference, the Derek de Solla Price Award of the International Journal Scientometrics was given to Mike Thelwall, Professor of Information Science at the University of Wolverhampton (UK), in a special session organized for this purpose. This award recognizes excellence through outstanding, sustained career achievements in the field of quantitative studies of science and their applications.

The satellite workshops of the conference reflected the diversity of the field. In "**Mining Scientific Papers: Computational Linguistics and Bibliometrics",** researchers in bibliometrics and computational linguistics were brought together to study the ways bibliometrics can benefit from large-scale text analytics and sense mining of scientific papers, thus exploring the interdisciplinarity of Bibliometrics and Natural Language Processing. The workshop on "**Grand challenges in data integration for research and innovation policy**" dealt with problems of big, open and linked data. The "**Forecasting science: Models of science and technology dynamics for innovation policy"** workshop discussed methodology for predicting the circumstances leading to scientific or technological innovation. "**Workshop on Bibliometrics Education**" brought together educational institutions, employers, professional societies, and Bibliometrics researchers and professionals to tackle this problem. Finally, "**Google Scholar and related products"** was a highly interactive workshop on the benefits and limitations of some of the most important citation tools.

All contributions for the conference were evaluated by at least two reviewers of the Scientific Program Committee. The papers that required additional reviews were discussed by the Program Chairs before a decision was reached. From 228 full and research in progress paper submissions, 123 papers were accepted for publication (54 percent acceptance rate). 82 of these papers were full papers, and 41 were research in progress. There was a large number of paper submissions on social media, technology transfer, science policy and research assessment. From 123 poster and ignite talk submissions, 68 posters and 13 ignite talks were accepted (66 percent). The ignite talks were to increase discussion of underrepresented topics and novel ideas. Because of the large number of papers, and to allow proper discussion for each paper, four parallel sessions were implemented. Several poster sessions were organized, each containing a relatively manageable number of posters. The conference brought together researchers from 42 countries and the works of 458 researchers were presented.

We thank all our contributors for their submissions, the members of the Organizing Committee for their work, the Scientific Program Committee for their reviewing effort, the ISSI board for their trust and guidance, the Rectorate and the Faculty of Engineering of Boğaziçi University for their constant assistance and support, as well as the sponsors for their generous financial contributions. We particularly thank Metin Tunç (Thomson Reuters), Elif Gürses (formerly of TÜBİTAK ULAKBİM), Juan Gorraiz (Universitat Wien), Figen Atalan (Boğaziçi University), Orçun Madran (Hacettepe University) and Büşra Şahin (DEKON Congress & Tourism) for their help in organizing ISSI2015.

Albert Ali Salah, Yaşar Tonta, Mirat Satoğlu, Alkım Almıla Akdağ Salah, Cassidy Sugimoto, Umut Al

TABLE OF CONTENTS

THEORY AND METHODS & TECHNIQUES **PAGE**

ISSI 2015

KEYNOTES

Increasing the Relevance of the ISSI Community in Today's Changing Scientific Landscape

Kevin W. Boyack

kboyack@mapofscience.com
SciTech Strategies, Inc., 8421 Manuel Cia Pl NE, Albuquerque, NM 87122 (USA)

The call for papers for the ISSI 2015 conference set forth a bold agenda by specifically asking for papers related to the "Future of Scientometrics". Many fields in science are what one might call primary fields in the same sense that there are primary colors. These are fields with a self-contained base and upon which other fields build. Scientometrics is not one of these primary fields, but rather operates on theories and data about the processes and outputs of science. We are, in essence, a service industry, and as a service industry we have the potential to exercise great influence on science as a whole. We also have the potential to flounder and die a slow death, or to be overtaken and replaced by another industry. In my opinion, our best opportunity to flourish as a field and community is to truly understand the structure, dynamics, and interactions of science as a whole and in parts, at multiple levels of detail, and to not only measure things but to develop predictive capacities. Opportunities exist for us to expand our view beyond traditional roles, if we can but see what they are.

In this talk I will propose that our opportunities to expand and flourish as a community can be enhanced in several ways. First, it is time for most of us to become far more acquainted with the work done by the pioneers in our field, and in related fields, than we currently are. Scientometrics is a melting pot in many ways, populated to a large degree by those trained in other fields – physics, chemistry, engineering, etc. Many of us are lacking in historical knowledge. We hear the names of Kuhn, Price, Merton, Crane, Latour, and many others, but how many of us are really familiar with not only their popular contributions, but also their smaller experiments that are less well known? There is much to be learned from the work started (and often abandoned due to lack of resources) by these giants that is perhaps even more relevant today than before.

Second, as a community we are highly focused on measuring the "arguments" (documentation) of science, whether using citation data or altmetrics. The science system, however, is comprised of far more than "arguments". Some of us do, to a lesser degree, address other parts of the science system – funding, workforce, data and instrumentation, research objects, and innovation. However, it is rare to see analyses that integrate multiple parts of the science system and explore their interactions. Our influence as a community can certainly be increased if we focus more on these interactions.

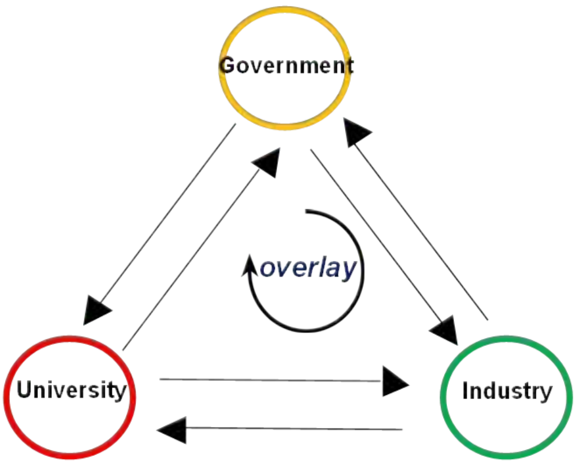
Third, and perhaps most controversially, I suggest that we seek to understand the effect of motivations on science. Perhaps the best way to do this is to start with ourselves, and reflect on "Why do we do what we do?" Are our motives aligned with the purest motives of society? Are we seeking, as individuals and as a community, to serve science and society, or are we seeking for self-aggrandizement? Each of us is many things in life, among which being a researcher or policy maker or scientometrician is only one facet. Often our choice of a career, and of the particular topics we research and for which we advocate are directly tied to these motives. Each of us has a story. Once we understand how our stories drive us to do what we do, then perhaps we can extend that knowledge to better understand science as a whole and how it is driven by the interacting motivations of researchers and institutions. Dick Klavans and I recently created a map of altruism, and were amazed at how much the motives in that map reflect why we do what we do. The parts of the science system mentioned above are all motivated differently. Do we consider this in our models and analyses? How would our analyses change if we were to consider motivation?

Although this talk will use some examples from my current research, it will be largely philosophical, and will raise far more questions than it will give answers. I fully expect many to disagree with much of what will be presented. Nevertheless, I submit that raising these questions at this time has the potential to cause us all to think critically, and that such critical thinking is the first step toward increasing our relevance as a community in the scientific world of the future.

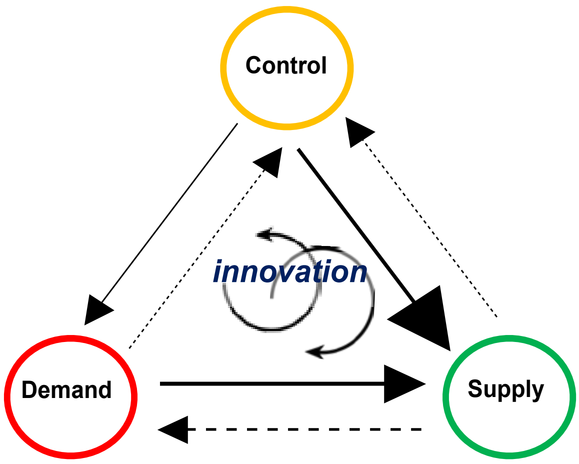
The Triple Helix of Knowledge Production, Wealth Generation, and Normative Control: A Neo-evolutionary Model of Innovation Ecosystems

Loet Leydesdorff

loet@leydesdorff.net
University of Amsterdam, Amsterdam School of Communication Research, Amsterdam, The Netherlands



The neo-institutional model of the Triple Helix of University-Industry-Government Relations (Etzkowitz & Leydesdorff, 1995).



The neo-evolutionary model of innovation in three dimensions (Leydesdorff & Zawdie, 2010; Lawton-Smith & Leydesdorff, 2014).

When three sub-dynamics can operate as selection environments on the variations among one another, a communication field can be generated that proliferates auto-catalytically using each third actor as a feedback or feed forward operating on mutual relations in clockwise or counter-clockwise rotations. This model improves on the neo-Schumpeterian models of innovation systems in evolutionary economics and technology studies, while these models assume a dialectics or co-evolution, for example, between trajectories and selection environments. By extending the Lotka-Volterra equations from two to three dimensions, Ivanova & Leydesdorff (2014) proved the possible emergence of a communication field ("overlay") as an emerging (fourth) subdynamic. In the communication field new options can be generated by sharing meaning provided to the events (Leydesdorff & Ivanova, 2014). This extension of innovative options can be measured as redundancy in terms of bits of information. Petersen, Rotolo & Leydesdorff (in preparation) analyzed Medicals Subject Headings (MEDLINE/PubMed) of approximately 100,000 articles in three research areas including technological breakthroughs in medical innovation (honored with Nobel Prizes in Physiology and Medicine) in terms of "Diseases" (demand), "Drugs and Chemicals" (supply), and "Techniques" (control). Periods of synergy (operationalized as redundancy) can be distinguished from periods in which outward exploration prevails. Innovation systems (e.g., at national or regional levels, but also sectorial ones such as in medicine) provide institutional mediation between wealth generation, knowledge production, and governance as different perspectives. In the case of China, Leydesdorff & Zhou (2014) found, for example, that the four municipalities play a mediating role above expectation between knowledge production and wealth generation. Note that the three dimensions can differently be operationalized depending on the research design (e.g., as “university,” “industry,” and “government”); but the dimensions have to be specified as analytically independent so that the three co-variations can be measured (Leydesdorff, Park, & Lengyel, 2014).

References:

Etzkowitz, H., & Leydesdorff, L. (1995). The Triple Helix---University-Industry-Government Relations: A Laboratory for Knowledge-Based Economic Development. *EASST Review 14*(1), 14-19.

Ivanova, I. A., & Leydesdorff, L. (2014). Redundancy Generation in University-Industry-Government Relations: The Triple Helix Modeled, Measured, and Simulated. *Scientometrics, 99*(3), 927-948. doi: 10.1007/s11192-014-1241-7

Lawton Smith, H., & Leydesdorff, L. (2014). The Triple Helix in the context of global change: dynamics and challenges. *Prometheus* (ahead-of-print), 1-16.

Leydesdorff, L., & Ivanova, I. A. (2014). Mutual Redundancies in Inter-human Communication Systems: Steps Towards a Calculus of Processing Meaning. *Journal of the Association for Information Science and Technology, 65*(2), 386-399.

Leydesdorff, L., Park, H. W., & Lengyel, B. (2014). A Routine for Measuring Synergy in University-Industry-Government Relations: Mutual Information as a Triple-Helix and Quadruple-Helix Indicator. *Scientometrics, 99*(1), 27-35. doi: 10.1007/s11192-013-1079-4

Leydesdorff, L., & Zawdie, G. (2010). The Triple Helix Perspective of Innovation Systems. *Technology Analysis & Strategic Management, 22*(7), 789-804.

Leydesdorff, L. & Zhou, P. (2014). Measuring the Knowledge-Based Economy of China in terms of Synergy among Technological, Organizational, and Geographic Attributes of Firms, *Scientometrics 98*(3), 1703-1713.

Petersen, A., Rotolo, D. & Leydesdorff, L. (in preparation), The Interaction of ‘Supply’, ‘Demand’, and ‘Technology’ in terms of Medical Subject Headings: A Triple Helix Model of Medical Innovations.

ISSI 2015

ALTMETRICS

WEBOMETRICS

Who Publishes, Reads, and Cites Papers? An Analysis of Country Information

Robin Haunschild[1], Moritz Stefaner[2], and Lutz Bornmann[3]

[1] *R.Haunschild@fkf.mpg.de*
Max Planck Institute for Solid State Research, Heisenbergstr. 1, 70569 Stuttgart (Germany)

[2] *moritz@stefaner.eu*
Eickedorfer Damm 35, 28865 Lilienthal (Germany)

[3] *bornmann@gv.mpg.de*
Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8, 80539 Munich (Germany)

Abstract
The research field of altmetrics has gathered increased attention within scientometrics. Here, we pay particular attention to the connection between countries of readers of papers (at Mendeley) and countries of authors as well as citers of papers (from Web of Science). This study uses the Mendeley application programming interface to gather Mendeley reader statistics for the comprehensive F1000Prime publication set (n_r=149,227 records, n_p = 114,582 papers). F1000Prime is a post-publication peer-review system for papers of the biomedical research. The F1000 papers are rated by experts as good, very good, or exceptional. We find no significant differences between authorship, readership, and authorship of citing papers broken down into countries across quality levels. Most authors, citers, and readers are located in the USA followed by UK and Germany. Except for a few cases, we find that percentages of readers, citers, and authors are rather well balanced. Although Russia and China host many large research groups with a large publication output, both countries are below the top 10 countries ordered according to readership percentages.

Conference Topic
Altmetrics

Introduction
Online reference managers can be seen as the scientific variant of social bookmarking platforms, in which users can save and tag web resources (e.g. blogs or web sites). The best known online reference managers with a social networking component are Mendeley (www.mendeley.com) and CiteULike (www.citeulike.org), which were launched in 2004 (CiteULike) and 2008 (Mendeley), and can be used free of charge (Li et al., 2012). Mendeley – in 2013 acquired by Elsevier (Rodgers and Barbrow 2013) – has developed since then into the most popular product among the reference managers (Haustein 2014), and most empirical studies involving reference managers have used data from Mendeley. Mendeley has obtained a rather unique position as an online reference manager with desktop and mobile app versions. Furthermore, Mendeley offers social networking services, which go beyond the capability of most reference managers.

The platforms allow users to save or organize literature, to share literature with other users, as well as to save keywords and comments on a publication (or to assign tags to them) (Bar-Ilan, et al., 2014, Haustein et al., 2014). Even if it is literature that is mainly saved by the users, they can also add to a library other products of scientific work (such as data sets, software and presentations). The providers of online reference managers make available a range of data for the use of publication by the users: The most important numbers are the user counts, which provide the number of readers of publications via the saves of publications (Li et al., 2012). The readers can be differentiated into different status and country groups as well as scientific sub-disciplines. The readers' data from Mendeley is also evaluated to make suggestions to the users for new papers and potential collaborators (Priem & Hemminger, 2010, Galloway et al., 2013). Although it is not quite known what Mendeley reader counts mean exactly, they can be viewed as citations to be. Many Mendeley users bookmark a paper in Mendeley with the intend to cite this paper in a forthcoming manuscript. As this is not the only reason to bookmark a publication in Mendeley, it is clear that Mendeley reader counts measure also something different than citations. This additional part of a publication's impact is another means to measure its usage.

In this study, the country information of Mendeley readers is used to compare the readers of papers with their authors as well as those authors who have cited the papers. We are interested in differences and similarities between the countries worldwide: Which are the countries in which the scientists read (or cite) more than publish and vice versa? In which countries are the numbers of authors, readers, and citers similar? As publication set, we used papers from the post-publication peer review system of F1000. It is an advantage of this dataset that each paper is classified according to its quality (based on expert scores). Thus, we are able to investigate the distribution of authors, readers, and citers for papers with different quality.

Literature review

Mendeley is used chiefly by science, technology, engineering and mathematics researchers (Neylon et al., 2014). According to a questionnaire in the bibliometric community (Haustein et al., 2014), 77% of those questioned know Mendeley. But Mendeley is actually used by only 26% of those questioned. However, with respect to the number of saved papers there are large differences between disciplines: Thus, for example, only about a third of the humanities articles indexed in the Web of Science (WoS) can also be found in Mendeley; however, in the social sciences, it is more than half (Mohammadi & Thelwall, 2013). Among the reference managers, Mendeley seems to have the best coverage of globally published literature (Haustein et al., 2014, Zahedi et al., 2014). The large user population and coverage result in Mendeley being seen as the most promising new source for evaluation purposes (among the online reference managers) (Haustein, 2014). Priem (2014) sees Mendeley already as a rival to commercial databases (such as Scopus and WoS).

With a view to the use of the data from online reference managers in research evaluation, bookmarks to publications (i.e. the saving of bibliographic data about publications in libraries) express the interest of a user in a publication (Weller & Peters 2012). But this interest is very variable; the spectrum extends from simple saving of the bibliographic data of a publication up to painstaking reading, annotation and use of a publication (Shema et al., 2014, Thelwall & Maflahi, in press). According to Taylor (2013), the following motives could play a role in the saving of a publication: "Other people might be interested in this paper … I want other people to think I have read this paper … It is my paper, and I maintain my own library … It is my paper, and I want people to read it … It is my paper, and I want people to see that I wrote it" (p. 20). The problem of the unclear meaning of the saving (or naming) of a publication is common to bookmarks in reference managers and also many other traditional and alternative metrics: Thus, for example, traditional citations can mean either simple naming citations in the introduction to a paper, as well as extensive discussions in the results or discussion sections (Bornmann & Daniel, 2008). Traditional citations can also be self-citations.

The data from online reference managers is seen as one of the most attractive sources for the use of altmetrics in research evaluation (Sud & Thelwall, in press). The following reasons are chiefly given for this:

- The collection of literature in reference managers is – similar to the way this is the case with citations and downloads of publications – a by-product of existing workflows (Haustein 2014). This is why saves are appropriate as an alternative metric chiefly for the measurement of impact in areas of work where literature is collected and evaluated (such as with researchers in academic and industrial research, students and journalists).
- Whereas the impact of classical papers can be measured very well via citations in databases (such as the WoS), this is hardly possible with other types of publication such as books or reports.
- According to Mohammadi and Thelwall (2014), usage data of literature may be partially available (i.e. from publishers); but there is a shortage of global and publisher-independent usage data.
- Data sets of online reference managing platforms are highly accessible. The data may be available via API or database dumps (Priem & Hemminger, 2010).

However, the use of data from online reference managers is not only seen as advantageous, but also as problematic:

- Since not everybody who reads and uses scientific literature works with an online reference manager (and Mendeley, particularly), there is the problem that the evaluation of saved data only takes into account a part of the actual readership. Among researchers this part is probably younger, more sociable and more technologically-oriented than average for researchers (Sud & Thelwall, in press).
- The data which are entered by users into the online reference managers are erroneous or incomplete. This can lead to saves not being able to be associated unambiguously with a publication (Haustein, 2014).

Similar to Twitter citations, readership counts can also be manipulated relatively simply (for example with artificially generated spam) (Bar-Ilan et al., 2014).

Many of the empirical-statistical studies into social bookmarking – according to Priem and Hemminger (2010) – deal with tags and tagging. Seen overall, the studies come to the conclusion that exact overlaps of tags and professionally created metadata are rare; most matches are found when comparing tags and title terms (Haustein & Peters, 2012). A large part of the studies into online reference managers has evaluated the correlation between traditional citations (from Scopus, Google Scholar and the WoS) and bookmarks in Mendeley and/or CiteULike. The meta-analysis of (Bornmann, 2015) shows that the correlation is medium to large (CiteULike pooled r=0.23; Mendeley pooled r=0.51).

Two studies have already investigated country information from Mendeley: (1) Haustein and Larivière (2014) analyzed the journal *Aslib Proceedings* (AP) with a set of indicators from several perspectives. The results show that the largest share of AP papers in the last eight years were written by authors affiliated to UK (58 %), Iran (6 %), South Africa and USA (both 5 %). In contrast, Mendeley readers of AP articles were mainly from the USA (14 %), UK (12 %), Spain (6 %), India (4 %), Canada (3 %), South Africa (3 %) and Malaysia (2 %). (2) For some WoS categories, Thelwall and Maflahi (in press) downloaded all article (article meta data) that were written in English from 2011. The country affiliation of the authors was extracted from the WoS affiliation field; each article was searched for in Mendeley to receive the number of readers from each country. The results of the study show that there is a tendency for articles to be more read in countries with a higher share of their authorship. Possible reasons for the tendency are that authors are often readers of their own articles and that the readers often know or have heard of the authors.

Methods

Peer ratings provided by F1000Prime

F1000Prime is a post-publication peer review system of the biomedical literature (papers from medical and biological journals). F1000 Biology was launched in 2002 and F1000 Medicine in 2006. The two services were merged in 2009 and today form the F1000 database. Papers for F1000Prime are selected by a peer-nominated global Faculty of leading scientists and clinicians who then rate them and explain their importance (F1000, 2012). This means that only a restricted set of papers from the medical and biological journals covered is reviewed, and most of the papers are actually not (Kreiman & Maunsell, 2011, Wouters & Costas, 2012).

The Faculty nowadays numbers more than 5,000 experts worldwide, assisted by 5,000 associates, which are organized into more than 40 subjects (which are further subdivided into over 300 sections). On average, 1,500 new recommendations are contributed by the Faculty each month (F1000, 2012). Faculty members can choose and evaluate any paper that interests them; however, the great majority pick papers published within the past month, including advance online papers, meaning that users can be made aware of important papers rapidly (Wets et al., 2003). Although many papers published in popular and high-profile journals (e.g. *Nature*, *New England Journal of Medicine*, *Science*) are evaluated, 85% of the papers selected come from specialized or less well-known journals (Wouters & Costas, 2012). Less than 18 months since Faculty of 1000 was launched, the reaction from scientists has been such that two-thirds of top institutions worldwide already subscribe, and it was the recipient of the Association of Learned and Professional Society Publishers (ALPSP) award for Publishing Innovation in 2002 (http://www.alpsp.org/about.htm) (Wets et al., 2003).

The papers selected for F1000Prime are rated by the members as good, very good, or exceptional, which is equivalent to recommendation scores (rs) of 1, 2, or 3, respectively. Since many papers are not rated by one member alone, but by several, we calculated a mean rs for every paper. In order to categorize the F1000 papers into three quality levels, papers with mean rs < 2 have been categorized as Q1 and papers with mean rs > 2.5 as Q3. Papers with rs in-between are categorized as Q2, then. This is not a categorization of low and high quality because all F1000Prime papers have a very high quality compared to other papers in their field. This is merely a further distinction between high quality papers, as papers with low quality do not get recommended into F1000Prime.

Data sets used from Mendeley and WoS

In January 2014, F1000 provided one of the authors with data on all recommendations (and classifications) made and the bibliographic information for the corresponding papers in their system (n_r=149,227 records, n_p = 114,582 papers). Each of these records with either a PubMed-ID or a DOI was used to retrieve the Mendeley usage statistics via the R (http://www.r-project.org, accessed October 14, 2014) API of Mendeley (https://github.com/Mendeley/mendeley-api-r-example, http://dev.mendeley.com/methods/, both accessed October 14, 2014). An example R script is available at http://dx.doi.org/10.6084/m9.figshare.1335688. In the summer of 2014, a new version of the API was released which we used for this study (Bonasio, 2014). The previous API had some limitations, such as providing only the information of the demographics for the top three categories as a percentage. Another problem (which has not been solved yet) is that most users do not record their country and so only some readership country location information is available (Thelwall & Maflahi, in press). We requested the actual numbers of Mendeley users for each F1000 record (and the result was not truncated after the top three categories). We observed several (probably random) connection problems. Overall, about 99% of the F1000 paper set was found on Mendeley, which implies a rather good coverage of scientific papers on Mendeley (Bornmann & Haunschild, 2015). We recorded a total of 5,885,534 Mendeley reader counts.

For bibliometric analysis in the current study, country information of the authors who published a F1000 paper or published a paper citing a F1000 paper were sought in an in-house database of the Max Planck Society (MPG) based on the WoS and administered by the Max Planck Digital Library (MPDL). Despite different meanings of (citing) authors' and readers' countries, we talk about countries of readers and (citing) authors in the same way in the following sections.

Technical limitations

Only about 17.6% of 5,885,534 Mendeley reader counts (n=1,038,449) provided were available with their country association. For only 1,064 records of the F1000 data set, we found that the sum over all reader's countries was equal to the total number of reader counts. Thus, in the majority of cases (99.3%) some Mendeley readers are missing in our statistic because many readers did not share their location.

In contrast to the Mendeley data (in which the country information is reader-specific), the country information for the (citing) authors is address-specific. If two authors have different addresses, the country information is counted twice. However, if the addresses are identical, they are counted once. This limitation is unavoidable using our current WoS data. A second limitation of the data is that papers with different publication years have been considered without time-normalization in the study. For different publication years, one can expect different numbers of readers and citers: The longer the reader and citation window, the more counts are expectable. Since the counts have not been time-normalized in the study, papers with longer windows will have a greater effect on the results than papers with smaller windows. However, the papers with longer and smaller windows are unsystematically distributed across the different quality levels of the papers. Thus, the missing time-normalization of the data won't influence the investigation of the relationship between the distribution of readers and (citing) authors across countries and quality levels.

Processing and visualization of the data

The Mendeley reader data, as well as the WoS author and citer data, were processed by Perl (http://www.perl.org/) and Gawk (http://awk.info/) scripts. Visualization of the data was carried out using Tableau (http://www.tableausoftware.com/). Plots of country and world maps use the Mercator projection.

Results

The results of the study including all F1000 papers with data from WoS and Mendeley are shown in Figures 1 and 2, as well as Table 1 (all papers). For each country, we calculated the percentage of authors, readers, and citers. In Figure 1, the percentage of authors (red colour), citers (blue colour), and readers (green colour) are visualized for all countries worldwide. Figure 2 shows a more detailed analysis of Europe as very many circles are overlapping in this region in Figure 1. The left panel of Figure 2 compares readers (green colour) and authors (blue colour) while the right panel compares citers (red colour) and authors (blue colour). The bigger the circle on the maps, the higher the percentage for a country is.

As the results in Figure 1 show most authors, readers, and citers are located in the USA. The results in Table 1 (all papers) point out that 29.2% of all readers, 38.3% of all authors, and 39.9% of all citers come from the USA. The USA is the country with the most readers, authors, and citers – significantly more than any other country. The high percentages of authors and citers point to a high level of research activity in the USA. The population and number of research groups in the USA are significantly higher compared to most other countries. In Table 1 (all papers), the USA is followed by the UK (all papers: readers=10.7%, citers=6.6%, and authors=9.3%). Further countries in the table (Germany, France, Japan, and Canada) show small differences in the percentages compared to the UK (less than 10 percentage points). Despite the rather large number of research groups in Russia and China, it is quite surprising that both do not appear in the top 10 list ordered by the number of Mendeley readers. In fact, we find China on rank 13 and Russia on rank 25, close to Poland and the Czech Republic.

As the results in Table 1 further show, many countries have different percentages of authors, readers, and citers. The US has a similar percentage of authors and citers (see e.g. the numbers for all papers), but the percentage of readers is lower than both other percentages. This result seems to reflect the fact that Mendeley is only one reference manager software among others in the USA. For other countries it is the other way around. For example, while 4.7% of all readers come from Brazil (all papers), less than 1% of all authors and citing authors are working in this country.

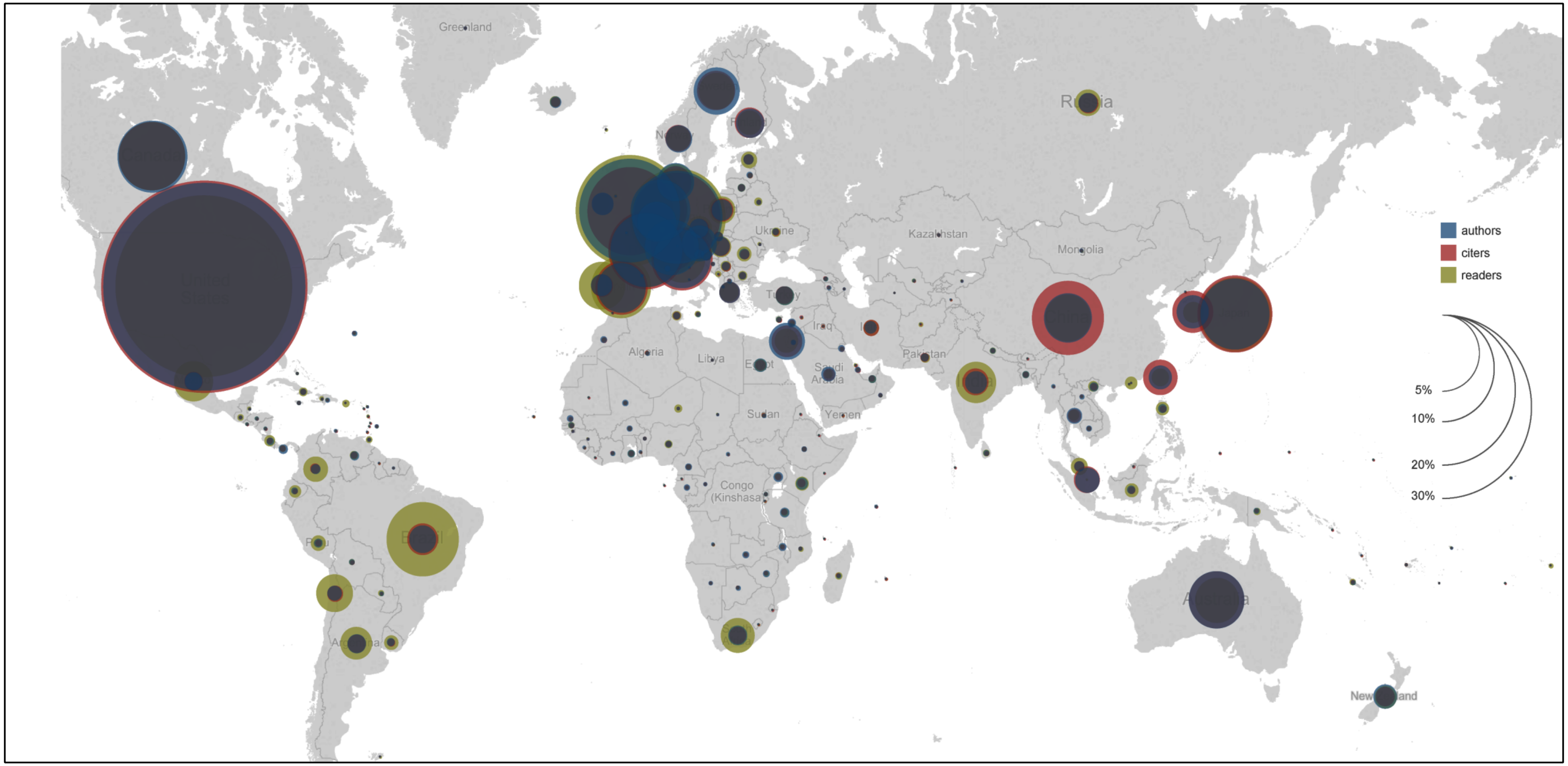


Figure 1. Percentage of authors (blue colour), citers (red colour), and readers (green colour). The circle sizes indicate the share of the country in the amount of readers, citers and authors, respectively. The map is based on all F1000 papers.

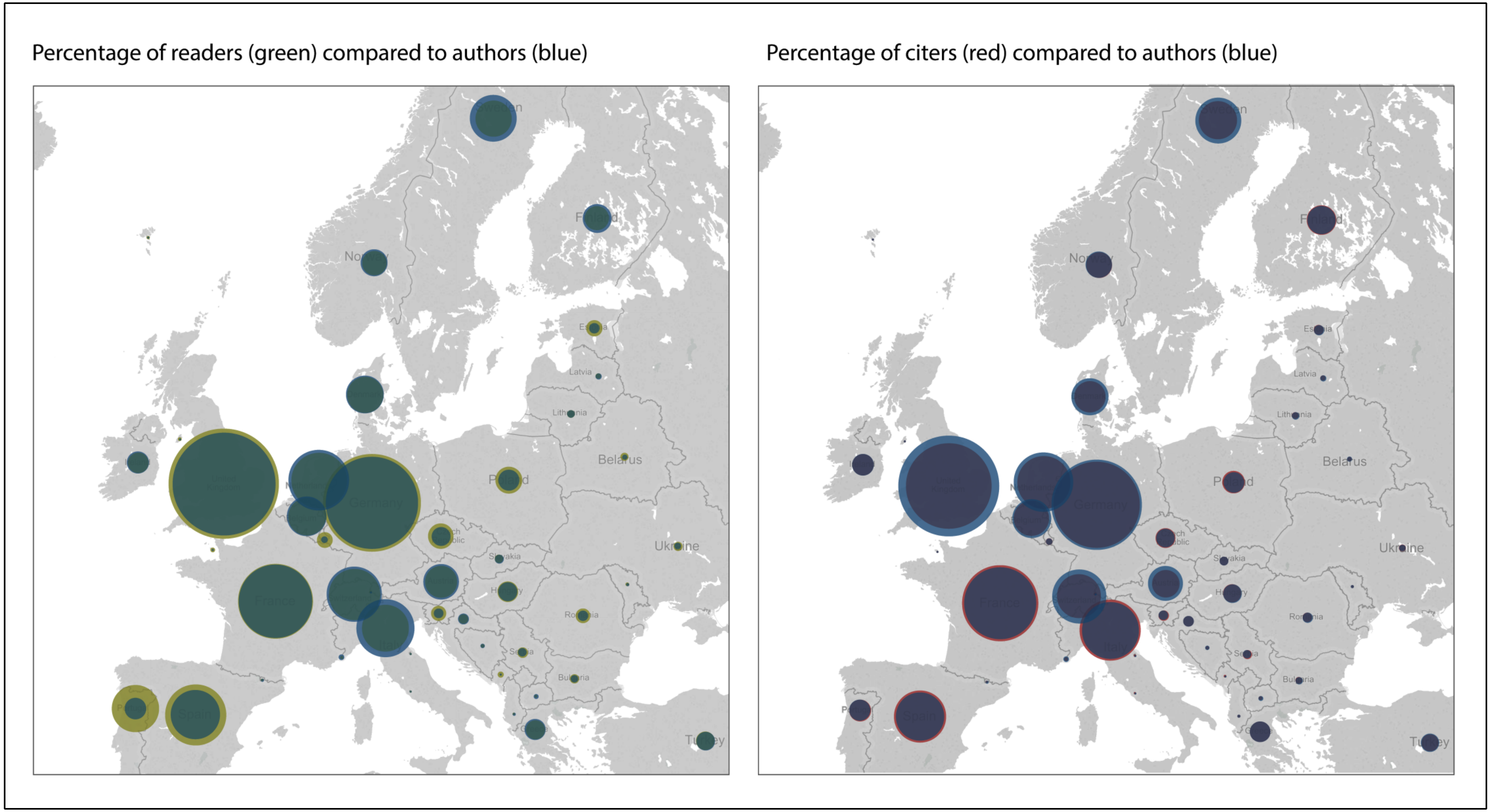


Figure 2. Percentage of readers (green colour) and authors (blue colour) on the left panel, as well as percentages of citers (red colour) and authors (blue colour) visualized on the right panel for European countries. The circle sizes indicate the share of the country in the amount of readers, citers and authors, respectively. The map is based on all F1000 papers.

Table 1. Percentage of authors, citers, and readers from different countries. The percentages are presented for all papers, as well as for papers with Q1 (rs<2), Q2 (rs>=2 and rs<=2.5), and Q3 (rs>2.5) quality. The ten countries are listed with the highest percentage of readers.

All papers	Authors	Citers	Readers	Q1	Authors	Citers	Readers
USA	38.3	39.9	29.2	USA	37.7	39.4	28.7
UK	9.3	6.6	10.7	UK	9.2	6.6	10.7
Germany	7.4	6.8	8.4	Germany	7.4	6.7	8.3
France	4.7	5.2	4.9	Brazil	0.6	0.8	5.0
Japan	4.3	5.1	4.7	France	4.7	5.3	4.9
Brazil	0.6	0.8	4.7	Japan	4.3	5.0	4.5
Canada	4.4	4.0	4.0	Canada	4.5	4.0	4.0
Spain	2.0	2.4	3.2	Spain	2.1	2.5	3.3
Netherlands	3.1	2.5	2.6	Netherlands	3.2	2.5	2.6
Switzerland	2.6	1.7	2.2	Switzerland	2.4	1.6	2.2
Q2	Authors	Citers	Readers	Q3	Authors	Citers	Readers
USA	39.0	40.4	29.4	USA	40.7	41.2	30.6
UK	9.5	6.7	10.7	UK	9.3	6.6	10.7
Germany	7.5	6.9	8.5	Germany	8.0	6.8	8.4
Japan	4.3	5.1	5.0	Japan	4.2	5.4	5.1
France	4.5	5.2	4.8	France	4.6	5.0	4.6
Brazil	0.5	0.7	4.4	Canada	4.2	3.7	4.0
Canada	4.3	3.9	3.9	Brazil	0.6	0.8	4.0
Spain	2.0	2.3	3.1	Spain	1.5	2.2	3.0
Netherlands	3.1	2.4	2.6	Netherlands	2.9	2.3	2.7
Switzerland	2.8	1.7	2.2	Switzerland	2.9	1.7	2.1

This result points out that Brazil rather receives than produces scientific results in the field of biomedical research: Since a low percentage of citing authors reflects a low number of subsequent published papers (following and basing on the F1000Prime papers), this percentage is not only an indicator of reception but also of productivity. Similar results as for Brazil are not only visible on the map in Figure 1 for other south-American countries (such as Argentina or Chile), but also for India and African countries.

From the European countries, Spain and Portugal receive more F1000 papers than they produce (c.f. left panel of Figure 2). Spain is located on rank 8 (see Table 1), and Portugal is located on rank 11. The northern European countries produce more F1000 papers than they cite (c.f. right panel of Figure 2). This is vice versa for most southern European countries.

Table 1 shows the percentage of authors, citers, and readers from different countries not only for all papers, but also for papers with different rs: Q1 (rs < 2), Q2 (2 <= rs <= 2.5), and Q3 (rs > 2.5) section. Comparing the numbers of authors, citers, and readers for different paper quality levels, we see only minor differences for most countries: Brazil shows a somewhat higher amount of readers in the Q1 section (5%) than in the Q3 section (4%), while the percentage of authors and citers does not differ at all between Q3 and Q1 section papers. The USA shows a somewhat higher amount of authors, citers, and readers in the Q3 section (40.7%, 41.2%, and 30.6%, respectively) than in the Q1 section (37.7%, 39.4%, and 28.7%, respectively). The UK shows a nearly constant percentage across quality levels for authors, citers and readers: 9.2%, 6.6%, and 10.7%, respectively for Q1, 9.5%, 6.7%, and 10.7%, respectively for Q2, and 9.3%, 6.6%, and 10.7%, respectively for Q3.

Discussion

By far the highest number of authors, citers, and readers are located in the USA. More F1000 papers are authored, cited, and read in western European countries than in eastern European countries. The amount of F1000 papers authored, cited, and read in China and Russia is small compared to the large number of research groups located there (rank 13 and 25, respectively, according to Mendeley readers). Other reference softwares might be more popular in these countries (or this kind of software is scarcely in use). Traffic data from Alexa.com can be used as an estimate for the Mendeley distribution. The top 5 countries where Mendeley is used seem to be USA (30.4%), India (20.7%), UK (4.3%), Pakistan (3.9%), and Malaysia (3.0%) (http://www.alexa.com/siteinfo/www.mendeley.com, visited on 19 December 2014). Roughly a year earlier, the top 5 countries were somewhat different: USA (16.1%), India (13.2%), Belgium (9.9%), Germany (6.2%), and UK (5.9%) (Thelwall and Maflahi, in press). This relative gain of Mendeley traffic from India, Pakistan, and Malaysia is different from our results, as they do not appear on our top 10 list of Mendeley readers. Within the F1000 readership on Mendeley, India is on rank 15, Malaysia on rank 38, and Pakistan on rank 59. Probably, scientists who use Mendeley in these countries are not that active in the bio-medical research. Belgium, which was in the top 5 list of Mendeley traffic a year ago, is on rank 17 according to our Mendeley readership results of the F1000 paper set.

We find only minor differences in the readership of papers with different quality levels Q1-Q3. The similarities of the results across paper quality levels can be explained with the very high standard of all publications in the F1000Prime set. Also, papers within the Q1 quality section in the F1000 publication set gather a rather high amount of citations (Bornmann 2014). Considering that all papers in the F1000 publication set are of a higher than average quality in the biomedical area, one probably cannot expect a clear difference between quality levels in the Mendeley readership.

Most countries show a quite good balance between consumption and production of F1000 papers. See for example in Table 1, the percentages of Germany are 7.4% authors, 6.8% citers, and 8.4% readers. Although scientists in Germany seem to consume somewhat more of the literature of the F1000 paper set, the difference between authors (citers) and readers can be neglected, considering the limitations of our study and the (necessary) counting of authors (citers) and readers on unequal footing. In contrast to Germany, the number of readers is significantly higher than the number of authors and citers in some south-American countries (e.g. Brazil, Mexico, Chile, and Argentina) and some European and Asian countries (e.g. Portugal and India).

It is important to keep in mind that we measure authors and citers based on their institutional affiliation and readers on a personal level.

Another problem in the interpretation of the results is that the distribution of the Mendeley software is probably different for each country. Mendeley is free of charge. Thus, one could expect a higher number of Mendeley users in countries with tight research budgets. However, scientists in countries with tight research budgets might not author, cite, or read many publications which got recommended into F1000Prime, as many F1000Prime papers were published in journals with rather high subscription fees.

A third problem in the interpretation of the results is that a rather small number of readers provide their country, as it is not mandatory information. While we found approximately 99% of the F1000 papers at Mendeley, country information were available only for nearly 18% of the reader counts. This is significantly less than the value reported in a previous study done using a much smaller amount of papers (Haustein and Larivière 2014). However, it is reasonable to expect that Mendeley users who do not provide their location are evenly distributed over the world and are reading all quality classes of the F1000 papers.

Acknowledgments

Lutz Bornmann would like to thank Adie Chan, Ros Dignon, and Iain Hrynaszkiewicz from F1000 for providing him with the F1000Prime data set

References

Bar-Ilan, J., Shema, H. & Thelwall M. (2014). Bibliographic References in Web 2.0. In B. Cronin and C. R. Sugimoto (Eds.), *Beyond bibliometrics: harnessing multi-dimensional indicators of performance* (pp. 307-325). Cambridge, MA, USA: MIT Press.

Bonasio, A. (2014). *A look at Mendeley Readership Statistics*. Retrieved October 14, 2014, from http://blog.mendeley.com/academic-features/a-look-at-mendeley-readership-statistics/.

Bornmann, L. (2014). Validity of altmetrics data for measuring societal impact: A study using data from Altmetric and F1000Prime. *Journal of Informetrics*, *8*(4), 935-950.

Bornmann, L. (2015). Alternative metrics in scientometrics: A meta-analysis of research into three altmetrics. *Scientometrics*, 103(3), 1123-1144

Bornmann, L. & Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45-80.

Bornmann, L. & Haunschild, R. (2015). Which people use which scientific papers? An evaluation of data from F1000 and Mendeley, *Journal of Informetrics*, 9(3), 477-487

F1000. (2012). What is F1000? Retrieved October 25, from http://f1000.com/about/whatis.

Galloway, L. M., Pease, J. L. & Rauh, A. E. (2013). Introduction to altmetrics for science, technology, engineering, and mathematics (STEM) librarians. *Science & Technology Libraries*, 32(4), 335-345.

Haustein, S. (2014). Readership metrics. In B. Cronin and C. R. Sugimoto (Eds.), *Beyond bibliometrics: harnessing multi-dimensional indicators of performance* (pp. 327-344). Cambridge, MA, USA: MIT Press.

Haustein, S. & Lariviѐe, V. (2014). A multidimensional analysis of Aslib Proceedings – using everything but the impact factor. *Aslib Journal of Information Management*, *66*(4), 358-380.

Haustein, S. & Peters, I. (2012). Using social bookmarks and tags as alternative indicators of journal content description. *Firstmonday*, *17*(11).

Haustein, S., Peters, I., Bar-Ilan, J., Priem, J., Shema, H. & Terliesner, H. (2014). Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics*, 1-19.

Kreiman, G. & Maunsell, J. H. R. (2011). Nine criteria for a measure of scientific output. *Frontiers in Computational Neuroscience*, 5.

Li, X., Thelwall, M., & Giustini, D. (2012). Validating online reference managers for scholarly impact measurement. *Scientometrics*, *91*(2), 461-471.

Mohammadi, E. & Thelwall, M. (2013). Assessing the Mendeley readership of social science and humanities research. *Proceedings of ISSI 2013 Vienna: 14th International society of scientometrics and informetrics conference*. J. Gorraiz, E. Schiebel, C. Gumpenberger & M. Ho. Vienna, Austria, Austrian Institute of Technology GmbH**:** 200-214.

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.

Neylon, C., Willmers, M. & King, T. (2014). Rethinking Impact: Applying Altmetrics to Southern African Research. Ottawa, Canada: International Development Research Centre.

Priem, J. (2014). Altmetrics. In B. Cronin and C. R. Sugimoto (Eds.), *Beyond bibliometrics: harnessing multi-dimensional indicators of performance*. Cambridge, MA, USA: MIT Press.

Priem, J. & Hemminger, B. M. (2010). Scientometrics 2.0: toward new metrics of scholarly impact on the social Web. *First Monday*, *15*(7).

Rodgers, E. P. & Barbrow, S. (2013). A look at altmetrics and its growing significance to research libraries. Ann Arbor, MI, USA, The University of Michigan University Library.

Shema, H., J. Bar-Ilan & Thelwall, M., (2014). Do blog citations correlate with a higher number of future citations? Research blogs as a potential source for alternative metrics. *Journal of the Association for Information Science and Technology*, *65*(5), 1018-1027.

Sud, P. & Thelwall, M. (in press). Not all international collaboration is beneficial: the Mendeley readership and citation impact of biochemical research collaboration. *Journal of the Association for Information Science and Technology.*

Taylor, M. (2013). Towards a common model of citation: some thoughts on merging altmetrics and bibliometrics. *Research Trends*, 35, 19-22.

Thelwall, M. & Maflahi, N. (in press). Are scholarly articles disproportionately read in their own country? An analysis of Mendeley readers. *Journal of the Association for Information Science and Technology*.

Weller, K. & Peters, I. (2012). Citations in Web 2.0. In A. Tokar, M. Beurskens, S. Keuneke et al. *Science and the Internet* (pp. 209-222). Germany, Düsseldorf: University Press.

Wets, K., Weedon, D. & Velterop, J. (2003). Post-publication filtering and evaluation: Faculty of 1000. *Learned Publishing,* 16(4), 249-258.

Wouters, P. & Costas, R. (2012). *Users, narcissism and control – tracking the impact of scholarly publications in the 21st century*. Utrecht, The Netherlands: SURFfoundation.

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*, 1-23.

Do Mendeley Readership Counts Help to Filter Highly Cited WoS Publications better than average citation impact of journals (JCS)?

Zohreh Zahedi[1,2], Rodrigo Costas[1] and Paul Wouters[1]

z.zahedi.2@cwts.leidenuniv.nl; rcostas@cwts.leidenuniv.nl; p.f.wouters@cwts.leidenuniv.nl
[1]CWTS, Leiden University, P.O. Box 905, Leiden, 2300 AX (The Netherlands)
[2]Department of Knowledge & Information Sciences (KIS), Faculty of Humanities, Persian Gulf University, Bushehr, 7516913817 (Iran)

Abstract
In this study, the 'academic status' of users of scientific publications in Mendeley is explored in order to analyse the usage pattern of Mendeley users in terms of subject fields, citation and readership impact. The main focus of this study is on studying the filtering capacity of Mendeley readership counts compared to journal citation scores in detecting highly cited WoS publications. Main finding suggests a faster reception of Mendeley readerships as compared to citations across 5 major field of science. The higher correlations of scientific users with citations indicate the similarity between reading and citation behaviour among these users. It is confirmed that Mendeley readership counts filter highly cited publications (PPtop 10%) better than journal citation scores in all subject fields and by most of user types. This result reinforces the potential role that Mendeley readerships could play for informing scientific and alternative impacts.

Conference Topic
Altmetrics

Introduction
Mendeley is a popular reference management tool and a rich source of readership metrics for scholarly outputs, used by more than 2.5 million users[1]. This platform collects a wide variety of different metadata[2] for each publication saved by the different types of users in their individual library. Among these metadata, statistics about 'academic status', 'discipline' and 'country' provide useful information on the typologies of users of scientific publications in Mendeley.
Mendeley has different coverage and presence across different fields of science (Zahedi, Costas & Wouters, 2014). A moderate correlation between Mendeley readership and citation counts has been observed for different sets of publications from different fields showing that Mendeley readership counts reflect similar but (perhaps) also other types of impact (Thelwall et al., 2013; Haustein et al., 2013; Zahedi, Costas & Wouters, 2014; Mohammadi & Thelwall, 2014). Also, a weak correlation among number of authors, departments, institutions and countries and readership and citation counts for WoS publications has been observed (Sud & Thelwall, in press; Thelwall & Maflahi, in press). Research on users showed that the majority of Mendeley users per publication are PhDs and students. However, one important limitation with Mendeley data on the analysis of users was the data restriction caused by the reporting of only the three most common user types per publication. Full data on users are necessary in order to properly determine the readership patterns among types of users (Zahedi, Costas & Wouters, 2013 & 2014; Haustein & Larivière, 2014; Mohammadi et al., 2014).
The new Mendeley API provides data on all typologies of readers per publication. This means that 100% of all the users per publication are now fully reported[3]. This study represents one of

[1] http://blog.mendeley.com/start-up-life/mendeley-has-2-5-million-users/
[2] See: http://apidocs.mendeley.com/home/user-specific-methods/user-library-document-details
[3] according to William Gunn in the 1:Am altmetrics conference in London (September 2014) www.altmetricsconference.com/

the first approaches to the analysis of Mendeley readerships based on statistics per publication from all users. We overcome the main limitation of previous studies which were limited to restricted Mendeley users statistics.

In this paper, the usage patterns of the different Mendeley users based on their 'academic status'[4] by fields, citation and readership impact are studied. Also, we analyse the extent to which Mendeley readerships correlate with the number of citations and across 5 major fields of science in the Leiden Ranking (LR). An important focus of this study is on studying the filtering capacity of Mendeley readerships compared to journal citation scores in detecting highly cited publications. Therefore, particular attention will be paid to the extent to which highly cited outputs can be distinguished by these different impact indicators. Similarly, potential differences among Mendeley users in detecting highly cited publications will be also explored. The concrete objectives and research questions of the paper are the following:

O1: To study the general distribution of Mendeley readerships over WoS publications

Q2. What is the distribution of Mendeley readerships across LR fields and by different users?

O3: To study the relationship of Mendeley readerships with bibliometric indicators

Q4. Are there any differences in correlation by different Mendeley users and across LR fields?

O5: To investigate the ability to identify highly cited publications by Mendeley readerships in contrast to journal citation impact indicators

Q6. Which one of these impact indicators can better filter the WoS highly cited publications across LR fields and by different users?

Data and Methodology

For this study, we used a dataset of 1,196,421 Web of Science (WoS) publications from the year 2011 with Digital Object Identifiers (DOI). DOIs were used as the basis to extract readership metrics through the Mendeley REST API in mid-October 2014. The data from Mendeley has been matched with the CWTS in house WoS to add citation data. Citations have been calculated up to 2014.

Although Mendeley has released the full statistics for all the typologies of the users per publications through its API, some Mendeley user statistics are still missing from some publications[5]. These publications were excluded from the analysis due to their unclear reader counts and types. Limiting the dataset to articles and reviews, a final set of 977,067 publications received 12,418,426 total readerships[6] and 6,882,632 total citations. Comparing the ratios of mean citation score per publication (MCS) and mean readerships per publication (MRS), we also find higher MRS (12.7) than MCS (7.04). The actual number of the different types of Mendeley users per publication has been calculated as well as several bibliometrics

[4] These are the different types of users in Mendeley (i.e. PhD students, Professors, Post doc, researchers, Students (under graduates and post graduates), Librarians, Lecturers, Other Professionals and Academic and non-Academic researchers) who have saved publications in their individual libraries. This information allows us to identify users of scientific publications but this information is not free of limitations. For example, it is not clear whether the academic status of the users is updated regularly or how to distinguish users who could belong to more than one category (e.g. a librarian who is also a PhD student).

[5] There are 144,8495 publications with missing readership statistics. These publications have been saved in Mendeley but since their readership counts are missing, they are excluded from the analysis.

[6] We have found some inconsistencies in the counts of readerships. There is a difference between the sum of total readership counts reported by Mendeley (i.e. as they come directly from the readership count provided by Mendeley) and the sum of the individual Mendeley readerships by the different users (calculated by ourselves). (12,418,426 - 12,412,305=6121 differences)

indicators. Precision-recall analysis (Waltman & Costas, 2014) has also been performed, considering 5 major fields of science as represented in the Leiden Ranking (LR)[7].

Analysis and Results

General distribution of Mendeley readerships by major fields of Science and by Mendeley users

Table 1 shows that Biomedical & health sciences (37%) have the highest share of publications with readerships while Mathematics and computer science (8%) have the lowest share. In terms of readership density (i.e. MRS scores) the Life & earth sciences have the highest values (17.5) followed by the Social science & humanities (17), Biomedical & health sciences (14.4) and Natural sciences & engineering (9.7). Mathematics and computer science (9.4) exhibit the lowest readerships density. Also, on average, all fields show higher MRS scores than MCS scores. This could be explained by the relative early publication year (2011) of publications, which could still need some time to get their optimum levels of citations, while in terms of social media, the uptake is normally faster (Haustein et al, 2013), although we still lack information on the obsolescence and time patters of readerships for publications.

Table 1. Mendeley readerships distribution across 5 major fields of science in LR.

LR Main fields of all Publications	P	%	TCS	%	MCS	TRS	%	MRS
Biomedical & health sciences	419,693	37	3617563	44	8,6	6051206	39	14,4
Natural sciences & engineering	322,009	28	2362700	29	7,3	3119704	20	9,7
Life & earth sciences	204,392	18	1469979	18	7,2	3572266	23	17,5
Social sciences & humanities	105,827	9	422046	5	4,0	1795194	12	17,0
Mathematics & computer science	90,813	8	332946	4	3,7	857319	6	9,4
Total		100		100			100	

Total Citation Score (TCS); Total Readership Score (TRS); Mean Citation Score (MCS); Mean Readership Score (MRS)

Figure 1 shows the proportion of readerships by the different types of Mendeley users across the LR fields. Although there are some differences across the fields, in general we find that PhD and students are the most common types of users while Lecturers and Librarian are the least common types of users across all LR fields.

[7] http://www.leidenranking.com/ranking/2013

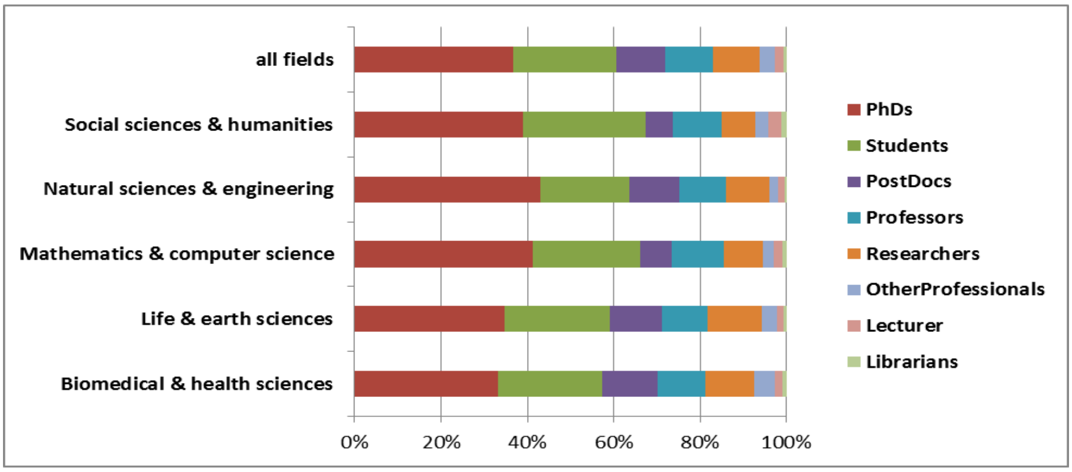


Figure 1. Distribution of Mendeley readerships by the different types of users across LR fields.

Relationship of Mendeley readerships with bibliometric indicators

Spearman correlation analysis among readerships and bibliometric indicators and by the different types of users and across LR Fields has been calculated. The focus here is to explore the extent to which the readerships for the publications saved by the different users in Mendeley are related to their citations and journal indicators. Overall correlation scores among total readerships and bibliometrics indicators are positive and moderate ranging from p=.41 to p=.52 (Table 2).

Table 2. Spearman Correlation analysis of bibliometrics and altmetrics variables.

n=977,067	CS	NCS	JCS	NJCS	RS
CS	1	.93	.57	.43	.52
NCS		1	.40	.46	.50
JCS			1	.75	.44
NJCS				1	.41
RS					1

Citation Score (CS); Normalized Citation Score (NCS); Journal Citation Score (JCS); Normalized Journal Citation Score (NJCS); Readership Score (RS)

Regarding the different types of users, citations have a higher correlation with PhD followed by Students, PostDocs, Researchers, Professors and Other Professionals; however, Librarians and Lecturers exhibit the lowest correlations with citations. These different patterns in terms of correlations among the different types of users might suggest that they have different readership patterns and potentially different readership interests. For example, readership scores for Students, PostDocs, Professors and Researchers correlate most with PhD readership as ‘Scientific users’, which may indicate their similar scholarly and research usage behaviour. On the other hand, scientific users correlate less with ‘other professionals’ and Librarians (i.e. suggesting a kind of ‘Professional users’) and Lecturers as the ‘Educational users’ (Zahedi, Costas & Wouters, 2013). The latter also correlate most among themselves which may suggest both their similar use of scientific outputs and usage for other purposes than citation such as for self-awareness, teaching and educational or practical and professional purposes (Table3).

Table 3. Spearman Correlation analysis of citation and readerships variables by types of Mendeley users.

n=977,067	CS	PhDs	Students	Post Docs	Professors	Researchers	Other Professionals	Lecturers	Librarians
CS	1	.46	.40	.41	.36	.37	.24	.18	.06
PhDs		1	.58	.49	.48	.47	.25	.27	.08
Students			1	.41	.44	.44	.31	.29	.12
PostDocs				1	.42	.43	.26	.21	.06
Professors					1	.39	.27	.26	.09
Researchers						1	.32	.23	.11
Other Professionals							1	.20	.12
Lecturers								1	.09
Librarians									1

In terms of LR fields, the correlation of citations and readerships is the highest for Social sciences and humanities (p=.61) followed by Natural sciecnes and engineering (p=.59), Life and earth sciences (p=.57), Biomedical and health sciences (p=.55) and the least for Mathematics and computer sciences (p=.45). Regarding the readership by user types and across fields, for most users the highest correlations are in Social sciences and humanities. The lowest correlation with citations is in the field of Mathematics and computer sciences for PhD, Students, PostDocs, Professors and Researchers while for Other Professionals, Lecturers and Librarians the field Natural sciences and engineering displays the lowest correlation with citations (Table 4). This may indicate a relatively stronger use of social media platforms such as Mendeley by scholars in Social science and humanities in their research process than other fields (Rowlands et al., 2011; Tenopir, Volentine & King, 2013).

Table 4. Spearman Correlation analysis of citation and readership by types of Mendeley users across 5 LR Fields.

LR Fields	Total CS and RS	PhD	Student	Post Doc	Professor	Researcher	Other Professional	Lecturer	Librarian
Biomedical & health sciences	.55	.47	.42	.42	.40	.39	.26	.19	.05
Natural sciences &engineering	.59	.51	.43	.39	.35	.33	.17	.18	.04
Life & earth sciences	.57	.53	.46	.43	.40	.39	.24	.22	.06
Mathematics & computer science	.45	.42	.34	.26	.26	.27	.18	.18	.05
Social sciences & humanities	.61	.54	.50	.41	.43	.42	.31	.27	.12

CS (Citation Score); RS (Readership Score)

Analyzing the filtering capacity of highly cited publications by Mendeley readerships

The focus here is to explore the potential use of Mendeley users for filtering highly cited publications compared to journal citation scores. For this purpose, the proportion of top 10% highly cited publications (PPtop 10%)[8] in the sample have been detected. The precision-recall analysis[9] has been performed for all publications in the sample and the 5 LR fields and the different Mendeley users have been explored. Figure 2 shows the general precision-recall analysis of total readership scores and Journal Citation Scores (JCS) for all the publications in the dataset. This figure shows that readerships perform better than JCS in identifying the PPtop 10% most cited publications. The figure indicates that for example a recall of 0.5 (50%) corresponds with a precision of 0.45 (45%) for readership and 0.25 (25%) for journal citation scores in identifying highly cited publications, that is, publications belonging to the top 10% of their field in terms of citations. This means that in order to select half of all highly cited publications we have an error rate of 55% when the selection is made based on readership and an error rate of 75% when the selection is made based on journal citation scores. Since readership outperforms journal citation scores at all levels of recall, we conclude that readership scores identify highly cited publications much better than JCS.

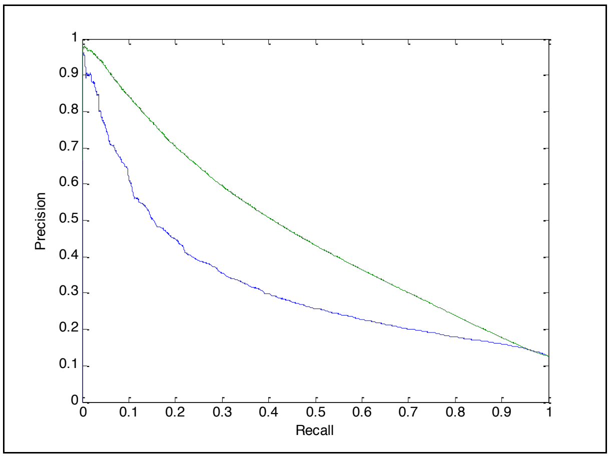
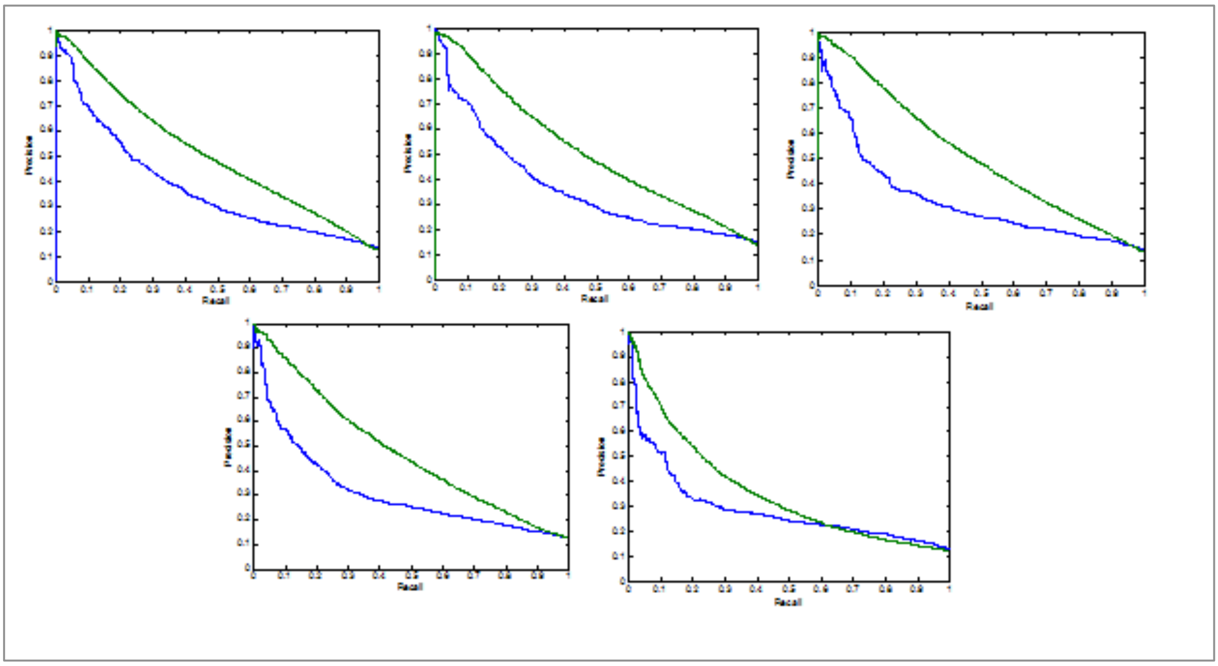


Figure 2. General Precision-recall curves for JCS (blue line) and total readerships (green line) for identifying PPtop10% most highly cited publications.

[8] PP(top 10%) (proportion of top 10% publications). Refers to the proportion of the publications that compared with other publications in the same field and in the same year, belong to the top 10% most frequently cited.

[9] following Waltman & Costas (2014), For a given selection of publications, "precision is defined as the number of highly cited publications in the selection divided by the total number of publications in the selection. Recall is defined as the number of highly cited publications in the selection divided by the total number of highly cited publications".



From left to right: Biomedical & health sciences, Life & earth sciences, Natural Sciences & engineering, Social sciences & humanities, Mathematics & computer science

Figure 3.Precision-recall curves for JCS (blue line) and LR Fields (green line) for identifying PPtop10% most highly cited publications.

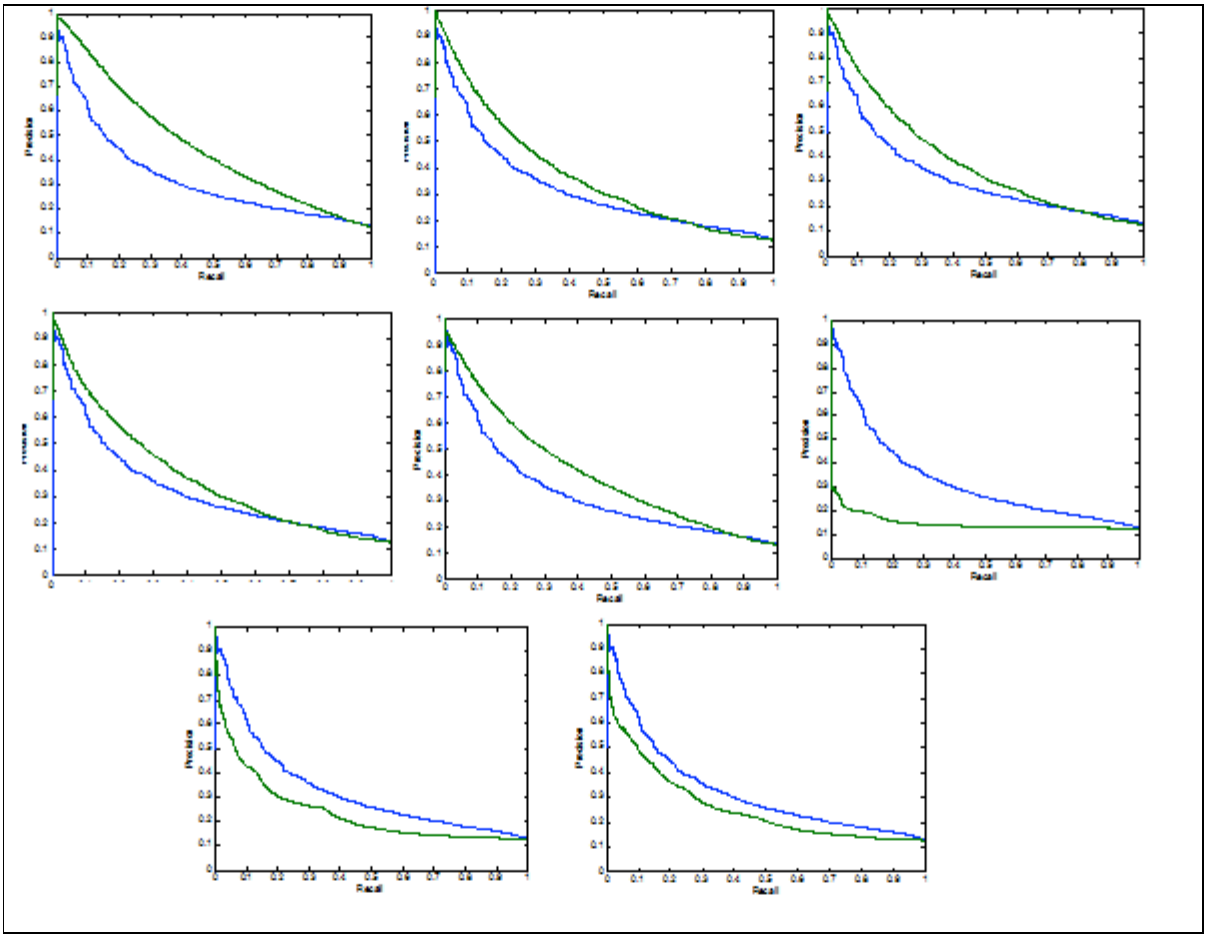
Precision-recall analysis of the different fields of science

The results of the precision-recall analysis for all fields of science again show that readership outperforms JCS scores in filtering highly cited publications. This result supports the idea that Mendeley readership counts filter highly cited publications better than average citation impact of journals (JCS) for all LR fields within our sample. All the figures are similar resembling the general pattern in figure 2 except the figure for Mathematics & computer science, which shows that from recall of 0.6 (60%), the two lines intersect each other and from that point onwards there is a small improvement of JCS over readership scores.

Precision-recall analysis of different types of Mendeley users

The same approach has been done based on the different Mendeley users. Figure 4 shows the results of the precision-recall analysis of readerships scores by the different types of users in Mendeley and Journal Citation Score (JCS). Again, readerships perform better than JCS for most types of users (PhDs, PostDocs, Professors, Researchers and Students vs. Other Professionals, Librarians and Lecturers) in identifying the PPtop10% most highly cited publications within our dataset thus resembling the general pattern in Figure 2. The only exceptions are observed for Librarians, Lecturers and other Professionals where JCS overlaps or outperforms Mendeley readerships. This is in line with the result of the correlation analyse in which these Mendeley user types exhibit less correlations with citations than other types.

Also, regarding the figures for PostDocs, Professors, Researchers and Students, from recall of 0.8 onwards two lines intersect each other and there is a slight improvement of JCS over readerships in the highest level of recall. However, in general, considering readership scores by most types of Mendeley users can help to detect highly cited publications.



From left to right: PhDs, PostDocs, Professors, Researchers, Students, Librarians, Lecturers and Other professionals

Figure 4. Precision-recall curves for JCS (blue line) and type of users readerships (green line) for identifying PPtop10% most highly cited publications.

Main results and discussion

Mendeley is a major multidisciplinary source of readership counts for scholarly publications (Zahedi, Costas & Wouters, 2014) and also it is one of the most promising tools for 'altmetrics' research (Li, Thelwall & Giustini, 2012; Wouters & Costas, 2012). The statistics about the 'Academic Status' of Mendeley users is a valuable source of information to learn more about the academic and non-academic positions of readers of scientific outputs, thus opening the possibility of studying the different types of impact that these different users may entail. Although Mendeley is now reporting the full data per publication, yet more clarity on how Mendeley users are defined is very important, as well as on how the typologies are chosen and updated by the users. For example, the relatively strong correlation between PhDs and Students could suggest that (some) students that become PhD do not update their profiles and therefore they 'read' like PhD students but without updating their 'Academic status' in Mendeley.

The current study has analysed and compared the readership and citation impact of the scholarly publications saved in Mendeley in terms of their types of users and across different LR fields, particularly focusing on the filtering capacity of readership and journal citation impact indicators in identifying highly cited publications. The findings showed that in terms of readership density across the 5 major LR fields, on average, all fields show higher MRS scores than MCS values. This suggests a faster reception of Mendeley readerships as compared to citations and encourages the need to study the temporality and pace of readership counts. Regarding the types of users, the most common types of users in Mendeley are PhDs and Students, for all LR fields. Correlation analysis shows relatively positive and moderate correlations among the different types of users and citations. The different correlations across users might support the idea that different users could be reading different publications, and thus justifying the use of 'Academic Status' to identify different reading behaviour and typologies of impact. For example, the higher correlations of scientific users with citations, supports their similar reading and citation behaviour vs. other more educational, teaching or professional patterns with lower correlations with citations. This may also be relevant in the analysis of the use of scientific publications in teaching or professional activities. Our results also suggest that readership counts really improve the filtering capacity of highly cited publications over JCS. This is one of the most promising results of this paper, showing the relevance of Mendeley readerships as a relevant filtering tool, something that has not been observed in the previous studies and for other altmetric sources (cf. Costas et al, 2014; Waltman & Costas, 2014). However, it should be taken into account that there are many scholars who don't use Mendeley or any other reference management tools in their scholarly process, so the act of using this type of tools may change in the future. Hence, the use of Mendeley readerships for evaluative purposes still needs careful consideration of its limitations and potential negative effects on the behaviour of individual scholars.

Acknowledgments

The authors are grateful to Erik van Wijk from CWTS for his support on the data collection of Mendeley information for this study. Also, special thanks to Ludo Waltman from CWTS for his fruitful comments on this paper.

References

Costas, R., Zahedi, Z. & Wouters, P. (2014). Do altmetrics correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, 1-31. DOI: 10.1002/asi.23309

Haustein, S., Peters, I., Bar-Ilan, J., Priem, J., Shema, H., & Terliesner, J. (2013). Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics*,101 (2): 1145-1163. DOI: 10.1007/s11192-013-1221-3

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

Li, X., Thelwall. M., & Giustini, D. (2012). Validating online reference managers for scholarly impact measurement. *Scientometrics*, 91(2), 461-471.

Mohammadi, E., Thelwall, M, Larivière, V., & Haustein, S. (2014). Who Reads Research Articles? An Altmetrics Analysis of Mendeley User Categories. *Journal of the Association for Information Sciences and Technology*. Available from:http://www.scit.wlv.ac.uk/~cm1993/papers/WhoReadsResearchArticlesPreprint.pdf

Rowlands, I., Nicholas, D., Russell, B., Canty, N., and Watkinson, A. (2011). Social media use in the research workflow. *Learned Publishing*, 24(3):183-195

Sud, P. & Thelwall, M. (in press). Not all international collaboration is beneficial: The Mendeley readership and citation impact of biochemical research collaboration. *Journal of the Association for Information Science and Technology*.

Tenopir, C., Volentine, R. & King, D.W. (2013).Social media and scholarly reading. *Online Information Review*, 37 (2), 193-216. DOI:10.1108/OIR-04-2012-0062

Thelwall, M,. Haustein, S. Larivière, V. & Sugimoto, C. (2013). Do altmetrics work? Twitter and ten other candidates. *PLOS ONE*, 8(5): e64841. doi:10.1371/journal.pone.0064841

Thelwall, M. & Maflahi, N. (in press). Are scholarly articles disproportionately read in their own country? An analysis of Mendeley readers. *Journal of the American Society for Information Science and Technology*.

Waltman, L., Van Eck, N.J., Van Leeuwen, T.N., Visser, M.S. & Van Raan, A.F.J. (2011). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*, 5(1): 37–47.

Waltman, L. & Costas, R. (2014). F1000 Recommendations as a Potential New Data Source for Research Evaluation: A Comparison with Citations. *Journal of the Association for Information Science and Technology*, *65*: 433–445. DOI: 10.1002/asi.23040

Wouters, P. & Costas, R. (2012). Users, Narcissism and control: Tracking the impact of scholarly publications in the 21st century. Utrecht: SURF foundation. Retrieved from: http://www.surffoundation.nl/nl/publicaties/Documents/Users%20narcissism%20and%20control.pdf

Zahedi, Z., Costas, R. & Wouters, P. (2013). What is the impact of the publications read by the different Mendeley users? Could they help to identify alternative types of impact? In PLoS ALM Workshop. 7-9 October. 2013. San Francisco. Available from:

https://openaccess.leidenuniv.nl/handle/1887/23579

http://article-level-metrics.plos.org/files/2013/10/Zahedi.pptx

Zahedi, Z., Costas, R. & Wouters, P. (2014). How well developed are Altmetrics? Cross-disciplinary analysis of the presence of 'alternative metrics' in scientific publications". Scientometrics, *101*(2): 1491-1513. DOI:10.1007/s11192-014-1264-0

Zahedi, Z., Costas, R. & Wouters, P. (2014). Assessing the impact of the publications read by the different Mendeley users: Is there any different pattern among users? Proceedings of the IATUL Conferences, Paper 4. Available from :http://docs.lib.purdue.edu/iatul/2014/altmetrics/4

Influence of Study Type on Twitter Activity for Medical Research Papers

Jens Peter Andersen[1] and Stefanie Haustein[2]

[1] *jepea@rn.dk*
Aalborg University Hospital, Medical Library, DK-9000 Aalborg (Denmark)

[2] *stefanie.haustein@umontreal.ca*
École de bibliothéconomie et des sciences de l'information, Université de Montréal, Montréal (Canada)

Abstract

Twitter has been identified as one of the most popular and promising altmetrics data sources, as it possibly reflects a broader use of research articles by the general public. Several factors, such as document age, scientific discipline, number of authors and document type, have been shown to affect the number of tweets received by scientific documents. The particular meaning of tweets mentioning scholarly papers is, however, not entirely understood and their validity as impact indicators debatable. This study contributes to the understanding of factors influencing Twitter popularity of medical papers investigating differences between medical study types. 162,830 documents indexed in Embase to a medical study type have been analysed for the study type specific tweet frequency. Meta-analyses, systematic reviews and clinical trials were found to be tweeted substantially more frequently than other study types, while all basic research received less attention than the average. The findings correspond well with clinical evidence hierarchies. It is suggested that interest from laymen and patients may be a factor in the observed effects.

Conference Topic

Altmetrics

Introduction

In the context of altmetrics, defined as "the study and use of scholarly impact measures based on activity in online tools and environments" (Priem, 2014, p. 266), Twitter has been identified as one of the most interesting and widely-used data sources (Costas, Zahedi, & Wouters, 2014; Thelwall, Haustein, Larivière, & Sugimoto, 2013). Although restricted by brevity—a tweet is limited to 140 characters—Twitter is at the heart of the altmetrics idea to enable a broader scope for impact assessment beyond citation impact. As Twitter is used widely and particularly outside of academia by currently 284 million monthly active users[1], tweets mentioning scientific papers are hoped to capture use by the general public and thus societal impact. Initially suggested as predictors of future citations and thus early indicators of scientific impact (Eysenbach, 2011), more recent large-scale empirical studies suggest that tweets are more likely to reflect online visibility including some social and scientific impact but also self-promotion and buzz (Costas et al., 2014; Haustein, Larivière, Thelwall, Amyot, & Peters, 2014; Haustein, Peters, Sugimoto, Thelwall, & Larivière, 2014). The most tweeted documents seem to attract a lot of online attention rather due to humorous or curious topics than their scientific contributions, often fitting "the usual trilogy of sex, drugs, and rock and roll" (Neylon, 2014, para. 6).

Various, mostly quantitative, studies have shown, with respect to scientific papers, that—after the reference manager Mendeley—Twitter is the altmetrics data source with the second-largest prevalence and it is constantly increasing to currently more than one fifth of 2012 papers being tweeted (Haustein, Costas, & Larivière, 2015). Correlation studies provide evidence that tweets and citations measure different things (for example, Costas et al., 2014;

[1] https://about.twitter.com/company

Haustein, Larivière, et al., 2014; Haustein, Peters, et al., 2014; Priem, Piwowar, & Hemminger, 2012; Thelwall et al., 2013; Zahedi, Costas, & Wouters, 2014). The latest research shows that Spearman correlations with citations for 2012 papers in Web of Science are low at ρ=0.194 for all 1.3 million papers and ρ=0.148 excluding untweeted papers. Beyond the particular differences of Twitter coverage and density between scientific disciplines, research fields and journals reported by various studies (Costas et al., 2014; Haustein, Larivière, et al., 2014; Haustein, Peters, et al., 2014; Zahedi et al., 2014), Haustein et al. (2015) also identified large variations between document types deviating from patterns known for citations. For example, news items and editorial material, which are usually considered non-citable items (Martyn & Gilchrist, 1968), are the most popular types of journal publications on Twitter, showing a tendency of increasing Twitter impact for brief and condensed document types. A study based on a random sample of 270 tweets to scientific papers found that the majority of tweets contained either the paper title or a summary, did not attribute authorship and had a neutral sentiment, while 7% were self-citations (Thelwall, Tsou, Weingart, Holmberg, & Haustein, 2013). Other findings suggest that automated diffusion of article links on Twitter plays a role as well (Haustein, Bowman, et al., 2015).

Although these findings provide more evidence that the mechanisms behind tweeting a paper are different from those citing it, the meaning of tweets to scientific papers as well as the role of Twitter in scholarly communication are still unclear, not in the least due to the difficulty to identify 'tweeter motivations' based on 140 characters. This study aims to contribute to a better understanding of tweets as impact metrics by analysing the type of content that is distributed on Twitter. We propose that certain types of articles appeal more to the public than others, for example, because of their potential impact on health issues and everyday life or due to the fact that they are written in a certain way. Previous research has suggested that certain medical study types have a larger citation potential than others (Andersen & Schneider, 2011; Kjaergard & Gluud, 2002; Patsopoulos, Analatos, & Ioannidis, 2005), likely because they are more useful to the research community. In the context of Twitter, medical papers are of particular interest, because, on the one hand, these are particularly relevant to general Twitter users—as opposed to, for example, physics research—and practicing physicians belong to early adopters of social media in their work practice (Berger, 2009). In a survey asking researchers about social media use in research, the uptake by health scientists was, however, slightly below average (Rowlands, Nicholas, Russell, Canty, & Watkinson, 2011).

The aim of this paper is thus to investigate whether there is a connection between different medical study types and the frequency of tweets per article. We hypothesize that some study types are more popular on Twitter due to their attractiveness for a broader audience such as applied medical research relevant to patients as well as meta-analyses summarizing research and condensing results. We will approach this hypothesis by first investigating the potential differences in tweet frequency for a range of medical study types. We argue that logically there should be a connection between the clinical evidence hierarchy (further explained below) and the types of studies patients might consider interesting to discuss or spread on social media, as the highest evidence levels are those which are most likely to affect clinical practice. We therefore expect differences in tweet frequency to be related to evidence levels.

Materials and Methods

Comparing the impact of medical research study types on Twitter requires two pieces of information per research article: a classification of the study type as well as the number of tweets received by each particular paper. Currently no database contains both pieces of information, so that it was necessary to combine data from different sources. For this purpose, the medical study type classifications from the Embase bibliographical database was used, enriched with metadata from PubMed and Web of Science and then matched to Twitter data from Altmetric.com. The datasets and the matching approach are described in further detail below. Following these descriptions is an account of the specific measurements and statistical tools employed as well as the limitations of this study.

Data collection and matching

Due to Twitter's 140 character limitation, mentions of a scientific paper in tweets are restricted to links to the publisher's homepage or unique document identifiers such as the Digital Object Identifier (DOI) or PubMed ID (PMID). As Twitter only provides access to the most recent tweets[2], it is necessary to constantly query various article identifiers to obtain a database of tweets to scientific papers. Altmetric LLP has been collecting tweets based on multiple document identifiers including the DOI, PMID and the publisher's URL since July 2011 and thus provides a valuable data source for the purposes of our study. To assure reliable and complete Twitter data, we focus our study on papers published 2012. In order to link all tweets to the bibliographic data and study type classification from Embase, the DOI and the PMID are needed.

The study type classifications (see below) for the analysis were retrieved from the Embase bibliographical database. Embase is a major database containing more bibliographical records than PubMed Medline; for example, 24%[3] more for documents published in 2012. It is unclear whether the study type classifications of either database outperforms the other, however, as the indexing of Embase is more exhaustive, we have chosen to use this database for our study. In order to identify relevant papers from Embase (and to be able to perform a citation analysis in the future), *Clinical Medicine* journals were selected from the Web of Science (WoS) based on the National Science Foundation (NSF) journal classification system. The Web of Science also provides bibliographic data and DOIs for the relevant papers, which were used to match Embase study types and tweets from Altmetric.

Embase was queried for the relevant journals using the journal name and various abbreviations as well as the ISSN. Limiting the results to papers published in 2012, the metadata of 593,974 records was retrieved from Embase. In order to obtain the PMID needed to match tweets, PubMed was queried in the same way resulting in 497,619 records. Embase, PubMed and Web of Science were matched using the DOI, PubMed as well as string matches of bibliographic information resulting in 238,560 documents in the final dataset, 94.9% of which with a PMID and 91.1% with a DOI.

The bibliographic metadata was matched to the Altmetric database using the DOI and PMID resulting in 80,116 records with at least one social media event as captured by Altmetric and 74,060 with at least one tweet at the time of data collection in August 2014. This amounts to 31% of the 238,560 being mentioned on Twitter at least once, which corresponds almost exactly to the Twitter coverage of biomedical & health sciences papers found by Haustein, Costas and Larivière (2015). To ensure comparability between tweets published in January and December 2012, we fixed the tweeting window to 18 months (546 days) for each of the tweeted documents, including tweets until 30 June 2013 for papers published on 1 January 2012 and until 30 June 2014 for papers published on 31 December 2012. The day of publication is based on the publication date provided by Altmetric. As this date is not available for all records and is sometimes incorrect, the dataset was further reduced to 52,911 documents, which had an Altmetric publication date in 2012 and not received a tweet before

[2] Twitter's REST API is limited to tweets from the previous week, while the Streaming API provides realtime data only.

[3] For the publication year 2012, Embase contains 1,334,356 records (search: "2012".yr) and PubMed Medline contains 1,072,384 (search: 2012[pdat]).

the publication date. Although these steps lead to an underestimate of the percentage of tweeted papers, they help to reduce biases induced by publication age when comparing the visibility of different medical study types on Twitter.

Medical study type classification

Embase indexes all articles using a controlled vocabulary (the Emtree thesaurus), which contains hierarchically ordered keywords in a classical thesaurus structure. Among these keywords are study type classifications, of which some are directly identifiable as such (e.g. randomised controlled trials), while others require some translation (e.g. "sensitivity and specificity" which is used for diagnostic accuracy studies). The Emtree thesaurus is designed for indexing and retrieval, and there is thus not a given connection between the hierarchical ordering of study type keywords and different levels of research methodology. This is particularly important, as one of the predominant approaches to Western medical research and practice is the so-called evidence based medicine (EBM). One of the cornerstones of EBM is the distinction between study types and their hierarchical ordering based on how much 'evidence' a study is assumed to contribute to the understanding of a given problem (Greenhalgh, 2010). Different hierarchies exist, e.g. the Oxford Centre for Evidence Based Medicine's "Levels of Evidence" (OCEBM Levels of Evidence Working Group, 2011).

Table 1. Medical study type classification system based on Röhrig et al (2009) and OECBM. Classifications with raised numerals have narrower terms, which are not shown here.

	Medical research							
	Primary research						Secondary research	
research_type	A. Basic research		B.Clinical research		C. Epidemiological research		D. Synthesising research	
class	A1. Theoretical	A2. Applied	B1. Experimental	B2. Observational	C1. Experimental	C2. Observational	D1. Meta-analysis	D2. Review
	Method development	Animal study; cell study; genetic engineering/sequencing; biochemistry; material development; genetic studies	Clinical study; phase I-IV	Therapy; prognostic; diagnostic; observational study with drugs; secondary data analysis; case series; case report	Intervention study; field study; group study	Cohort (prospective/historical); case control; cross-sectional; ecological; monitoring, surveillance; Description with registry data		Systematic; narrative
study_type *embase_keyword*	**A1.1 Theoretical study** *Theoretical study* **A1.2 Method development** -	**A2.1 Ex vivo study** *Ex vivo study* **A2.2 In vivo studies** *Animal experiment* **A2.3 In vitro study** *Animal tissue, cells or cell components*[i] *Cell, tissue or organ culture*[ii] *Human tissue, cells or cell components*[iii] **A2.4 Genetic engineering** *Genetic engineering and gene technology* *Genetic engineering* *Gene sequence* **A2.5 Biochemistry** *Biochemistry* *Neurochemistry* *Phytochemistry*	**B1.1 Clinical trial** *Clinical trial* *Clinical trial (topic)* *Controlled clinical trial* *Multicenter study* *Phase 1 clinical trial* *Phase 2 clinical trial* *Phase 3 clinical trial* *Phase 4 clinical trial* *Randomized controlled trial*	**B2.1 Case study** *Case report* *Case study* **B2.2 Prognostic study** *Prognosis* **B2.3 Diagnostic study** *Diagnosis* *Diagnostic test* *Sensitivity and specificity* **B2.4 Therapy** - **B2.5 Observational study with drugs** *Observational study AND (major clinical study OR controlled study OR clinical article)*	**C1.1 Intervention study** *Intervention study* **C1.2 Field study** *Field study* **C1.3 Group study** -	**C2.1 Case control study** *Case control study*[iv] **C2.2 Cohort study** *Cohort study* *Longitudinal study* *Retrospective study* *Prospective study* **C2.3 Cross sectional study** *Cross-sectional study* **C2.4 Ecological study** - **C2.5 Monitoring** *Patient monitoring* **C2.6 Surveillance** - **C2.7 Registry study** -	**D1.1 Meta-analysis** *Meta-analysis*	**D2.1 Review** *Review* *Systematic review*

We have chosen to use a particular hierarchy, which allows a classification of study types on their level of research (Röhrig et al., 2009). We have added to the classification of Röhrig et al. (2009) by adding classification codes and the corresponding keywords in Emtree. The resulting system has been validated by two field-experts, and is displayed in Table 1. As can be seen, the classification system allows direct translation between specific Emtree keywords (we have added the broadest terms as well as their relevant narrower terms) and our classification codes on the third level (study_type). The system allows grouping of study types into classes and research types (levels 2 and 1), thus allowing us to analyse the connection between tweets and the specific study types as well as the broader categories.

Of the entire population of 238,560 records, 162,830 records can be classified using our study type classification system. Of these, 36,595 (22.5%) receive at least one tweet within the fixed 18 months tweet window. Of the remaining 75,730 records without a classification, 16,316 (21.5%) receive at least one tweet. These data delimitations will be used to control for systematic errors in our main dataset (records with classifications). Among those that were classified, 55% had only one classification, 26% had two, 12% had three and the remaining 7% had four or more classifications. References with *n* classifications are treated as *n* observations, thus resulting in more than 162,830 observations on either classification level. Some classes in our classification system were not observed at all in the dataset. These classes are omitted in the results section.

Statistical methods and indicators

For each study type classification level we report several statistics for all documents (referred to by $*_A$, e.g. N_A) as well as the subset that has received at least one tweet ($*_T$). The included statistics are number of articles per classification (N), mean tweets per article (μ), the standard deviation from the mean (σ), percentage of articles with at least one tweet (N_T/N_A), and the mean normalised tweets ($\hat{\mu}$) defined as the ratio between μ for a specific classification and μ for the entire population.

As the distributions of tweets for any classification are extremely skewed (see results) similar to citations, the adequacy of the mean as an indicator of average activity is debatable (Calver & Bradley, 2009). However, while the median might be a methodologically more sound choice, the distributions are so extremely skewed that for study type level classification, medians are all 0 when all papers are included and either 1 or 2 if only tweeted papers are included. The corresponding means range from 0.35 to 1.74 and 2.02 to 5.01, providing considerably more information, especially as the scales for the mean are continuous. We therefore use the mean for comparisons, with due care and inclusion of standard deviations and percentage of tweeted articles to provide further information on differences in means. As we have large sample sizes, we expect any major differences in means to be real and not due to chance. However, to test this assumption, all classifications are tested pairwise and against the background population using the independent sample, unpaired Mann-Whitney test.

Limitations

The most obvious error source in this study is the proportion of papers included in the final analysis, compared to the overall population of papers published in 2012. Our background population of 162,830 classified papers only represents 27.4% of the 593,974 records downloaded from Embase. However, it still represents 68.3% of the 238,560 matchable records. This is a fairly high number of papers that could be classified, and if it is possible to improve the matching algorithms, it should also be possible to increase the total number of classified papers comparably. The only systematic error in this regard is the omission of particular documents based on lacking or erroneous DOI's. However, as missing DOI's are also an issue in collecting tweets, this error is not likely to affect the tweet counts with the limitations to tweet-collection that currently exist.

To test if there is a systematic error in the number of tweets per paper, with regard to whether a paper has been classified with a study type or not, we compare the percentage of papers with tweets for classified papers with unclassified papers. For the 162,830 papers with a classification, 36,595 (22.5%) received at least one tweet, while the 75,730 unclassified papers received tweets on 16,316 (21.5%) papers. These values also corroborate findings by Haustein, Costas & Larivière (2015). For the classified papers, mean tweets were 0.67, while the mean was 0.71 for the unclassified papers. These differences are not random ($p = 2.7e-14$, using independent two-sample t-test), however, the effect size is also extremely small (Cohen's $d = 0.018$). We should therefore not consider the lack of study types as confounders for the number of tweets.

While the classification system we have used here was validated by two domain experts, it is only one possible system. Other classifications could have been created, in particular with regard to the translation from Emtree keywords to our classification system. The choices made in this regard will affect the results as presented here. However, when we compare the pairwise scores within a research class, we find high consistency between what could be considered "similar" research types. The only study type, which varies greatly from the other study types in their class is the non-systematic review. This is meaningful, as non-systematic reviews are regarded by medical researchers as much less evidential as their systematic counterparts.

Results

We analysed the classified papers on the three levels present in our classification system: research type, research class and study type. In Tables 2 to 4 we report summary statistics for the three levels, for all papers as well as limited to tweeted papers to determine differences between the share of tweeted papers as well as intensity of (re)use. Results are visualized in Figure 1. In Figures 2 to 4 we provide the results of the pairwise comparison to determine the statistical significance of differences between study types including binary and continuous statistical significance as well as Cohen's d to estimate effect size.

Summary statistics

As can be seen from Tables 2 to 4, there are large differences in the mean tweets per classification, regardless of classification level, although the largest differences are observable in the study types. The differences are clear from the means (μ_A and μ_T), but even more obvious when regarding the relative means ($\hat{\mu}_A$ and $\hat{\mu}_T$). This is also where we find the largest standard deviations, likely due to the smaller N per classification. Meta-analyses and systematic reviews receive considerably more tweets than other study types, which makes the synthesizing research type stand out as well. Overall, a generally increasing interest of the Twitter community can be observed from basic (A) over clinical (B) and epidemiological (C) to synthesizing research (D) papers. Larger variations per research type can be observed for clinical research, where clinical trials are much more tweeted than other study types. In fact, case studies (B2.1) have the lowest mean number of tweets per paper (μ_A), which also reflects in the low mean of observational clinical research (B2) on the research class level. Epidemiological research also performs above average of the entire sample, while basic research (A) consequently performs below, although with somewhat higher scores for genetic engineering (A2.4) than the papers classified as ex vivo (A2.1), in vivo (A2.2) and in vitro (A2.3) studies.

Table 2. Summary statistics for research type.

Research type	N_A	μ_A	σ_A	N_T	N_T/N_A	μ_T	σ_T	$\hat{\mu}_A$	$\hat{\mu}_T$
A. Basic research	130,171	0.434	1.491	25,992	0.200	2.172	2.712	0.642	0.743
B. Clinical research	70,262	0.766	2.699	16,623	0.237	3.238	4.773	1.133	1.108
C. Epidemiological research	43,733	0.963	3.201	12,132	0.277	3.472	5.313	1.425	1.188
D. Synthesising research	38,558	1.005	3.223	10,641	0.276	3.640	5.295	1.486	1.245

Table 3. Summary statistics for research class.

Research class	N_A	μ_A	σ_A	N_T	N_T/N_A	μ_T	σ_T	$\hat{\mu}_A$	$\hat{\mu}_T$
A2. Applied basic research	130,171	0.434	1.491	25,992	0.200	2.172	2.712	0.642	0.743
B1. Experimental clinical research	28,343	1.219	3.495	8,949	0.316	3.860	5.337	1.803	1.321
B2. Observational clinical research	41,919	0.460	1.928	7,674	0.183	2.511	3.894	0.680	0.859
C2. Observational epidemiological research	43,733	0.963	3.201	12,132	0.277	3.472	5.313	1.425	1.188
D1. Meta-analyses	1,883	1.742	4.488	655	0.348	5.009	6.448	2.577	1.714
D2. Reviews	36,675	0.967	3.139	9,986	0.272	3.550	5.199	1.430	1.215

Table 4. Summary statistics for study type.

Study type	N_A	μ_A	σ_A	N_T	N_T/N_A	μ_T	σ_T	$\hat{\mu}_A$	$\hat{\mu}_T$
A2.1. Ex vivo study	1,061	0.425	1.285	223	0.210	2.022	2.155	0.629	0.692
A2.2. In vivo study	52,127	0.437	1.435	10,676	0.205	2.135	2.536	0.647	0.731
A2.3. In vitro study	75,287	0.427	1.519	14,699	0.195	2.190	2.821	0.632	0.749
A2.4. Genetic engineering	1,696	0.606	1.951	394	0.232	2.607	3.345	0.896	0.892
B1.1. Clinical trial	28,343	1.219	3.495	8,949	0.316	3.860	5.337	1.803	1.321
B2.1. Case study	21,788	0.348	1.847	3,204	0.147	2.367	4.292	0.515	0.810
B2.2. Prognostic study	6,618	0.525	1.842	1,407	0.213	2.469	3.341	0.776	0.845
B2.3. Diagnostic study	13,513	0.608	2.081	3,063	0.227	2.682	3.680	0.899	0.917
C2.1. Case control study	2,428	0.975	3.547	664	0.273	3.566	6.065	1.443	1.220
C2.2. Cohort study	34,822	0.943	3.163	9,585	0.275	3.424	5.276	1.394	1.171
C2.3. Cross sectional study	4,891	1.106	3.300	1,440	0.294	3.756	5.201	1.636	1.285
C2.5. Monitoring	1,592	0.956	3.163	443	0.278	3.436	5.242	1.414	1.175
D1.1. Meta-analysis	1,883	1.742	4.488	655	0.348	5.009	6.448	2.577	1.714
D2.1. Review	32,962	0.885	2.909	8,694	0.264	3.354	4.878	1.309	1.147
D2.2. Systematic review	3,713	1.695	4.653	1,292	0.348	4.871	6.839	2.507	1.666

The distributions of tweets per classification are shown in Figure 1, illustrating the highly skewed nature of these distributions, but also the large differences between some categories. The results shown in these boxplots are directly comparable to the summary statistics, and the same classifications stand out as being particularly often tweeted.

From previous research we know that meta-analyses, systematic reviews and clinical trials are also the most highly cited study types (Andersen & Schneider, 2011). However, whether there is a connection between the citedness and tweetedness of medical study types is not obvious from the present data, and will require further research.

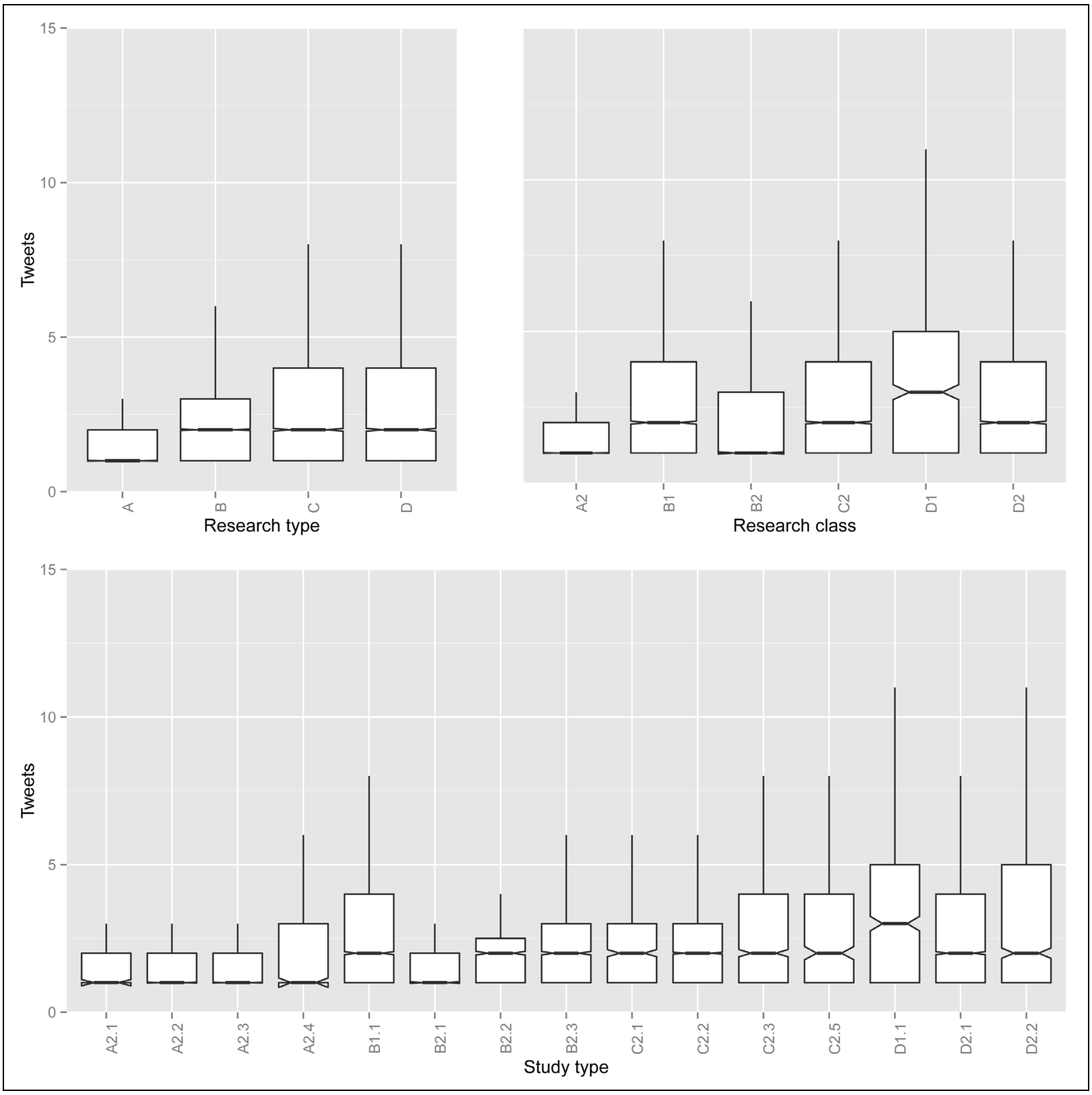


Figure 1. Notched boxplots showing tweet distributions for A) Research type, B) Research class and C) Study type.

Pairwise comparison

In order to analyse the magnitude of differences in classifications further, pairwise comparisons were made on each level. The independent two-sample Mann-Whitney test was used to test whether differences in sample means were due to random effects, and Cohen's *d* was used to estimate the effect size of varying means. There is of course a connection between the *p*-values of the Mann-Whitney tests and Cohen's *d*, to the extent that non-significant differences will also have very small effect sizes, as our sample sizes are quite large. In Figures 2 to 4 these pairwise comparisons are plotted as heatmaps, in which the diagonal and lower half have been omitted. The statistical significance of differences in mean are plotted as both binary maps (*p* below or above 0.05) and as continuous values. On the research type level, basic research stands out the most from the other types, with a lower mean of tweets per paper. For research classes, meta-analyses stand out with very large effect sizes, but overall the effect sizes are somewhat larger on this level than the broader research types. On the study type level, meta-analyses and systematic reviews stand out, but also clinical trials and epidemiological study types have fairly large effect sizes, compared to other study types.

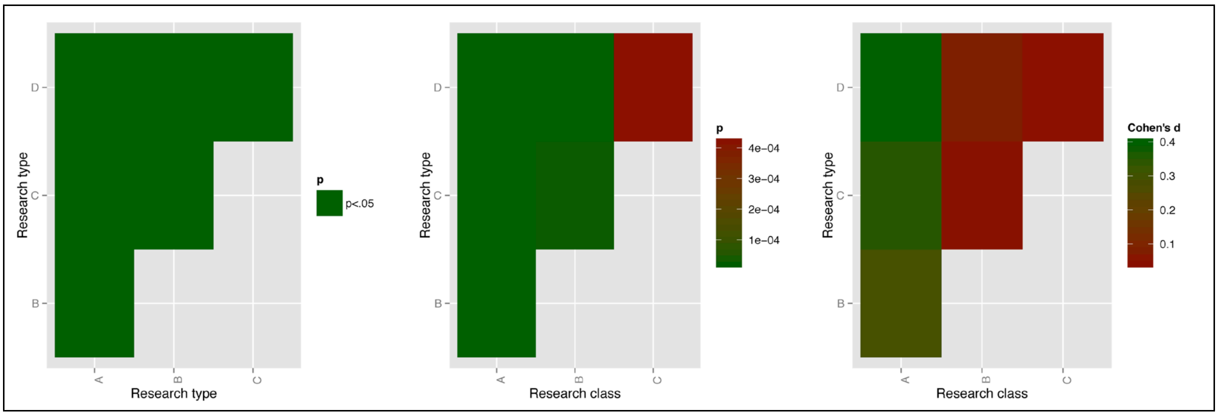


Figure 2. Heatmaps of pairwise comparisons showing A) binary statistical significance, B) continuous statistical significance and C) Cohen's *d* as effect size estimate. All figures are grouped on the research type level.

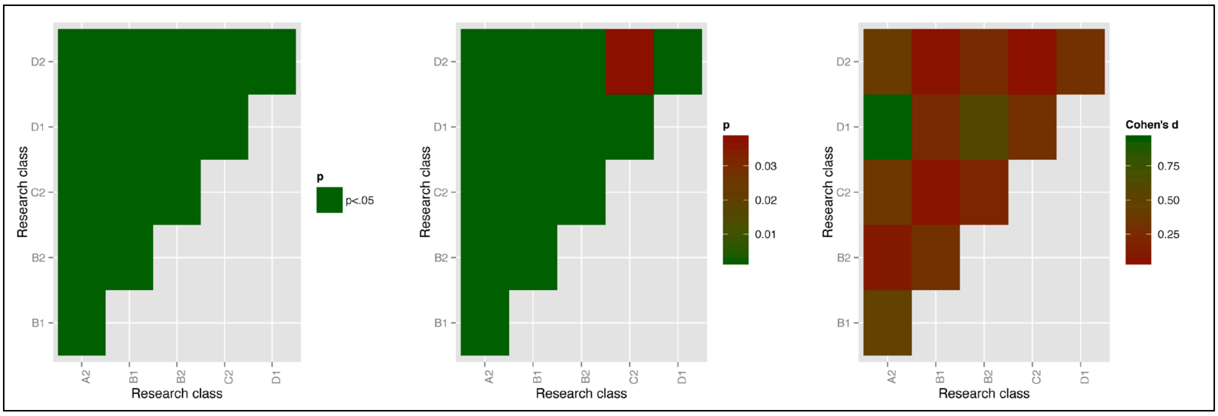


Figure 3. Heatmaps of pairwise comparisons grouped on the research class level. See figure 2 for legend.

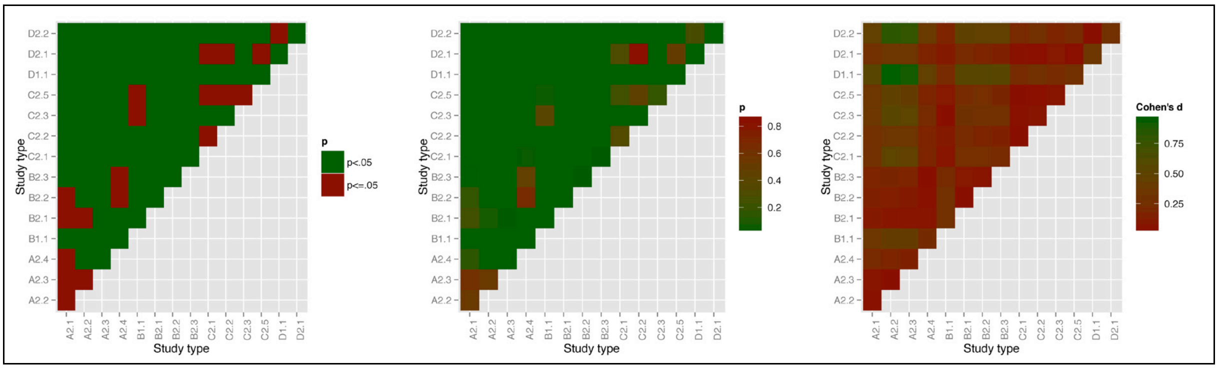


Figure 4. Heatmaps of pairwise comparisons grouped on the study type level. See figure 2 for legend.

Discussion and Outlook

We have analysed the frequency of tweets for medical research papers, distinguished by their specific study type. Our hypothesis was that some study types would be more frequently tweeted, because they were interesting to a wider audience (e.g., patients and other laymen) than other types. It has not been possible to identify literature on which types of research are actually more useful to laymen, or even which types are most often used. We therefore assume that research, which is close to clinical practise and may contribute to changes in treatments would be more interesting to patients, as they might see a specific benefit to themselves. Based on findings by Haustein, Costas and Larivière (2015) that briefer and condensed document types received more tweets than research articles, we also assumed that synthesising research papers would be more popular on Twitter than basic research.

On the broadest classification level, the results fit well with this assumption, as basic research stands out as the least frequently tweeted research type on average. Basic medical research is also furthest removed from the actual treatment of diseases—so much that some physicians consider it irrelevant to their clinical practise (Andersen, 2013)—which makes them less interesting for the general public of medical laymen and patients active on Twitter. When fine-tuning the analysis to study types, meta-analyses and systematic reviews stand out particularly, followed by clinical trials and epidemiologic study types. This corresponds with typical evidence hierarchies and reflects similar patterns found for citations (Andersen & Schneider, 2011; Kjaergard & Gluud, 2002; Patsopoulos et al., 2005). While this might indicate a relationship between tweets and citations, other studies on a broader level have found this is not the case (Costas et al., 2014; Haustein et al., submitted; Haustein, Larivière, et al., 2014; Zahedi et al., 2014). Other explanations may be that physicians are more likely to tweet about high-evidence studies or that these are also the same types of studies which are most interesting to patients. The latter appears obvious, as high-evidence studies are also more likely to be included in clinical practice guidelines and thus have a greater potential for changing practice. Moreover, results indicating the uptake of social media to be lower among health researchers (Rowlands et al., 2011), while the frequency of tweets per paper in this area is high (Haustein, Peters, et al., 2014), provide some evidence, that the large effect size found for these study types cannot be explained purely by large Twitter-activity from medical researchers. Patients, patient groups and laymen interested in research or other factors may thus play an important role in this observation.

While factors such as entertaining topics may play a role (Neylon, 2014) when looking at the the top per mille most frequently tweeted papers, it is unlikely that all 1,883 meta-analyses, 3,713 systematic reviews and 28,343 clinical trials should have a higher tweet count than other study types due to entertainment value, especially as these are also the most highly regarded study types by the researchers as measured through citations. The mean may of course be affected by single high-scoring studies, however, as can be seen from Figure 1, it is the entire distribution rather than merely the mean, which is increased for these study types. In fact, the maximum tweets per study type is 46 for meta-analyses and 59 for systematic reviews, while it is 65 for two of the basic research study types and 62 for clinical trials. The lowest maximum tweet frequency of a study type is 25 (an in vivo study) and the highest is 67 (a cohort study). It can thus be concluded that medical study types are one of the factors determining popularity of scientific papers on Twitter but they are certainly not the only ones. Apart from factors explored by previous studies and known also from the citation context—such as discipline, publication age, number of authors etc.—Twitter-specific effects should also be investigated. This includes the effect of the number of followers and affordance use as well as the extent to which scientific papers receive tweets due to author and journal self-promotion as well as automated Twitter accounts (Haustein, Bowman, et al., 2015).

Acknowledgements

The authors would like to thank Euan Adie and Altmetric.com for access to their Twitter data. SH acknowledges funding from the Alfred P. Sloan Foundation, grant no. 2014-3-25.

References

Andersen, J. P. (2013). *Conceptualising research quality in medicine for evaluative bibliometrics*. University of Copenhagen. Retrieved June 10, 2015 from http://vbn.aau.dk/files/119316655/JensPeterAndersenThesis.pdf

Andersen, J. P., & Schneider, J. W. (2011). Influence of study design on the citation patterns of Danish, medical research. In *Proceedings of the ISSI 2011 Conference* (pp. 46–53).

Berger, E. (2009). This sentence easily would fit on Twitter: Emergency physicians are learning to "tweet." *Annals of Emergency Medicine*, *54*(2), A23–A25. doi:10.1016/j.annemergmed.2009.06.002

Calver, M. C., & Bradley, J. S. (2009). Should we use the mean citations per paper to summarise a journal's impact or to rank journals in the same field? *Scientometrics*, *81*(3), 611–615. doi:10.1007/s11192-008-2229-y

Costas, R., Zahedi, Z., & Wouters, P. (2014). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, n/a–n/a. doi:10.1002/asi.23309

Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research*, *13*(4), e123. doi:10.2196/jmir.2012

Greenhalgh, T. (2010). How to read a paper: The basics of evidence-based medicine (4th ed.). Oxford: BMJ Books.

Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (2015). Tweets as impact indicators: Examining the implications of automated bot accounts on Twitter. *Journal of the Association for Information Science and Technology*. Retrieved from http://arxiv.org/abs/1410.4139

Haustein, S., Costas, R., & Larivière, V. (2015). Characterizing social media metrics of scholarly papers: the effect of document properties and collaboration patterns. *Submitted to PLOS ONE*.

Haustein, S., Larivière, V., Thelwall, M., Amyot, D., & Peters, I. (2014). Tweets vs. Mendeley readers: How do these two social media metrics differ? *It - Information Technology*, *56*(5), 207–215. doi:10.1515/itit-2014-1048

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

Kjaergard, L. L., & Gluud, C. (2002). Citation bias of hepato-biliary randomized clinical trials. *Journal of Clinical Epidemiology*, *55*(4), 407–10.

Martyn, J., & Gilchrist, A. (1968). *An Evaluation of British Scientific Journals*. London: Aslib.

Neylon, C. (2014). Altmetrics: What are they good for? *PLOS Opens*. Retrieved from http://blogs.plos.org/opens/2014/10/03/altmetrics-what-are-they-good-for/

OCEBM Levels of Evidence Working Group. (2011). *The Oxford 2011 Levels of Evidence*. Oxford: Oxford Centre for Evidence-Based Medicine.

Patsopoulos, N. a, Analatos, A. a, & Ioannidis, J. P. A. (2005). Relative citation impact of various study designs in the health sciences. *JAMA : The Journal of the American Medical Association*, *293*(19), 2362–6. doi:10.1001/jama.293.19.2362

Priem, J. (2014). Altmetrics. In B. Cronin & C. R. Sugimoto (Eds.), *Beyond bibliometrics: harnessing multi-dimensional indicators of performance* (pp. 263–287). Cambridge, MA: MIT Press.

Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. *arXiv*, 1–17. Digital Libraries. doi:http://arxiv.org/abs/1203.4745v1

Röhrig, B., du Prel, J.-B., Wachtlin, D., & Blettner, M. (2009). Types of study in medical research: part 3 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International*, *106*(15), 262–8. doi:10.3238/arztebl.2009.0262

Rowlands, I., Nicholas, D., Russell, B., Canty, N., & Watkinson, A. (2011). Social media use in the research workflow. *Learned Publishing*, *24*(3), 183–195. doi:10.1087/20110306

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PloS One*, *8*(5), e64841. doi:10.1371/journal.pone.0064841

Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013). Tweeting links to academic articles. *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics*, 1–8. Retrieved from http://cybermetrics.cindoc.csic.es/articles/v17i1p1.html

Zahedi, Z., Costas, R., & Wouters, P. (2014). How well developed are altmetrics? cross-disciplinary analysis of the presence of "alternative metrics" in scientific publications. *Scientometrics*. doi:10.1007/s11192-014-1264-0

Is There a Gender Gap in Social Media Metrics?

Adèle Paul-Hus[1], Cassidy R. Sugimoto[2], Stefanie Haustein[1], and Vincent Larivière[3]

[1] *adele.paul-hus@umontreal.ca; stefanie.haustein@umontreal.ca*
Université de Montréal, École de bibliothéconomie et des sciences de l'information, C.P. 6128, Succ. Centre-Ville, H3C 3J7 Montreal, Qc. (Canada)

[2] *sugimoto@indiana.edu*
Indiana University Bloomington, School of Informatics and Computing, 1320 East 10th St., 47405 Bloomington, IN (USA)

[3] *vincent.lariviere@umontreal.ca*
Université de Montréal, École de bibliothéconomie et des sciences de l'information, C.P. 6128, Succ. Centre-Ville, H3C 3J7 Montreal, Qc. (Canada) and Université du Québec à Montréal, Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), Observatoire des Sciences et des Technologies (OST), CP 8888, Succ. Centre-Ville, H3C 3P8 Montreal, Qc. (Canada)

Abstract
The gender gap in science has been the focus of many analyses which have, for the most part, documented lower research productivity and citation impact for papers authored by female researchers. Given the rise of scholarly use of social media to disseminate scientific production and the healthy proportion of women on these sites, further investigation of potential gender disparities in social media metrics are warranted. Comparing event counts from Twitter, blogs, and news with citations, this study examines whether publications with male and female authors differ regarding their visibility on the social web and whether gender disparities can be observed in terms of social media metrics. Findings demonstrate increased gender parity using social media metrics than when considering scientific impact as measured by citations. It is acknowledged that this could be the results of the different impact communities, as the scientific community constituting the citing audience is more male-dominated than the social media environment. The implications for the use of social media metrics as measures of scientific quality are discussed.

Conference Topic
Altmetrics

Introduction
Early Internet use was heavily male-dominated—to the point of being considered a “boy toy” (Morahan-Martin, 1998; Weiser, 2000)—and promises of gender equity in computer-mediated communication were left unrealized (Herring & Stoerger, 2013). However, recent transformations in both the function and functionalities of the Internet have led to increased participation of women, particularly in the use of social networking sites (Kimborough et al., 2013). As of September 2014, slightly more women are using social networking sites than men (Duggan, Ellison, Lampe, Lenhart & Madden, 2015). However, although men and women now both employ social media, the ways in which they use them remain gendered (Correa, Hinsley, de Zuniga, 2010; Koenig, 2015; Muscanell & Guadagno, 2012; Piazza Technologies, 2015).

Twitter—an online social networking service for microblogging—is one of the top websites in the world (Alexa.com). However, despite equality in other social media sites, there appears to be a growing gender disparity in Twitter, with men using the platform at higher rates than women (24 vs 21%) (Duggan et al., 2015). Moreover, the gender gap in Twitter usage has been increasing in the last two years (Duggan & Brenner 2013; Duggan et al., 2015). Gender bias is also reflected by journalism’s practices on Twitter, where reporters’ tweets severely underrepresent women in quotes (Artwick, 2013). This speaks to women’s underrepresentation as authorial voices—that is, voices that can speak as experts and authority on matters of merit. Given the rise of scholarly use of Twitter (Costas, Zahedi & Wouters, 2014; Haustein, Costas & Larivière, 2015; Holmberg, Bowman, Haustein & Peters, 2014; Pscheida et al., 2013; Rowlands et al., 2011), further investigation of potential gender disparities in scholarly communication and measures of impact from this site are warranted.

Microblogging is not the only web space with demonstrated gender disparities. Given the underrepresentation of women in science (Larivière, Ni, Gingras, Cronin & Sugimoto, 2013; West, Jacquet, King, Correll & Bergstrom, 2013), many studies have sought to examine whether the web might provide a democratizing space for female academics. These studies have shown that men tend to have greater web presence than women (van der Weijden & Calero Medina, 2014) and blog at a greater rate (Puschmann & Mahrt, 2012; Shema, Bar-Ilan & Thelwall, 2012). Bar-Ilan and van der Weijden (2014) recently investigated whether gender specific differences could be found when considering Mendeley (a social bookmarking service) readership counts. Using the gender of one of the co-authors of astrophysics papers—a field where *hyperauthorship* is commonplace (Cronin, 2001), thus making it difficult to distinguish papers attributed to female researchers from male researchers—they showed that the share of papers, to which at least one male contributed were found more often on the platform that those to which at least one women contributed. On the other hand, women attract more profile view in Academia.edu (an academic social networking site) in certain disciplines (Thelwall & Kousha, 2014). Many of these social media sites are associated with less formal ways of discussing and sharing research results with a wider audience (Shema, Bar-Ilan & Thelwall, 2012; 2014). The degree to which this engagement is gender-neutral begs further investigation.

This study builds on these analyses and seeks to examine whether publications with male and female authors differ regarding their visibility on the social web, and whether gender disparities can be observed in terms of social media metrics. Comparing event counts from Twitter, blogs and news with citations, this study aims to answer the following research questions:

- Does the gender gap in scholarly communication observed for publications and citations extend to social media?
- Does the visibility of male and female authored papers differ among Twitter, blogs, and mainstream news media?
- Does the gender gap in social media visibility of scholarly journal articles differ by scientific discipline?

There has been a growing call for researchers to demonstrate social impact (e.g., Force 11, 2011; REF, 2014). Social media metrics have been promoted as a source of such impact measures (Priem, 2014). However, the degree to which gender inequalities exist on such platforms must be investigated prior to wide-scale adoption and use of social media metrics.

Methods

Data were drawn from Thomson Reuters' Web of Science (WoS), which includes the Science Citation Index Expanded, the Social Science Citation Index and the Arts and Humanities Citation Index. These databases index annually documents published in over 12,000 journals across all scholarly disciplines. To determine differences between scientific disciplines, the NSF field classification of journals (National Science Foundation, 2006) was used instead of WoS categories in order to avoid possible double counting of papers by classifying, as the NSF classification assigns each journal to only one specialty.

Only papers published in 2012 were considered, as this year provides the best compromise between the length of the citation window—citations to papers take time to accumulate—and the recent uptake of social media activity (Thelwall, Haustein, Larivière & Sugimoto, 2014). Citations to 2012 papers were counted until the end of 2013, which allows for a citation window of at least one complete year for all papers. Selecting 2012 publications also has the advantage of guaranteeing complete coverage of social media data for the whole year, as Altmetric.com started data collection mid-2011 (Costas, Zahedi & Wouters, 2014).

Altmetric.com was chosen as the data source for social media and mainstream media counts, as it is the most comprehensive source of social media data associated with scientific papers (Robinson-García, Torres-Salinas, Zahedi & Costas, 2014). News items, tweets and scientific blogs entries were selected for the analysis. Mainstream media and news sources captured by Altmetric.com include online mentions of scientific papers in more than 1,000 mainstream media and news outlets such as the Washington Post, Süddeutsche or CNN[1], giving insight on the visibility of a paper among the general public. The audience of Twitter and scientific blogs covered by Altmetric.com may reflect the overlap between the scientific community and the general public as both are widely used outside of academia but also by scholars. These metrics were selected because they represent three different types of social media events and levels of engagement from users, ranging from the one end of the spectrum with an engagement limited to 140 characters on Twitter, to the redaction of whole blog entries or newspaper articles, at the other end. Altmetric.com data includes counts collected up to August 2014. Given the quick uptake of social media-based indicators (excluding Mendeley) reported by Thelwall et al. (2014), we consider that the social media activity window of more than a full year considered in this study is long enough to cover the vast majority of social media activity around papers published in 2012.

The link between WoS papers and the Altmetric.com list of indicators was made using the Digital Object Identifier (DOI). Hence, papers that did not have DOIs were excluded from the analysis. As one might expect, the proportion of papers with DOIs is not distributed evenly across scientific disciplines. While, for most fields, the proportion of journals with publications with a DOI is very high (e.g., above than 70%), a substantial share of journals (30%), particularly in the Social Sciences and Humanities, do not use DOIs (Haustein, Costas & Larivière, 2015). Hence, for papers published in the latter group of journals, results from Altmetric.com are more likely to underestimate their actual online visibility, which represents a limitation of this study (as well as the great majority of social media metrics analyses). Arts and Humanities papers were thus excluded of the analysis because of the low number of papers and of citations. The gender of authors was attributed using the authors' given names, following the method developed in Larivière et al. (2013). The method allowed to assign a gender to the first author of 67.7% (N=696,186) of all 2012 papers that had a DOI (N=1,028,382). The analysis is, thus, based on this dataset of papers, and the gender of the first author is used to categorize the paper as female or male.

The prevalence of social media metrics is measured through intensity, which indicates the mean number of events for papers that show at least one of the particular events (non-zero counts) and coverage, percentage of papers with at least one event. While coverage reflects the probability of a document to be cited or mentioned on the particular platform, the intensity indicate rate aims to measure the frequency or popularity with which documents are (re)used once they are on the platform and remains independent of the coverage and zero values (Haustein, Costas & Larivière, 2015).

The scientific impact of male and female researchers is compared using the average of relative citations (ARC). The ARC provides a field-normalization and thus allows the

[1] http://www.altmetric.com/sources-news.php

comparison of citation impact between the different specialities that have otherwise different citation practices. More specifically, the number of citations received by a given paper is divided by the average number of citations received by articles in the same NSF research specialty published in the same year. An ARC greater than 1 indicates that an article is cited above the world average for the same field, and an ARC below 1 means that it is cited below the world average.

Results

Figure 1 compares the ARC of papers first authored by women and men, respectively, in order to assess whether a gender gap can be found in the dataset of papers used. Figure 1 confirms the widespread gender disparities observed in science (Larivière et al., 2013) in terms of scientific impact. More specifically, in each discipline, papers first authored by male researchers have higher citation impact, with the only exception of Engineering and Technology where papers first authored by female researchers have a slight advantage (ARC value of 1.18 for women and 1.17 for men). Biomedical Research (0.95 for women and 1.11 for men), Professional Fields (1.11 for women and 1.26 for men), Mathematics (1.03 for women and 1.19 for men) and Psychology (0.97 for women and 1.12 for men) show the greatest gender differences regarding citation impact.

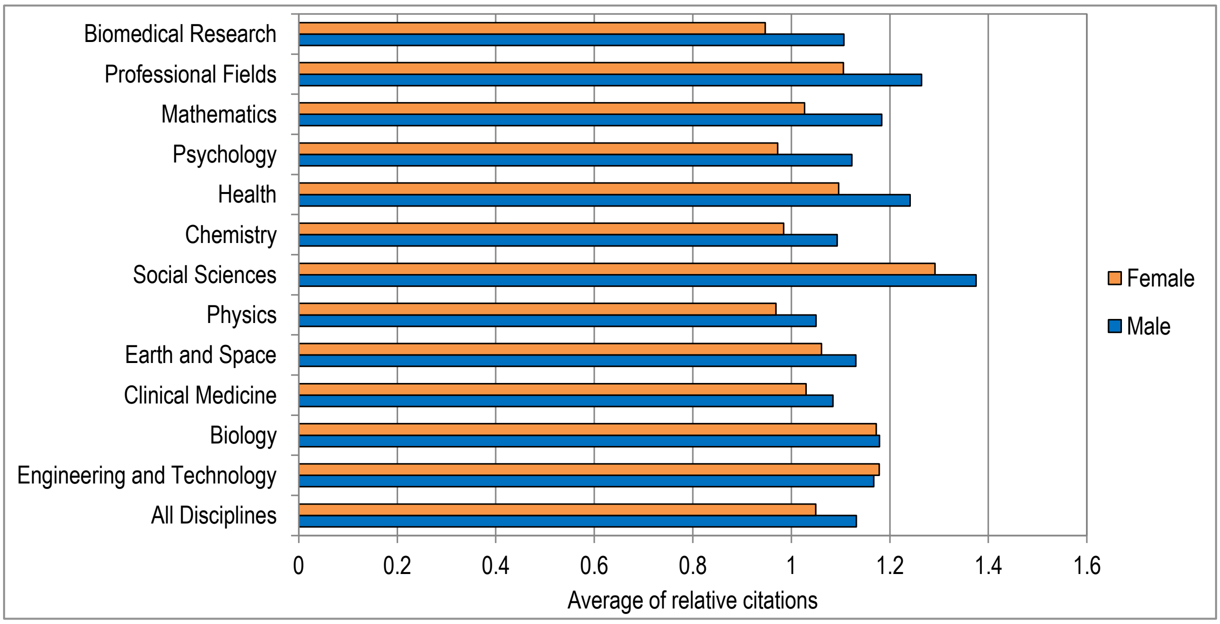


Figure 1. Average of relative citations of papers first authored by female and male researchers, by discipline and ordered by gender gap, 2012.

Figure 2 compares papers first authored by female and male researchers, in terms of intensity of news items (i.e., the mean number of events for all documents with at least one event) and coverage by news items (i.e., the percentage of papers with at least one event). All disciplines taken together, the intensity and the coverage of news items is gender-balanced, with an intensity difference of less than 0.07 event and a coverage difference of less than 1%. Physics (mean number of 1.04 for women and 1.34 for men) and Biomedical Research (1.63 for women, 1.87 for men) are the disciplines showing the strongest gender gap in terms of intensity of news items, in favour of papers first authored by men, corroborating the gender gap found in terms of citation impact (Figure 1). Coverage by news items of papers published in Biomedical Research (1.20% for women, 1.49% for men), Earth and Space (1.17% for women, 1.42% for men), Chemistry (0.59% for women, 0.84% for men) and Psychology

(1.26% for women, 1.50% for men) also confirm the gender gap found in terms of citation impact. However, papers first authored by female researchers in Health (1.32 for women, 1.26 for men), Clinical Medicine (1.39 for women, 1.33 for men) and Professional Fields (1.47 for women, 1.17 for men) have higher mean numbers of news items than that of male researchers while in Biology (0.73% for women, 0.62% for men), Engineering and Technology (0.60% for women, 0.55% for men) and Clinical Medicine (0.67% for women, 0.52% for men) they have a greater coverage.

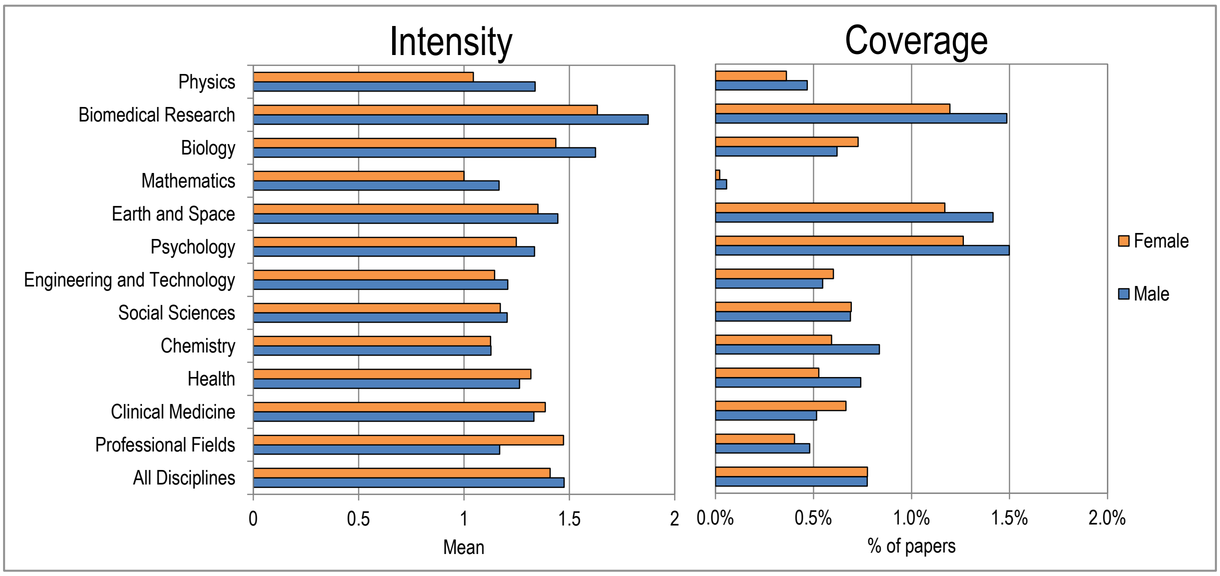


Figure 2. Intensity and coverage of news items of papers first authored by female and male researchers, by discipline, 2012.

Figure 3 provides the average numbers of tweets for all papers with at least one tweet (intensity for non-zero event items) and the percentage of papers with at least one tweet (coverage) by gender. It clearly shows that Twitter is the most popular platform among the three social media and mainstream media metrics analysed here, with an intensity of almost 3 tweets for papers tweeted at least once and coverage of almost 20% of papers (all genders and disciplines taken together). Gender analysis shows that, for all disciplines, papers first authored by female researchers are more intensely tweeted (2.98 tweets for women, 2.94 for men) and have a higher probability of being tweeted than papers first authored by male researchers (21% for women and 18% for men). Consistent with what has been found in terms of citations (Figure 1) and news items (Figure 2), Psychology and Biomedical Research show the highest gap in favour of men in terms of mean numbers of tweets.

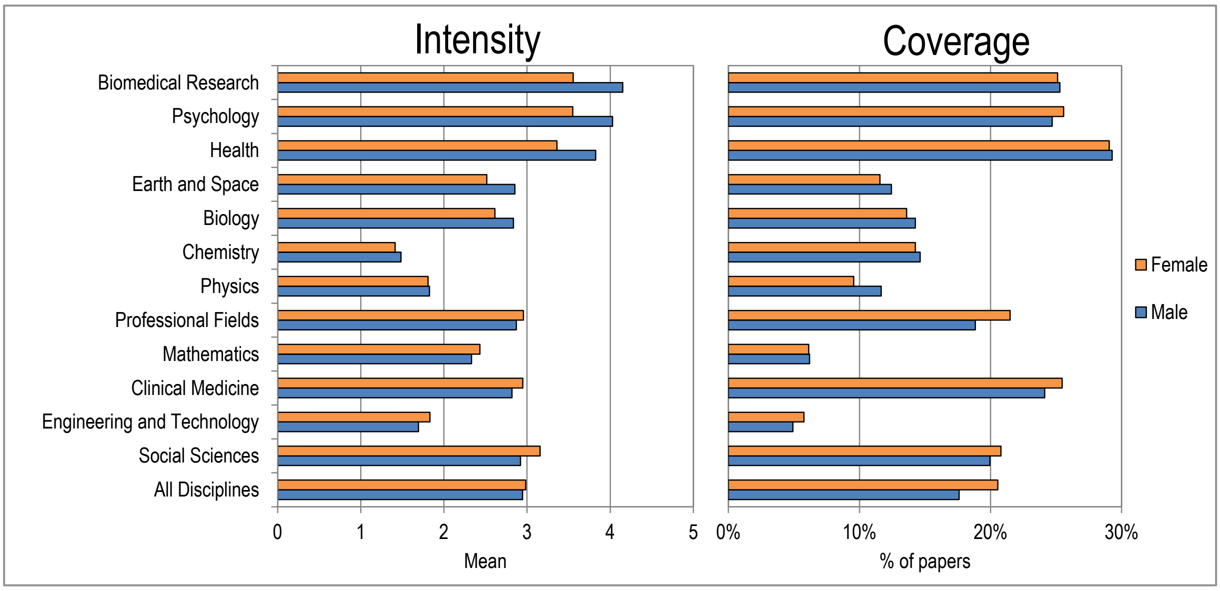


Figure 3. Intensity and coverage of tweets of papers first authored by female and male researchers, by discipline, 2012.

Figure 4 presents intensity and coverage by blog entries of papers first authored by women and men. All disciplines taken together, papers first authored by male researchers show a slightly higher intensity in terms of mean number of blog entries (1.33 for women, 1.40 for men) and higher coverage (1.68% for women, 1.78% for men). As previously shown, Psychology and Biomedical Research present important gender gaps, both in terms of intensity and coverage of blog entries. With respect to intensity, the average of blog entries of papers first authored by female and male researchers are equivalent in Health, Physics and Chemistry and papers authored by women have a slight advantage in Engineering and Technology. Papers authored by female researchers have stronger blog coverage in Clinical Medicine (1.30 % for women, 1.23% for men), Professionals Fields (1.08% for women, 1.02% for men) and Engineering and Technology (0.95% for women, 0.89% for men). However, the extreme gender gap in blog authors—both Puschmann and Mahrt (2012) and Shema, Bar-Ilan and Thelwall (2012) showed that about three quarters of bloggers where male—seems to transfer to the authors cited in blogs as confirmed by the coverage of papers authored by male researchers.

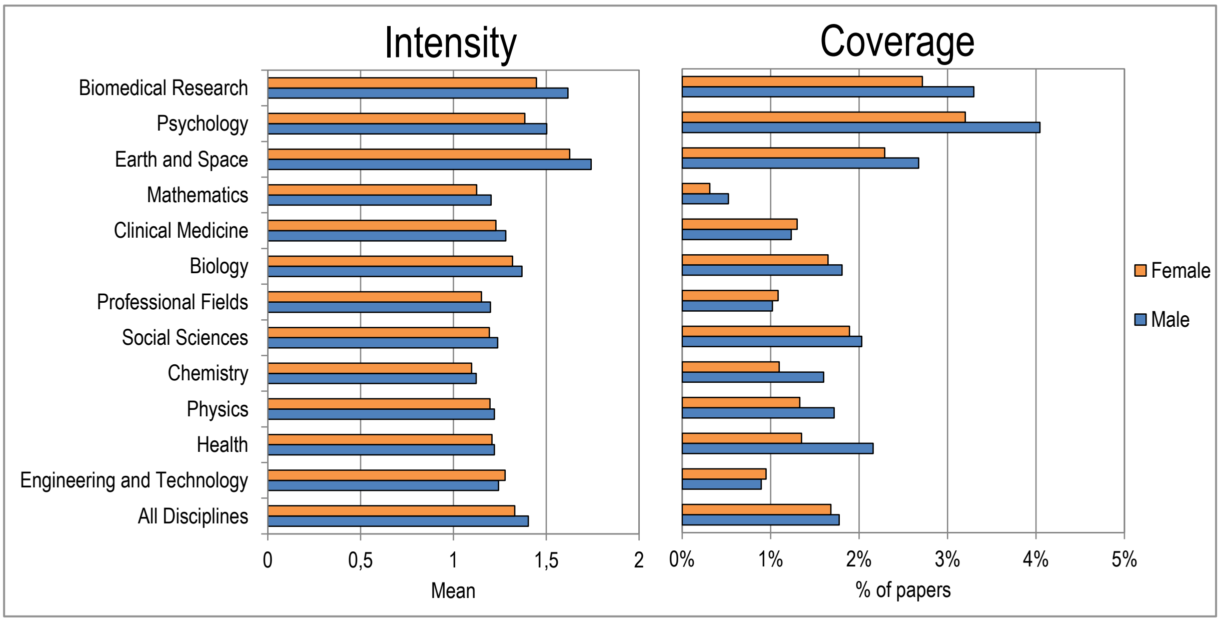


Figure 4. Intensity and coverage of blog entries of papers first authored by female and male researchers, by discipline, 2012.

Discussion and conclusion

Our findings demonstrate a more gender-balanced portrait when considering social media and mainstream media metrics (Figures 2 to 4), than when considering scientific impact as measured by citations (Figure 1). This could be explained by the fact that the impact communities contributing to these metrics are different: the scientific community which constitute the citing audience is more male-dominated than the social media environment (Kimbrough et al., 2013).

However, there is uniformity in the results neither by discipline nor platform. Coverage varied significantly by discipline, as did the mean impact score by gender. Furthermore, gender differences were found when examining microblogging, blogging, and news coverage. This suggests more information is needed before conclusive evidence on gender equality or inequality in social media metrics can be determined.

It could be argued that the diversity of the social media audience gives a broader audience an ability to respond to scholarly communication and therefore these measures of impact are a more honest metric of the absolute value of the work. However, lacking adequate validation of the meaning of social media metrics (Wouters & Costas, 2012), it is perhaps pre-emptive to make such a claim, as many tweets are actually made by bots (Haustein et al., in press). Further research on the nature of highly tweeted research will thus be necessary to assess the underlying mechanisms underneath the observed trends.

Acknowledgments

The authors acknowledge funding from the Alfred P. Sloan Foundation, grant no. 2014-3-25 and would like to thank Euan Adie and Altmetric.com for access to their data.

References

Artwick, C. G. (2014). News sourcing and gender on Twitter. *Journalism*, *15*(8), 1111–1127.

Bar-Ilan, J., & van der Weijden, I. (2014). Altmetric gender bias? Preliminary results. Presented at the 19th International Conference on Science and Technology Indicators, Leiden.

Correa, T., Hinsley, A. W., & de Zuniga, H. G. (2010). Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in Human Behavior*, *26*(2), 247–253.

Costas, R., Zahedi, Z., & Wouters, P. (2014). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, n/a–n/a.

Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology*, *52*(7), 558–569.

Duggan, M., & Brenner, J. (2013). *The Demographics of Social Media Users — 2012*. Retrieved from http://www.lateledipenelope.it/public/513cbff2daf54.pdf

Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., & Madden, M. (2015, January). Social Media Update 2014. *Pew Research Center's Internet & American Life Project*. Retrieved January 19, 2015, from http://www.pewinternet.org/2015/01/09/social-media-update-2014/

Force 11. (2011). *About Force 11*. Retrieved January 20, 2015, from https://www.force11.org/about

Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (in press). Tweets as impact indicators: Examining the implications of automated bot accounts on Twitter. *To be published in Journal of the Association for Information Science and Technology*. Retrieved April 15, 2015, from http://arxiv.org/abs/1410.4139

Haustein, S., Costas, R., & Larivière, V. (2015). Characterizing Social Media Metrics of Scholarly Papers: The Effect of Document Properties and Collaboration Patterns. *PLoS ONE*, 10(3), e0120495. doi: 10.1371/journal.pone.0120495

Herring, S. C., & Stoerger, S. (2013). Gender and (A)nonymity in Computer-Mediated Communication. In J. Holmes, M. Meyerhoff, & S. Ehrlich (Eds.), *The Handbook of Language and Gender* (2nd ed., pp. 567–586). Hoboken, NJ: Wiley-Blackwell Publishing. Retrieved from http://info.ils.indiana.edu/~herring/herring.stoerger.pdf

Holmberg, K., Bowman, T. D., Haustein, S., & Peters, I. (2014). Astrophysicists' Conversational Connections on Twitter. *PLoS ONE*, *9*(8), e106086.

Kimbrough, A. M., Guadagno, R. E., Muscanell, N. L., & Dill, J. (2013). Gender differences in mediated communication: Women connect more than do men. *Computers in Human Behavior*, *29*(3), 896–900.

Koenig, R. (2015, January 6). In STEM Courses, a Gender Gap in Online Class Discussions. *The Chronicle of Higher Education Blogs: Wired Campus*. Retrieved January 7, 2015, from http://chronicle.com/blogs/wiredcampus/in-stem-courses-a-gender-gap-in-online-class-discussions/55399#disqus_thread

Larivière, V., Ni, C. C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Global gender disparities in science. *Nature*, *504*(7479), 211–213.

Morahan-Martin, J. (1998). The Gender Gap in Internet Use: Why Men Use the Internet More Than Women—A Literature Review. *CyberPsychology & Behavior*, *1*(1), 3–10.

Muscanell, N. L., & Guadagno, R. E. (2012). Make new friends or keep the old: Gender and personality differences in social networking use. *Computers in Human Behavior*, *28*(1), 107–112.

National Science Foundation. (2006). Science and Engineering Indicators. Chapter 5: Academic Research and Development. Data and Terminology. Retrieved from http://www.nsf.gov/statistics/seind06/c5/c5s3.htm#sb1

Piazza Technologies. (2015). *STEM Confidence Gap*: *Piazza Blog*. Retrieved January 10, 2015 from http://blog.piazza.com/stem-confidence-gap/

Priem, J. (2014). Altmetrics. In B. Cronin & C. R. Sugimoto (Eds.), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Schorlarly Impact* (pp. 263–288). Cambridge, MA: MIT Press.

Pscheida, D., Albrecht, S., Herbst, S., Minet, C., & Köhler, T. (2013). *Nutzung von Social Media und onlinebasierten Anwendungen in der Wissenschaft. Erste Ergebnisse des Science 2.0-Survey 2013 des Leibniz-Forschungsverbunds „Science 2.0"*. Retrieved from http://www.qucosa.de/fileadmin/data/qucosa/documents/13296/Science20_Datenreport_2013_PDF_A.pdf

Puschmann, C., & Mahrt, M. (2012). Scholarly blogging: A new form of publishing or science journalism 2.0? In A. Tokar, M. Beurskens, S. Keuneke, M. Mahrt, I. Peters, C. Puschmann, & T. van Treeck (Eds.), *Science and the Internet* (pp. 171–182). Düsseldorf: Düsseldorf University Press.

REF. (2014). *REF 2014*. Retrieved from http://www.ref.ac.uk/about/

Robinson-García, N., Torres-Salinas, D., Zahedi, Z., & Costas, R. (2014). New data, new possibilities: exploring the insides of Altmetric.com. *El Profesional de La Informacion*, *23*(4), 359–366.

Rowlands, I., Nicholas, D., Russell, B., Canty, N., & Watkinson, A. (2011). Social media use in the research workflow. *Learned Publishing*, *24*(3), 183–195.

Shema, H., Bar-Ilan, J., & Thelwall, M. (2012). Research Blogs and the Discussion of Scholarly Information. *PLoS ONE*, *7*(5), e35869.

Shema, H., Bar-Ilan, J., & Thelwall, M. (2014). How is research blogged? A content analysis approach. *Journal of the Association for Information Science and Technology*, n/a–n/a.

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do Altmetrics Work? Twitter and Ten Other Social Web Services. *PLoS ONE*, *8*(5), e64841.

Thelwall, M., & Kousha, K. (2014). Academia.edu: Social network or Academic Network? *Journal of the Association for Information Science and Technology*, *65*(4), 721–731.

van der Weijden, I., & Calero Medina, C. (2014). *Gender effects on evaluation indicators*. Leiden: CWTS ACUMEN Deliverable.

Weiser, E. B. (2000). Gender Differences in Internet Use Patterns and Internet Application Preferences: A Two-Sample Comparison. *CyberPsychology & Behavior*, *3*(2), 167–178.

West, J. D., Jacquet, J., King, M. M., Correll, S. J., & Bergstrom, C. T. (2013). The Role of Gender in Scholarly Authorship. *PLoS ONE*, *8*(7), e66212.

Wouters, P., & Costas, R. (2012). *Users, narcissism and control – tracking the impact of scholarly publications in the 21st century.* SURFfoundation.

PubMed and ArXiv vs. Gold Open Access: Citation, Mendeley, and Twitter Uptake of Academic Articles of Iran

Ashraf Maleki

Malekiashraf@ut.ac.ir
Department of Library and Information Science, University of Tehran, Enqelab sq. Tehran (Iran)

Abstract
Despite contradicting evidence that open access (OA) articles might have greater citation advantage, there is less case studies in developing countries showing whether their global publication availability pattern advantages scientific impact metrics. Also, by addition of altmetrics to the world scientific evaluation system it is less known how different research access channels such as OA publishers, PubMed database and arXiv repository help altmetric indicators. Therefore, this paper investigates the case of WoS publications of Iran (2001-2012) for impact of mentioned publication availability models on citation, Mendeley readership, and tweet counts across four broader disciplines. Findings on 98,453 articles show that gold OA papers (5%) do not benefit significantly more metric counts, except in tweets linking to OA medical publications. Articles in PubMed Central (3%) significantly advantage the three investigated metrics, whereas arXiv preprints (2%) had higher readership advantage only. Different from PubMed publications, tweets to OA medical research were not significantly correlated with citations, suggesting their social impact rather than scientific. Additionally, OA publications are not significantly read by Mendeley users in developing countries, but developed ones, only in life science and biomedicine. Therefore, repository availability appears to be highly impactful in terms of citation and readership, whereas OA publications tend to receive rather high social impact through tweets.

Conference Topic
Altmetric

Introduction
Although traditional citation analysis helps countries to assess academic aspects of research impact and to fund them, so far wider aspects of impact including social and educational influence of research publications have been mainly ignored. However, by developing models of science assessment it seems that there will be better tools to assess influential aspects of research perhaps advantageous for public society rather than academic communities (Bornmann, 2012). Therefore, to improve aspects of wider impact, open access movement encourages researchers to make their research available online using various solutions. The open access (OA) availability of publications was a substantial addition to scholarly communication that enhanced science availability to a wider social audience and the researchers who had no access to subscription-based scientific data sources, especially those in developing countries (Contreras, 2012). With the advent of social networking sites and an access to free and open science, wider audience are now encouraged to publicly distribute science and give feedback about the scientific outputs. Extensive bookmarking of students and academics in research networks such as Mendeley (Mohammadi & Thelwall, 2013; Zahedi, Costas & Wouters, 2014; Haustein & Larivière, 2014) and prevalent reflection of the users' interest in online social networking sites such as Twitter (Haustein et al., 2013; Maleki, 2014) are evidence of wider impact of scientific publications beyond formal citations. Therefore, freely available publications not only advantage more citations (Lawrence, 2001; Gargouri et al., 2010; Laakso & Bjork, 2013), but also there is evidence they benefit from early reflection of impact in online media metrics in a way seemingly different from non-OA. In this respect, many of the top papers with higher altmetric scores in *Altmetric.com* were open access (Van Noorden, 2012). However, in spite of these evidence, there is less case

studies showing whether OA advantage is available for publishing pattern in developing countries, as in this research for Iranian WoS (Web of Science) publications.

The evidence suggests that developing countries have more OA journals than even some distinguished European countries (Bayry, 2013) and institutional repository growth since 2010 (Pinfield et al., 2014), however their journals are less internationally recognized or listed in scientific databases such as PubMed (Bayry, 2013). There are also barriers such as language, lack of knowledge about how OA publishing systems work (Salager-Meyer, 2014), and less funding for the researchers in these countries to contribute in high quality OA journals. Hence, it is less known how availability of their publications advantage citation and altmetric indicators. Therefore currents research aim to test OA impact on formal citations, Mendeley readerships and Twitter mentions (more below) to scholarly publications with Iranian authors, because this country in recent years had a rather noticeable scientific publication growth (e.g. Moin, Mahmoudi & Rezaei, 2005; Brown, 2011).

Furthermore, a fundamental challenge as Moed discussed (2012) is that along with OA journals (gold OA), self-archiving forms of publications (green OA) come a wide variety. There are about 80% of publishers that permit self-archiving (Laakso, 2014) in institutional homepages, subject repositories and web portals that excluding them might decline accuracy of OA advantage analyses (Moed, 2012). Amongst the online repositories, PubMed and arXiv have the highest web presence and impact according to Webometrics ranking (Cybermetrics Lab 2015, see more at http://repositories.webometrics.info), however it is less known how they advantage citations compared to OA journals, which is the subject of current research.

It is necessary to recognize the differences between OA journal and these repositories. PubMed refers to an important search engine for peer-reviewed medical research and has a significant role in research uptake in related fields, whereas arXiv is a preprint repository in *Cornell University* for self-archiving papers even before peer-review, mostly in physical sciences. The gold open access is a widespread solution across disciplines. However, a restricted number of publications in the world currently are published in journals with a free online version, as Harnad estimated gold open access articles about 5% in 2004; and without a considerable change in 2009, this proportion was 5.9% as covered in WoS (Laakso, 2009). However, there were better improvement in green OA reaching to about 12% in 2011 (Björk et al., 2014).

Among altmetric indicators, Mendeley readership and Twitter mentions to articles are known for their prevalent users (Thelwall et al., 2013; Zahedi, Costas & Wouters, 2014). However, evidently the two metrics are different in terms of aspects of impact. Majority of the online users in Mendeley are students (Mohammadi et al., in press; Zahedi et al., 2013; Haustein & Larivière, 2014), but in Twitter are the public audience (Maleki, 2014). They also are different from citation in terms of aspects like statistical distribution pattern (Thelwall & Wilson, in press; Eysenbach, 2011), and incidence, as tweets are fast and immediate (Eysenbach, 2011; Shuai et al., 2012) but Mendeley readerships and citations gradually increase. Also their prevalence is different, as tweets are linking to less publications than Mendeley readerships and citations (Thelwall et al., 2013). Thus, they individually reveal aspects of impact in different ways.

Background Literature

Citation advantage of open access publications

Various studies have reported that OA availability increases citation rate to articles in various fields. The premiere signs of OA citation advantage was reported from conference papers in computer science (Lawrence, 2001). More recently, Gargouri et al. (2010) found both self-selective self-archiving and mandatory self-archiving highly cited. In addition, Laakso and

Bjork (2013) observed that delayed OA policy for 2011 publications with about 78% available within the first year and about 85% within the two year after the publication, increased journal citation rate twice as much as non-OA journals and three times more than immediate OA journals.

In contrast, there are other studies that did not support a citation advantage for OA publications, some of them reviewed in Craig et al. (2007). Amongst more recent evidence Davis did several studies finding no OA citation advantage. He did a randomized control of 11 journals of American Physiological Society, finding no OA advantage after 9-12 month (Davis et al., 2008). His other study included 11 biology and medicine journals among which citations to OA articles fell from 32% in 2003 to 11% in 2007 (Davis, 2011). Gaule and Maystre (2011) also found 17% OA articles in PNAS during 2004 to 2006, where they found no OA diffusion advantage, but rather an author self-selection advantage after adjustment for confounders.

Studies report various evidence that online repositories increase citation advantage of articles, whereas subject repositories are more known to researchers than institutional ones (Cullen & Chawner, 2011). For instance, a study on articles in four math journals deposited in the arXiv indicated 35% more citation on average (Davis & Fromerth, 2007). Wren (2005) also showed that from both OA and non-OA journals with higher Journal Impact Factor (IF) over a third had OA reprints in non-journal websites of which over half had educational domains (.edu), providing a wider access to open research. Furthermore, Jeong and Huh (2014) showed that listing non-OA, non-Medline journals in the open access database of PubMed Central has over years led to an increase in their citation rate and impact factor in comparison with non-OA, non-listed journals.

Wider impact of open access publications

The OA publications were one of the premiere resources of online impact studies of scholarly publications, which revealed aspects of wider impact beyond traditional citations (Kousha & Thelwall, 2006; Vaughan & Shaw, 2007). For instance, Kousha and Thelwall (2006) studied URLs linking to OA publications of library and information science, which were demonstrative of 43% of their formal and 18% informal impact. In another study, Google Scholar unique citation to a sample of articles in 39 WoS OA journals in biology, chemistry, physics and computing was studied finding non-journal Google Scholar citations to OA publications indicator of their wider impact (Kousha & Thelwall, 2008). Other studies revealed usage advantage of online OA publications. Davis (2011) indicated that OA publications had more reader than subscription-based publications but not more citation advantage, for 89% more full-text downloads, 42% more PDF downloads, and 23% more unique visitors.

Only very recently a few studies compared altmetrics across OA publications. Adie (2014) reported that in the *Nature Communication* OA articles attract significantly more Mendeley readers and tweets. Also, Alhoori et al. (2015) displayed that OA papers have 60% more readers and 7% more tweets than non-OA, although non-OA articles were relatively highly covered in both Mendeley and Twitter.

Online Readership Impact assessment in Mendeley

The number of users who bookmarked publications in Mendeley reference sharing site is known as Mendeley readership metric for majority (55%) of users who add papers to their Mendeley libraries for reading or with the intention to read (Mohammadi, Thelwall & Kousha, in press). There is various evidence that Mendeley readerships can be indicative of scientific impact of research and predictor of correlates formal citations (Bar-Ilan, 2012; Thelwall, Haustein, Larivière & Sugimoto, 2013), moderately and weakly in social sciences, and humanities, respectively (Mohammadi & Thelwall, in press) and strongly in many fields in medical research (Thelwall & Wilson, in press). Wang et al. (2014) reports correlations of Mendeley and citation in a range of 0.36 to 0.61 with 1% significance level in seven PLoS journals and increased html views in correlation with altmetric scores of the articles. A study on arXiv repository examined impact of European astrophysics preprints on Mendeley readerships, finding that 47% of the publications in Scopus are in arXiv, whereas there were more arXiv papers (40%) in Mendeley than Scopus publications (27%) (Bar-Ilan, 2013). Furthermore, Mendeley metric had larger correlation with citations and Journal Impact Factor (IF) than Faculty of 1000 article factors for Genomics and Genetics articles (Li & Thelwall, 2012).

Social Impact Assessment via Twitter mentions

Studies had shown that Twitter is a promising social media to examine social popularity of articles (Thelwall et al., 2013) where tweets linked to about 10% of 1.4 million PubMed articles; and were a fast metric to track comments on arXiv preprints (Shuai et al., 2012). In another study, Wee and Chia (2014) showed that among 20 highly cited WoS articles citations were significantly correlated with altmetric scores in some subject categories including general and internal medicine (Pearson correlation significant in 0.36 level), applied physics (0.39), sociology (0.49), literature (0.62), and music (0.67). The correlation turned out to be significant among articles with highest altmetric scores in multidisciplinary engineering (0.35) and communication (0.31), whilst majority of altmetric scores in various fields coming from Twitter mentions (65% to 89%) rather than Facebook (1% to 11%), news (0 to 19%), and blogs (2% to 11%). Current research is a further exploration into the previous study on Twitter uptake of WoS publications with Iranian authors (Maleki, 2014). The study suggested 5% of publications in 2011-2012 with positive Twitter mentions with the highest uptake was in life science and biomedicine (10%) where links were often created by public society rather than scientific communities (*ibid*).

Research Questions

1. The extent to which are OA, PubMed and arXiv publications by Iranian authors tweeted, read and cited?
2. How do readerships and tweets correlate with formal citations when studies are available through the three above channels across disciplines?
3. Do OA publications advantage more readers in developing countries than developed ones?

Method

As a follow-up study to the previous research on Twitter mentions (Maleki, 2014), the dataset is the same as in the previous research, confined to publications in 2001 to 2012. WoS citations are based on the data available from May 2013 for 98,455 articles with DOIs. Twitter mentions are available according to results in July 2013 through *Altmetric.com* - a subscription based altmetric data provider (see the reasons for choosing *Altmetric.com* in Maleki, 2014); Mendeley readerships are examined via DOI submission to *ImpactStory.org*, another subscription based altmetric data provider which was free at the time of gathering data, in July 2013. *ImpactStory.org* was used because it provided attributes of Mendeley users and because it was different from *Altmetric.com* which provided readers only if papers had social media buzz. However choosing *ImpactStory.org* it was possible to gather a sample of about 30,000 papers rather than all the data.

DOAJ (Directory of Open Access Journals), WOS and Scopus journal datasets are consulted for OA availability of journals and papers OA status is modified based on journals' *Start year* in DOAJ. Data about PubMed archival of the articles was gathered by using DOIs of the publications on the full publication dataset available from PubMed Central. Publications were available via PubMed across four broader research areas for 2,978 papers (3%) the most in life science and biomedicine (2132 papers, 7%). ArXiv preprints of papers were examined using arXiv API, via DOI submission. For this purpose a custom-built program was used to submit 100 DOIs each query to arXiv. The data from arXiv might be not accurate because DOIs are available in arXiv if the authors have provided them for the publications. Results showed that there was overall 489 publication with preprints in arXiv consisting 1.3% of physical science article in 2001 to 2012 and very small proportion in technology (0.1%).

As altmetrics are faster than WoS citations, to learn if tweet and Mendeley uptakes are predictive of later WoS citations the dataset is tested in two time periods. Therefore, an interval is required to be considered for the publications to provide the opportunity to get citations. In case of Twitter, because the reliable and available data is confined to the most recent years (2011 onwards) citations are checked for publications in 2011-2012 in two time intervals after the publication year, the first in July 2013 and the second in December 2014. In Mendeley the data from July 2013 for both recent and older publications could be reliably used, thus the data is compared for recent publications in 2011-2012 and for older publications in 2001-2010. A signed-rank Mann-Whitney test is used to examine differences in medians and means of counts for OA, PubMed and arXiv publication against their counterparts (non-OA, non-PubMed, non-arXiv, respectively) inside each publication period.

A zero inflated negative binomial regressions analysis model is used to assess whether citation, readership and tweet counts dependend on publication access channels. Therefore, articles available via open access journals, PubMed, and arXiv are individually taken as nominal explanatory dummy variables coded as 1, and all the other cases not available in the corresponding availability model coded as 0. The 0 is the reference variable, which is also redundant because OA, PubMed and arXiv are true for minority of the cases. The reason for choosing this model is the overdispersion in the counts or the exceeding variance of the three metric counts from their means.

The analyses were supplemented with users' nationality data on the Mendeley readership counts for the publications. The results are compared across development status of countries for difference in readership of OA, PubMed and arXiv articles in Mendeley. Some articles in Mendeley were recorded with multiple variations, to avoid duplicates the ones with higher readership counts were considered.

Results

The main results of study suggest that out of 98,453 articles in 2001-2012 which had DOIs, 4,772 articles (4.7%) were published in 449 (6%) gold OA journals. There also were 3,043 articles (3%) listed in PubMed Central and 1,489 articles (0.5%) with preprints in arXiv. The articles which were linked by at least one tweet appeared in 1,067 journals, among which there were 116 gold OA journals (11%), 202 journals (19%) with articles indexed in PubMed Central, and 55 journals (5%) with article preprints in arXiv. As mentioned in method a smaller set of publications (35% of all above) were tested for readerships including all articles in 2,522 journals, comprising 273 (11%) gold OA journals, 307 journals (12%) available in PubMed list, and 56 journals (2%) with preprints in arXiv.

The OA journal *PLoS One* with 102 articles all available via PubMed Central had the most articles with tweets (36 papers) and readership counts (83 papers). The following two checked journals with articles available via PubMed with more articles in Mendeley were *Journal of Assisted Reproduction and Genetics* (48 out of 63 papers with readership, and 2 tweeted papers) and *International Journal of Nanomedicine* (38 out of 47 papers with readership, and 3 tweeted papers). Additionally, the results suggested that tweets link to more articles with preprints in arXiv in the journals *Astrophysics and Space Science* (with 35 tweeted articles and only 20 with preprints in arXiv), *Physical Review D* (27 tweeted articles whereas 75 with preprints in arXiv), and *Physical Review E* (17 tweeted articles, 27 preprints in arXiv) both former journals in astronomy and astrophysics and the latter one in soft-matter physics. However, there were journals with many papers in Mendeley, but poorly available preprints in arXiv; for instance there were 54 articles with readership counts in *International Journal of Theoretical Physics* out of 249 articles whereas only 6 with preprints in arXiv. Other OA journals with numerous articles with both citations and readerships, were *Analytical Science* (84 with readership and 116 with citations out of 118 papers) and *Molecules* (51 articles with readerships and 81 with citations out of 93 and 2 tweeted articles.

Table 1. Spearman correlation between Mendeley readership counts and WoS citations across years in terms of four broader research areas and of OA, PubMed, and arXiv availabilities of articles.

Disciplines / Availability		2012	2011	2010	2009	2008	2001-2007
Life science and biomedicine	OA[a]	.314** 218	.364** 209	.378** 120	.415** 96	.337** 96	.388** 87
	NOA[b]	.236** 1402	.274** 1317	.275** 1020	.296** 803	.339** 609	.302** 1157
	PubMed	.371** 100	.325** 176	.460** 109	.486** 85	.358** 75	.552** 42
	Non-PubMed	.258** 568	.204** 959	.220** 854	.279** 708	.309** 570	.296** 1202
Physical sciences	OA	.159 94	.060 85	.060 76	.194 49	-.016 29	.057 119
	NOA	.293** 838	.229** 816	.237** 691	.275** 539	.282** 470	.167** 1216
	arXiv	.217 35	.291 42	.418* 25	-193 20	-.232 23	-.232 23
	Non-arXiv	.220** 397	.187** 677	.248** 652	.236** 525	.220** 470	.156** 1333
Technology	OA	.160 39	.403 15	-.019 13	-.189 11	-.315 7	.173 19
	NOA	.154** 840	.259** 833	.325** 702	.289** 609	.328** 349	.358** 75
Social sciences and humanities	OA	.304* 52	.188 31	.266 9	.947* 5	.500 3	.293** 4482
	NOA	.363** 56	.259 33	.454* 26	.061 19	.815** 14	.462** 39

Correlation between altmetrics and citations in terms of availability models

Tables 1 and 2 show the correlation between Mendeley readerships and tweets with citations. The readerships of OA articles in life science and biomedicine are appropriately in moderate correlation with citations, and likewise, PubMed publications are correlated, but in stronger levels (correlation coefficients ranging from 0.31 to 0.55). However, the correlations in non-OA and non-PubMed papers are in lower levels (ranging from 0.20 to 0.34) - all correlations are significant in $p < 0.001$. This advantage were not available for the other three broader research areas, where the correlations were significant about non-OA publications rather than OA. The findings suggest that readership of publications with scientific impact have enhanced over years by OA and PubMed availability of life science and biomedicine articles, since older publications are in stronger correlation with citations than newer ones, although they are less numerous.

The figures in Table 2 suggest that there is a weak and significant correlation between tweets and later WoS citations in life science and biomedicine and physical sciences. Different from PubMed articles, tweet to OA publications did not have significant correlation with citations, perhaps for their social impact rather than scientific. On the other hand, correlations between tweets and citations are usually weak and significant after the interval for articles to receive citations in life science and biomedicine (correlations ranging from 0.07 to 0.17 significant in $p < 0.01$) and physical sciences (correlation significant in 0.13, $p < 0.001$). Correlations in all the fields does not show an OA advantage. Instead, there were weak and significant correlation in PubMed and non-OA publications in life science and biomedicine, and non-arXiv and non-OA articles in physical sciences after the interval.

Table 2. Spearman correlation between Twitter mentions and WoS citations in 2011-2012 in terms of four broader research areas and of OA, PubMed, and arXiv availabilities of articles.

Research areas / availability model		2012 Early citation[c]	2012 Later citation[d]	2011 Early citation	2011 Later citation
Life science and biomedicine	OA[a]	.015	.131	.071	.209
		159	159	74	74
	NOA[b]	.072*	.063	.059	.153*
		801	801	256	256
	PubMed	.087	.169*	.002	.143
		200	200	92	92
	Non-PubMed	.049	.034	.056	.147*
		760	760	238	238
Physical sciences	OA	.090	.094	-.178	-.045
		41	41	10	10
	NOA	.074	.130**	.054	.068
		405	405	86	86
	arXiv	-.001	-.009	7	7
		28	28		
	Non-arXiv	.078	.126**	.011	.024
		418	418	89	89
Technology	OA	10	-.048	2	2
			10		
	NOA	-.023	.131	-.017	-.130
		135	135	51	51
Social sciences and humanities	OA	.500	.866	1	1
		3	3		
	NOA	.521**	.345		-.487
		25	25	6	6

[a]. OA: Open Access; [b]. NOA: Non-Open Access; [c]2012 Early citations: citations to 2012 publications in July 2013; [d] citations to 2012 publications in Dec. 2014; Correlation significant at the 0.05 level (*);0.01 level (**).

Metrics dependencies to OA, PubMed and ArXiv publications

As figures in Table 3 show, tweeted gold OA publications (301 papers, 0.8%) are less than non-OA (1,975, 4.4%), whereas in fact more OA articles (11% of all OAs) tend to be tweeted than non-OA (5% of all non-OAs). This happens across the four broader fields with the highest occurrence in life science and biomedicine (15% OA vs. 10% non-OA). Also, findings suggest that tweets tend to link to significantly more PubMed publications in life science and biomedicine (24%), whereas this proportion is higher than tweeted OA publications (15%). The same is observed in physical sciences where arXiv preprints (55%) tend to receive tweets more than OA articles (7%). A Mann-Whitney test suggests that tweets to arXiv (206 tweets to 136 articles) were not significantly more than tweets to publications without arXiv preprint (472 tweets to 406 papers).

Also, tweets to PubMed (1,118 tweets to 293 papers vs. 2,105 tweets to 972 non-PubMed papers), and OA articles in life science and biomedicine (778) are significantly higher than their relative counterparts (i.e. non-PubMed and non-OA, respectively) (p<0.001). There were no significant difference between tweets to OA and non-OA in other fields, however. Additionally non-OA publications significantly advantage more citations to tweeted articles in 2011-2012 either in the early stage after publication (3.8 mean tweets to non-OA vs 1.6 tweets to OA) or later stage (10.3 vs. 6). This observations is in line with the correlations above which were significant in cases the publications were non-OA rather than OA in all fields excluding life science and biomedicine.

Table 3. Mean and median tweets and citations to articles with at least one tweet across publication availability models.

Source/Publication year		Median Mean: OA	Non-OA	PubMed	Non-PubMed	arXiv	Non-arXiv
Twitter mentions	2011-2012	1 2.9**	1 2.0	2 3.6**	1 1.8	1 1.3	1 1.2
Early citations	Jul. 2013	0 1.6	1 3.8**	1 5.1	1 3.2	1 1.5	1 3.5
Later citations	Dec. 2014	3 6.0	4 10.3**	4 13.1	4 9.1	4 6.7	4 9.7
Total articles	2011-2012	315 (14%)	1975 (84%)	336 (15%)	1954 (85%)	35 (6%)	532 (94%)

*significantly more than its counterpart category (OA vs. non-OA; PubMed vs. non-PubMed; arXiv vs. non-arXiv) in p < 0.01 level.**significantly more than its counterpart category in p < 0.001 level.

Table 4. Mean and median readerships and citations to articles with at least one Mendeley readership across publication availability models.

Source/Publication year		Median Mean: OA	Non-OA	PubMed	Non-PubMed	arXiv	Non-arXiv
Mendeley readers	2001-2010	3 4.3	2 4.1	3 6.9**	1 2.4	5 6.1*	2 3.5
	2011-2012	2 4.2	2 3.3	3 5.8**	2 3.2	4 5.3**	2 2.8
Later citations	2001-2010	5 9.0	5 8.9	4 11.4	4 7.5	9 12.6	6 10.7
Early citations	2011-2012	2 3.0	2 3.3**	1 2.2	1 1.8	1 2.2	1 2.3
Total articles	2001-2010	737 (8%)	8850 (92%)	374 (4%)	9213 (96%)	10 (0.3%)	3211 (99.7%)
	2011-2012	743 (11%)	6079 (89%)	558 (8%)	6264 (92%)	75 (4%)	1758 (96%)

*significantly more than its counterpart category (OA vs. non-OA; PubMed vs. non-PubMed; arXiv vs. non-arXiv) in p < 0.01 level.**significantly more than its counterpart category in p < 0.001 level.

Table 4 shows proportion of publication with positive Mendeley readership 5% OA (1,480 papers) and 52% non-OA (14,929 papers), with the highest article uptake in life science and biomedicine (9% OA and 61% non-OA) and the least in physical sciences (4% OA and 44% non-OA). Further results show that users tend to read non-OA publications (58%) rather similar to OA (55%) while there is no significant difference in their readership patterns across four broader research areas. However, despite in less papers than OA, PubMed publications (932 papers) tend to have higher readerships (5,566 PubMed vs. 4,675 OA readerships), with the highest occurring in life science and biomedicine (for 76% PubMed vs. 67% OA papers) ($p < 0.05$). The same was seen in arXiv preprints as their read articles (85 papers) tend to have significantly higher readership counts than non-archive ($p < 0.01$).

The OA publications in the two time periods (8% in 2001-2010 and 11% in 2011-2012) are more than PubMed (4% and 8%) and arXiv (0.3% and 4%). The mean PubMed readerships were significantly more than non-PubMed for the publications in older time period of 2001 to 2010 (6.9 PubMed vs. 2.4 non-PubMed) and for articles in 2011-2012 (5.8 vs. 3.2) ($p < 0.001$). ArXiv preprints in Physical science on average also had higher readerships than non-arXiv in both publication periods (significant in $p < 0.01$ in 2001-2010 and $p < 0.001$ in 2011-2012). There were no significant *citation* advantage for OA, PubMed and arXiv papers with Mendeley readerships, neither in the early nor the later stage after the publication year in none of the four research areas, although non-OA publications in social science and humanities and life science and biomedicine had significantly more readerships than OA.

Table 5 shows results of zero-inflated negative binomial regression analysis. The significance of alpha values in Table 5 identifies overdispersions for the three metrics. Voung statistics being above the critical value of 1.96 approves the overdispersion and the need for the zero inflated method. The estimates of the regression coefficients are shown by the values b and the estimated standard errors are the ratios of the coefficients. Therefore, b values show how much the availability of the articles by various models increases metric counts.

The results in Table 5 suggest that PubMed articles significantly advantage the three metric counts. However, (gold) open access were not significant indicator of neither citations nor the two altmetric counts. In addition, publications with preprints in arXiv had significantly more readership counts only.

Table 5. Zero inflated negative binomial regression analysis for citations, readerships and Twitter mentions by variables of availability channels.

	Citations (2001-2012)		Mendeley Readerships (2001-2012)		Tweets (2011-2012)	
Variables	*b*	*Standard error*	*b*	*Standard error*	*b*	*Standard error*
Open Access	-0.26**	0.02	-0.30**	0.03	-0.39**	0.09
PubMed	0.14**	0.03	0.79**	0.04	0.96**	0.09
ArXiv	-0.64**	0.06	0.37*	0.09	-0.21*	0.09
Constant	2.06**	0.01	1.31**	0.01	0.64**	0.03
Alpha	1.05	0.01**	0.58	0.01**	0.52	0.02**
Vuong Statistics	330.9**		254.9**		64.79**	
Log Likelihood	-200924.8		-39551.92		-3830.57	
Rest Log Liklihood $\chi2$ (3)	229.2**		579.1**		181.59**	
Publications	98,454		28,758		39,119	

Publication readership across countries development status

An important limitation of statistics about nationality attributes of users is that Mendeley suggests only top three countries with higher readership counts per paper. Based on these data, users were recognized from 141 countries, including 28,966 readerships from developed countries for 16,472 papers and 21,848 readerships from developing countries for 12,699 papers. Median readerships were more in papers with readers from developed countries rather than developing ones (4 vs. 3 readers per paper). The OA life science and biomedicine publications (excluding other field) had significantly more readers in developed countries ($p<0.05$). PubMed publications also had significantly more readerships in developed countries than developing ones ($p<0.001$), whereas there were no such difference about readership of arXiv preprints. In addition, users in developing countries significantly read more non-OA articles in technology (3,483, mean users = 1.77 vs. 1.68) and physical sciences (3,628, mean users = 1.73 vs. 1.66) ($p<0.001$). All tests were significant in a signed-rank Mann-Whitney test.

Discussions

A main limitation in this research is that it does not include other potential sources of publication availability such as homepages and institutional repositories and social networking sites for self-archiving. Also, a problem may associate with the regression analysis for which the research is very optimistically focused on direct impact of publication access patterns, whereas results might be affected by other correlates of the metrics such as Journal Impact Factor or Immediacy Index. Therefore, designing more complex models for assessment of availability impact might be subject of future studies.

Regarding the first research question results suggest that there are more OA articles (5%) than PubMed listed articles (3%) and arXiv preprints (2%). Also, there are more OA publication with readers (9%) than PubMed (6%) and arXiv (2%), whereas tweets link to relatively more PubMed (15%) papers than OA (14%) and arXiv (6%). Regarding the second question of research there were a significant correlation between tweets and citations to PubMed articles, indicating their scientific impact. However, tweeted OA publications seem to be reflective of social impact rather than scientific since they do not appear correlated with citations neither in early nor later year. In addition, publications in 2012 are more correlated than 2011, suggesting an overtime increasing publication uptake via tweets. A moderately significant and across years decreasing correlation between readerships and citations to OA and PubMed availability of articles in life science and biomedicine (excluding other fields) suggest that older publication had the opportunity to get higher citations.

The mean tweets to both OA (3.3) and PubMed (3.7) life science and biomedicine papers were significantly more than non-OA and non-PubMed, respectively. These publication strategies have obviously enhanced various aspects of research impact. The difference between the mean tweets to arXiv preprints (1.3) and non-arXiv physical science papers (1.2) is statistically significant, however these tweets are very low and does not reflect an aspects of impact, while generally arXiv papers are regularly tweeted for classification and dissemination purposes. The finding from previous study supports this, as papers in physical science are mainly tweeted by subject specific tweeters for classificatory reasons rather than scientific or social impact (Maleki, 2014). In contrast to OA advantage on Twitter mentions of articles (only in life science and biomedicine), Mendeley readerships was not significantly different across gold open access and non-OA publications in the four field.

The regression models for the three metrics also had results in line with the results from previous section. There is a significant citation advantage only for PubMed publications. Both PubMed and arXiv papers advantage Mendeley readerships. The only difference is in tweets where similar to above results show significantly more tweets to PubMed publications, however unlike the above non-OA advantage significantly more tweets than OA, which shows the effect of other hidden variables.

The expected higher readership of OA papers in developing countries failed to be true. A noteworthy result suggests that Iranian OA medical publication readerships by developed countries were significantly higher than developing countries, whereas this connection was vice versa in technology and physical sciences for non-OA articles. This can be connected to development and competitive abilities in research in these areas and/or the distribution of Mendeley users in various fields across countries. In this respect, the inferences need to be made with caution. However, it seems that Iranian medical research tend to get higher uptake by developing countries by appearing in PubMed index.

Conclusions

An important result of the study suggests that PubMed and arXiv strategies of publication availability can enhance the metric counts especially Mendeley readerships. Citations were mainly influenced by PubMed availability of broader field of life science and biomedical research, whereas tweets mainly link by publications available via gold OA journals. Furthermore, nationality of Mendeley readers appear to be informative about publication uptake patterns worldwide. Also, regarding results in this research with the ones from previous study on tweets it seem that Twitter has the potentials to reflect social impact of medical research for which OA availability and PubMed will help. In addition, subject repositories get higher readerships and tweets chance than papers out of them. Future studies might bring more variables associating these metrics for more realistic look at OA advantage in publication and research impact assessment.

Acknowledgment

The author would like to thank Dr. Kayvan Kousha of the Statistical Cybermetrics Research Group, in University of Wolverhampton, for his inspirations and very useful comments.

References

Adie, E. (2014). Attention! A study of open access vs non-open access articles. *Figshare*, doi:http://dx.doi.org/10.6084/m9.figshare.1213690.

Alhoori, H., Ray Choudhury, S., Kanan, T., Fox, E., Furuta, R., & Giles, C. L. (2015). On the Relationship between Open Access and Altmetrics. *iConference 2015 Proceedings*.

Bar-Ilan, J. (2013). Astrophysics publications on arXiv, Scopus and Mendeley: a case study. *Scientometrics*, 1-9.

Bayry, J. (2013). Journals: Open-access boom in developing nations. *Nature, 497*(7447), 40-40.

Björk, B.-C., Laakso, M., Welling, P., & Paetau, P. (2014). Anatomy of green open access. *Journal of the Association for Information Science and Technology, 65*(2), 237-250, doi:10.1002/asi.22963.

Bornmann, L. (2012). Measuring the societal impact of research. *EMBO reports, 13*(8), 673-676.

Brown, M. (2011). China, Turkey and Iran emerge as scientific giants. Wired.co.uk. Accessed 2 April 2014.

Contreras, J. L. (2012). Open access scientific publishing and the developing world. *St. Antony's International Review, 8*(2012), 43-69.

Craig, I., Plume, A., McVeigh, M., Pringle, J., & Amin, M. (2007). Do open access articles have greater citation impact?A critical review of the literature. *Journal of Informetrics, 1*(3), 239-248, doi:10.1016/j.joi.2007.04.001.

Cullen, R., & Chawner, B. (2011). Institutional repositories, open access, and scholarly communication: a study of conflicting paradigms. *The Journal of Academic Librarianship, 37*(6), 460-470.

Davis, P. M. (2010). Access, readership, citations: A randomized controlled trial of scientific journal publishing. (Vol. 0723, pp. n/a-n/a).

Davis, P. M. (2011). Open access, readership, citations: a randomized controlled trial of scientific journal publishing. *The FASEB Journal, 25*(7), 2129-2134, doi:10.1096/fj.11-183988.

Davis, P. M., & Fromerth, M. J. (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? *Scientometrics, 71*(2), 203-215, doi:10.1007/s11192-007-1661-8.

Davis, P. M., Lewenstein, B. V., Simon, D. H., Booth, J. G., & Connolly, M. J. (2008). Open access publishing, article downloads, and citations: randomised controlled trial. [Randomized Controlled Trial Research Support, Non-U.S. Gov't]. *BMJ, 337*, a568, doi:10.1136/bmj.a568.

Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research, 13*(4).

Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., et al. (2010). Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research. *PloS one, 5*(10), e13636, doi:10.1371/journal.pone.0013636.

Gaulé, P., & Maystre, N. (2011). Getting cited: Does open access help? *Research Policy, 40*(10), 1332-1338, doi:10.1016/j.respol.2011.05.025.

Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., et al. (2004). The Access/Impact Problem and the Green and Gold Roads to Open Access. *Serials Review, 30*(4), 310-314, doi:10.1016/j.serrev.2004.09.013.

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*, 656-669, doi:10.1002/asi.23101.

Jeong, G.-H., & Huh, S. (2014). Increase in frequency of citation by SCIE journals of non-Medline journals after listing in an open access full-text database. *Sci Ed, 1*(1), 24-26, doi:10.6087/kcse.2014.1.24.

Kousha, K., & Thelwall, M. (2006). Motivations for URL citations to open access library and information science articles. *Scientometrics, 68*(3), 501-517.

Kousha, K., & Thelwall, M. (2008). Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines. *Scientometrics, 74*(2), 273-294, doi:10.1007/s11192-008-0217-x.

Laakso, M., Welling, P., Bukvova, H., Nyman, L., Björk, B.-C., & Hedlund, T. (2011). The Development of Open Access Journal Publishing from 1993 to 2009. *PloS one, 6*(6), e20961, doi:10.1371/journal.pone.0020961.

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature, 411*(6837), 521-521.

Maleki, A. (2014). *Twitter Users in Science Tweets Linking to Articles: The Case of Web of Science Articles with Iranian Authors.* Paper presented at the American Society for Information Science and Technology, presented at SIG/MET post conference workshop, Seattle, USA,

Moed, H. (2012). Does open access publishing increase citation or download rates. *Research Trends, 28*.

Mohammadi, E., Thelwall, M., Haustein, S., & Larivière, V. (2015). Who Reads Research Articles? An Altmetrics Analysis of Mendeley User Categories. *Journal of the Association for Information Science and Technology,* doi: 10.1002/asi.23286.

Mohammadi, E., Thelwall, M., & Kousha, K. (in press). Can Mendeley Bookmarks Reflect Readership? A Survey of User Motivations. *Journal of the Association for Information Science and Technology*, doi: 10.1002/asi.23286.

Moin, M., Mahmoudi, M., & Rezaei, N. (2005). Scientific output of Iran at the threshold of the 21st century. *Scientometrics, 62*(2), 239-248.

Mounce, R. (2013). Open access and altmetrics: Distinct but complementary. *Bulletin of The American Society for Information Science and Technology, 39*(4), 14-17.

Pinfield, S., Salter, J., Bath, P. A., Hubbard, B., Millington, P., Anders, J. H. S., et al. (2014). Open-access repositories worldwide, 2005–2012: Past growth, current characteristics, and future possibilities. *Journal of the Association for Information Science and Technology, 65*(12), 2404-2421, doi:10.1002/asi.23131.

Salager-Meyer, F. (2014). Writing and publishing in peripheral scholarly journals: How to enhance the global influence of multilingual scholars? *Journal of English for Academic Purposes,* 13, 78-82.

Shuai, X., Pepe, A., & Bollen, J. (2012). How the Scientific Community Reacts to Newly Submitted Preprints: Article Downloads, Twitter Mentions, and Citations. *PloS one,* 7, doi:10.1371/journal.pone.0047523.

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web. *PloS one,* 8(5), doi:10.1371/journal.pone.0064841.

Thelwall, M., & Wilson, P. (in press). Mendeley Readership Altmetrics for Medical Articles: An Analysis of 45 Fields. *Journal of the Association for Information Science and Technology*, doi: 10.1002/asi.23501.

Vaughan, L., & Shaw, D. (2007). A new look at evidence of scholarly citation in citation indexes and from web sources. *Scientometrics, 74*(2), 317-330, doi:10.1007/s11192-008-0220-2.

Wang, X., Liu, C., Fang, Z., & Mao, W. (2014). From Attention to Citation, What and How Does Altmetrics Work? *arXiv preprint arXiv:*1409.4269.

Wren, J. D. (2005). Open access and openly accessible: a study of scientific publications shared via the internet. [Research Support, U.S. Gov't, Non-P.H.S.]. *BMJ, 330*(7500), 1128, doi:10.1136/bmj.38422.611736.E0.

Zahedi, Z., Costas, R., & Wouters, P. (2013). What is the impact of the publications read by the different Mendeley users? Could they help to identify alternative types of impact? *Proc. PLoS ALM Workshop.*

Alternative Metrics for Book Impact Assessment: Can *Choice* Reviews be a Useful Source?

Kayvan Kousha[1] and Mike Thelwall[2]

[1]*k.koushal@wlv.ac.uk*

Statistical Cybermetrics Research Group, School of Mathematics and Computer Science, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY (UK)

[2] *m.thelwall@wlv.ac.uk*
Statistical Cybermetrics Research Group, School of Mathematics and Computer Science, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY (UK)

Abstract
This article assesses whether academic reviews in *Choice: Current Reviews for Academic Libraries* could be systematically used for indicators of scholarly impact, uptake or educational value for scholarly books. Based on 451 *Choice* book reviews from 2011 across the humanities, social sciences and science, there were significant but low correlations between Choice ratings and citation and non-citation impact metrics. The highest correlations found were with Google Books citations (.350) in science and with WorldCat library holdings counts in the humanities (.304). Books recommended by Choice reviewers for undergraduates were mentioned more often in online course syllabi than were other recommended books. Similarly, books recommended for researchers, faculty members and professionals or graduates tended to receive more Google Books citations than did books recommended for undergraduates. In conclusion, metrics derived from Choice academic book reviews can be used as indicators of different aspects of the value of books but more evidence is needed before they could be used as proxies for peer judgements about individual books.

Conference Topic
Webometrics; Altmetrics

Introduction
Impact assessment in book-based subject areas is more challenging than for article-oriented fields because the major current citation indexes are dominated by academic journal articles, and are therefore inadequate for assessing the research impact of books (Hicks, 1999, Archambault, Vignola-Gagné, Côté, Larivière, & Gingras, 2006, Nederhof, 2006; Huang & Chang, 2008). In recognition of the need to include citations from books (Garfield, 1996), the Thomson Reuters Book Citation Index (BKCI) and Scopus now index selected books, but their coverage seems to be too low to make a difference for impact assessment and they are restricted to just a few publishers and books that are mainly in English (Torres-Salinas et al., 2014). The way that the books are indexed also creates other issues for book impact assessment (Leydesdorff & Felt, 2012; Gorraiz, Purnell, & Glänzel, 2013).
Another important issue is that some academic books, such as textbooks and introductory science books, are primarily written for teaching (Gurung, Landrum, & Daniel, 2012) and other books, such as novels and literary works, may have cultural influence (White, Boell, Yu et al., 2009) or play a public engagement role (Kousha & Thelwall, in press). Moreover, education may be seen as particularly important in the humanities and a core part of its value to society (e.g., Nussbaum, 2012). All of these are unlikely to be reflected by citation counts.
Peer review can be used to evaluate the impact of books but it is time-consuming. For instance, in some book-based fields (e.g., history and law) in the 2008 UK Research Assessment Exercise (RAE) reviewers had to assess the research merits of up to 100 books each (Kousha, Thelwall, & Rezaie, 2011). Hence not all of the submitted books may have been examined in detail (Taylor & Walker, 2009). Peer review is also subjective, perhaps

most strongly in the humanities where books are most common. Although critical evaluation is a core skill in the humanities (Small, 2013), it also seems to thrive on controversy and disagreements (Bauerlein, 2002). Moreover, the opinions of reviewers could be more subjective about the teaching or cultural benefits of books than about their research contributions (Weller, 2001).

In response to the weakness of citations for book impact assessment, there have been attempts to assess wider impacts of books (see below), using scholarly book reviews, library holdings statistics, and publisher prestige as well as with altmetrics (Priem, Taraborelli, Groth, & Neylon, 2011). Book reviews are somewhat similar to post-publication reviews for academic articles in systems like Faculty of 1000 (Hunter, 2012; Li & Thelwall, 2012; Mohammadi & Thelwall, 2013; Waltman & Costas, 2014), and both could be useful as additional quality control mechanisms for the critical analysis of published works (Crotty, 2012). The current study explores an alternative source for book impact assessment, *Choice: Current Reviews for Academic Libraries*, which is owned by the American Library Association, and compares it with citation and non-citation metrics. Choice has published reviews of academic books by editors, experts and librarians across different subject areas for about 50 years and is therefore a substantial and successful source of book reviews aimed at librarians making library purchasing decisions. Despite publishing about 7,000 book reviews per year that are relevant to academic libraries, it appears to be an untapped resource in terms of book impact assessment.

Metrics for Book Impact Assessment

Citation Metrics

Web of Science (WoS) and Scopus: Citations to books can be manually extracted from article reference lists (e.g., Cullars, 1998; Krampen, Becker, Wahner & Montada, 2007) or through cited reference searches in WoS (e.g., Bar-Ilan, 2010; Butler & Visser, 2006) or Scopus (Kousha, Thelwall, & Rezaie, 2011), which now includes tens of thousands of books. However, these methods are time-consuming and do not include many citations from books to books. Book to book citations can give different results from article to book citations, especially in book-based fields such as in the humanities and some social sciences (Cronin, Snyder, & Atkins, 1997, Archambault, et al., 2006).

Book Citation Index: The Thomson Reuters Book Citation Index now indexes the references in about 60,000 books and monographs (Book Citation Index, 2014) and is an optional addition to WoS. Nonetheless, only about 3% of BKCI-indexed books are in non-English languages and about 75% of their publishers are from the USA and England (Torres-Salinas et al., 2014). Added to the absence of aggregated citation counts for edited volumes, its use for evaluative purposes would be problematic (Leydesdorff & Felt, 2012; Gorraiz, Purnell, & Glänzel, 2013).

Google Books: Although Google Books (GB) is not a citation index, it can be used to extract citations from digitised books for book impact assessment. GB citations to academic books are more plentiful than citations in traditional citation databases (Scopus and BKCI) in the humanities and in some social sciences but not in science (Kousha & Thelwall, 2009; Kousha, Thelwall, & Rezaie, 2011; Kousha & Thelwall, 2014). For instance, in one study the median number of GB citations was three times higher than the median number of Scopus citations to 1,000 books in the 2008 UK RAE in seven fields (Kousha, Thelwall, & Rezaie, 2011).

Non-Citation Metrics

Book Reviews: Scholarly book reviews are significant academic outputs (Hartley, 2006), especially in some humanities fields, such as history, literature and philosophy (Zuccala &

Van Leeuwen, 2011). One early study found a high correlation (0.620) between the number of reviews in the Book Review Index and the number of library holdings in the OCLC database for 200 novels (Shaw, 1991), suggesting that both indicators may reflect a common factor, such as the popularity of the novels. Another study found that sociology books with more positive reviews tended to attract more citations (Nicolaisen, 2002), although the strength of association between the number of book reviews and citations varies between disciplines (Gorraiz, Gumpenberger, & Purnell 2014). Low but significant Spearman correlations have also been found between the numbers of Amazon book reviews and citation metrics (Kousha & Thelwall, in press).

Libcitations: National or international library holdings statistics can give useful information about potential usage of, or interest in, books (Torres-Salinas & Moed, 2009; White, Boell, Yu et al., 2009). White, Boell, Yu et al. (2009) argued that libcitation statistics could be used as an indication of the cultural benefit of books, especially in the social sciences and humanities. Several follow up studies have found significant, but low, correlations between library holdings statistics and citation metrics for books (Linmans, 2010; Zuccala & Guns, 2013; Kousha & Thelwall, in press), suggesting that library holdings reflect diverse kinds of influence, such as teaching and cultural impacts, that cannot be traced through citations.

Publisher Prestige:

In the absence of credible citation-based indicators for the impact assessment of books, publisher prestige has been proposed as an alternative (Donovan & Butler, 2007). Attempts to estimate the prestige of publishers through surveys of academics have shown that the perception of prestige varies by field (Garand & Giles, 2011; Giménez-Toledo, Tejada-Artigas & Mañana-Rodríguez, 2013). In addition to reputational surveys, BKCI indicators (Torres-Salinas et al., 2012), Scopus citations and matching library holdings data from WorldCat.org (Zuccala, Guns, Cornacchia, & Bod, in press) have also been used to rank academic book publishers.

Syllabus Mentions:

Academics may write textbooks for teaching or monographs that are widely used in teaching rather than, or in addition to, research (Gurung, Landrum, & Daniel, 2012). This kind of teaching contribution may be undervalued or unrewarded (Boyer, 1990; Jenkins, 1995; Healey, 2000) but evidence of inclusion in academic syllabi can reflect some aspects of teaching scholarship success (Albers, 2003; Thompson, 2007). In response, an attempt has been made to capture citations from online course syllabi for WoS-indexed articles across multiple fields, with the results suggesting that online syllabus mentions can be a useful indicator in some social sciences fields (Kousha & Thelwall, 2008).

Research Questions

The following research questions are designed to assess whether ratings and recommendation information in *Choice: Current Reviews for Academic Libraries* could be useful for the impact assessment of academic books.

1. Do Choice book ratings correlate with citation metrics or with other non-citation metrics for books?
2. Are Choice audience recommendations reflected in citation and non-citation metrics? For instance, do books recommended for undergraduates have more syllabus mentions than books recommended for researchers?

Methods

Choice Reviews

The recommendations for 451 book reviews from a free sample issue of *Choice Reviews Online* published in 2011 were extracted from the *Humanities*, *Social & Behavioral Sciences*, and *Science & Technology* categories but omitting reviews for the *Reference* section. The books were selected, with permission of *Choice*, from the collection of free sample reviews. The recommendation levels assigned to Choice reviews (see http://www.ala.org/acrl/choice/about) were converted into a number, from 1 for 'Not recommended' to 5 for 'Essential'.

- *Essential:* A publication of exceptional quality for academic audiences and a core title for academic libraries supporting programs in relevant disciplines.
- *Highly recommended:* A publication of high quality and relevance for academic audiences.
- *Recommended:* A publication containing good content and coverage and suitable for academic audiences.
- *Optional:* A publication that, due to limited value or deficiencies, is marginal for academic audiences.
- *Not recommended:* A poor quality publication or one not suitable for academic audiences.

Choice reviewers include extra information about usefulness for different academic audiences, such as undergraduates, researchers, faculty members and, professionals (Table 1). This information was used for further analyses.

Table 1. Examples of audience recommendations in Choice book reviews.

Audience recommendations	*Examples*
Mainly for undergraduates	*Essential.* Upper-division undergraduates through faculty. *Highly recommended.* Lower-division undergraduates through faculty. *Recommended.* Undergraduate and graduate studies. Optional. Upper-division undergraduates and above.
Mainly for graduates, researchers, professionals and academics	*Essential.* Graduate students, faculty, and professionals. *Highly recommended.* Research libraries and scholars. *Recommended.* All academic and professional audiences. *Optional.* Graduate students, researchers, and faculty.

Google Books Citations

For GB citations, Google Books API searches were used in the previously developed and tested software *Webometric Analyst* (http://lexiurl.wlv.ac.uk, "Books" tab) to extract citations from digitised books indexed by Google Books (for method details see: Kousha & Thelwall, 2014). To locate GB citations in other digitised books, we searched for the first author last name and the first (up to) ten terms of the book title as a phrase search, combined with the publication year.

Lurz "Mindreading animals: The debate over what animals know about other" 2011

For books with three or less words in their titles we added the publisher to the query:

Benford "Performing mixed reality" 2011 "MIT Press"

Syllabus Mentions

For syllabus mentions, an automatic method was used to search for mentions of the 451 books in public online course syllabi indexed by the Bing search engine. *Webometric Analyst* software and a set of rules were used to identify the syllabus mentions in academic websites and to exclude false matches in order to give accurate, although not comprehensive, results. This method was developed to capture academic syllabus mentions for books rather than articles (cf. Kousha & Thelwall, 2008). The first author last name was combined with the book title as a phrase search and either "syllabus" or "course description", with the results of the two combined and false matches automatically filtered out. The automatic syllabus citation extraction method applied in this study seems to give high accuracy (over 90%), although it misses results from non-academic institutions and syllabi stored in password protected databases and systems (see also Kousha & Thelwall, in press).

Barnett "Empire of humanity a history of humanitarianism" "course description"|Barnett "Empire of humanity a history of humanitarianism" "syllabus"

WorldCat Library Holdings

For library holdings, we manually searched for the 451 books in WorldCat online (http://www.worldcat.org) and recorded the number of library holdings for each one.

Mendeley Readers

For Mendeley reader counts, we used the Mendeley API in *Webometric Analyst* with queries combining the last name of the first author, the book title and the publication year for 451 books in the data set (for method details see: Mohammadi & Thelwall, 2014). This returns the number of users of the social reference sharing site Mendeley that have added the book to their personal library.

Amazon.com Reviews

The numbers of customer reviews were automatically extracted from the main Amazon.com URLs for each of the 451 books via *Webometric Analyst* (for method details see: Kousha & Thelwall, 2014 in press).

Sources not used

Not all book impact metrics were collected for the books in the data set. Publisher prestige was not collected because there is not a recognised source of this evidence and it varies by field. WoS/BKCI and Scopus citations were also not collected because Google Books citations have been shown to be superior for book impact assessment in most fields (Kousha & Thelwall, 2009; Kousha, Thelwall, & Rezaie, 2011; Kousha & Thelwall, 2014).

Results

Roughly three-quarters of books with Choice reviews had at least one GB citation (Table 2), and this is higher in the social sciences (80%, median: 3) than in science (68%, median: 2). Moreover, about 45% of the books had one or more academic syllabus mentions and the median number of syllabus mentions is higher in science (1) compared to the humanities (0) and the social sciences (0). About 30% of the Choice books had at least one Amazon review and all 451 books had at least one WorldCat library holding (median: 394). Nevertheless, only 1.5% of books had at least one Mendeley reader. Follow-up manual investigations with Mendeley searches confirmed that this very low number was not a technical artefact but genuinely reflected the virtual absence of the Choice books from this site. The low Mendeley coverage confirms previous results that, although academic journal articles often have many Mendeley readers (e.g., 78% with one or more readers in the medical sciences, see Thelwall & Wilson, in press), the same is not true for books and monographs (Kousha & Thelwall, in press; see also: Hammarfelt, 2014), suggesting that Mendeley is currently not useful for book impact assessment.

Overall, it seems that GB citations are plentiful enough for book citation impact assessment and academic syllabus mentions, libcitations and Amazon reviews may be common enough to be used to indicate different types of impact, such as teaching, cultural or public interest.

Table 2. Google Books citations, syllabus mentions, libcitation, Amazon reviews and Mendeley reader counts for 451 books with Choice reviews published in 2011 in three broad fields.

Choice subjects	*No. of books*	*Google Books No. (% with GB cites*) median (mean)*	*Syllabus No. (% with syllab.*) median (mean)*	*Libcitation No. (% with holdings*) median (mean)*	*Amazon Rev. No. (% with reviews*) median (mean)*	*Mendeley No. (% with readers*) median (mean)*
Human.	136	474 (69.8%) 2 (3.5)	120 (39.7%) 0 (0.9)	62098 (100%) 356 (456.6)	105 (35.2%) 0 (0.8)	31 (3.7%) 0 (0.2)
Social Sci.	234	1278 (79.9%) 3 (5.5)	349 (45.7%) 0 (1.5)	130018 (100%) 442 (555.6)	951 (34.2%) 0 (4.1)	90 (3.4%) 0 (0.4)
Sci. & Tech	81	367 (67.9%) 2 (4.5)	149 (50.6%) 1 (1.8)	41585 (100%) 391 (513.4)	174 (27.2%) 0 (2.15)	194 (3.7%) 0 (2.4)
Total	451	2119 (74.7%) 2 (4.7)	618 (44.8%) 0 (1.4)	233701 (100%) 394 (518.2)	1230 30.8%) 0 (2.7)	315 (1.5%) 0 (0.7)

*% of books with at least one Google Books citation, academic syllabus mention, WorldCat libcitation, Amazon review and Mendeley reader.

Table 4 compares the metrics between those for books with Choice reviews claiming teaching utility (mainly for undergraduates) and those for books with reviews claiming benefits for graduates, researchers, faculty members and professionals. Books with research or other academic relevance have higher GB citation impact (median 3) than books with benefits for undergraduates (GB median 2). In contrast, books with more teaching utility for undergraduate studies tended to have more academic syllabus mentions (median 1 and 55% with one or more syllabus mentions) than books for academic audiences (median zero and 34% with one or more syllabus mentions). Hence, it seems that Choice reviews are broadly capable of distinguishing between the different types of audiences for books.

Table 3. A comparison of book metrics based on Choice book reviews with different rating recommendation levels.

Recommendation	*No. of books*	*Google Books No. (% with GB cites*) median (mean)*	*Syllabus No. (% with syllab.*) median (mean)*	*Libcitations No. (% with holdings*) median (mean)*	*Amazon Rev. No. (% with reviews*) median (mean)*	*Mendeley No. (% with readers*) median (mean)*
Essential/highly recommended	150	768 (88%) 3 (5.1)	186 (48.6%) 0 (1.2)	85256 (100%) 482.5 (568.4)	440 (40%) 0 (2.9)	51 (5.3%) 0 (0.34)
Other	301	1351 (68.1%) 2 (4.5)	432 (42.8%) 0 (1.4)	148445 (100%) 359 (493.2)	790 (26.2%) 0 (2.6)	264 (2.9%) 0 (0.9)

Table 4. A comparison of book metrics based on Choice recommendations for undergraduates and other academic audiences (graduates, researchers, faculty).

Audience recommendation	*No. of books+*	*Google Books No. (% with GB cites) median (mean)*	*Syllabus No. (% with syllab.) median (mean)*	*Libcitation No. (% with holdings) median (mean)*	*Amazon Rev. No. (% with reviews) median (mean)*	*Mendeley No. (% with readers) median (mean)*
Undergraduates	240	1098 (70.1%) 2 (4.7)	420 (55%) 1 (1.7)	122497 (100%) 394.5 (510.4)	649 (29.6%) 0 (2.7)	267 (5.4%) 0 (1.1)
Graduates, faculty, researchers, profess.	203	1006 (79.8%) 3 (4.9)	197 (34%) 0 (0.9)	108260 (100%) 405 (533.3)	579 (33%) 0 (2.85)	48 (2%) 0 (0.2)

+.Eight books with "Not recommended" Choice reviews were excluded.

There are low but significant positive Spearman correlations between Choice ratings and various citation and non-citation indicators (Table 5). Thus, in general, books with more GB citations, academic syllabus mentions, library holdings or Amazon reviews tended to be recommended more highly by book reviewers. The correlation is highest between Choice ratings and libcitations (0.201). This may reflect academic libraries ordering books based on Choice reviews and recommendations, especially in the United States (About Choice magazine, 2015).

Table 5. Spearman correlations between Choice ratings and other metrics across all fields (n=451).

Metrics	*Choice rating score*	*GB citations*	*Syllabus mentions*	*Libcitations*	*Amazon reviews*
Choice rating score	1	**.142****	**.103***	**.201****	**.141****
GB citations		1	.171**	.189**	.196**
Syllabus mentions			1	.121*	.073
Libcitations				1	.222**
Amazon reviews					1

**. Significant at p=0.01

*. Significant at p=0.05

There are disciplinary differences in the strength of association between Choice ratings and the other metrics (Tables 6-8). The highest correlation is between Choice ratings and GB citations in Science & Technology (0.350), but this correlation is much lower in Social & Behavioural Sciences and in the Humanities category. Hence, it seems that science books with more positive reviews tend to be more cited in other books and so Choice reviews may be a useful indicator for assessing the research contribution of scientific books. This is a surprising finding given that books are not as highly valued in science as in the humanities.

In the Humanities category there is a low and statistically insignificant correlation between Choice ratings and GB citations but this may reflect the weak association between citations and research quality in the humanities more than a lack of correlation between Choice ratings and research value or impact. The higher association between Choice ratings and libcitations (0.304) suggests that books with higher review ratings tend to be more often acquired by academic libraries but that this does not translate into citations. This may represent 'cultural benefits' of humanities books (Belfiore & Upchurch, 2013; White, Boell, Yu et al. 2009) and supports a previous finding that Outstanding Academic Titles in Choice are more likely to be purchased by academic libraries and have slightly higher library usage than non-Choice books (Levine-Clark, & Jobe, 2007). In Humanities there is also a low but significant correlation between Choice ratings and academic syllabus mentions (0.131), suggesting that in some teaching based fields, Choice reviews may reflect the educational merits of books. In Social & Behavioural Sciences, however, there is no relationship between Choice ratings and either citation or non-citation metrics. A possible explanation is that in the social sciences books have very different patterns of scholarly usage in research and teaching and the relationship between the number of book reviews and citations could therefore differ between subject areas (Gorraiz, Gumpenberger, & Purnell 2014).

Table 6. Spearman correlations between Choice rating scores and other metrics in Science & Technology (n=81).

Metrics	*Choice rating score*	*GB citations*	*Syllabus mentions*	*Libcitations*	*Amazon reviews*
Choice rating score	1	**.350****	.090	**.274****	**.297****
GB citations		1	.097	.326**	.250*
Syllabus mentions			1	.196	-.019
Libcitations				1	.028
Amazon reviews					1

Table 7. Spearman correlations between Choice rating scores and other metrics in Humanities (n=136).

Metrics	*Choice rating score*	*GB citations*	*Syllabus mentions*	*Libcitations*	*Amazon reviews*
Choice rating score	1	.144	**.131***	**.304****	.089
GB citations		1	.145	.193*	.170*
Syllabus mentions			1	.045	.025
Libcitations				1	.118
Amazon reviews					1

Table 8. Spearman correlations between Choice rating scores and other metrics in Social & Behavioural Sciences (n=234).

Metrics	*Choice rating score*	*GB citations*	*Syllabus mentions*	*Libcitations*	*Amazon reviews*
Choice rating score	1	.081	.095	.123	.123
GB citations		1	.193**	.127	.179**
Syllabus mentions			1	.117	.116
Libcitations				1	.314**
Amazon reviews					1

Limitations

This study tested only 451 books with Choice reviews from a free issue of *Choice Reviews Online* published in 2011 and a larger data may give different results. The sample of 451 is from the most public part of Choice, its free samples, and so is atypical in that regard. The small sample size was also not enough for a fine grained analysis of individual subject areas and this is an important limitation for the correlation tests because citation practices and educational norms (e.g., typical class sizes and the role of textbooks) can vary substantially between fields in a way that would systematically reduce correlation results when the fields are grouped together. Another limitation is that the data only included GB citations from books to books and so would miss citations from articles to books. Hence, a future study might use cited reference searches in WoS or Scopus order to check whether stronger relationships can be found.

Discussion and Conclusions

This study seems to be the first to assess whether the book reviews in *Choice: Current Reviews for Academic Libraries* reflect the value of books and could be used for indicators of value or impact. The analysis of a small sample of 451 books published in 2011 found weak but often significant relationships with other indicators, suggesting that Choice should be particularly helpful for books that have uses that do not necessarily attract citations.

In answer to the first research question, books that were highly rated in Choice received more GB citations, academic syllabus mentions, libcitations and Amazon reviews than did lower rated books. In answer to the second research question, books recommended for undergraduates (e.g., textbooks) received more academic syllabus mentions, reflecting teaching influence of books, and books recommended for researchers, faculty and professionals received more citations than did books recommended for undergraduates, indicating the ability of Choice reviews to distinguish between the different audiences for books.

The low (but statistically significant) Spearman correlations between Choice ratings and all citation and non-citation indicators suggest that Choice reviews are either somewhat subjective, or (more likely) do not reflect exactly the same aspects of the value of a book (e.g., teaching, research, cultural or social impacts) as any of the other indicators. Hence, the evidence presented here is insufficient to claim that Choice recommendations are reliable indicators of audience or value at the individual book level. Nevertheless, the correlations will be weakened by the broad categories used (e.g., 200 library holdings might be a spectacular success for a monograph on Old Norse but a failure for one on Shakespeare's women). In addition, the correlations will also be weakened by the fact that the other indicators are not direct *measures* of anything (e.g., educational value) but are indirect (not cause-and-effect) reflections and so strong correlations should not be expected. Hence, the low correlations are not evidence that Choice book reviews have little value but probably reflect the complex multifaceted nature of the value of books and the difficulty in finding indicators to effectively reflect those values. In this context, Choice book reviews are a promising new source of post-publication peer review evidence of the value of books. They are a welcome additional source of evidence for the particularly challenging task of book impact assessment and when positive reviews are used for impact assessments of scholarly outputs by evaluators, funders or perhaps even national research assessments (e.g., PBRF, 2013).

References

About Choice magazine, 2015. http://www.ala.org/acrl/choice/about

Albers, C. (2003). Using the syllabus to document the scholarship of teaching. *Teaching Sociology*, *31*(1), 60-72.

Archambault, É., Vignola-Gagné, É., Côté, G., Larivière, V., & Gingras, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics*, *68*(3), 329-342.

Bar-Ilan, J. (2010). Citations to the "Introduction to informetrics" indexed by WOS, Scopus and Google Scholar. *Scientometrics*, *82*(3), 495-506.

Bauerlein, M. (2002). Disagreements in the humanities. *Knowledge, Technology & Policy*, *15*(1), 188-195.

Belfiore, E., & Upchurch, A. (Eds.). (2013). *Humanities in the Twenty-first Century: Beyond Utility and Markets*. Basingstoke, UK: Palgrave Macmillan.

Boyer, E. L. (1990). Scholarship revisited. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching. http://184.168.109.199:8080/jspui/bitstream/123456789/2134/1/ED326149.pdf

Butler, L., & Visser, M. (2006). Extending citation analysis to non-source items. *Scientometrics*, *66*(2), 327–343

Cronin, B., Snyder, H. & Atkins, H. (1997). Comparative citation rankings of authors in monographic and journal literature: a study of sociology. *Journal of Documentation*, *53*(3), 263-273.

Crotty, D. (2012). Life after publication - post-publication peer review. *Biochemist*, *34*(4), 26-28. http://www.biochemist.org/bio/03404/0026/034040026.pdf

Cullars, J. (1998). Citation characteristics of English-language monographs in philosophy. *Library & Information Science Research*, *20*(1), 41–68.

Donovan, C., & Butler, L. (2007). Testing novel quantitative indicators of research "quality," esteem and "user engagement:" An economics pilot study. *Research Evaluation*, *16*(4), 231-242.

Garand, J.C., & Giles, M.W. (2011). Ranking scholarly publishers in political science: An alternative approach. *PS: Political Science and Politics*, *44*(2), 375-383.

Garfield, E. (1996). Citation Indexes for Retrieval and Research Evaluation. Consensus Conference on the Theory and Practice of Research Assessment, Capri.

Giménez-Toledo, E., Tejada-Artigas, C., & Mañana-Rodríguez, J. (2013). Evaluation of scientific books' publishers in social sciences and humanities: Results of a survey. *Research Evaluation*, *22*(1), 64-77.

Gorraiz, J., Gumpenberger, C., & Purnell, P. J. (2014). The power of book reviews: A simple and transparent enhancement approach for book citation indexes. *Scientometrics*, *98*(2), 841-852.

Gorraiz, J., Purnell, P. J., & Glänzel, W. (2013). Opportunities for and limitations of the book citation index. *Journal of the American Society for Information Science and Technology*, *64*(7), 1388-1398.

Gurung, R. A. A., Landrum, R. E., & Daniel, D. B. (2012). Textbook use and learning: A North American perspective. *Psychology Learning and Teaching*, *11*(1), 87-98.

Hammarfelt, B. (2014). Using altmetrics for assessing research impact in the humanities. *Scientometrics*, *101*(2), 1419-1430.

Hartley, J. (2006). Reading and writing book reviews across the disciplines. *Journal of the American Society for Information Science and Technology*, *57*(9), 1194-1207.

Hicks, D. (1999). The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. *Scientometrics*, *44*(2), 193−215.

Huang, M., & Chang, Y. (2008). Characteristics of research output in social sciences and humanities: from a research evaluation perspective. *Journal of the American Society for Information Science and Technology*, *59*(11), 1819–1828.

Hunter, J. (2012, August 30). Post-publication peer review: Opening up scientific conversation. *Frontiers in Computational Neuroscience*, doi:10.3389/fncom.2012. 00063

Jenkins, A. (1995). The research assessment exercise, funding and teaching quality. *Quality Assurance in Education*, *3*(2), 4-12.

Kousha, K., & Thelwall, M. (2008). Assessing the impact of disciplinary research on teaching: An automatic analysis of online syllabuses. *Journal of the American Society for Information Science and Technology*, *59*(13), 2060-2069.

Kousha, K., & Thelwall, M. (2009). Google Book Search: Citation analysis for social science and the humanities. *Journal of the American Society for Information Science and Technology*, *60*(8), 1537-1549.

Kousha, K., & Thelwall, M. (2014). An automatic method for extracting citations from Google Books. *Journal of the Association for Information Science and Technology*, *66*(2), 309-320.

Kousha, K., & Thelwall, M. (in press, 2015). Can Amazon.com reviews help to assess the wider impacts of books? *Journal of the Association for Information Science and Technology*. http://www.koosha.tripod.com/AmazonReviewstoAssessBooks-Preprint.pdf

Kousha, K., & Thelwall, M. (in press). An automatic method for assessing the teaching impact of books from online academic syllabi. *Journal of the Association for Information Science and Technology*, http://www.scit.wlv.ac.uk/~cm1993/papers/SyllabiBookCitations.pdf

Kousha, K., Thelwall, M., & Rezaie, S. (2011). Assessing the citation impact of books: The role of Google Books, Google Scholar, and Scopus. *Journal of the American Society for Information Science and Technology*, *62*(11), 2147-2164.

Krampen, G., Becker, R., Wahner, U., & Montada, L. (2007). On the validity of citation counting in science evaluation: Content analyses of references and citations in psychological publications. *Scientometrics*, *71*(2), 191–202.

Levine-Clark, M., & Jobe, M. M. (2007). Do reviews matter? An analysis of usage and holdings of choice-reviewed titles within a consortium. *Journal of Academic Librarianship*, *33*(6), 639-646.

Leydesdorff, L., & Felt, U. (2012). Edited volumes, monographs and book chapters in the Book Citation Index (BKCI) and Science Citation Index (SCI, SoSCI, A&HCI). *Journal of Scientometric Research*, *1*(1), 28–34.

Li, X., & Thelwall, M. (2012). F1000, Mendeley and traditional bibliometric indicators. *17th International Conference on Science and Technology Indicators* (Vol. 3, pp. 1–11). http://sticonference.org/Proceedings/vol2/Li_F1000_541.pdf

Linmans, A. J. M. (2010). Why with bibliometrics the humanities does not need to be the weakest link. Indicators for research evaluation based on citations, library bindings and productivity measures. *Scientometrics*, *83*(2), 337–354.

Mohammadi, E. & Thelwall, M. (2013). Assessing non-standard article impact using F1000 labels. *Scientometrics*, *97*(2), 383-395.

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.

Nederhof, A. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, *66*(1), 81−100.

Nicolaisen, J. (2002). The scholarliness of published peer reviews: A bibliometric study of book reviews in selected social science fields. *Research Evaluation*, *11*(3), 129-140.

Nussbaum, M. C. (2012). *Not for profit: Why democracy needs the humanities*. Princeton, NJ: Princeton University Press.

PBRF. (2013). Performance-Based Research Fund Quality Evaluation Guidelines 2012. http://www.tec.govt.nz/Documents/Publications/PBRF-Quality-Evaluation-Guidelines-2012.pdf

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2011). altmetrics: a manifesto. http://altmetrics.org/manifesto.

Shaw, D. (1991). An analysis of the relationship between book reviews and fiction holdings in OCLC. *Library and Information Science Research*, *13*(2), 147-154.

Small, H. (2013). *The Value of the Humanities*. Oxford: Oxford University Press.

Taylor, J. & Walker, I. (2009). Peer assessment of research: How many publications per staff? Lancaster University Management School, Working Paper 2009/035. http://eprints.lancs.ac.uk/31757/1/006236.pdf

Thelwall, M. & Wilson, P. (in press). Mendeley readership Altmetrics for medical articles: An analysis of 45 fields. *Journal of the Association for Information Science and Technology*. http://www.scit.wlv.ac.uk/~cm1993/papers/MendeleyInScienceAltmetricsPreprint.pdf

Thompson, B. (2007). The syllabus as a communication document: Constructing and presenting the syllabus. *Communication Education*, *56*(1), 54-71.

Torres-Salinas, D. & Moed, H. F. (2009). Library catalog analysis as a tool in studies of social sciences and humanities: An exploratory study of published book titles in economics. *Journal of Informetrics*, *3*(1), 9-26.

Torres-Salinas, D., Robinson-García, N., Campanario, J. M., & López-Cózar, E. D. (2014). Coverage, field specialisation and the impact of scientific publishers indexed in the book citation index. *Online Information Review*, *38*(1), 24-42.

Torres-Salinas, D., Robinson-García, N., Jiménez-Contreras, E., & Delgado López-Cózar, E. (2012). Towards a 'Book Publishers Citation Reports'. First approach using the 'Book Citation Index'. *Revista Española de Documentación Científica*, *35*(4), 615-620.

Torres-Salinas, D., Rodríguez-Sánchez, R., Robinson-García, N., Fdez-Valdivia, J., & García, J. A. (2013). Mapping citation patterns of book chapters in the Book Citation Index. *Journal of Informetrics*, *7*(2), 412-424.

Waltman, L. & Costas, R. (2014). F1000 Recommendations as a potential new data source for research evaluation: a comparison with citations. *Journal of the Association for Information Science and Technology*, *65*(3), 433-445.

Weller, A. C. (2001). *Editorial Peer Review: Its Strengths and Weaknesses*. Medford, N.J: Information Today.

White, H. D., Boell, S. K., Yu, H., Davis, M., Wilson, C. S., & Cole, F. T. H. (2009). Libcitations: A measure for comparative assessment of book publications in the humanities and social sciences. *Journal of the American Society for Information Science and Technology*, *60*(6), 1083-1096.

Zuccala, A., & Guns, R. (2013). Comparing book citations in humanities journals to library holdings: Scholarly use versus "perceived cultural benefit" (RIP). In J. Gorraiz, E. Schiebel, C. Gumpenberger, M. Hörlesberger, & H. Moed (Eds.) *Proceedings of 14th International Conference of the International Society for Scientometrics and Informetrics* (pp.353–360). Vienna, Austria: AIT Austrian Institute of Technology GmbH Vienna.

Zuccala, A. & Van Leeuwen, T. (2011). Book reviews in humanities research evaluations. *Journal of the American Society for Information Science and Technology*, *62*(10), 1979–1991.

Zuccala, A., Guns, R., Cornacchia, R., & Bod, R. (in press). Can we rank scholarly book publishers? A bibliometric experiment with the field of history. *Journal of the Association for Information Science and Technology*. http://www.illc.uva.nl/evaluating-humanities/RankingPublishers(Preprint_2014).pdf

A Longitudinal Analysis of Search Engine Index Size

Antal van den Bosch[1], Toine Bogers[2], and Maurice de Kunder[3]

[1] *a.vandenbosch@let.ru.nl*
Centre for Language studies, Radboud University, PO Box 9103, 6500 HD Nijmegen (The Netherlands)

[2] *toine@hum.aau.dk*
Department of Communication, Aalborg University Copenhagen, A.C. Meyers Vænge 15, 2450 Copenhagen (Denmark)

[3] *maurice@dekunder.nl*
De Kunder Internet Media, Toernooiveld 100, 6525 EC, Nijmegen (The Netherlands)

Abstract
One of the determining factors of the quality of Web search engines is the size of their index. In addition to its influence on search result quality, the size of the indexed Web can also tell us something about which parts of the WWW are directly accessible to the everyday user. We propose a novel method of estimating the size of a Web search engine's index by extrapolating from document frequencies of words observed in a large static corpus of Web pages. In addition, we provide a unique longitudinal perspective on the size of Google and Bing's indexes over a nine-year period, from March 2006 until January 2015. We find that index size estimates of these two search engines tend to vary dramatically over time, with Google generally possessing a larger index than Bing. This result raises doubts about the reliability of previous one-off estimates of the size of the indexed Web. We find that much, if not all of this variability can be explained by changes in the indexing and ranking infrastructure of Google and Bing. This casts further doubt on whether Web search engines can be used reliably for cross-sectional webometric studies.

Conference Topic
Webometrics

Introduction
Webometrics (or cybermetrics) is commonly defined as the study of the content, structure, and technologies of the World Wide Web (WWW) using primarily quantitative methods. Since its original conception in 1997 by Almind & Ingwersen, researchers in the field have studied aspects such as the link structure of the WWW, credibility of Web pages, Web citation analysis, the demographics of its users, and search engines (Thelwall, 2009). The size of the WWW, another popular object of study, has typically been hard to estimate, because only a subset of all Web pages is accessible through search engines or by using Web crawling software. Studies that attempt to estimate the size of the WWW tend to focus on the surface Web—the part indexed by Web search engines—and often only at a specific point in time.

In the early days of search engines, having the biggest index size provided search engines with a competitive advantage, but a changing focus on other aspects of search result quality, such as recency and personalization, has diminished the importance of index size in recent years. Nevertheless, the size of a search engine's index is important for the quality of Web search engines, as argued by Lewandowski and Höchstötter (2008). In addition, knowledge of the size of the indexed Web is important for webometrics in general, as it gives us a ceiling estimate of the size of the WWW that is accessible by the average Internet user.

The importance of index sizes in the early days of Web search resulted in several estimation methods, most of which used the overlap between different Web search engines to estimate the size of the indexed Web as a whole. Bharat and Broder (1998) used an overlap-based method to estimate the size of the WWW at around 200 million pages. Lawrence & Giles

(1998, 1999) produced higher estimates of 320 and 800 million pages in 1998 and 1999 using a similar method, and Gulli and Signorini (2005) updated these estimates to 11.5 billion pages. The last decade has seen little work on index size estimation, but a general problem with all of the related work so far is that all the analyses have been cross-sectional. There has been no analysis of index size on a longer time scale that sheds light on the robustness of the different estimation methods. The handful of studies that have taken a longer-term perspective have typically focused on Web page persistence (Koehler, 2004) or academic link structure (Payne & Thelwall, 2008), but never search engine index size.

In this paper we present a novel method of estimating the size of a Web search engine's index by extrapolating from document frequencies of words observed in a large static corpus of Web pages. In addition, we provide a unique longitudinal perspective on our estimation method by applying it to estimate the size of Google and Bing's[1] indexes over a period of close to nine years, from March 2006 until January 2015.[2]

We find that index size estimates of these two search engines tend to vary wildly over time, with Google generally possessing a larger index than Bing. This considerable variability has been noted in earlier work (e.g., Rousseau, 1999; Payne & Thelwall, 2008), which raises doubts about the reliability of previous one-off estimates of the size of the indexed Web. In our analysis, we find that much of this variability can be explained by changes in the indexing and ranking infrastructure of Google and Bing. This casts further doubt on whether Web search engines can be used reliably for one-off Webometric studies, confirming similar sentiments expressed by, for instance, Payne and Thelwall (2008), and Thelwall (2012).

The remainder of this paper is organized as follows. The next section contains a review of related work in webometrics and on estimating the size of the indexed WWW. We then explain our estimation method in more detail, followed by the results of our estimation method and an analysis of the variability we uncover. We then discuss our findings and draw our conclusions.

Related work

Since its inception, researchers have studied many different aspects of the Web. This section provides a brief overview of some of the key studies on measuring different properties of Web search engines and the WWW, in particular work on estimating their size.

Measuring the Web

Over the past two decades many aspects of the WWW have been studied, such as the link structure of the Web that emerges from the hyperlinks connecting individual Web pages. Broder et al. (2000) were among the first to map the link structure of the WWW. They showed that the Web graph can be visualized as a bow-tie structure with 90% of all pages being a part of the largest strongly connected component, which was confirmed in 2005 by Hirate et al. (2008). Payne and Thelwall (2008) performed a longitudinal analysis of hyperlinks on academic Web sites in the UK, Australia and New Zealand over a six-year period. They found that the inlink and outlink counts were relatively stable over time, albeit with large fluctuations at the individual university level. As a result, they concluded that such variability could create problems for the replicability and comparability of webometrics research. Other related work on analyzing the link structure of the Web includes Kleinberg et al. (1999) and Björneborn (2004).

[1] Formerly known as Microsoft Live Search until May 28, 2009.
[2] Recent daily estimates produced by our method can be accessed through http://www.worldwidewebsize.com/. The time series data displayed in Figure 1 are available online at http://toinebogers.com/?page_id=757.

Web search engines are an essential part of navigating the WWW and as a result have received much attention. Many different aspects of Web search have been investigated, such as ranking algorithms, evaluation, user behavior, and ethical and cultural perspectives. Bar-Ilan (2004) and Zimmer (2010) provide clear, multi-disciplinary overviews of the most important work on these aspects.

From a webometric perspective the hit counts, search engine rankings, and the persistence of the indexed URLs are highly relevant for the validity and reliability of webometric research using Web search engines. Rousseau (1999) was among the first to investigate the stability of search engine results by tracking the *hit counts*—the number of results indicated for a query—for three single-word search terms in Altavista and NorthernLight over a 12-week period in 1998. Altavista exhibited great variability over a longer time period, even with only three anecdotal query words. Rousseau attributed this to changes in Altavista's infrastructure with the launch of a new version in 1998. Thelwall (2008) also performed a cross-sectional, quantitative comparison of the hit counts and search engine results of Google, Yahoo!, and Live Search. He extracted 1,587 single-word queries from English-language blogs "based purely on word frequency criteria" (Thelwall, 2008, p. 1704), found strong correlations between the hit count estimates of all three search engines, and recommended using Google for obtaining accurate hit count estimates. Uyar (2009) extended Thelwall's work by including multi-word queries. He found that the number of words in the query significantly affects the accuracy of hit counts, with single-word queries providing nearly double the hit count accuracy as compared to multi-word queries. Finally, Thelwall and Sud (2012) investigated the usefulness of the Bing Search API 2.0 for performing webometric research. They examined, among other things, the hit count estimates and found that these can vary by up to 50% and should therefore be used with caution in webometric research.

Bar-Ilan et al. (2006) compared the rankings of three different Web search engines over a three-week period. They observed that the overlap in result lists for textual queries was much higher than for image queries, where the result lists of the different search engines showed almost no overlap. Spink et al. (2006) investigated the overlap between three major Web search engines based on the first results pages and found that 85% of all returned top 10 results are unique to that search engine.

The issue of Web page persistence in search engine indexes—how long does a Web page remain indexed and available—was first examined by Bar-Ilan (1999) for a single case-study query during a five-month period in 1998. She found that for some search engines up to 60% of the results had disappeared from the index at the end of the period. She hypothesized that the distributed nature of search engines may cause different results to be served up from different index shards at different points in time. Koehler (2004) reported on the results of a six-year longitudinal study on Web page persistence. He also provided an overview of different longitudinal studies on the topic and concluded, based on the relatively small number of studies that exist, that Web pages are not a particularly persistent medium, although there are meaningful differences between navigation and content pages.

Index size estimation

In the last two decades, various attempts have been directed at estimating the size of the indexed Web. Some approaches focus on estimating the index size of a single search engine directly, while a majority focuses on estimating the overlap to indirectly estimate the size of the total indexed Web.

Highly influential work on estimating index size was done by Bharat & Broder (1998), who calculated the relative sizes of search engines by selecting a random set of pages from one engine, and checking whether each page was indexed by another engine. They used 35,000 randomly generated queries of 6 to 8 words selected at random from a Web-based lexicon and sent these queries to four search engines. One of every top-100 results pages was randomly selected, after which they calculated the relative sizes and overlaps of search engines by selecting this random set of pages from one engine, and checking whether the page was indexed by another engine. By combining their method with self-reported index sizes from the commercial search engines, they estimated the size of the WWW to be around 200 million pages. Gulli et al. (2005) extended the work of Bharat and Broder by increasing the number of submitted queries by an order of magnitude, and using 75 different languages. They calculated the overlap between Google, Yahoo!, MSN Live, and Ask.com, and updated the previous estimates to 11.5 billion pages in January 2005. Most approaches that use the work of Bharat and Broder as a starting point focus on improving the sampling of random Web pages, which can be problematic because not every page has the same probability of being sampled using Bharat and Broder's approach. Several researchers have proposed methods of near-uniform sampling that attempt to compensate for this ranking bias, such as Henzinger et al. (2000), Anagnostopoulos et al. (2006), and Bar-Yossef and Gurevich (2006, 2011).

Lawrence and Giles (1998) estimated the indexed overlap of six different search engines. They captured the queries issued by the employees of their own research institute and issued them to all six engines. The overlap among search engines was calculated on the aggregated result sets, after which they used publicly available size figures from the search engines to estimate the size of the indexed Web to be 320 million pages. Lawrence and Giles updated their previous estimates to 800 million Web pages in July 1999. Dobra et al. (2004) used statistical population estimation methods to improve upon the original 1998 estimate of Lawrence and Giles. They estimated that Lawrence and Giles were off by a factor of two and that the Web contained around 788 million Web pages in 1998. Khelghati et al. (2012) compared several of the aforementioned estimation methods as well as some proposed modifications to these methods. They found that a modified version of the approach proposed by Bar-Yossef et al. (2011) provided the best performance.

Estimating the Size of a Search Engine Index through Extrapolation

On the basis of a textual corpus that is fully available, both the number of documents and the term and document frequencies of individual terms can be counted. In the context of Web search engines, however, we only have reported hit counts (or document counts), and we are usually not informed about the total number of indexed documents. Since it is the latter we are interested in, we want to estimate the number of documents indexed by a search engine indirectly from the reported document counts.

We can base such estimates on a training corpus for which we have full information on document frequencies of words and on the total number of documents. From the training corpus we can extrapolate a size estimation of any other corpus for which document counts are given. Suppose that, for example, we collect a training corpus T of 500,000 web pages, i.e. $|T| = 500{,}000$. For all words w occurring on these pages we can count the number of documents they occur in, or their document count, $d_T(w)$. A frequent word such as *are* may occur in 250,000 of the documents, i.e., it occurs in about one out of every two documents; $d_T(are) = 250{,}000$. Now if the same word *are* is reported to occur in 1 million documents in another corpus C, i.e., its document count $d_C(are) = 1{,}000{,}000$, we can estimate by extrapolation that this corpus will contain about $|C| = \frac{d_C(are)\times|T|}{d_T(are)}$, i.e., 2 million documents.

There are two crucial requirements that would make this extrapolation sound. First, the training corpus would need to be representative of the corpus we want to estimate the size of. Second, the selection of words[3] that we use as the basis for extrapolation will need to be such

[3] We base our estimates on words rather than on multi-word queries based on the findings of Uyar (2009).

that the extrapolations based on their frequencies are statistically sound. We should not base our estimates on a small selection of words, or even a single word, as frequencies of both high-frequency and low-frequency words may differ significantly among corpora. Following the most basic statistical guidelines, it would be better to repeat this estimation for several words, e.g., twenty times, and average over the extrapolations.

A random selection of word types is likely to produce a selection with relatively low frequencies, as Zipf's second law predicts (Zipf, 1995). A well-known issue in corpus linguistics is that when any two corpora are different in genre or domain, very large differences are likely to occur in the two corpora's word frequencies and document frequencies, especially in the lower frequency bands of the term distributions. It is not uncommon that half of the word types in a corpus occur only once; many of these terms will not occur in another disjoint corpus, even if it is of the same type. This implies that extrapolations should not be based on a random selection of terms, many of which will have a low frequency of occurrence.

The selection of words should sample several high-frequency words but preferably also several other words with frequencies spread across the document frequency bands.

It should be noted that Zipf's law concerns word frequencies, not document frequencies. Words with a higher frequency tend to recur more than once in single documents. The higher the frequency of a word, the more its document frequency will be lower than its word frequency. A ceiling effect thus occurs with the most frequent words if the corpus contains documents of sufficient size: they tend to occur in nearly all documents, making their document frequencies about the same and approaching the actual number of documents in the corpus, while at the same time their word token frequencies still differ to the degree predicted by Zipf's law (Zipf, 1995). This fact is not problematic for our estimation goal, but it should be noted that this hinges on the assumption that the training corpus and the new corpus of which the frequencies are unknown, contain documents of about the same average size.

As our purpose is to estimate the size of a Web search engine's index, we must make sure that our training corpus is representative of the web, containing documents with a representative average size. This is quite an ambitious goal. We chose to generate a randomly filtered selection of 531,624 web pages from the DMOZ[4] web directory. We made this selection in the spring of 2006. To arrive at this selection, first a random selection was made of 761,817 DMOZ URLs, which were crawled. Besides non-existing pages, we also filtered out pages with frames, server redirects beyond two levels, and client redirects. In total, the DMOZ selection of 531,624 documents contains 254,094,395 word tokens (4,395,017 unique word types); the average DMOZ document contains 478 words.

We then selected a sequence of DMOZ words by their frequency rank, starting with the most frequent word, and selecting an exponential series where we increase the selection rank number with a low exponent, viz. 1.6. We ended up with a selection of the following 28 words, the first nine being high-frequency function words and auxiliary verbs: *and*, *of*, *to*, *for*, *on*, *are*, *was*, *can*, *do*, *people*, *very*, *show*, *photo*, *headlines*, *william*, *basketball*, *spread*, *nfl*, *preliminary*, *definite*, *psychologists*, *vielfalt*, *illini*, *chèque*, *accordée*, *reticular*, *rectificació*. The DMOZ directory is multilingual, but English dominates. It is not surprising that the tail of this list contains words from different languages.

Our estimation method then consists of retrieving document counts for all 28 words from the search engine we wish to estimate the number of documents for, obtaining an extrapolated estimate for each word, and averaging (taking a mean) over the 28 estimations. If a word is not reported to occur in any document (which hardly happens), it is not included in the average.

[4] DMOZ is also called the Open Directory Project, http://www.dmoz.org/.

To stress-test the assumption that the DMOZ document frequencies of our 28 words yield sensible estimates of corpus size, we estimated the size of a range of corpora: the New York Times part of the English Gigaword corpus[5] (newspaper articles), the Reuters RCV1 corpus[6] (newswire articles), the English Wikipedia[7] (encyclopedic articles, excluding pages that redirect or disambiguate), and a held-out sample of random DMOZ pages. Table 1 provides an overview of the estimations on these widely different corpora. The size of the New York Times corpus is overestimated by a large margin of 126%, while the sizes of the other three corpora are underestimated. The size of the DMOZ sample—not overlapping with the training set, but drawn from the same source—is relatively accurately estimated with a small underestimation of 1.3%. Larger underestimations, for Reuters RCV1 and Wikipedia, may be explained by the fact that these corpora have shorter documents on average.

The standard deviations in Table 1, computed over the 28 words, indicate that the different estimates are dispersed over quite a large range. There seems to be no correlation with the size of the difference between the actual and the estimated number of documents. Yet, the best estimate, for the small DMOZ held-out sample (–1.3% error), coincides with the smallest standard deviation.

Table 1. Real versus estimated numbers (with standard deviations) of documents on four textual corpora, based on the DMOZ training corpus statistics: two news resources (top two) and two collections of web pages (bottom two). The second and third column provides the mean and median number of words per document.

	Words per document					
Corpus	*Mean*	*Median*	*# Documents*	*Estimate*	*St. dev.*	*Difference*
New York Times '94-'01	837	794	1,234,426	2,789,696	1,821,823	+126%
Reuters RCV1	295	229	453,844	422,271	409,648	– 7.0%
Wikipedia	447	210	2,112,923	2,024,792	1,385,105	– 4.2%
DMOZ test sample	477	309	19,966	19,699	5,839	– 1.3%

After having designed this experiment in March 2006, we started to run it on a daily basis on March 13, 2006, and have continued to do so. Each day we sent the 28 DMOZ words as queries to two search engines: Bing[i] and Google[8]. We retrieve the reported number of indexed pages on which each word occurs as it is returned by the web interface of both search engines, not their APIs. This number is typically rounded: it retains three or four significant numbers, the rest being padded by zeroes. For each word we use the reported document count to extrapolate an estimate of the search engine's size, and average over the extrapolations of all words. The web interfaces to the search engines have gone through some changes, and the time required to adapt to these changes sometimes caused lags of a number of days in our measurements. For Google 3,027 data points were logged, which is 93.6% of the 3,235 days between March 13, 2006 and January 20, 2015. For Bing, this percentage is 92.8% (3,002 data points).

Results

Figure 1 displays the estimated sizes of the Google and Bing indices between March 2006 and January 2015. For visualization purposes and to avoid clutter, the numbers are unweighted

[5] https://catalog.ldc.upenn.edu/LDC2003T05.
[6] http://trec.nist.gov/data/reuters/reuters.html.
[7] Downloaded on October 28, 2007.
[8] We also sent the same 28 words to two other search engines that were discontinued at some point after 2006.

running averages of 31 days, taking 15 days before and after each focus day as a window. The final point in our measurements is January 20, 2015; hence the last point in this graph is January 5, 2015. Rather than a linear, monotonic development we observe a rather varying landscape, with Google usually yielding the larger estimates. The largest peak in the Google index estimates is about 49.4 billion documents, measured in mid-December 2011. Occasionally, estimates are as low as under 2 billion pages (e.g. 1.96 billion pages in the Google index on November 24, 2014), but such troughs in the graph are usually short-lived, and followed by a return to high numbers (e.g., to 45.7 billion pages in the Google index on January 5, 2015).

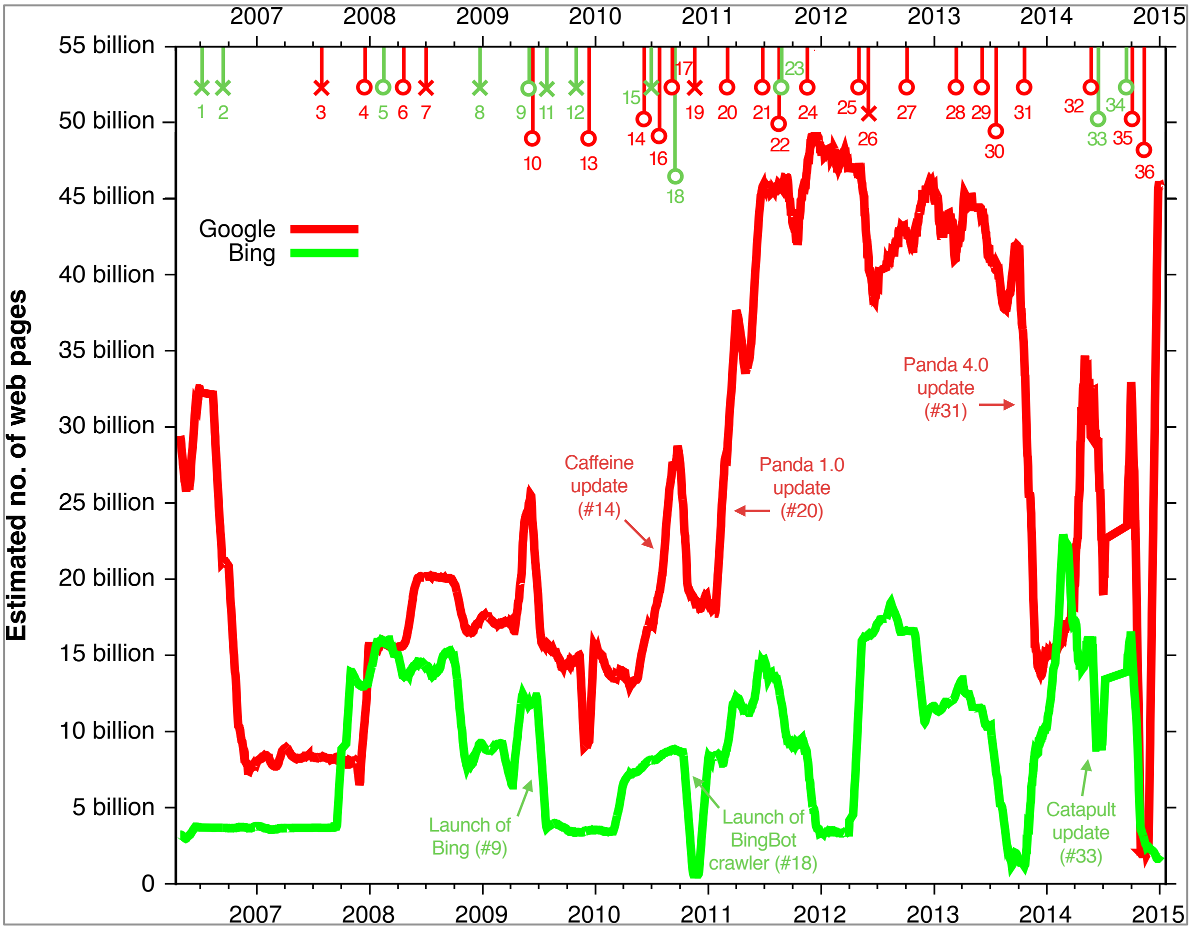


Figure 1. Estimated size of the Google and Bing indexes from March 2006 to January 2015. The lines connect the unweighted running daily averages of 31 days. The colored, numbered markers at the top represent reported changes in Google and Bing's infrastructure. The colors of the markers correspond to the color of the search engine curve they related to; for example, red markers signal changes in Google's infrastructure (the red curve). Events that line up with a spike are marked with an 'O', other events are marked with an '✕'.

Extrinsic variability

The variability observed in Figure 1 is not surprising given the fact that the indexing and ranking architectures of Web search engines are updated and upgraded frequently. According to Matt Cutts[9], Google makes "roughly 500 changes to our search algorithm in a typical year", and this is likely the same for Bing. While most of these updates are not publicized,

[9] http://googleblog.blogspot.com/2011/11/ten-algorithm-changes-on-inside-search.html.

some of the major changes that Google and Bing make to their architectures are announced on their official blogs. To examine which spikes in Figure 1 can be attributed to publicly announced architecture changes, we went through all blog posts on the Google Webmaster Central Blog[10], the Google Official Blog[11], the Bing Blog[12], and Search Engine Watch[13] for reported changes to their infrastructure. This resulted in a total of 36 announcements related to changes in the indexing or ranking architecture of Google and Bing[14]. The colored, numbered markers at the top of Figure 1 show how these reported changes are distributed over time.

For Google 20 out of the 24 reported changes appear to correspond to sudden spikes in the estimated index size, and for Bing 6 out of 12 reported changes match up with estimation spikes. This strongly supports the idea that much of the variability can be attributed to such changes. Examples include the launch of Bing on May 28, 2009 (event #9), the launch of Google's search index Caffeine on June 8, 2010 (event #14), the launch of the BingBot crawler (event #18), and the launches of Google Panda updates, and Bing's Catapult update (events #20, #31, and #33).

Of course not all spikes can be explained by reported events. For example, the spike in Bing's index size in October 2014 does not match up with any publicly announced changes in their architecture, although it is a likely explanation for such a significant change. In addition, some changes to search engine architectures are rolled out gradually and would therefore not translate to spikes in the estimated size. However, much of the variation in hit counts, and therefore estimated index size, appears to be caused by changes in the search engine architecture—something already suggested by Rousseau in his 1999 study.

Discussion and Conclusions

In this paper we presented a method for estimating the size of a Web search engine's index. Based on the hit counts reported by two search engines, Google and Bing, for a set of 28 words, the size of the index of each engine is extrapolated. We repeated this procedure and performed it once per day, starting in March 2006. The results do not show a steady, monotonic growth, but rather a highly variable estimated index size. The larger estimated index of the two, the one from Google, attains high peaks of close to 50 billion web pages, but occasionally drops to small indices of 2 billion pages as well. Are we measuring the extrinsic variability of the indices, or an intrinsic variability of our method? Our method is fixed: the same 28 words are sent to both search engines on every day. The frequencies of our test words are unlikely to change dramatically in a corpus as big as a crawl of the indexed Web; especially the document counts for our high-frequent words in our list should approximate (or at least be in the same order of magnitude as) the total number of documents in the index. We therefore believe that the variability we measure is largely, if not entirely attributable to the variability of the index of Google and Bing. In other words, what we are measuring is the genuine extrinsic variability of the indices, caused by changes (e.g., updates, upgrades, overhauls) of the indices. In Figure 1 we highlighted several publicly announced changes to both search engines' indices, many of which co-occur with drastic changes in index size as estimated by our method (20 out of the 24 reported changes in the Google index, and 6 out of 12 changes in Bing's index).

This variability, noted earlier also by Rousseau (1999), Bar-Ilan (1999), and Payne and Thelwall (2008), should be a cause for concern for any non-longitudinal study that adopts

[10] http://googlewebmastercentral.blogspot.com/.
[11] http://googleblog.blogspot.com/.
[12] http://blogs.bing.com/.
[13] http://searchenginewatch.com.
[14] A complete, numbered list of these events can be found at http://toinebogers.com/?page_id=757.

reported hit counts. It has been pointed out that "Googleology is bad science" (Kilgariff, 2007), meaning that commercial search engines exhibit variations in their functioning that do not naturally link to the corpus they claim to index. Indeed, it is highly unlikely that the real indexable Web suddenly increased from 20 to 30 billion pages in a matter of weeks in October 2014; yet, both the Bing and Google indices report a peak in that period. It is important to note, however, that the observed instability of hit counts does not automatically imply that measuring other properties of search engines for use in webometric research, such as result rankings or link structure, suffer from the same problem.

Our estimates do not show a monotonic growth of Web search engines' indices, which was one of the hypothesized outcomes at the onset of this study in 2006. The results could be taken to indicate that the indexed Web is not growing steadily the way it did in the late 1990s. They may even be taken to indicate the indexed Web is not growing at all. Part of this may relate to the growth of the unindexed Deep Web, and a move of certain content from the indexed to the Deep Web.

The unique perspective of our study is its longitude. Already in 1999, Rousseau remarked that collecting time series estimates should be an essential part of Internet research. The nine-year view visualized in Figure 1 shows that our estimation is highly variable. It is likely that other estimation approaches, e.g. using link structure or result rankings, would show similar variance if they were carried out longitudinally. Future work should include comparing the different estimation methods over time periods, at least of a few years. The sustainability of this experiment is non-trivial and should be planned carefully, including a continuous monitoring of the proper functioning. The scripts that ran our experiment for nearly nine years, and are still running, had to be adapted to changes in the web interfaces of Google and Bing repeatedly. The time required for adapting the scripts after the detection of a change caused the loss of 6-7% of all possible daily measurements.

Our approach, but also the different approaches discussed in the section on related research introduce different kinds of biases. We list here a number of possible biases and how they apply to our own approach:

Query bias. According to Bharat and Broder (1998), large, content-rich documents have a better chance of matching a query. Since our method of absolute size estimation relies on the hit counts returned by the search engines, it does not suffer from this bias, as the result pages themselves are not used.

Estimation bias. Our approach relies on search engines accurately reporting the genuine document frequencies of all query terms. However, modern search engines tend to not report the actual frequency, but instead estimate these counts, for several reasons. One such reason is their use of federated indices: a search engine's index is too large to be stored on one single server, so the index is typically divided over many different servers. Update lag or heavy load of some servers might prevent a search engine from being able to report accurate, up-to-date term counts. Another reason for inaccurate counts is that modern search engines tend to use document-at-a-time (DAAT) processing instead of term-at-a-time (TAAT) processing (Turtle & Flood, 1995). In TAAT processing the entire postings list is traversed for each query term in its entirety, disregarding relevant documents with each new trip down the postings list. In contrast, DAAT processing the postings list is traversed one document at a time for all query terms in parallel. As soon as a fixed number of relevant documents—say 1,000—are found, the traversal is stopped and the resulting relevant documents are returned to the user. The postings list is statically ranked before traversal (using measures such as PageRank) to ensure high quality relevant documents. Since DAAT ensures that, usually, the entire postings list does not have to be traversed, the term frequency counts tend to be incomplete. Therefore, the term frequencies are typically estimated from the section of the postings list that was traversed.

Malicious bias. According to Bharat and Broder (1998, p. 384), "a search engine might rarely or never serve pages that other engines have, thus completely sabotaging our approach". This unlikely scenario is not likely to influence our approach negatively. However, if search engines were to maliciously inflate the query term counts, this would seriously influence our method of estimating the absolute index sizes.

Domain bias. By using text corpora from a different domain to estimate the absolute index sizes, a domain bias can be introduced. Because of different terminology, term statistics collected from a corpus of newswire, for instance, would not be applicable for estimating term statistics in a corpus of plays by William Shakespeare or corpus of Web pages. We used a corpus of Web pages based on DMOZ, which should reduce the domain bias considerably. However, in general the pages that are added to DMOZ are of high quality, and are likely to have a higher-than-average PageRank, which might introduce some differences between our statistics and the ideal statistics.

Cut-off bias. Some search engines typically do not index all of the content of all web pages they crawl. Since representative information is often at the top of a page, partial indexing does not have adverse effect on search engine performance. However, this cut-off bias could affect our term estimation approach, since our training corpus contains the full texts for each document. Estimating term statistics from, say, the top 5 KB of a document can have a different effect than estimating the statistics from the entire document. Unfortunately, it is impractical to figure out what cut-off point the investigated search engines use so as to replicate this effect on our training corpus.

Quality bias. DMOZ represents a selection of exemplary, manually selected web pages, while it is obvious that the web at large is not of the same average quality. Herein lies a bias of our approach. Some aspects of the less representative parts of the web have been identified in other work. According to Fetterly et al. (2005), around 33% of all Web pages are duplicates of one another. In addition, in the past about 8% of the WWW was made up of spam pages (Fetterly et al., 2004). If this is all still the case, this would imply that over 40% of the Web does not show the quality nor the variation present in the DMOZ training corpus.

Language bias. Our selection of words from DMOZ are evenly spread over the frequency continuum and show that DMOZ is biased towards the English language, perhaps more than the World Wide Web at large. A bias towards English may imply an underestimation of the number of pages in other languages, such as Mandarin or Spanish.

This exploratory study opens up at least the following avenue for future research that we intend to pursue. We have tacitly assumed that a random selection of DMOZ pages represents "all languages". With the proper language identification tools, by which we can identify a proper DMOZ subset of pages in a particular language, our method allows to focus on that language. This may well produce an estimate of the number of pages available on the Web in that language. Estimations for Dutch produce numbers close to two billion Web pages[ii]. Knowing how much data is available for a particular language, based on a seed corpus, is relevant background information for language engineering research and development that uses the web as a corpus (Kilgariff & Grefenstette, 2003).

References

Almind, T.C. & Ingwersen, P. (1997). Informetric analyses on the World Wide Web: Methodological approaches to 'Webometrics'. *Journal of Documentation*, *53*, 404–426.

Anagnostopoulos, A., Broder, A. & Carmel, D. (2006). Sampling search-engine results. In *Proceedings of WWW '06*, (pp. 397–429). New York, NY, USA: ACM Press.

Bar-Ilan, J. (1999). Search engine results over time: a case study on search engine stability. *Cybermetrics*, *2*, 1.

Bar-Ilan, J. (2004). The use of Web search engines in information science research. *Annual Review of Information Science and Technology*, *38*, 231–288.

Bar-Ilan, J., Mat-Hassan, M. & Levene, M. (2006). Methods for comparing rankings of search engine results. *Computer Networks*, *50*, 1448–1463.

Bar-Yossef, Z. & Gurevich, M. (2006). Random sampling from a search engine's index. In *Proceedings of WWW '06* (pp. 367–376). New York, NY, USA: ACM Press.

Bar-Yossef, Z. & Gurevich, M. (2011). Efficient search engine measurements. *ACM Transactions on the Web*, *5*, 1–48.

Bharat, K. & Broder, A. (1998). A technique for measuring the relative size and overlap of public web search engines. In *Proceedings of WWW '98* (pp. 379–388). New York, NY, USA: ACM Press.

Björneborn, L. (2004). *Small-World Link Structures across an Academic Web Space: A Library and Information Science Approach*. PhD thesis, Royal School of Library and Information Science.

Dobra, A. & Fienberg, S.E. (2004). How large is the World Wide Web? In *Web Dynamics* (pp. 23–43). Berlin: Springer.

Fetterly, D., Manasse, M. & Najork, M. (2005). Detecting phrase-level duplication on the World Wide Web. In *Proceedings of SIGIR '05* (pp. 170–177). New York, NY, USA: ACM Press.

Fetterly, D., Manasse, M. & Najork, M. (2004). Spam, damn spam, and statistics: Using statistical analysis to locate spam Web pages. In *Proceedings of the 7th International Workshop on the Web and Databases: colocated with ACM SIGMOD/PODS 2004* (pp. 1–6).

Gulli, A. & Signorini, A. (2005). The indexable Web is more than 11.5 billion pages. In *Proceedings of WWW '05* (pp. 902–903). New York, NY, USA: ACM Press.

Henzinger, M., Heydon, A., Mitzenmacher, M. & Najork, M. (2000). On near-uniform URL sampling. *Computer Networks*, *33*, 295–308.

Hirate, Y., Kato, S. & Yamana, H. (2008). Web structure in 2005. In Aiello, W., Broder, A., Janssen, J. & Milios, E. (Eds.) *Algorithms and Models for the Web-Graph*, vol. 4936, Lecture Notes in Computer Science, (pp. 36–46). Berlin: Springer.

Khelghati, M., Hiemstra, D. & Van Keulen, M. (2012). Size estimation of non-cooperative data collections. In *Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services* (pp. 239–246). New York, NY, USA: ACM Press.

Kilgarriff, A. & Grefenstette, G. (2003). Introduction to the special issue on Web as corpus. *Computational Linguistics*, *29*.

Kilgarriff, A. (2007). Googleology is bad science. *Computational Linguistics*, *33*, 147–151.

Kleinberg, J.M., Kumari, R., Raghavan, P., Rajagopalan, S. & Tomkins, A.S. (1999). The Web as a Graph: Measurements, Models, and Methods. In *COCOON '99: Proceedings of the 5th Annual International Conference on Computing and Combinatorics* (pp. 1–17). Berlin: Springer.

Koehler, W. (2004). A longitudinal study of Web pages continued: a report after six years. *Information Research*, 9. http://www.informationr.net/ir/9-2/paper174.html.

Lawrence, S. & Giles, C.L. (1998). Searching the World Wide Web. *Science*, *280*, 98–100.

Lawrence, S. & Giles, C.L. (1999). Accessibility of Information on the Web. *Nature*, *400*, 107–109.

Lewandowski, D. & Höchstötter, N. (2008). Web searching: a quality measurement perspective. In Spink, A. & Zimmer, M. (Eds.) *Web Search*, vol. 14, Information Science and Knowledge Management (pp. 309–340). Berlin: Springer.

Payne, N. & Thelwall, M. (2008). Longitudinal trends in academic Web links. *Journal of Information Science*, *34*, 3–14.

Rousseau, R. (1999). Daily time Series of common single word searches in AltaVista and NorthernLight. *Cybermetrics*, *2*, 1.

Spink, A., Jansen, B.J., Kathuria, V. & Koshman, S. (2006). Overlap among major Web search engines. *Internet Research*, *16*, 419–426.

Thelwall, M. (2008). Quantitative comparisons of search engine results. *Journal of the American Society for Information Science and Technology*, *59*, 1702–1710.

Thelwall, M. (2009). Introduction to Webometrics: Quantitative Web Research for the Social Sciences. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 1, pp. 1–116.

Thelwall, M. & Sud, P. (2012). Webometric research with the Bing search API 2.0. *Journal of Informetrics*, *6*, 44–52.

Turtle, H. & Flood, J. (1995). Query evaluation: Strategies and optimizations. *Information Processing & Management*, *31*, 831–850.

Uyar, A. (2009). Investigation of the accuracy of search engine hit counts. *Journal of Information Science*, *35*, 469–480.

Zimmer, M. (2010). Web Search Studies: Multidisciplinary Perspectives on Web Search Engines. In Hunsinger, J., Klastrup, L. & Allen, M. (Eds.) *International Handbook of Internet Research* (pp. 507–521). Berlin: Springer.

Zipf, G.K. (1949). *Human Behaviour and the Principle of Least Effort*; Reading, MA: Addison-Wesley.

Online Attention of Universities in Finland: Are the Bigger Universities Bigger Online too?

Kim Holmberg[1]

[1] *kim.j.holmberg@utu.fi*
Research Unit for the Sociology of Education, University of Turku, 20014 Turku (Finland)

Abstract
As universities have entered a time of increased demand for public outreach and measurable impact, the universities are also exploring social media for student recruitment and science communication. Because many of the popular social media sites are free to use they could provide more democratic channels for organizational communication and marketing efforts. This research in progress investigates the social media presences of 14 universities in Finland and studies whether the offline performances of the universities are reflected in social media. The results suggest that while the RG score from ResearchGate and the Google Trends score for relative search volume correlate well with both productivity of the universities and university rankings, some of the other social media sites do not reflect the institutional characteristics as well. This is assumed to be a result of different types of usage and different purposes of the different social media sites.

Conference Topics
Webometrics; Altmetrics; Country-level studies

Introduction
Universities have entered a time of increased demand for public outreach and measurable impact. While competing for students the Humboldtian research universities try their best to conduct high quality research for the benefit of the society and to create a foundation for the research based education. At the same time social media has become mainstream in organizational communication (e.g., Badea, 2014; Huang, Baptista & Galliers, 2013; Lovejoy & Saxton, 2012). Organizations use social media for various purposes, both internally and externally, and for universities social media would seem to be an especially efficient tool for public outreach and for recruiting students. Social media are particularly efficient for sharing information through the online social networks, an aspect that would allow universities to efficiently reach their audiences. As the most popular social media sites are free to use, they may provide a more democratic way for universities to reach out to the various audiences and interest groups. This research in progress investigates whether this is true in the case of 14 Finnish universities: are smaller universities taking full advantage of the more democratic ways of communication or are the bigger universities with more resources also “bigger” in social media?

Literature review
Forkosh-Baruch and Hershkovitz (2012) investigated the use of social media sites Twitter and Facebook for scholarly purposes among higher education institutes in Israel. Their findings showed how the social media sites were extensively used for sharing academic or professional news. The authors suggest that use of these social media sites could therefore promote knowledge sharing and informal learning. Based on a content analysis of the messages shared in social media by the group of Israeli HEIs, the authors also discovered that the social media usage patterns followed similar offline usage patterns. The similar patterns here being the perception that colleges are more open and social, while universities tend to focus more on research and involvement in the research community; characteristics that were discovered in the content of the analyzed social media messages. Because of this lack of socializing and

interactivity among the universities, the authors conclude that "*the potential of SNS [social networking sites] as means of sharing academic knowledge in higher education institutes in Israel has not been actualized yet, but is indeed being explored by these organizations…*" With this the authors emphasize the importance of interactivity and audience involvement in organizational communication in social media.

In addition to social media visibility, interest towards universities, as measured by search volume on Google Trends, has also been discovered to have a connection with academic reputation (Vaughan & Romero-Frías, 2014). Vaughan and Romero-Frías (2014) used Google Trends to collect the relative search volume of the top 50 universities in the QS ranking from the US and the 56 Spanish universities included in the ARWU ranking. Their findings indicate that highly ranked universities attracted also more attention, as measured by search volume. In Google Trends the results can also be focused on searches within specific countries; one could for instance look up the search volume for "Kate Upton" in the UK or "Justin Bieber" in Norway. Vaughan and Romero-Frías (2014) discovered that while a great amount of searches for the US universities came from outside the US, only a few searches for the Spanish universities came outside of Spain, which according to the authors also reflects the international positions of the two sets of universities. As searches in English in general and for universities in English in particular may be assumed to be relatively low in non-English speaking countries, it may not make sense in all cases to focus on the country-level search volume in English. For instance in the case of Finnish universities we can assume that searches for them from Finland would mainly use their Finnish names, while the volume of searches in English would mainly reflect the international attention and interest.

Thelwall and Kousha (in press) took another approach to study universities' online presences and investigated whether the usage of ResearchGate and the publications uploaded to it by researchers has a connection with the "academic hierarchies" of different university rankings. ResearchGate is a scholarly social networking site where scholars can create their own profile pages and upload their publications to it, network with other researchers, and find possibly relevant and interesting publications, based partly on their own interests (as indicated on their profile pages) and partly on the interests of those in their social network. Based on researchers' activity on the site and their publications (both number of publications and the journal impact factor of the journals where the papers have been published in) ResearchGate calculates RG scores as a measure of individual researchers' "scientific reputation". The exact formula with which the RG score is calculated is, however, not revealed by ResearchGate. This approach can also be criticized because use of journal impact factors to evaluate or rank individual researchers has increasingly been criticized and condemned (e.g., DORA, 2013). Collectively the RG scores for researchers from a specific institution can give an institutional RG score, supposedly indicating institutional reputation. This is the score that Thelwall and Kousha (in press) used to compare to different university rankings. Their findings showed a moderate correlation between the rankings on ResearchGate and the other university rankings (The Higher Education ranking, QS world university rankings, Academic Ranking of World Universities, CWTS Leiden ranking, and the ranking on Webometrics.info). Because the rankings on ResearchGate are based on researchers' activities on the site and their research work, the findings by Thelwall and Kousha (in press) suggest that the usage of ResearchGate "broadly reflects traditional academic capital."

The current university rankings do place somewhat different weight on different things. For instance the ranking provided by the Webometrics.info measure online visibility, presence and impact, weighting most on visibility as measured by hyperlinks, while the other rankings use more traditional measures of research productivity and impact, i.e. publications and citations, and give them different weights (Aguillo, Bar-Ilan, Levene & Ortega, 2010). Still the different university rankings tend to give similar results, which would suggest that universities performing well in one area also perform well on other areas. In other words, a university that is performing well when assessed with publications and citations seems to also perform well online. But whether this is reflected to the universities usage of social media and the attention they receive there is unclear. Attention and visibility in social media, as measured with various social media metrics, has been suggested to be a potential indicator of research impact (e.g., Bollen, Van De Sompel, Hagberg & Chute, 2009; Priem & Hemminger, 2010; Lin & Fenner, 2013). These new social media metrics, the so called altmetrics, could potentially give a more nuanced view of the attention towards research outputs. It has also been suggested that altmetrics could provide indicators for the societal impact of research (Bornmann, 2014) or provide knowledge about the interest towards research from a wider audience outside academia (Haustein, 2014). Although not yet extensively studied, altmetrics may also be able to provide country-level indicators of research impact, as Alhoori et al. (2014) have discovered significant correlations between bibliometric data and some altmetrics when aggregated to the country-level.

The research in progress presented here investigates the social media presence of 14 universities in Finland and with that opens research for institutional altmetrics.

Data and methods

The 14 universities in Finland all have online presences in social media. All have profiles, pages or groups on the most popular social media sites Facebook, Twitter, YouTube and LinkedIn, and some also have accounts on Instagram, Flickr or Pinterest. These are usually linked to from the university's webpage. The goal of this research is to 1) study how universities are using social media, 2) how much attention they have attracted, and 3) whether this attention is connected to other offline descriptive metrics about the universities' resources and performance.

Descriptive statistics were manually collected by visiting the universities' official social media profiles, as linked to from the universities' websites. The data consists of the number of tweets, followers and following on Twitter, "likes" on Facebook, subscriptions to and views on the universities' YouTube channel, followers on LinkedIn, and the universities RG score on ResearchGate. In addition to this universities' relative search volumes, as indicated by Google Trends, were retrieved. As the Google Trends score is a score relative to the search volume of the other words searched at the same time (maximum of five different terms compared in one search), we retrieved the scores for the universities' names in English by keeping the two universities with the highest scores included in the search for reference. This way all the scores were relative to those universities with the biggest search volume. The descriptive data about the universities and their performance were retrieved from the report of the State of Scientific Research in Finland, commissioned by the Academy of Finland (http://www.aka.fi/en-GB/A/Decisions-and-impacts/The-state-of-scientific-research-in-Finland/). This performance data consists of variables from 2012; the number of PhDs awarded, total person-years of the teaching and research staff, research funding, and number of publications. In addition to these the rankings of the Finnish universities were retrieved from the following university rankings; CWTS Leiden, ARWU, QS, THE, and Webometrics.info. Only Webometrics.info could provide the rankings for all but one of the 14 universities: the ranking of the fairly new University of the Arts (the former Academy of Fine Arts, Sibelius Academy and Theatre Academy merged to the University of the Arts in 2013). Nine of the 14 universities were found on QS ranking, seven on the CWTS ranking and on THE ranking, and five on the ARWU ranking. Only rankings from Webometrics.info and the QS were used in further analysis.

Spearman rank correlations between the social media metrics and offline data about the universities' performance were investigated to discover whether social media usage would follow the academic capital at these universities. In addition to this, connections between the social media metrics and university rankings were also tested to see whether the universities reputation and performance was reflected in social media attention and usage.

Results

The different offline university specific metrics are clearly associated, showing how number of students and faculty, funding and publications are all very tightly connected (Table 1). This naturally means that universities with more funding have bigger faculty, more students and produce more publications. As some of these metrics are also used for university rankings it is only natural that the rankings correlate well with these (0.830, n=13, between publications and Webometrics.info; 0.867, n=9, between publications and QS ranking, both Spearman rank correlations significant at level 0.05). The universities that were omitted from the analysis due to non-existent data on Webometrics.info and QS were the universities with the least publications, a probable explanation why they were not covered by the university rankings.

Table 1. Spearman rank correlations between the social media metrics and offline metrics of the 14 universities in Finland. Correlations in bold are significant at the 0.05 level, two-tailed. (RG = RG score; GT = Google Trends score; Tw = Tweets in Twitter; Tw.a = Followers on Twitter; Tw.b = Following on Twitter; FB = Facebook likes; YTs = YouTube subscriptions; YTv = YouTube views; LI = LinkedIn followers; Phd. = PhDs awarded in 2012; Fa. = Faculty in 2012; Fu. = Research funding in 2012; Pu. = Peer-reviewed publications in 2012).

	RG	GT	Tw	Tw.a	Tw.b	FB	YTs	YTv	LI	PhD.	Fa.	Fu.	Pu.
RG	**1**	**0,679**	0,473	0,367	0,046	0,389	0,337	0,204	0,385	**0,923**	**0,938**	**0,952**	**0,969**
GT		**1**	0,444	0,435	0,251	0,266	0,316	0,342	0,160	**0,750**	**0,690**	**0,648**	**0,746**
Tw			**1**	**0,776**	0,516	0,345	**0,579**	**0,587**	**0,618**	**0,670**	**0,604**	0,534	**0,543**
Tw.a				**1**	0,499	0,059	**0,557**	**0,613**	**0,749**	**0,551**	0,468	0,393	0,420
Tw.b					**1**	-0,099	0,233	0,314	0,196	0,192	0,143	0,064	0,116
FB						**1**	0,260	0,015	-0,178	0,463	**0,574**	**0,604**	0,389
YTs							**1**	**0,871**	**0,700**	0,397	0,414	0,392	0,317
YTv								**1**	**0,754**	0,333	0,266	0,231	0,284
LI									**1**	0,423	0,349	0,323	0,380
PhD.										**1**	**0,974**	**0,949**	**0,960**
Fa.											**1**	**0,987**	**0,947**
Fu.												**1**	**0,943**
Pu.													**1**

Overall the number of tweets and Facebook 'likes' correlated moderately with the performance metrics of universities (Table 1), with tweets giving somewhat higher correlations on average than Facebook. While the number of followers on Twitter had some connection to the offline metrics, the number of followed accounts only had a very weak connection. This suggests that larger universities are not necessarily more active on Twitter, but that they still generate more attention.

Our findings indicate that research productivity (and the other offline metrics), as measured by the number of peer-reviewed publication from 2012, did correlate almost perfectly with the RG score on ResearchGate (0.969 Spearman, significant at the 0.05 level). The RG score did, however, not correlate well with many of the other social media metrics. Search volume on Google Trends also correlated well with the offline metrics, with the Spearman rank correlation between Google Trends score and number of publications being 0.746, significant at 0.05 level. The relationships of these two cases are illustrated in figures 1 and 2. In both cases the University of Helsinki, the largest university in Finland, appear as an outlier due to its size. In figure 2 we can see a bit more scattering and how the University of Jyväskylä, and to some extent University of Eastern Finland and Aalto University, although not having exceptionally many publications still have managed to attract significant interest as measured by search volume on Google.

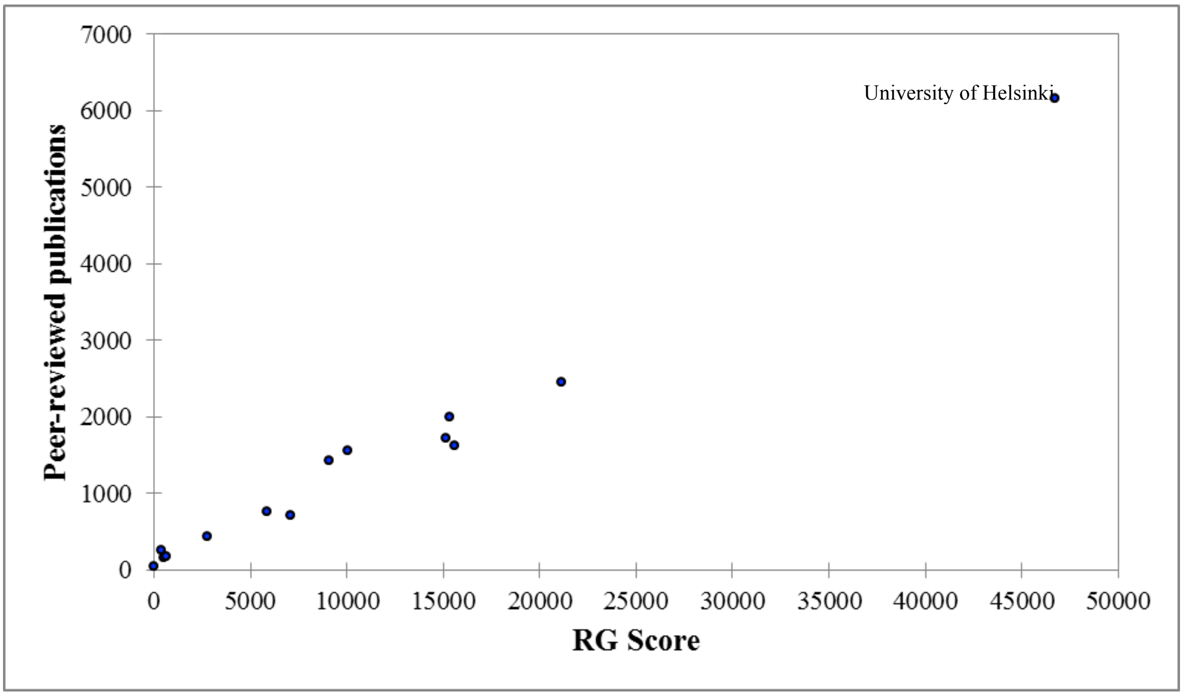


Figure 1. Correlation between the RG score (from ResearchGate) and the number of peer reviewed publications in 2012 at the Finnish universities (0.969 Spearman).

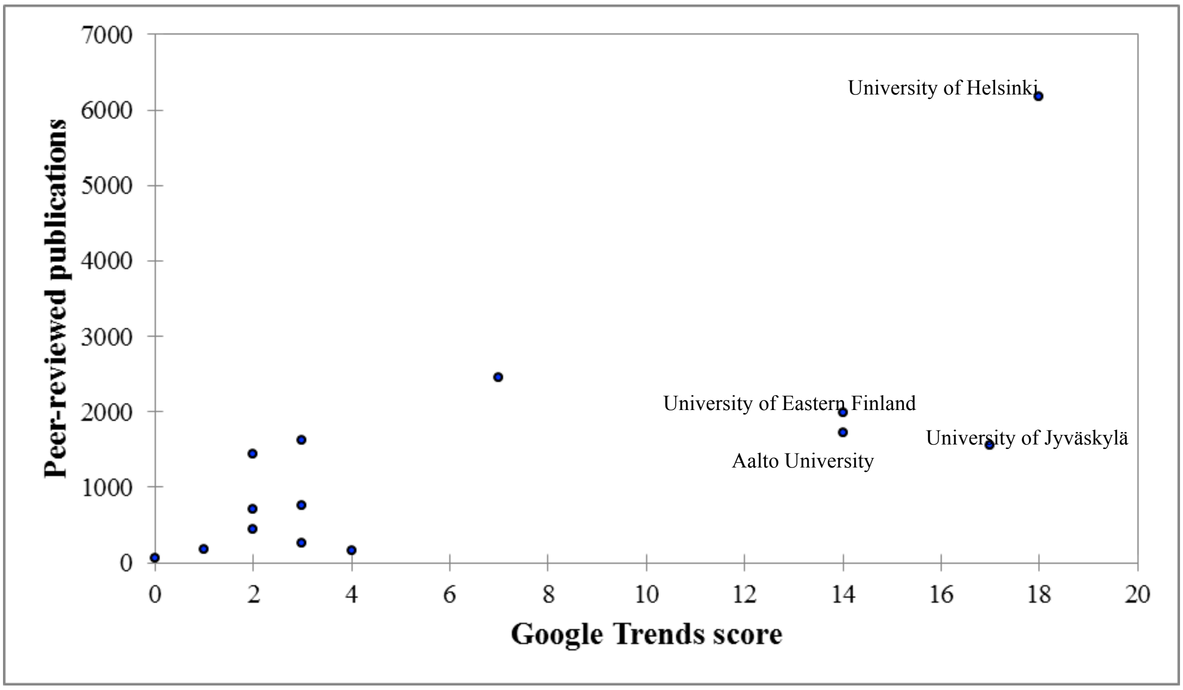


Figure 2. Correlation between the search volume as measured by Google Trends and the number of peer reviewed publications in 2012 at the Finnish universities (0.746 Spearman).

Discussion and conclusions

We set out to investigate the social media presences of 14 universities from Finland and the attention they have received in social media. Our results show that while in many cases the larger and more productive universities are also more active or receive more attention in social media; this is not always the case (Table 1). This suggests that the smaller universities, at least in this small sample, are benefitting from the more democratic channels of social media. Our findings also suggest, in line with the findings by Thelwall and Kousha (in press), that the institutional RG scores and the RG scores for individual researchers on ResearchGate, may be a promising source for altmetrics at institutional and possibly even country level. Due to the uncertainty of how the RG score exactly is calculated and because of the use of journal impact factors in that calculation more research into the topic is clearly needed.

The next step of this research in progress will be a content analysis of the universities social media accounts. This will provide new knowledge about how the universities are represented in social media, for what purposes they use social media, and how attention in social media is created. This will provide important background information for institutional altmetrics.

References

Aguillo, I., Bar-Ilan, J., Levene, M. & Ortega, J.L. (2010). Comparing university rankings. *Scientometrics*, *85*, 243-256. DOI: 10.1007/s11192-010-0190-z

Alhoori, H., Furuta, R., Tabet, M., Samaka, M. & Fox, E.F. (2014). Altmetrics for country-level research assessment. *Lecture Notes in Computer Science*, *8839*, 59-64. DOI: 10.1007/978-3-319-12823-8_7

Badea, M. (2014). Social media and organizational communication. *Procedia Social and Behavioral Sciences*, *149*, 70-75. DOI: 10.1016/j.sbspro.2014.08.192

Bollen, J., Van De Sompel, H., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures, *PLoS One*, *4*(6). DOI: 10.1371/journal.pone.0006022.

Bornmann, L. (2014). Validity of altmetrics data for measuring societal impact: a study using data from Altmetric and F1000 Prime. *Journal of Informetrics, 8*(4), 935-950. DOI: 10.1016/j.joi.2014.09.007

DORA (2013). San Francisco Declaration on Research Assessment. Retrieved on April 15, 2015, from http://www.ascb.org/dora-old/files/SFDeclarationFINAL.pdf.

Forkosh-Baruch, A. & Hershkovitz, A. (2012). A case study of Israeli higher-education institutes sharing scholalrly information with the community via social networks. *Internet and Higher Education*, *15*, 58-68. DOI: 10.1016/j.iheduc.2011.08.003

Haustein, S., Peters, I., Sugimoto, C.R., Thelwall, M. & Lariviére, 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

Huang, J., Baptista, J. & Galliers, R.D. (2013). Reconceptualizing rhetorical practices in organizations: The impact of social media on internal communications. *Information & Management*, *50*(2-3), 112-124. DOI: 10.1016/j.im.2012.11.003

Lin, J. & Fenner, M. (2013). The many faces of article-level metrics. *Bulletin of the Association for Information Science and Technology*, *39*(4). DOI: 10.1002/bult.2013.1720390409

Lovejoy, K. & Saxton, G.D. (2012). Information, community, and action: How nonprofit organizations use social media. *Journal of Computer-Mediated Communication*, *17*(3), 337-353. DOI: 10.1111/j.1083-6101.2012.01576.x

Priem, J., & Hemminger, B. M. (2010). Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web. First Monday, vol. 15, no. 7. Retrieved on January 20, 2015 from http://firstmonday.org/article/viewArticle/2874/2570.

Thelwall, M. & Kousha, K. (in press). ResearchGate: disseminating, communicating and measuring scholarship? *Journal of the Association for Information Science and Technology*. DOI: 10.1002/asi.23236.

Ranking Journals Using Altmetrics

Tamar V. Loach[1,2] and Tim S Evans[2]

[1] *t.loach @ digital-science.com*
Digital Science, 2 Trematon Walk, Wharfdale Road, London, N1 9FN (U.K.)

[2] *{t.loach, t.evans} @ imperial.ac.uk*
Imperial College London, Centre for Complexity Science, South Kensington Campus, London SW7 2AZ (U.K.)

Abstract
The rank of a journal based on simple citation information is a popular measure. The simplicity and availability of rankings such as Impact Factor, Eigenfactor and SciMago Journal Rank based on trusted commercial sources ensures their widespread use for many important tasks despite the well-known limitations of such rankings. In this paper we look at an alternative approach based on information on papers from social and mainstream media sources. Our data comes from altmetric.com who identify mentions of individual academic papers in sources such as Twitter, Facebook, blogs and news outlets. We consider several different methods to produce a ranking of journals from such data. We show that most (but not all) schemes produce results, which are roughly similar, suggesting that there is a basic consistency between social media based approaches and traditional citation based methods. Most ranking schemes applied to one data set produce relatively little variation and we suggest this provides a measure of the uncertainty in any journal rating. The differences we find between data sources also shows they are capturing different aspects of journal impact. We conclude a small number of such ratings will provide the best information on journal impact.

Conference Topic
Altmetrics

The background and purpose of the study
Journal metrics, such as the Thomson Reuters Journal Impact Factor, were originally developed in response to a publisher need to demonstrate the academic attention accorded to research journals. Over the intervening 50 years since Garfield's work in the field, the Impact Factor and other metrics, such as Eigenfactor (Bergstrom, 2007), have been used and misused in a variety of contexts in academia. An oft-discussed perception is that a journal-level metric is a good proxy for the quality of the articles contained in a journal.

In the evaluation and bibliometrics communities citation counting is generally understood not to be an appropriate proxy for quality but rather a measure of attention. The type of attention being measured in this case is quite specific and has particular properties. What is being measured is the attention to a paper of peers in related fields. The bar for registration of this attention is relatively high – the researcher or researchers making the citation must deem the target article to be of sufficient value that they include a citation in a work of their own that in turn is deemed publishable (e.g. see Archambault & Lariviére, 2009, and references therein). The timescale associated with citations is also long – typically being limited by the review and publication process associated with particular fields. Additionally, it is accepted that journal-level metrics say little regarding the merit of particular articles in the journal since journal-level metrics are often calculated based on thousands of articles and are often biased by the performance of the tails of the distribution of citations. These realisations have led to the recent growth in popularity of article-level metrics or altmetrics.

Altmetrics have broadened the range of types of attention that we can measure and track for scholarly articles. Mostly based in social and traditional media citations, the altmetric landscape is one that is constantly changing with the introduction of different data sources all the time. While, one the one hand, altmetrics suffer from all the unevenness of traditional citations, they occur over different timescales, which provides us with a more nuanced view of the lifecycle of a scholarly work. Aggregating alternative metrics at a journal level will complement Journal Impact Factor, giving us new insights into different facets of attention.

Traditional citation-based metrics are difficult to calculate since they are based on the bibliometric journal databases, such as Thomson Reuters' Web of Science. Conversely, Altmetrics are conglomerates of disparate sources of references to research output derived from non-traditional sources, primarily modern electronic sources characterised by fast response times (see Bornmann, 2014, for a recent overview). The lack of any systematic peer review is another characteristic of most altmetric data. The open and electronic nature of much altmetric data offers the prospect of alternative paper and journal metrics, which may be more accessible to stakeholders. The rapid response of such data to innovations suggests such metrics might offer improvements over metrics based on slower traditional sources.

This paper considers a number of approaches to the aggregation of altmetric data in order to create a robust journal-level metric that complements the existing citation-based metrics already in use across the academic community. The aim is not to create a contender for a single metric to quantify journal output but instead to create a useful measure that gives "the user" a sense of the non-citation attention that a journal attracts in the same way that Journal Impact Factor, Eigenfactor and other related metrics give this sense for citation attention.

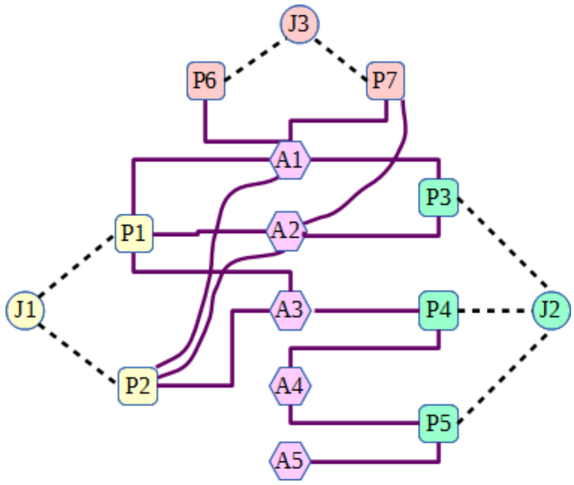


Figure 1. The relationships recorded in our altmetric.com data. The raw data illustrated here contains fifteen "mentions" (solid lines) by five "authors" (hexagons A1 to A5) of seven papers (squares P1 to P7). We also know the journal (circles), which published a paper (dashed lines).

Data Sources

In this paper we use the 2013 IF (Impact Factor) and EF (Eigenfactor) as examples of traditional sources of journal ratings. Our altmetric data comes from 20 months of data from altmetric.com, a commercial company. For each mention about a paper we had the journal in which it was published, the source (twitter, Facebook, etc.) and the account (here termed an 'author'), as shown in Figure 1. In our case, a 'paper' has to be an article coming from a known journal. A single 'author' for us is a single account (e.g. one twitter account) or a single source (a news outlet such as a newspaper). In some cases several different authors may be responsible for one site or one author could provide information to many different sites or accounts (a twitter account, a facebook account, a blog, etc) but in our data such an author appears as many distinct authors.

Methods

The simplest type of journal altmetric is one based on basic counts where each mention of a paper in a journal adds one to that journal's count. We collected counts for social media '*sbc*', non-social media '*nsbc*' (e.g. downloads) and combined scores '*bc*' (for blind count i.e. with no weighting for different sources). We also obtained the current journal rating produced by altmetric.com (denoted '*ca*'), which is a weighted count rating in which different sources are given different weights (blogs and news sources get highest weighting).

Network Definitions

A criticism of simple count based methods, such as Impact Factor or our altmetric counts discussed above, is that some citations or some altmetric authors are more important than others. Eigenfactor is an illustration of a response to these criticisms in the realm of traditional data (Bergstrom, 2007), as it uses a network based view to arrive at a PageRank style measure. We will also turn to a network-based view in order to look at a wide range of measures, which probe the relationships between journals on a much larger scale.

There are many possible network representations of our data. In this paper we will focus only on networks in which the nodes represent journals. The central idea in our construction of the relationship between two journals is that we only want to consider activity from authors who mention both journals because only these authors are making an implicit comparison between journals. The activity of each author is used to define their own "field of interest" in a self-consistent manner and so the activity of authors is used to make comparisons between journals in the same field as defined by each author's interests. This ensures that at a fundamental level we avoid the much discussed problem of making comparisons between papers or journals from different fields. An author only interested in medical issues will only contribute to the evaluation of Nature, Science and so forth in terms of their interest in these multidisciplinary journals relative to Cell or other specialised journals.

A useful analogy here is that each journal is a team and an author who mentions articles published in two journals represents one game between these journals – our pairwise comparison. The score in each game is the number of mentions so in comparing two journals j and l, the score for journal j from the game represented by author a is recoded as the entry J_{ja}.in a rectangular matrix. In Figure 1 the game between J1 and J2 represented by author A2 has the result 2-1, a 'win' for journal 1 over journal J2 suggesting that we should rate journal J1 more highly than journal J2 given the activity of this one author.

We shall consider three different ways of quantifying the journal relationships, the network edges. Our first approach gives us an adjacency matrix S where the entry S_{jl} gives the weight of the edge from journal j to journal l, and this is given by $S_{jl} = \frac{1}{|A_{jl}|}\sum_{a\in A_{jl}} J_{ja}$, where $A_{jl} = \{a|J_{ja} > 0, J_{la} > 0\}$. Here j and l represent different journals and a is one author. J_{ja} is a matrix, which is equal to the number of papers mentioned by author a which were published in journal j. The expression for S_{jl} is counting the number of times papers published in journal j are mentioned by authors who also mention papers in journal l, with the total normalised by the number of such authors. Note that this defines a sparse, weighted and directed network. In our conventions if journal j is better than journal l we will have $S_{jl} > S_{lj}$.

Our second definition gives us an adjacency matrix P where $P_{jl} = \frac{1}{|A_{jl}|}\sum_{a\in A_{jl}} \theta(J_{ja} - J_{la})$. Here $\theta(x) = 1$ if $x > 0$ otherwise this function gives 0. This definition counts how many authors mention more papers in journal j than they do papers in journal l., normalising again by the number of authors who are able to make this pairwise comparison. Again $P_{jl} > P_{lj}$ if journal j is better than journal l.

Finally we define an adjacency matrix Q where $Q_{jl} = \frac{1}{|A_{jl}|}\sum_{a\in A_{jl}} \Theta(J_{ja} - J_{la})$. Here $\Theta(x) = 1$ if $x > 0$, $\Theta(0) = 0.5$ while for negative values this function gives 0. This definition counts how many authors mention more papers in journal j than they do papers in journal l

but when this is balanced gives an equal weighting to both side. This definition has the useful property that $Q_{jl} + Q_{lj} = 1$ (not generally true for matrix P).

Network Measures

Once we have our network with journals as nodes, we need to find ways to use this structure to define which nodes are the most important. Measures which quantify the importance of a node are known as centrality measures in social network analysis. Unfortunately, many standard measures do not take into account the weights or directions of edges, both of which carry crucial information in our case. We used two well-known network centrality measures to illustrate our approach: PageRank and HITS (e.g. see Langville & Meyer, 2012). Both may be cast as eigenvector problems and there are fast algorithms for large networks which are readily available. We apply these two methods to all three networks, giving six different ratings e.g. 'qpr' indicates a PageRank rating derived from a Q matrix while 'ph' indicates a HITs rating derived using a P matrix.

We also tried a different type of measure known as Points Spread Rating (denoted 'psr') (p.117-120, Langville & Meyer, 2012) where the rating r_j for journal j is $r_j = \sum_l (S_{jl} - S_{lj}) / n_j$, (similarly for the P and Q matrices) and n_j is the number of journals. This expression ensures that the differences $(r_j - r_l)$ in the rating of any two journals j and l are as close as possible to the actual differences in the number of average mentions of papers.

Comparing Ratings

Once we have obtained different ratings, the final task is to make a comparison. The simplest approach is to make a qualitative comparison of the top ranked journals in each case. For a more quantitative approach we used standard methods of multivariate statistics. First we found a correlation matrix whose entries express the similarity of two rating methods: the Pearson correlation matrix based on the numerical values of the ratings obtained, Spearman's matrix which based on the ranking of journals, and finally Kendall's tau. These were analysed using principle component analysis or hierarchical clustering methods.

Findings

In terms of the altmetric data we found typical fat-tailed distributions, both for the number of mentions of a paper from different sources and in terms of the number of mentions put out by a single author. Some sources, such as twitter, are significantly larger than others.

When comparing different journal rating schemes, some results were found only with Spearman and Kendall tau correlation measures (which are based on the ranks of journals). The Pearson measure (based on actual rating values gave slightly different results in some cases. However in most cases there good agreement. Some typical results are shown in Figure 2 and numbers for ranking schemes in the following text refer to the labels in Figure 2.

The variation between different rating schemes for the same altmetric data source gives relatively little variation, roughly on the same scale as the difference we find between IF and EF. The four different methods shown for ratings based on Facebook mentions (6,12,16,19) are a typical example. Clearly our Points Spread Rating scheme (psr, 21,22,23) and our simple counts of non-social media mentions (nsbc, 6) produces outliers.

Some sources, such as Facebook and News, were also noticeably different from IF and EF, but the difference was much smaller than that found with the psr rating. One source, which gave ratings well correlated with IF and EF was blogs (8, 11, 15, 18).

Likewise, most of our simple count based ratings were just as close to IF (3) or EF (5) as these two rating schemes were to each other. This includes our unweighted count of all mentions (bc, 1), the number of times papers are mentioned (pc, 7), counts of just social media mentions (sbc, 14), and in particular the more sophisticated weighted journal ranking produced by altmetric.com (ca, 2).

Most of our work focused on statistics for the whole collection. A look at the top journals, see Table 1, confirmed that at an individual level our new altmetric network ratings were giving sensible results, but with variations which indicate the uncertainty in such rankings.

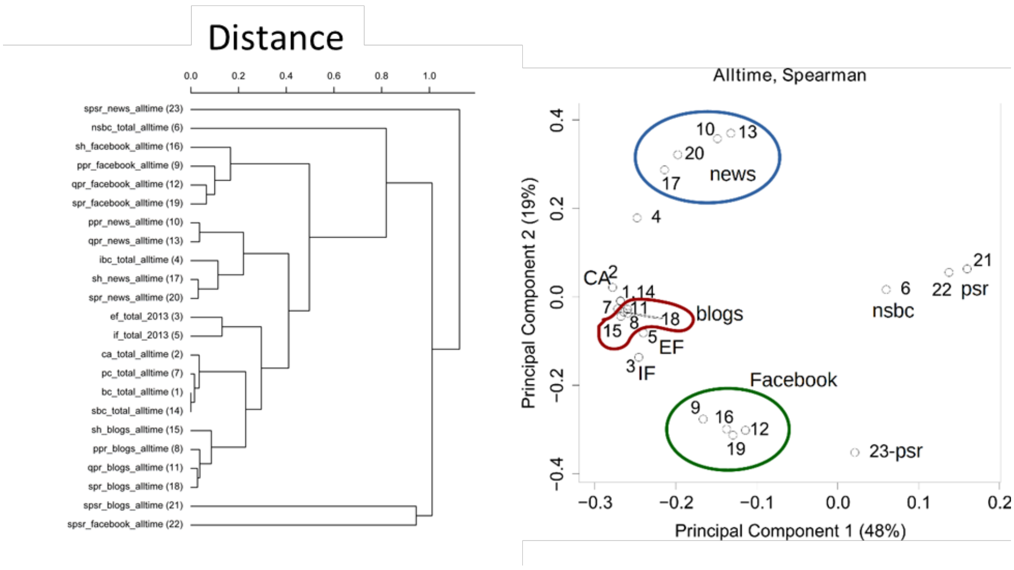


Figure 2. A comparison of some of the different ranking schemes using a Spearman correlation matrix. On the left a dendrogram and on the right a scatter plot using the first two principle components of PCA. For clarity, only a limited subset of our ratings were used in these plots.

Discussion

Given our differences between ranking based comparisons (Spearman and Kendall Tau) and results based on Pearson correlation matrices, this suggests that ratings are dominated by the measurement of the few journals, which have most of the mentions (fat tails). This is one reason we favour Spearman correlation matrices in Figure 2 and would suggest this makes sense in most journal ranking contexts.

Our Points Spread Rating scheme (psr, 21, 22, 23) seems to be reflecting very different patterns in the data from those found using other approaches. Given that the other approaches include Impact Factor, widely accepted as a measure of journal attention, we think it is hard to see a role for PSR to rank journals. Likewise, the simple blind counts of non-social media mentions (nsbc, 6) does not appear to be useful.

The remaining different altmetric sources and rating methods do show enough similarity to suggest that they are all an acceptable measure of journal importance. At the same time there are some interesting differences indicating that our altmetric based schemes are capturing different features of the impact of journals. At the very least this diversity will indicate the level of uncertainty in rating schemes. Two possible reasons for the close correlation of blogs and IF are as follows. Perhaps papers in high IF journals are of intrinsic interest to blog writers. Alternatively blog authors may read a limited number of journals but these tend to be those with high IF. Probably both factors are important, each reinforcing the other to produce the strong correlation we find.

Another interesting feature is that most of our simple count based ratings, which are not normalised by the number of articles per journal, are also well correlated with IF (3) which does use normalised counts. This can be explained if there is a correlation between the number of papers in a journal and its impact, something we can see in of count of number of papers (pc, 7). We will be looking at normalised altmetric counts in the future but it appears normalisation may not be essential. In particular, we note the altmetric.com journal rating (ca, 2) is well correlated and so provides a good handle on the impact of journals.

Table 1. Top ten journals based on various network based altmetric measures.

Rank	*Q, HITS, Blogs*	*Q, HITS, News*	*S, PageRank, Google+*
1	Nature	Nature	Nature
2	PNAS	PNAS	PLoS ONE
3	Science	PLoS ONE	Science
4	PLoS ONE	Science	PNAS
5	New England J. of Med.	New England J. of Med.	New England J. of Med.
6	British Medical J.-C.R.Ed.	British Medical J.-C.R.Ed.	British Medical J.-C.R.Ed.
7	The Lancet (British Ed.)	Nature Communications	Scientific Reports
8	JAMA	JAMA	JAMA
9	Proc. Royal Soc. B:	The Lancet (British Ed.)	The Lancet (British Ed.)
10	Current Biology	Pediatrics	PLoS Biology

The fact that we tried many different rating methods and that (with the exception of psr based measures) they showed variations on scales no bigger than those found between IF and EF, suggests that no one method is optimal in any sense. However we can use such a suite of metrics to get a handle on the uncertainty associated with any measure. This would be of great utility for users and a contrast to the three decimal point 'accuracy' associated with IF results.

Conclusions

We have shown how to use altmetric data to provide a reasonable journal ranking. Most types of altmetric data appear to give useful information in the sense that the correlation with IF is acceptable. At the same time altmetric data can be sufficiently different that it might reflect different types of impact. Our results suggest that different rating methods can provide a measure of the uncertainty of any journal ranking. Confirming these patterns over longer periods and producing a better understanding of the social reasons for the patterns we have found are future directions for our work. It would also be interesting to compare our results with journal attention measures derived from journal usage patterns, see for example Bollen et al 2009, an aspect not included in our data.

Acknowledgments

We would like to thank Euan Adie (altmetric.com) for providing us with the altmetric data, and for useful discussions along with Jonathan Adams and Daniel Hook (Digital Science).

References

Archambault, É., & Lariviére, V. (2009). History of the journal impact factor: Contingencies and consequences. Scientometrics, *79*(3), 635–649.

Bergstrom, C. (2007). Eigenfactor: Measuring the value and prestige of scholarly journals. *College and Research Libraries News*, *68*, 314-316.

Bollen, J., Van de Sompel, H., Hagberg, A. & Chute, R. (2009). A Principal Component Analysis of 39 Scientific Impact Measures. *PLoS ONE, 4*, e6022.

Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. *Journal of Informetrics*, *8*, 895-903.

Langville, A. N. & Meyer, C. D. (2012) *Who's #1?: The science of rating and ranking*, Princeton: Princeton University Press.

Who Tweets about Science?

Andrew Tsou, Tim Bowman, Ali Ghazinejad, and Cassidy Sugimoto

atsou@umail.iu.edu, ali.ghazinejad@gmail.com, sugimoto@indiana.edu, tim.bowman@gmail.com
School of Informatics and Computing, Indiana University Bloomington (Bloomington, IN, USA)

Abstract
Twitter is currently one of the primary venues for online information dissemination. Although its detractors portray it as nothing more than an exercise in narcissism and banality, Twitter is also used to share news stories and other information that may be of interest to a person's followers. The current study sampled tweeters who had tweeted at least one link to an article in one of four leading journals, with a focus on studying who, precisely, these tweeters were. The results showed that approximately 76% of the sampled accounts were maintained by individuals (rather than organizations), 67% of these accounts were maintained by a single man, and 34.4% of the individuals were identified as possessing a Ph.D, suggesting that the population of Twitter users who tweet links to academic articles does not reflect the demographics of the general public. In addition, the vast majority of students and academics were associated with some form of science, indicating that interest in scientific journals is limited to individuals in related fields of study.

Conference Topic
Altmetrics

Introduction

Twitter is currently one of the primary venues for online information dissemination. Nearly a quarter of adult Internet-users take advantage of Twitter (Pew Research, 2014), and according to Alexa (2015), as of January 22, 2015, Twitter is ranked as the 8th most visited site on the Web (and the 7th most visited in the United States). Although its detractors portray it as nothing more than an exercise in narcissism and banality, Twitter is also used to share news stories and other information that may be of interest to a person's followers. Amidst much vapidity can be found discussions or links of genuine merit, and indeed, it has been found that "academic articles are now frequently tweeted and so Twitter seems to be a useful tool for scholars to use to help keep up with publications and discussions in their fields" (Thelwall, Tsou, Weingart, Holmberg, & Haustein, 2013, p. 1). Previous research has discussed the content of such tweets, their sentiments (Thelwall, Tsou, Weingart, Holmberg, & Haustein, 2013), tweeting behaviour across venues and disciplines (Haustein, Peters, Sugimoto, Thelwall, & Larivière, 2014), the use of Twitter for altmetrics (Thelwall, Haustein, Larivière, & Sugimoto, 2013), and the effect that automated bots have on the legitimacy of using tweets to assess academic impact (Haustein et al., 2014). However, the demographics of tweeters who post links to academic articles have not yet been investigated. This study proposes to address this gap.

Methods

Sampling frame.

The initial sampling frame was a list of individuals who had provided a link to an academic article in a tweet. These tweets were gathered by running a Twitter query approximately every hour from March 17, 2012 to March 17, 2013 for each of a number of URLs of journals (Table 1). The journals were selected as leading journals that were widely tweeted (based on a manual examination of the data) and had a simple URL format for articles that could be collected by a query. Collecting tweets in this way was a practical step because many people link to articles if they mention them and it is easy to search for articles by part of URL. In each case article URLs had a common starting text, such as a domain name, and queries for this common part matched all articles in the site. Although Twitter shortens almost all URLs in tweets, it is possible to use URL-based queries because Twitter search returns matches for the original URLs rather than the shortened versions.

Table 1. Queries for links to academic articles in Twitter.

Source	Twitter query
Nature journal	"go.nature.com"
PLOS ONE journal	"plosone.org/article"
PNAS journal	"pnas.org/content"
Science journal	"scim.ag"

This method does not retrieve all tweets of academic articles published in the selected journals. In particular, it does not capture links to copies of the articles elsewhere (e.g., self-archived preprints) and does not capture articles mentioned by name rather than by link. Also, Twitter does not guarantee comprehensive matches to all searches so it is likely that not all URLs matching the above set of queries were found. Some data was also lost due to power cuts and an enforced shutdown at Wolverhampton in December 2012. However, this provides an authoritative list of scholarly tweets.

Sample

From this sampling frame, a list of all unique twitter accounts was generated. From this list, a sample of 500 unique tweeters for each journal was randomly selected. Duplicate accounts were removed and replaced so that the sample represented 2,000 unique accounts (this was necessary as some accounts tweeted articles from more than one journal).

Survey

The initial plan was to directly survey the journal tweeters and, accordingly, a survey was set up in Qualtrics and a separate DID Cascades Twitter account was established for the purpose of tweeting a link to the survey to all 2000 account. We set up an automated system to send out invitations to the survey to the identified twitter handles in batches small enough to not violate Twitter's mass tweeting policies. However, even working within these parameters, our account was suspended immediately upon our first batch of survey invitations. We mention this failure here as it is relevant to conducting research in this environment. Although some modes of inquiry (e.g., large-scale survey research) may be more appropriate for answering certain questions, they are untenable due to the current affordances of the platform. These limitations should be taken into consideration for future analyses.

Codebook construction

Given that obtrusive research was not possible, we turned to unobtrusive measures (i.e., content analysis) to analyse the identities of those who tweet about science. The codebook was developed inductively through several iterative explorations with four researchers. Variables such as gender, academic affiliation, and (in the case of non-individuals) organization type were collected. Iterative coding led to refining of the initial categories (e.g., the "Finance" category originally proposed was expanded to "Business/Finance", "Freelance" was incorporated into the coding due to the high frequency of this position, and "Non-profit" was added in the organizational category).

One of the initial desires was to be able to tag those who were "affiliated with science." This was intended to distinguish between the "layperson" and the "scientists". This seemingly simple distinction proved to be overwhelmingly difficult to code unobtrusively. Those who explicitly identified with academic institutions and were readily associated with science departments within those institutions were easy to identify. However, many of the non-academics were also affiliated with science in some form (e.g., government positions in science and technology). This also led to the issue of determining what constitutes science (e.g., are humanists, entrepreneurs, and technologists scientists?). This was equally difficult for organizations. For example, an online consumer or financial corporation might not have science as the main objective, but have an arm of the organization that conducts research. This question was further complicated by false negatives—that is, instances where we could not provide evidence that the individual was associated with science, but also could not provide evidence that they were not.

The issue of false positives and false negatives on other questions was addressed by adding an "unknown" option in addition to "yes/no" options. For example, one question asked whether the individual was a student. As it was frequently impossible to definitively state whether or not an individual was not a student (i.e., the lack of information regarding a person's reenrolment in a university would not, in itself, extinguish the possibility of their academic involvement at the student level). However a "no" option remained available for those situations in which it could be ascertained with a high degree of certainty that the individual was not or no longer a student (e.g., from a detailed LinkedIn profile or online curriculum vita).

Coding

Initial coding began in May 2013 and was completed on December 15, 2013. Coding was done by two coders for whom a high interrater reliability was ascertained. The twitter handles were used as the initial point of departure for the search. Coders determined what they could from the information provided in the short biographical information on twitter. If a url was provided on twitter, this was followed. Google searches were also employed, using as a seed the person's first name and/or twitter handle and limiting searches to the first three pages of results. Where there was a dispute between sources, the more contemporary source was used.

The first coding variable asked the coder to distinguish whether the account was held by an individual or an organization. Although most accounts are technically managed by a single person, a distinction was made between people who represented themselves and people who represented a company or organization. If a person simply affiliated with an organization, they were still coded as an individual.

Research centers at universities were coded as university. Research centers outside of a university setting were coded as non-profits. Although universities could be considered "government" or "non-profit" (and in some rare cases a corporation), all academic institutions were coded as universities.

Results

Approximately three-quarters of the sampled accounts could be identified as belonging to individuals (n=1520), while slightly under 23% belonged to organizations (n=459) (Figure 1). Of the accounts belonging to people, the majority were associated with a male tweeter (Figure 2). Nearly 12% of the individuals were identified as students (either undergraduate, master's, or doctoral). Of the students, 67.2% were doctoral students or candidates. It should be noted that, for some codes, a failure to mark a quality as "present" does not necessarily indicate that the reverse is true. For example, it is likely not the case that 88.2% of the individuals are *not* students; rather, all that we can say is that we were able to identify 11.8% of the individuals as students.

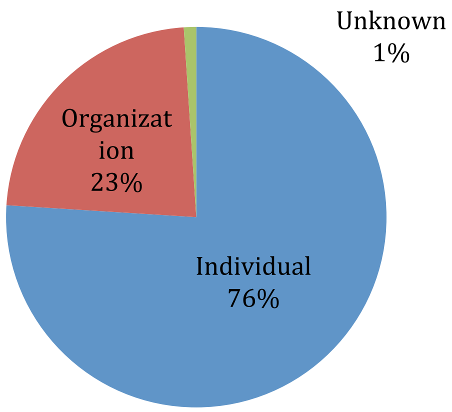


Figure 1. Twitter accounts by type.

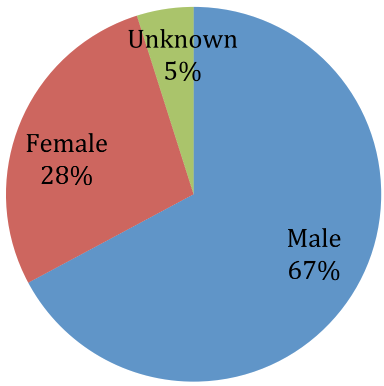


Figure 2. Individual accounts by gender.

In terms of the entire population of individuals, 34.4% were identified as possessing a Ph.D (this discounts the students who were working towards a Ph.D), suggesting that the population of Twitter users who tweet links to academic articles does not reflect the demographics of the general public. STEM fields were dominant both within the group of users identified as students and within the group of users identified as working in academe.

In terms of the students, 52.4% were affiliated with general science, 15.1% were associated with health/medical study, and 10.8% were associated with technology/engineering. In terms of the academics, 62% were associated with general science, 10.4% were affiliated with health/medical study, 8.1% were associated with the social sciences, and 7.5% were affiliated with technology/engineering (Figure 3).

Of the organizations, 41.6% were identified as non-profits, 29.2% were identified as corporations, and 13.1% were identified as universities. 18.9% were classified as news/media/outreach institutions (note that this was considered a non-exclusive category independent of the earlier classifications).

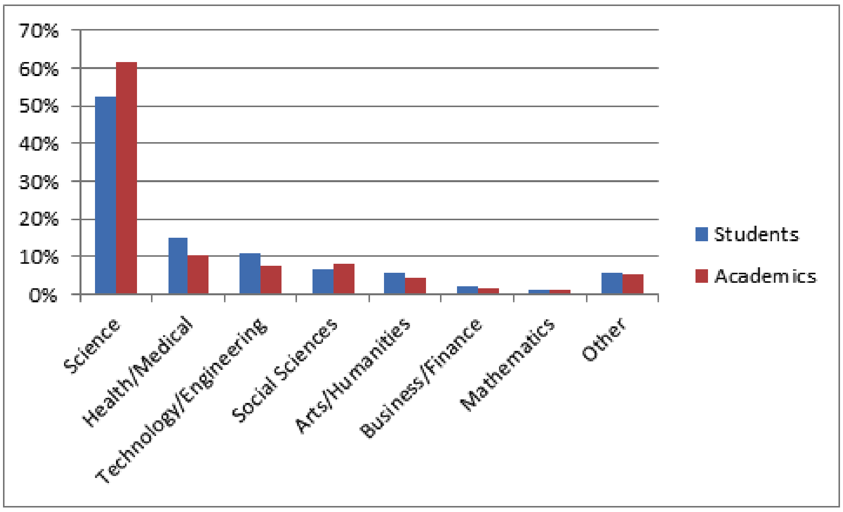


Figure 3. Proportion of twitter accounts by disciplinary domain.

Discussion and Conclusion

The demographics of the individual tweeters did not reflect the general population of Twitter users. Whereas women are overall slightly more likely to take advantage of social networking sites than men are (Kimbrough, Guadagno, Muscanell, & Dill, 2013; Pew Research, 2014), men use Twitter slightly more (24% of male Internet users, compared to 21% of female Internet users). Our study was much more male-baised, with nearly 70% of individual accounts maintained by men. This percentage is in keeping with male to female ratios found in the scientific workforce and scholarly publishing (Larivière et al., 2013).

A growing body of literature seeks to validate social media metrics, or "altmetrics" as valid forms of the social (i.e., public) impact of scholarly research. However, this research indicates that a large portion (i.e., nearly half) of those who tweet about science already have a doctoral degree or are in pursuit of one. This proportion far exceeds the 1% of the US population, for instance, holding a doctoral degree (Petersons, 2014). This suggests caution when utilizing social media metrics as an indication of the value of the work for the public. Rather, this emphasizes the strong use of these tools for dissemination and discussion of scholarship *among scholars*. Acknowledgement of the scholarly context of social media metrics must be taken into account in evaluative uses of these metrics.

Limitations

The study only considered journals that were frequently tweeted. It is possibly that the demographics of users who tweet articles from less popular journals might differ from those of tweeters who share links to the highest echelon of scientific journals. In addition, the information that could be gathered about the tweeters was limited to what was readily available online. Accordingly, the percentages generated by the study represent conservative estimates rather than absolute figures.

Future research might consider a wider variety of journals, as well as employing other methods to ascertain tweeter demographics (e.g., studying the users' tweets in an attempt to ascertain gender, academic affiliation, etc. for those users for whom such information was not publicly available). In addition, it is theoretically possible to directly survey the tweeters who shared links to academic articles, although such an approach would likely rely on publicly available contact information (primarily e-mail addresses), and would most likely face the same issues that were encountered in this study.

Acknowledgments

Funding for this project was provided by the Alfred P. Sloan Foundation ("Understanding the use and meaning of social media metrics in scholarly communication") and the National Science Foundation (grant SMA-1208804) as part of the Digging into Data initiative. We would like to thank Mike Thelwall for providing data for this project.

References

Alexa (2015). Alexa top 500 global sites. Retrieved from http://www.alexa.com/topsites.

Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (2014). Tweets as impact indicators: Examining the implications of automated bot accounts on Twitter. arXiv preprint arXiv:1410.4139.

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.

Kimbrough, A. M., Guadagno, R. E., Muscanell, N. L., & Dill, J. (2013). Gender differences in mediated communication: Women connect more than do men. Computers in Human Behavior, *29*(3), 896 900. doi:10.1016/j.chb.2012.12.005

Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C.R. (2013). Global gender disparities in science. *Nature, 504*(7479), 211-213.

Petersons (2014). Ph.D. programs are rigorous educational experiences. Retrieved from http://www.petersons.com/graduate-schools/phd-programs-rigorous-educational.aspx

Pew Research Center (2014). Social media update 2014. Retrieved from http://www.pewinternet.org/files/2015/01/PI_SocialMediaUpdate20144.pdf

Pew Research Center (2014). Social networking fact sheet. Retrieved from http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. PloS one, *8*(5), e64841.

Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013). Tweeting links to academic articles. Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics, (17), 1-8.

Classifying altmetrics by level of impact

Kim Holmberg

kim.j.holmberg@utu.fi
Research Unit for the Sociology of Education, University of Turku, 20014 Turku (Finland)

Introduction

In the light of current knowledge we can conclude that altmetrics do not present an alternative for traditional citation-based analysis of research impact (e.g., Haustein *et al*., 2014). Altmetrics have instead the potential to show some other aspects of research activities and provide a more nuanced view of the impact research has made on various audiences (Liu & Adie, 2013; Piwowar, 2013). Altmetrics come in many forms and from many different sources, all of which can represent different aspects of the online activity or of the different levels of impact that various research products have made on different audiences. What exactly the different altmetrics represent we do not yet know, but the greatest advantage of altmetrics may be exactly in this diversity.

Aggregating all altmetrics to a single indicator would remove this advantage. With aggregation of different altmetrics we are just creating another impact factor, another indicator that in the worst case is used for something that it is neither designed for nor capable of indicating. However, because of the wide variety of different sources for altmetrics, some form of aggregation or classification is needed and different types of classifications are already used by some service providers. Here we present another approach, one based on the level of impact. With this we hope to stimulate further discussion about the actual meaning of altmetrics.

Diversity of altmetrics

The diversity of altmetrics has two interesting dimensions; the diversity of people creating the altmetrics, and the diversity of the impact they indicate. In any research assessment what we want to measure is value or quality of research. Quality is of course very subjective and difficult to quantify. Because we cannot evaluate quality directly, particularly not at large scale, we use volume of impact as a proxy for value (i.e. number of citations or more recently number of online mentions).

The different data sources and different data types collected from the mentions of research products in various social media sites can represent a wide spectrum of different levels of impact. For instance, while a tweet does not necessarily hold any indication of impact other than awareness, a blog entry or a Wikipedia citation reflect some level of influence or impact. The people creating the altmetrics then again range from researchers and practitioners to the public.

Aggregating altmetrics

In social media analytics the mentions of brands and products in various social media are often placed and grouped together on a spectrum according to level of engagement, ranging from visibility to influence and finally reaching engagement as the most desired level of reaction. In the context of altmetrics, Piwowar and Priem (2013) write about the different "flavours" of impact that altmetrics could potentially reflect, referring to the diversity of altmetrics and possibility to group similar metrics into these "flavours". This is in line with the ideas presented at PLoS too, with different sources and different timings of altmetrics reflecting engagement from different audiences and possibly also that of different purposes for the engagement (Lin & Fenner, 2013).

This approach has already been taken by some of the altmetrics service providers as they group the data collected from various sources into what reflects different types of activities. PLoS for instance groups the metrics they use into views, saves, mentions, and citations. These do roughly translate to what we can assume to be different levels of impact, reflecting the variety of actions and interactions that one can have with the research products. Saving a research product suggests that the research product have made a bigger impact than just viewing it suggests, mentioning it suggests additionally increased level of impact, and citing it suggests what could perhaps be considered as the ultimate level of impact, at least when the goal is to investigate scientific impact.

Aggregation by the level of impact

Indicators of impact come in many diverse forms on the web and in social media and the different social media sites and the different activities within them can provide various metrics of different levels of impact. A potential approach to aggregating altmetrics would be to use these different levels of impact as they are and to not try to combine them according to source or type of activity they represent.

When the metrics indicate low impact we cannot really be sure whether the research has made any impact at all as evidence of it is usually not clear; a page view, clicking on a tweet button next to the article, or sharing a research article on Facebook, all indicate that the user has seen what they are sharing but nothing indicates that it has made any impact on them, that they would have been influenced by it, or that they would have changed their behaviour because of it. Metrics indicating a medium level of impact would already come attached with at least some information that the research has made an impact, that it has in some way influenced the user. Whether the research product has been mentioned somewhere online or been bookmarked with the intent to use it later, the metrics generated from the activities at this level suggest that the users have been influenced some way, that the research has made at least some impact. Metrics indicating a high level of impact usually come attached with some additional, perhaps more qualitative data that we can use to investigate how the research has influenced the user and confirm what kind of impact it has made. A rough classification of different types of altmetrics that indicate different levels of impact could follow the one presented in Table 1. Besides impact, we can also measure reach with altmetrics; how many people have become aware of the research and how many of them have been influenced by it in some way.

Table 1. Levels of impact.

	Altmetrics		
Level of impact	Low	Medium	High
Reach	High	Medium	Low
Example activities	Awareness, visibility	Influence, interaction	Usage
Example metrics	Tweets, 'likes', shares, …	Mentions, downloads, bookmarks, …	Blog posts, …

More research is needed and both quantitative and qualitative methods are needed to confirm what level of impact different types of actions in different social media reflect and how they relate to each other.

Benefits of the proposed approach

Focusing future research on the level of impact has a couple of benefits compared to other approaches. First of all, impact is what we want to measure, hence grouping different metrics based on the level of impact they reflect makes sense. Second, using all the unique metrics (e.g., tweets, retweets, blog mentions, link in blogroll, Facebook shares, "likes", and mentions) would create a massive number of different metrics that would be difficult to a) keep track of, b) present, and c) control. Third, aggregating the different metrics by type of activity they represent may not give an accurate picture of the impact they represent, as similar types of activities on for instance different social media sites may be reflecting different levels of impact and/or different types of users. And fourth, aggregating all the metrics into a single indicator would just be creating another impact factor, but this time from a much wider diversity of different metrics indicating different aspects and which probably should not be aggregated at all because of that. And finally, focusing on the different indicators for different levels of impact instead of some specific sites would not be such a vulnerable approach relying on the continued existence and goodwill of the social media sites to allow access to their data.

Conclusions

We propose the classification of altmetrics based on the level of impact reflected by the specific altmetrics. This approach would have some clear benefits compared to aggregations based on activity or source of altmetrics. More research is, however, needed to establish the different levels. The key challenges for future altmetric research are a) identifying the groups of people that create different altmetrics, and b) mapping the different levels of impact the different metrics reflect. This line of research would bring us again one step closer to fully understand what altmetrics indicate, and with that, the meaning of altmetrics. It is nevertheless important to recognize that the true meaning of any altmetrics lies in the stories behind the numbers. Hence it is important that any altmetrics are presented together with the accompanied stories to give the full context in which they have been generated.

References

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 American Society for Information Science and Technology*, *65*(4), 656-669.

Lin, J. & Fenner, M. (2013). Altmetrics in evolution: defining and redefining the ontology of article-level metrics. *Information Standards Quarterly*, *25*(2), 20-26. http://dx.doi.org/10.3789/isqv25no2.2013.04.

Liu, J. & Adie, E. (2013). New perspectives on article-level metrics: developing ways to assess research uptake and impact online. *Insights: the UKSG Journal*, *26*(2), 153-158.

Piwowar, H. (2013). Altmetrics: what, why and where? *Bulletin of the Association for Information Science and Technology*, 39(4).

Piwowar, H. & Priem, J. (2013). The power of altmetrics on a CV. *Bulletin of the Association for Information Science and Technology*, 39(4).

Characterizing In-text Citations using N-gram Distributions

Marc Bertin[1] and Iana Atanassova[2]

[1] bertin.marc@gmail.com
Centre Interuniversitaire de Rercherche sur la Science et la Technologie (CIRST),
Université du Québec à Montréal (UQAM), Canada

[2] iana.atanassova@univ-fcomte.fr
Centre Tesniere, University of Franche-Comte, France

Introduction

This article focuses on a Natural Language Processing (NLP) approach for the analysis of citation functions in scientific papers. Bibliometric studies traditionally rely on citation metadata and count the number of times a publication has been cited. However, some recent studies rely also on full text processing on papers, e.g. (Boyack et al., 2013), (Bertin et al., 2013, 2014). The full text content of papers and more specifically the sentences containing citations provide valuable information on the functions of citations that can be exploited through NLP. To study citation acts, we need to consider full text papers and their rhetorical structure.

The main question that we want to answer here is whether the most frequent citation patterns are correlated to the rhetorical structure of scientific papers. We investigate the properties of the linguistic patterns that appear in citation contexts. For this, we study the distribution of n-gram classes containing verb forms, and we show the existence of three different types of distributions according to the rhetorical structure.

Method

By analyzing a large corpus of articles, we propose a quantitative study of the linguistic patterns around in-text citations. Some words or sets of words in n-grams are more frequent than others (Cavnar & Trenkle, 1994), and this idea is consistent with Zipf's Law (Zipf, 1949). The difficulty is that the calculation of n-grams in contexts results in a combinatorial explosion. We propose several filters to reduce the number of patterns.

The rhetorical structure of scientific papers is typically organized around a standardized pattern, known as the IMRaD structure (Introduction, Methods, Results and Discussion). We identify the four main section types of this structure by analysing section titles. Then, we consider the set of sentences containing citations and belonging to each section type.

We represent citation contexts by using sequences of words of length n called n-grams where 2<n<=5. In our approach we consider only n-grams within sentence boundaries because sentences are natural building blocks of the text. For each n-gram we observe its frequencies in the four section types of the IMRaD structure. For our study, we select only the n-grams that contain at least one verb form. In this way, the number of n-grams to process is much smaller and we eliminate word patterns containing only nominal groups like: "In this paper", "the present article", "the result of" etc. for 3-grams.

Dataset

We performed an automatic analysis of the seven peer-reviewed academic journals published in Open Access by the Public Library of Science (PLOS). The corpus contains about 85,660 research articles. Most of the articles are in the biomedical domain, but the corpus covers all fields of Human and Natural Sciences, as the publisher's main journal, PLOS ONE, is multidisciplinary. Around 98% of the articles in the corpus follow the IMRaD structure, which is imposed by editorial requirements.

Results

We select the most frequent verb forms in order to construct n-gram classes from in-text citation contexts. This data will be used to obtain a first typology of the distribution of n-grams depending on the rhetorical structure of articles.

The following figures present distributions of n-grams classes for the IMRaD sections. We can distinguish between three different type of classes, and we give one example of each. The horizontal axis presents the text progression of the section from 0% to 100%. The vertical axis gives the percentage of occurrences of each class relative to its occurrences in citation contexts in the entire article.

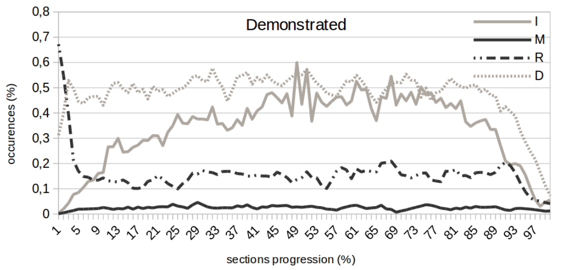


Figure 1. Demonstrated.

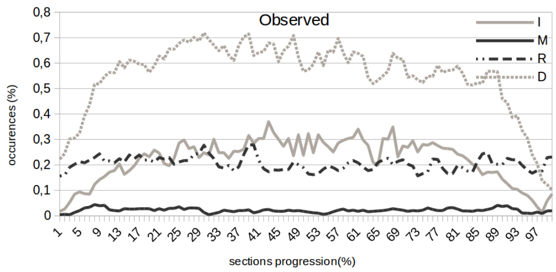


Figure 2. Observed.

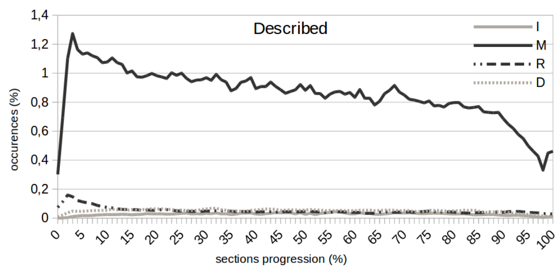


Figure 3. Described.

Discussion

Figure 1 shows the first class, which includes n-grams containing the verb *Demonstrated*. These n-grams appear with roughly equivalent frequencies in the sections Results and Discussion, but, at the same time the Methods section contains much lower frequencies of these patterns.

Figure 2 shows the second class type, which includes n-grams with the verb *Observed*. We can observe another type of distribution, with relatively very high frequencies in the Discussion section.

Figure 3 shows the distribution of n-grams with the verb *Described*. We can observe that the structure of the Methods section is unique, as the class *Described* is present with a very high frequency in this section and especially at the beginning of the section. Moreover, Figures 1 and 2 show that on the distributions for the other classes, the Methods section contains relatively few occurrences. In other words, the class *Described* is characteristic of the Methods section, where it appears with very high frequency, and it is very rare in all the other sections. The Methods section displays very low frequencies for all classes except *Described*.

These results imply that each section, depending on its nature, authorizes more or less easily the usage of specific patterns containing verbs. The Methods section is rather closed in nature, where we find a very small number of high frequency verbs. At the same time, the Discussion section is open to different forms and allows a larger number of variations in terms of the linguistic means that authors use in citation contexts.

Conclusion

The purpose of this study is to demonstrate the existence of frequent n-gram patterns in citation contexts and their strong relation with the rhetorical structure of scientific articles. Studying the n-gram classes containing verb forms, we show the existence of three different types of distributions according to the rhetorical structure. From our point of view, the problem of the automatic annotation of citation contexts is strongly related to identifying significant surface patterns for the annotation process.

Acknowledgments

We thank Benoit Macaluso of the Observatoire des Sciences et des Technologies (OST), Montreal, Canada, for harvesting and providing the PLOS data set.

References

Bertin, M., & Atanassova, I. (2014). A Study of Lexical Distribution in Citation Contexts through the IMRaD Standard. *Proceedings of the First Workshop on Bibliometric-enhanced Information Retrieval co-located with 36th European Conference on Information Retrieval* (ECIR 2014). Amsterdam, The Netherlands.

Bertin, M., Atanassova, I., Larivière, V., & Gingras, Y. (2013). The Distribution of References in Scientific Papers: an Analysis the IMRaD Structure. *Proceedings of the 14th International Conference of the International Society for Scientometrics and Informetrics* (pp. 591–603). Vienna.

Boyack, K. W., Small, H., & Klavans, R. (2013). Improving the accuracy of co-citation clustering using full text. *Journal of the American Society for Information Science and Technology, 64*(9),1759–1767.

Cavnar, W. B., & Trenkle, J. M. (1994). N-gram-based text categorization. Ann Arbor MI, 48113(2), 161–175.

Small, H. (1982). Citation context analysis. *Progress in Communication Sciences, 3*, 287–310.

Zipf, G. K. (1949). Human behavior and the principle of least effort.

Can Book Reviews Be Used to Evaluate Books' Influence?

Qingqing Zhou[1] and Chengzhi Zhang[2,*]

[1]*breeze7zhou@163.com*

[2] *zhangcz@njust.edu.cn*.

Department of Information Management, Nanjing University of Science and Technology,Nanjing, China

Introduction

Citation frequency has become a popular index for quality evaluation of academic publications, e.g. articles, journals or books. Traditional altmetrics researches pay less attention to book-level evaluation, and they do not make use of content information. In this paper, we present a novel method, reviewmetrics, namely altmetrics to evaluate academic books based on reviews. We combine star and reviews with the information of helpfulness which is given by readers reflecting the degree of how helpful this review is (Yin, Bond, & Zhang, 2014). Correlation analysis was also conducted with citation frequencies of academic books, so as to prove the validity of reviewmetrics.

Methodology

Framework

The purpose of the study is to evaluate the influence of academic books by mining book reviews. We conduct correlation analysis between citation frequencies and academic book scores calculated by reviewmetrics to prove the validity. Reviewmetrics includes combinations of factors like numbers of positive and negative reviews, star values and aspect values. Every combination has two schemes. Scheme 1 does not take information of helpfulness into consideration; Scheme 2 will consider information of helpfulness. The details are shown in Figure 1.

Data

We collected citation frequencies of academic books from three disciplines, including economics, management and literature, from reports on the academic influence of Chinese humanity and social science books (Su, 2011). We chose books that were cited more than 10 times as candidate books. We checked every candidate book in Amazon, and if it had more than 10 reviews, it would be selected as a final research book. In total, we have selected 182 books, including 40 economics books, 44 management books and 98 literature books. The corpora were collected in October, 2014. They cover citation frequencies, reviews, stars and helpfulness of the books.

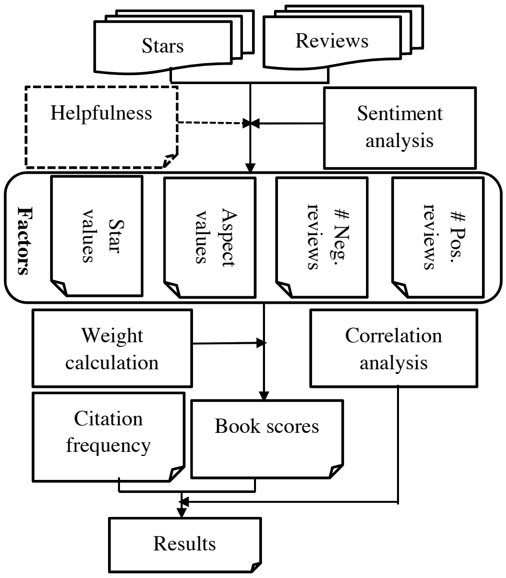


Figure 1. Frameworks ofcorrelation analysis.

Factor calculations

Calculations of numbers of positive reviews and negative reviews

We identify the sentiment polarities of reviews by conducting document-level sentiment analysis. Specifically, SVM (Hearst et. al, 1998) is used as a classification model, and TF-IDF (Salton & McGill, 1983) is used to select features and calculate their weightings. After sentiment classification, we get sentiment polarity of each review, and then we get numbers of positive reviews and negative reviews of each book.

Calculations of aspect values and star values

In the pre-processing step of calculations of aspect values, it has two subtasks: aspect extraction and aspect sentiment classification. Frequent nouns method is used to extract aspects. Frequent nouns are chosen as candidate aspects after POS (Part-Of-Speech) tagging; and top 10 of them are chosen as real aspects. For aspect sentiment classification, we use method proposed in (Ding et al, 2008) to calculate sentiment polarity sp_{ij} of aspect s_i in review r_j.

* *Corresponding author: Chengzhi Zhang, Tel: +86-25-84315963*.

As we have got the aspects and their sentiment polarities in every review, we can calculate the aspect values and star values of each book. The details are shown in Table 1.

Table 1.Calculations of book scores.

aspect values	$VAB_{it} = \sum_{j=1}^{N} sp_{ij} \Big/ \sum_{j=1}^{N} \lvert sp_{ij} \rvert$ $i = 1,2,\dots,10, t = 1,2,\dots M$
	$VAB'_{it} = \sum_{j=1}^{N} (sp_{ij} * h_j) \Big/ \sum_{j=1}^{N} \lvert sp_{ij} \rvert$
star values	$VSB_{jt} = \sum_{j=1}^{N} star_j / \mathrm{N}$
	$VSB'_{jt} = \sum_{j=1}^{N} (star_j * h_j) / \mathrm{N}$

For aspect values, VAB_{it} denotes aspect values of aspect s_i about book b_t without considering the information of helpfulness ($VAB_{it}{'}$ means with helpfulness), N means number of reviews with aspect s_i about book b_t; i denotes the numbers of aspects; M means the numbers of books of each discipline, $h_j$means helpfulness score of review r_j.
For star values, VSB_{jt}denotes star values of review $r_j$about book $b_t$without considering the information of helpfulness (VSB'_{jt} means with helpfulness), $star_j$ means star score of review r_j, it range from 1 to 5, N denotes the numbers of reviews about book b_t.

Calculations of book scores

We use the entropy method to calculate factor weightings (Hongzhan et al., 2009), and then get book scores. The details are shown in Table 2.

Table 2.Calculations of book scores.

Steps	Formulas
(1) Normalization	$p_{ij} = \frac{v_{ij}}{\sum_{i=1}^{N} v_{ij}}$ $i = 1,2,\dots,N, j = 1,2,\dots,m$
(2) Factors entropies	$e_j = -\frac{1}{\ln(n)} \sum_{i=1}^{N} p_{ij} \ln(p_{ij})$
(3) Factor weightings	$w_j = 1 - e_j / m - \sum_{j=1}^{m} e_j$
(4) Book scores	$SB_i = \sum_{j=1}^{m} p_{ij} * w_j$

where, p_{ij} denotes proportion of book b_i in factor f_j , v_{ij} denotes value of book b_i in factor f_j , N means the numbers of books, m means the numbers of factors.e_j denotes entropy of factorf_j.w_j denotes weighting of factorf_j, SB_i denotes book scores of book b_i.

Experimental result analysis

We conduct correlation analysis between citation frequency and book scores calculated by reviewmetrics about three disciplines, including consider the information of helpfulness or not. The results are shown in Table 3.

On the whole, with the information of helpfulness, reviewmetrics of three disciplines have significant Pearson correlations with citation frequency ($p < 0.1$).

Table 3. Results of correlation analysis.

Domains	Without H.	With H.
Economics	**0.383***	**0.378***
Management	**0.401****	**0.417****
Literature	0.197	**0.240***

Conclusions

In this paper, we propose a novel altmetrics method: reviewmetrics on the basis of book reviews to evaluate its influence. We prove reliability of our method by conducting correlation analysis between our method and citation frequencies. Two main conclusions can be drawn according to our above mentioned analysis: **WH (with helpfulness) conclusion**: the information of helpfulness is really useful to filter low quality reviews. **OC (overall correlation) conclusion**: It is reliable to use reviewmetrics to evaluate influences of academic books.

Acknowledgments

This work is supported by Major Projects of National Social Science Fund (13&ZD174), National Social Science Fund Project (No.14BTQ033) and the Opening Foundation of Alibaba Research Center for Complex Sciences, Hang-zhou Normal University (No. PD12001003002003).

References

Ding, X., Liu, B., & Yu, P. S. (2008). *A holistic lexicon-based approach to opinion mining*. Proceedings of the 2008 International Conference on Web Search and Data Mining.

Hearst, M. A., Dumais, S., Osman, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *Intelligent Systems and their Applications, IEEE, 13*(4), 18-28.

Hongzhan, N., Lü Pan, Q. Y., & Yao, X. (2009). Comprehensive fuzzy evaluation for transmission network planning scheme based on entropy weight method. *Power System Technology, 33*(11), 60-64.

Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval*.

Su, X. (2011). *A report on the academic influence of Chinese humanity and social science books: China Social Science Press* (In Chinese).

Yin, D., Bond, S. D., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *Mis Quarterly, 38*(2), 539-560

Adapting sentiment analysis for tweets linking to scientific papers

Natalie Friedrich[1], Timothy D. Bowman[2], Wolfgang G. Stock[1] & Stefanie Haustein[2]

[1] *natalie.friedrich@hhu.de, stock@phil.hhu.de*
Heinrich Heine University Düsseldorf, Institute of Linguistics and Information, Department of Information Science, Düsseldorf (Germany)

[2] *stefanie.haustein@umontreal.ca, timothy.bowman@umontreal.ca*
École de bibliothéconomie et des sciences de l'information, Université de Montréal, Montréal (Canada)

Introduction

In the context of "altmetrics", tweets have been discussed as potential indicators of immediate and broader societal impact of scientific documents (Thelwall et al., 2013a). However, it is not yet clear to what extent Twitter captures actual research impact. A small case study (Thelwall et al., 2013b) suggests that tweets to journal articles neither comment on nor express any sentiments towards the publication, which suggests that tweets merely disseminate bibliographic information, often even automatically (Haustein et al., in press). This study analyses the sentiments of tweets for a large representative set of scientific papers by specifically adapting different methods to academic articles distributed on Twitter. The aim is to improve the understanding of Twitter's role in scholarly communication and the meaning of tweets as impact metrics.

Dataset and Methods

Tweets and research articles

The study is based on all articles and reviews published in 2012 in the Web of Science (WoS) linked to tweets via the Digital Object Identifier (DOI) as captured by Altmetric.com until 30 June 2014. The dataset consists of 663,547 original tweets (i.e., excluding retweets) mentioning 238,281 documents.

Sentiment tools

A sentiment represents an emotion expressed by a person based on their opinion towards a subject. Text-based sentiment analysis focuses largely on identifying positive and negative, as well as the absence of, sentiments using linguistic algorithms (Thelwall et al., 2010). For our purposes the sentiment expressed in a tweet linking to a scientific paper is assumed to reflect the opinion of the tweeting user towards the paper. SentiStrength[1] (s_1) and Sentiment140[2] (s_2) were selected to automatically detect sentiments. SentiStrength assigns values from -5 to +5 to certain terms in a lexicon. Each processed tweet receives a negative and a positive value. To assign each tweet to exactly one category (positive, negative, neutral), the stronger value determines the sentiment. Sentiment140 provides one sentiment value per tweet on a scale from 0 (negative) to 4 (positive). For better comparison values are converted to obtain three sentiment categories positive, negative, and neutral. While SentiStrength has been developed for short online texts and Sentiment140 was particular implemented to analyse tweets, none of the tools seem suited to analyse tweets related to scientific topics. In contrast to SentiStrength, which provides options to change the lexicon, Sentiment140 is less transparent and only allows insight into the training corpus.

Intellectual coding of sentiments

The text from 1,000 random tweets was analysed and compared to the title of the papers the tweets linked to in order to gain an understanding of the discussions of scientific papers on Twitter and to determine their sentiment intellectually s_i. A second intellectual assessment is undertaken with regard to the capabilities of the sentiment analysis tools. For example, Natural Language Processing (NLP) tools are not able to detect irony. The results of these assessments function as the ground truth s_0, to which sentiments detected by the tools are compared.

Cleaning tweets

A tweet consists of 140 characters including text, hashtags (following the # sign), user names (following the @ sign), and/or links to websites. As user names, URLs, and the # sign are not considered to be part of the tweet content regarding the sentiment analysis, they were removed from the tweet. Hashtag terms are kept as they are assumed to carry meaning and sentiment. The tweets without specific affordances are called t_0.

The intellectual analysis revealed that many tweets contained the title of the scientific paper to which they linked, which influences the sentiment analysis—even though it does not reflect the users emotion and opinion towards the paper. As the sentiment tools are not adapted to scientific

[1] http://sentistrength.wlv.ac.uk/
[2] http://help.sentiment140.com/home

language, certain research topics are assigned positive or negative sentiments. For example, in SentiStrength the term 'cancer' receives the value -4 and 'disease' -3. As this influences the outcome of the sentiment analysis, tweets t_0 were further adapted by removing all title terms from the particular paper to which they link (using regular expressions in PHP) to derive tweets adapted for sentiment analysis t_a.

In addition to removing title words from tweets to avoid false positives regarding the sentiment detection, the lexicon was adapted to the scientific context for SentiStrength by identifying the terms leading to disagreement between s_0 and s_1. Overall, 51 terms (e.g., 'cancer', 'disease' or 'obesity' for negative sentiments, 'baby' or 'care' for positive sentiments) were removed from the lexicon. Results for SentiStrength after the lexicon changes are denoted as s'_1. The lexicon for Sentiment140 was not accessible and thus could not be adapted.

Results obtained by SentiStrength (s_1 and s'_1) and Sentiment140 s_2 are compared to the ground truth s_0 for cleaned tweets t_0 and t_a using percentage overlap and Cohen's Kappa κ.

Preliminary Results

The intellectual assessment of the tweet content s_i identified 4.3% of the 1,000 random tweets to contain positive, 0.9% negative, and 94.8% neutral sentiment, which is in agreement with findings by Thelwall et al. (2013b).

Table 1. Intellectual (s_0) and automated (s_1, s'_1, s_2) sentiment detection for 1,000 tweets.

		Sentiments (%)			Agreement w/ s_0	
		+	–	*n*	%	*Κ*
	s_i	4.3	0.9	94.8	n/a	
	s_0	4.1	0.6	95.3	n/a	
t_0	s_1	12.2	33.8	54.0	56.8	0.10
	s_2	0.6	1.6	97.8	94.3	0.16
t_a	s_1	8.2	11.2	80.6	83.8	0.29
	s'_1	8.0	2.8	89.2	92.9	0.52
	s_2	0.7	1.0	98.3	94.6	0.14

Results for SentiStrength (s_1, s'_1) and Sentiment140 (s_2) compared to the ground truth s_0 are shown in Table 1. Removing paper title terms from the tweets increases the accuracy in particular for neutral and positive tweets and raises agreement with s_0 from 56.8% to 83.8% for s_1, representing fair agreement according to Cohen's Kappa (κ=0.29). The process of adapting the lexicon (s'_1) leads to an additional increase to 92.9% (κ =0.52, moderate agreement). 90.2% of 41 positive tweets and 93.2% of 953 neutral tweets are detected correctly by s'_1 for t_a. However, the detection of negative sentiments decreases from 100% (s_1) to 66.7% (s'_1), as only 4 of 6 negative tweets were identified by s'_1.

Although the overall agreement between s_2 and s_0 for t_0 represents 94.3%, only 14.6% positive sentiments and none of the 6 negative sentiments were detected correctly by Sentiment140. The high overall agreement arises from the agreement of neutral sentiment that yields 937 tweets. Removing the title words from tweets leads to a small increase of the overall percentage agreement for Sentiment140 to 94.6%, however the percentage of identified positive tweets decreases to 12.2%.

Discussion and Future Work

Our analysis shows that current sentiment tools are not able to accurately detect sentiments for the specific context of tweets discussing academic papers. While SentiStrength overestimates sentiments of tweets about scientific papers, Sentiment140 is not able to detect any negative tweets and only 14.6% of positive tweets leading to slight agreement (κ=0.16). As it does not allow access to the lexicon, Sentiment140 remains a black box.

Automatic sentiment detection was significantly improved for SentiStrength by adjusting tweets (removing title terms) and lexicon leading from slight (κ=0.10) to moderate agreement (κ=0.52). However, the detection of negative sentiments remains problematic.

Future work will focus on improving negative sentiment detection by analyzing specific cases of false positives. The aim is to develop an adapted lexicon in order to perform an sentiment analysis the 663,547 tweets linking to 238,281 documents.

References

Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., & Larivière, V. (in press). Tweets as impact indicators: Examining the implications of automated bot accounts on Twitter. *Journal of the Association for Information Science and Technology*. Retrieved from http://arxiv.org/abs/1410.4139

Thelwall, M., Haustein, S., Lariviére, V., & Sugimoto, C.R. (2013a). Do Altmetrics work? Twitter and Ten Other Social Web Services. *PLoS ONE 8(5)*: e64841.

Thelwall, M., Tsou, A., Weingart, S., Holmberg, K., & Haustein, S. (2013b). Tweeting links to academic articles. *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics*, 1–8. Retrieved from http://cybermetrics.cindoc.csic.es/articles/v17i1p1.html

Thelwall, M., Buckley, K., Paltoglou, G. Cai, & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal oft he American Society for Information Science und Technology, 61*(12), 2544-2558.

Mendeley Readership Impact of Academic Articles of Iran

Ashraf Maleki[1]

[3] *malekiashraf@ut.ac.ir*
Faculty of Library and Information Science, University of Tehran, Enqelab sq., Tehran (Iran)

Introduction

By means of formal citation analysis, although scientific impact of research was measured, so far other influential aspects of research such as readership and educational impact was simply ignored. Now online reference management tools such as Mendeley allow creating collections of digital paper holdings, and collaborative filtering of scientific publications, whose data proved to predict future formal citations (Li, Thelwall & Giustini, 2012). Mendeley metric obtains credit by measuring readership, for majority of users who add papers to their Mendeley libraries to read, although they may save them to cite or use in professional, educational, or teaching activities (Mohammadi, Thelwall & Kousha, in press). Mendeley readership also has potentials to present knowledge flow across fields (Mohammadi & Thelwall, 2012), and popularity of papers among users from within various countries (Maflahi & Thelwall, 2014) and academic career stages (Haustein & Larivière, 2014). Although this metric is studied for patterns of impact in various fields, its application for research impact assessment practice in developing countries is less known. Therefore, this research assessed WoS (Web of Science of Thomson Scientific) publications of Iran (2000-2012) for users in Mendeley across four broader research areas. In addition, career stages and nationalities of Mendeley users are also analysed for patterns of interested users in papers. The results may help to understand how and to what extent Mendeley readership metric is applicable to assess publications of authors in Iran.

Method

To assess the extent to which publications are included in Mendeley libraries of users a random sample of 31,629 WoS-indexed papers with Iranian authors in 2000-2012 were selected, which comprise about 31% of all publications with DOIs, including 11,030 (35%) in broader field of life science and biomedicine, 11,618 (32%) in physical sciences, 8,462 (27%) in technology, and 519 (20%) in social science. Mendeley readership counts are gathered by submitting DOIs to *ImpactStory.org*, in July 2013. Some articles were recorded in Mendeley with multiple variations, then to avoid duplicates the ones with higher readership counts were considered.

There is a limitation regarding the data available for analysing users' career stage and nationality, which is also observed in previous studies (Mohammadi & Thelwall, 2014; Haustein & Larivière, 2014). Statistics are suggested in Mendeley for top three countries and career stages of users. For this reason, although there is a 100% contribution of users in about 67% of publications, rest of the papers include nationalities or academic stages for 24% to 94% of total users. Therefore, although a high extent of users' career stage and nationality were available, findings are not a full reflection of user properties.

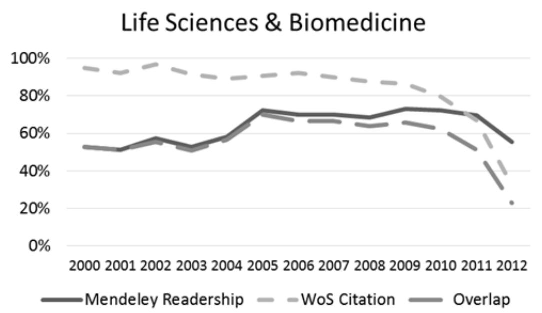
Results

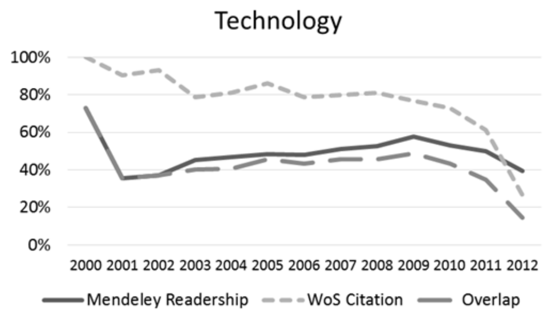
Overall results suggest that about 53% of papers (16,667) had at least one user in Mendeley. The field of life science and biomedicine (65%) had the highest coverage in terms of the papers included in Mendeley libraries; and it is followed by social sciences (50%), technology (48%) and physical sciences (44%). The figures 1 to 4 over years show proportion of publications with WoS citations, Mendeley readerships, and both of them (overlap) in four broader research areas. They show that although there are relativly less papers in recent years with WoS citations for the natural publication delay, readership uptake of publications follow a slighter decrease, where in the most recent years there are more papers read than cited. The findings suggest that 21% of publications in social sciences in 2012 only have readers whereas they do not receive citations; and this proportion is higher than the extent of publications which only receive citations (16%). By contrast, in other three fields the extent of papers only with citations are higher in proportion than the ones only with readers - 19% vs. 15% in life sciences and biomedicine, 27% vs. 14% in technology, and 36% vs. 8% in physical sciences. Therefore, uptake of publications highly vary in the most recent papers by the two metrics.

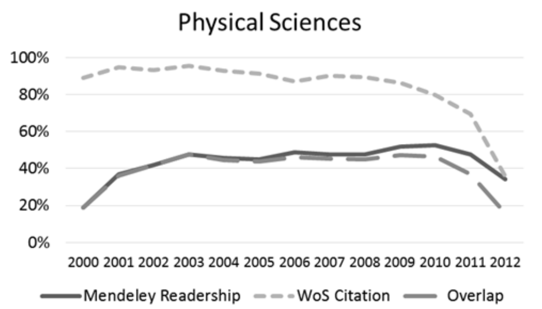
Career stages and nationalities of Mendeley users

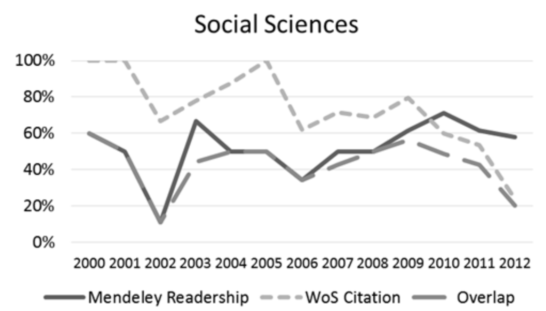
Results suggest that 31,629 readerships are mainly associated with the engagement of 30% (9,641) Ph.D students, 17% (5,233) master students, 9% (2,895) post docs, and 7% (2,325) researcher at academic institutions, whereas professors (4%), lecturers (2%), and senior lecturers (1%) are in minority.

Further results suggest that 79% of articles had at least one Mendeley user in the top 10 countries whereas other users are in 118 other countries. The papers with US readers are in majority (3,974 articles, 24%) in all fields except in technology where papers with Indian readers are high (3,025 articles mainly in physical sciences and technology, 18%). Also, UK readers include more papers (2,840 papers mainly in life science and biomedicine, 17%) than Iranian readers (11%, 1,897 papers with higher proportions in physical sciences).









Figures 1-4. Trend of relative proportion of publication uptake via formal WoS citations, Mendeley readerships and both of them (overlap) across four broader research areas- Y-axis shows percent of publications in each year.

Discussions and Conclusions

The main findings of study suggested that trend of publications' online readership is not only faster than WoS citations, but also is different from it. Many of the papers with Mendeley readers exclude WoS citations. They are often papers that might be read rather than cited, mostly in social sciences. This seems to be the advantage of online readership metric for evaluation of research in social sciences, and seems to be applicable for publications of Iran. However, in other field a considerable extent of papers also seem to get readers faster that citations, often in life sciences and biomedicine.

The results about career stages of the users are in line with previous observations in Haustein and Larivière (2014) and Zahedi, Costas and Wouters (2014) as they also found the highest inclusion of papers by Ph.D. students and the lowest by the lecturers and librarians. However the results about nationality of the readers differ from Thelwall and Maflahi (2014), since Iranian users of Mendeley are not excessively adding publications to their libraries but US, India and UK readers, which may reflects distribution of Mendeley users in various countries, than potential readers worldwide. Ultimately, it seems that Mendeley readership metric may help to assess impact of the publications, especially in fields, which tend to receive citations late.

Acknowledgments

The author would like to thank Dr. Kayvan Kousha, Statistical Cybermetrics Research Group, for his very useful comments.

References

Haustein S. & Larivière, V. (2014). Mendeley as a Source of Readership by Students and Postdocs? Evaluating Article Usage by Academic Status. *Proceedings of the IATUL Conferences*.

Li, X., Thelwall, M., & Giustini, D. (2012). Validating online reference managers for scholarly impact measurement. *Scientometrics*, *91*(2), 461-471.

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.

Mohammadi, E., Thelwall, M., & Kousha, K. (in press). Can Mendeley Bookmarks Reflect Readership? A Survey of User Motivations. *Journal of the Association for Information Science and Technology*.

Thelwall, M., & Maflahi, N. (2014). Are scholarly articles disproportionately read in their own country? An analysis of Mendeley readers. *Journal of the Association for Information Science and Technology*, doi:10.1002/asi.23252.

Does the Global South have Altmetrics? Analyzing a Brazilian LIS Journal

Ronaldo F. Araújo[1], Tiago R. M. Murakami[2], Jan L. de Lara[3] and Sibele Fausto[4]

[1] ronaldfa@gmail.com
Federal University of Alagoas and Federal University of Minas Gerais, Librarianship Dept, Av. Lourival Melo Mota, s/n, Tabuleiro dos Martins, Maceió, AL, CEP 57072-900 (Brazil)

[2] tiago.murakami@dt.sibi.usp.br, [3] jan.lara@sibi.usp.br [4] sifausto@usp.br
University of São Paulo, Rua da Biblioteca, s/n, Complexo Brasiliana, São Paulo, SP, CEP 05508-050 (Brazil)

Introduction

As a new emerging field, Altmetrics has become a trendsetter, and received a good deal of attention by researchers involved in the evaluation of scientific research. Moreover, it has led to a notable growth in the related academic literature. The international landscape has displayed an exponential growth in the field of scholarly publishing with several studies exploring altmetrics (both their potential benefits and limitations) in the last 3 years. However, in the Global South this subject is still not widespread, with a few empirical works. Alperín (2014) explored altmetrics measurements from articles in South American journals retrieved from sources such as SciELO, Redalyc and Latindex. This author also carried out an analysis of 21,560 articles published by the Brazilian journals in SciELO. This explored its altmetrics data with the Altmetric.com tool, and showed that these new measurements in the region are still in their early stages. Alperín (2014) also believed that the spread of science on the Internet and social networks in Brazil seems to have been limited in scope. This is because there are few or no sources of alternative performance metrics such as Blogs, Wikipedia, videos and social media like Google Plus, LinkedIn, Reddit, Pinterest, and others. The only media that appears to have significant data is Twitter, with 6.03% of mentions, followed by Facebook, with only 2.81%.

Nascimento & Oddone (2014) also used Altmetric.com to conduct an analysis of altmetrics indicators in 2 Brazilian journals in Library and Information Science (LIS). This showed that out of a total of 55 articles, 35 (63%) recorded mentions of Twitter, 22 (40%) of Mendeley, 19 (34%) of Facebook and 1 (1%) of Pinterest. Similarly, Araújo (2014) analyzed the altmetrics data of Brazilian LIS journals either through Altmetrics.com, with the cut-outs of 121 articles published in the last 3 editions of 4 core national journals in this area. From this total sample, only 6 articles of 3 different journals returned altmetrics data. Apart from the limited amount of altmetrics data in the source, it is clear that all of the data were from Twitter, with no mentions on Facebook, or on blog posts. Araújo (2014) argues that these meagre results in the use of Altmetrics.com may have been caused by (1) a limitation of the tool due to the issues already considered such as DOI and, others; and (2) the coverage provided by other social media services.

It has been suggested that this drawback in the use of social media (such as Twitter, Facebook and LinkedIn) can be overcome through the use of an API (Application Programming Interface) that once parametrized, can provide more precise altmetrics indicators from articles (Araújo, 2014). Following this suggestion, we performed an altmetrics analysis of a Brazilian LIS journal (DataGramaZero) through the use of APIs of the two largest social media in Brazil in terms of active users: Facebook and Twitter. DataGramaZero (DGZ) is a pioneer publishing venture in the area of the Brazilian LIS and has had an entirely digital format since its inception, as well as being among the core journals in LIS in the nation. However, the absence of a DOI precludes this journal from obtaining results from the use of tools for altmetrics data collection e.g. Altmetrics.com. In addition, as well as not being indexed in international databases, it is not included in the citation results of Web of Science (WoS). This study seeks to conduct an empirical analysis to check the altmetrics measurements in the DGZ articles as an example of the lack of altmetrics in the Global South.

Methods

This exploratory research study carried out an altmetrics analysis of the DGZ journal through the use of APIs of Facebook[i] and Twitter[ii]. The first difficulty in obtaining altmetrics data is how to establish the WWW by using URLs as a database, since the same content may have different URLs. Consultations were parametrized on June 21, 2014, to obtain the URL of all the articles in the journal, together with their quantitative and numerical representation in social media in terms of shared opinions, likes and comments to Facebook and tweets to Twitter, with parameter data output in a JSON format.

Results

Table 1. Mentions per year.

Year	Articles	Mentions	(%)
1999	6	22	1,89
2000	23	30	2,58
2001	26	29	2,49
2002	29	30	2,58
2003	27	23	1,98
2004	29	109	9,36
2005	24	31	2,66
2006	27	56	4,81
2007	26	85	7,30
2008	31	77	6,62
2009	34	68	5,84
2010	34	96	8,25
2011	39	112	9,62
2012	43	119	10,22
2013	32	79	6,79
2014	11	198	17,01
Total	**441**	**1164**	**100**

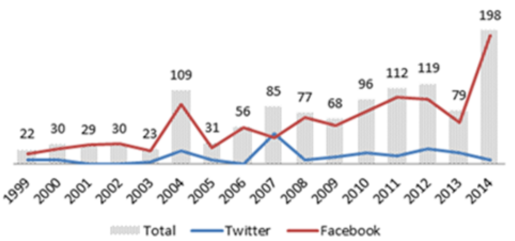


Figure 1. Mentions by Social Media.

Discussion

The DataGramaZero journal provided a total of 441 articles for analysis, published between 1999 to 2014. We identified 1,164 altmetrics data, which are shown on a year-by-year basis in Table 1. The URL <www.dgz.org> has the most widespread altmetrics data with 995 mentions, followed by URL <www.datagramazero.org> with 169 mentions, with an average of 2.63 mentions per article. A total of 211 articles obtained one or more mentions, and 230 did not provide any altmetrics data. Out of the 1,164 total sample, 15.72% of the mentions came from Twitter and 84.28% from Facebook. This result is quite different from those obtained by Alperín (2014), Nascimento & Oddone (2014), and Araújo (2014), where in a comparison made between the two social media, only a low number of mentions were obtained from Facebook or no mentions at all. Figure 1 shows the distribution of the mentions received annually, indicated by the total value (bar) and by the number of occurrences (line) in each social media. With regard to the differences in performance between each social media, the only year in which the mentions in Twitter exceeded the altmetrics data from Facebook was in 2007. In this year, Twitter provided 45 mentions, and Facebook, 40. In the other years Facebook leads the preference for the dissemination of journal articles.

Conclusions

Altmetrics is a relatively new field and has the potential to analyse the information flow from research publications and measure the amount of attention they receive in the social web. However, as Alperín (2014) points out, it seems that there remains an inherent bias within the altmetrics tools which can be attributed to the fact that social media is used to a greater extent by countries in the North, with less representation in the Southern hemisphere. The fact that a large amount of scientific output from the Global South is not indexed in international databases such as WoS, PubMed, Scopus and others, prevents the majority of those journals (including Brazilians) from being included in citation services as well as the default absence found in the journals, e.g. a DOI number also reduces their chances of obtaining altmetrics data in the current scenario, by using available tools.

The purpose of this research is to overcome these barriers by analysing a Brazilian LIS journal with the use of APIs in some social media and conducting an analysis of the individual URLs for each journal article. The altmetrics results showed that the use of APIs can represent an answer to this problem (since the search for URLs is applicable regardless of whether or not the journal has a DOI). This suggests that there is a much higher coverage than is shown by Altmetric.com, in either absolute terms or even individual numbers (for each social media), especially when looking at the performance of Facebook. Although the value of the altmetrics data represents a challenge for researchers who are involved in data collection through APIs, it is an alternative that should be considered.

References

Alperín, J. P. (2014). Open Access Indicators: Assessing Growth and Use of Open Access Resources from Developing Regions: The Case of Latin America, In J.P. Alperín, D. Babini & G. Fishman (Eds.) *Open Access and Scholarly Communications Indicators in Latin America*. (pp 15-78). Buenos Aires: CLACSO. Retrieved January 18, 2015 from http://biblioteca.clacso.edu.ar/clacso/se/20140917054406/OpenAccess.pdf.

Araújo, R. F. (2014). Cientometria 2.0, visibilidade e citação: uma incursão altmétrica em artigos de periódicos da C. Info, In 4º EBBC. Recife: UFPE. Retrieved January 20, 2015 from http://dx.doi.org/10.6084/m9.figshare.1047057.

Nascimento, A. G. & Oddone, N. (2014). Uso de indicadores altmétricos na avaliação de periódicos científicos brasileiros em C. Info, In 4º EBBC. Recife: UFPE. Retrieved January 20, 2015 from http://www.brapci.inf.br/_repositorio/2014/05/pdf_154dd0df78_0014317.pdf.

[i] http://graph.facebook.com
[ii] https://dev.twitter.com

Tweet or publish: A comparison of 395 professors on Twitter.

Timothy D. Bowman

tdbowman@indiana.edu
Department of Information and Library Science, Indiana University Bloomington, Indiana (USA)

Introduction

Twitter is increasingly accepted as a venue to consume and disseminate information (Gruzd et al., 2012) and is used by scholars to share information about (a) professional discussions, (b) network with others, (c) offer help/request help, (d) call attention to other social media involvement, (e) personal discussions, and (f) impression management (Veletsianos, 2012). It is also seen as one of the most promising sources to measure broader research impact in the context of "altmetrics" (Priem et al., 2010)

The idea of examining scholars' interactions and output on the web to understand how events affected societal impact and influence of scholarly work was discussed by Cronin (Cronin, 2005, p. 196) early on, who argued that there would "soon be a critical mass of web-based digital objects and usage statistics on which to model scholars' communication behaviours… and with which to track their scholarly influence and impact."

It is unclear what types of effect tweets have on scholarly production and scholarly impact. To examine whether there is an impact, this work contrasts the tweeting behaviour with the publication activity of 395 professors on Twitter.

Dataset and Methods

Survey of Professors

A survey was sent to 16,862 assistant, associate, and full professors from eight disciplines (Physics, Biology, Chemistry, Computer Science, Philosophy, English, Sociology, and Anthropology) at 62 Association of American Universities-member institutions. The survey asked professors about their a) Twitter use, b) type of account, c) affordance use, and d) demographics. Affordance (Gibson, 1977) is a term used to identify the functional attributes of an object. The primary affordances available in tweets are: mentions, hashtags, URLs, and re-tweets.

Data from 1,910 respondents was collected. It was found that 32% (613) of the respondents reported having at least one Twitter account. Of the 615 scholars with a Twitter account, 445 account handles were verified for 391 of the professors.

Tweet Collection

A sample of tweets from each account was collected using a PHP script on May 19, 2014. A total of 289,934 tweets were collected. Information retrieved included the tweet text, affordance use, the number of total tweets, followers, friends, profile information, and when the account was created.

Research Article Collection

In order to compare tweeting to publication behaviour, the names of the 391 professors with Twitter accounts were used to search a local Web of Science (WoS) database to retrieve their publication and average citation rates. Using a query based on author last name and first name initial(s), 321,033 publication records published during a five-year period from 2009-2013 were retrieved. A final set of 7,734 articles published by the 391 scholars was retained after a manual author name disambiguation was performed.

Results

Comparison of Survey Results

Professors having a Twitter account (n=613; 32%) were compared against those without an account by department, academic age, academic title, ethnicity, and gender. Results show that there were statistically significant relationships between all of these factors. Professors from computer science (50%) had the highest proportion of scholars with account, as compared to those from chemistry (21%) who had the lowest.

Professors who had been at their faculty position from nine to seven years had the highest proportion (41%) and those reporting being at their position six years or less were just below at 39%, whereas only 25% of professors at their positions 10 years or more reported having a Twitter account.

There were 24% of white/Caucasian professors with accounts compared to only 8% for non-whites, and 42% of full professors had an account as compared to 29% of both assistant and associate professors. Gender comparisons found that 28% of males reported being on Twitter compared with 33% of females.

Twitter Use Type

Personal, professional, and mixed use (personal and professional) of Twitter did not differ significantly by ethnicity, academic age, gender, and academic title, however, it was found that there was a significant relationship between Twitter account type and both age and department. Philosophy professors (44%) had the highest number of personal-only accounts, while English professors (60%) had the highest number of mixed accounts. Sociology and computer science professors reported the highest number of professional-only accounts (34%). Professors who identified their age as 35 and under had more professional accounts than expected and professors in the 36 to 45 age range chose the mixed accounts more than expected. Professors who identified as over 46 years old had a higher number of personal accounts than expected.

Tweet Analysis

English professors were found to have a higher median of friends (150), followers (294), and total tweets (410) than all others. Philosophy professors had the lowest median number of total tweets (39), Chemistry professors had the lowest median number of followers (43), and physics professors had the lowest median number of friends (33).
Sociology professors had the most occurrences of hashtags (7.4%) and user mentions (20%) in their tweets, whereas professors from philosophy had the highest use of URLs (1.7%). English professors had the highest number of retweets (291). Philosophy professors (1.96) had the highest average of mean tweets-per-day (TPD) as compared to professors from chemistry (0.52) and physics (0.52) who were found to have the lowest.

Tweet and Publication Activity Comparison

Professors who have a high number of publications had a very low TPD average, whereas those who had a high TPD average tended not to have many publications. In addition, the average citation impact was compared with the mean TPD per scholar (as shown in Figure 1) and there was no relationship found between the two activities.

Discussion and Future Work

Twitter use between scholars in the natural science and social science domains differed. There were also differences in tweet activity by academic title, department, academic age, gender, and age. Looking at impact on publication behaviour, it was found that those professors who had a higher average TPD tended to not publish and those who published quite a bit tended to not tweet very often. Tweeting seemed to have little impact on the citation rate of publications.

Future work should focus on identifying other indicators of scholarly communication and metrics on Twitter and examine the affordance use in tweets in order to better understand how scholars are using the functionality of Twitter to communicate in a professional manner.

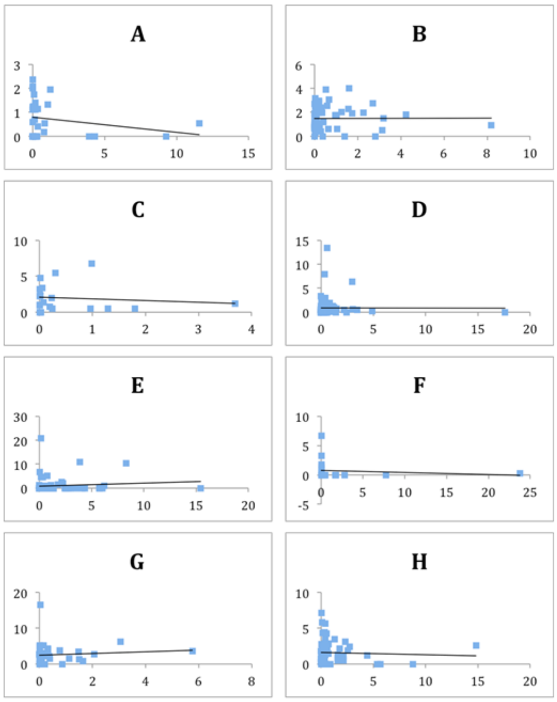


Figure 1. Average citation impact [y-axis] and average mean tweets-per-day [x-axis] for 395 professors in Anthropology [A], Biology [B], Chemistry [C], Computer Science [D], English [E], Philosophy [F], Physics [G], & Sociology [H].

References

Cronin, B. (2005). The hand of science: Academic writing and its rewards. Scarecrow Press.

Gibson, J.J. (1977). The Theory of Affordances, in Shaw, R. and Bransford, J. (Eds.), *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, Lawrence Erlbaum, Hillsdale, NJ, pp. 127–143.

Gruzd, A., Goertzen, M., & Mai, P. (2012). Survey results highlights: trends in scholarly communication and knowledge Dissemination in the Age of Social Media. Halifax, NS, Canada.

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: a manifesto. available at: http://altmetrics.org/manifesto/.

Veletsianos, G. (2012), Higher education scholars' participation and practices on Twitter, *Journal of Computer Assisted Learning*, 28(4), 336–349.

Stratifying Altmetrics Indicators Based On Impact Generation Model

Qiu Junping[1] and Yu Houqiang[2]

[1] *jpqiu@whu.edu.cn*
Research Centre for Chinese Science Evaluation, Wuhan University, Wuhan, 430072, Hubei (China)

[2] *yuhouq@yeah.net*
School of Information Management, Wuhan University, Wuhan, 430072, Hubei (China)

Introduction

Altmetrics has been a shelter for all possible alternative indicators corresponding to traditional citation-based indicators, with extra focus on online indicators. Altmetrics has been discussed in variety of contexts, such as open science (Mounce, 2013), institutional depositories (Adie, Francois, & Nixon, 2014), publishing industry (Piwowar, 2013) and scholarly communication reform (Priem, 2013) etc. Despite the wide recognition and adoption of altmetrics, it has been criticized that stakeholders get confused by so many altmetrics indicators and the exact meaning of each indicator is unclear.

We need a methodology with which the existing altmetrics indicators and future potential indicators can be incorporated and interpreted in a manifest and logical way. To reach this goal, this study will:

(1) firstly, tap into the meaning of impact by demonstrating the multi-faceted nature of it.

(2) secondly, based on multiple empirical researches, introduce an impact generation model that describe how impact becomes perceivable and measurable.

(3) thirdly, making use of the impact generation model, explore the different role that each altmetrics indicator plays in the impact generation process. Combined with the level of engagement theory, altmetrics indicators are stratified and logically ordered.

(4) fourthly, discuss the merits of the stratification based on impact generation model.

Exploring the meaning of impact

To make the idea of scholars' impact more intuitive, Figure 1 was created to demonstrate the composition.

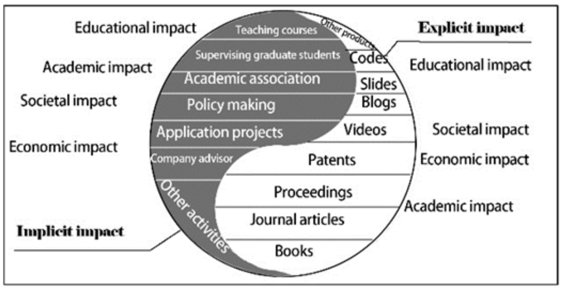


Figure 1. The composition of scholars' impact.

From Figure 1, we see scholars' impact is composed of two parts, the explicit impact derived from scientific products which is usually made public and thus well known by the academia, and the implicit impact brought by non-scientific activities that are often neglected or not well measured by the administrators. In order to achieve explicit impact, scholars keep active in manufacturing various types of scientific products. The major type is publications such as currently prevailing journal articles, books and proceedings. Meanwhile, in the web-native age, novel types thrive. Popular ones include talk videos, slides, codes and blogs. Different types of products are likely to yield different forms of impact. For example, journal articles and proceedings bring more academic impact although they can be used for developing technologies as well. Patents and codes usually benefit to societal or economic impact, and slides and videos will contribute to educational impact.

Impact Generation Model

Inspired by Priem's (Priem & Costello, 2010) theory of capturing the trace of invisible college using altmetrics indicators, and empirical studies (Wang et al., 2014) on exploring the quantitative relationship between different altmetrics data, an impact generation model was proposed to illustrate the process, as shown in Fig. 2. To keep the model as concise as possible, only three principal modules are preserved.

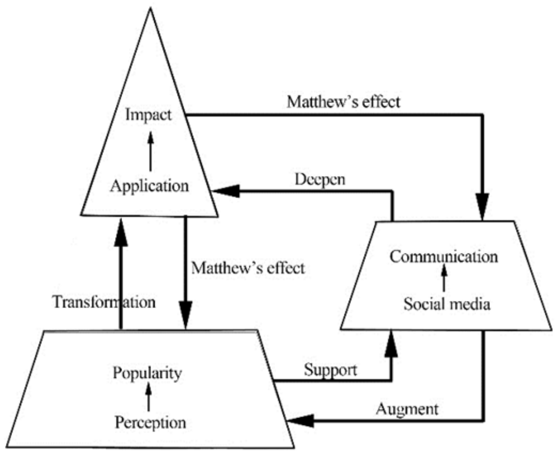


Figure 2. Impact generation model.

The basic philosophy in the model is transformation, which means that the higher level is transformed from the lower level, and the explicit level is transformed from the underlying level. The model has four basic features.

(1) Parallel relationship between the underlying world and the explicit world. Behind popularity is perception. The more scientific products are perceived by people, the greater popularity they gain. Behind impact is application. Whatever the application form is, the more scientific products are used and adopted by the others, the higher impact they obtain. Similarly, behind communication is social media. The more efficient and intelligent the social media is, the more active communication will become.

(2) Transformation from the lower level to the higher level. Only when scientific products get used, or adopted and become sensible, can it be claimed that the scientific products have generated real impact.

(3) Matthew's effect from the higher level to lower level. Once scientific products are used, especially when used successfully, they are likely to be propagated more widely.

(4) Social media (Communication) plays an important role in the model. Social media connects between perception level and application level.

Stratifying altmetrics indicators

An economic analysis of level of engagement phenomenon

It is argued that every type of altmetrics indicator is conveying certain degree of recognition, which is reflected in the level of engagement. It is observed that different altmetric indicators have different difficulty in accumulating data, because of the different cost for users to generate the data. Users' generation cost mainly includes three parts: (1) the time cost; (2) and the reputation cost; (3) and the energy cost. For example, it is much easy for a user to click a paper, but not so easy to read the full-text; It is a little hard for a user to download a paper and save it into his own library, because it takes his future time to deal with it; And it is harder for him to share it with his colleagues, because he is only willing to share those that he think his colleagues will also highly appreciate, in this case, the paper represents his judgment and influence his reputation. The hardest thing to do, perhaps, is citing one's work, because citation is a formal acknowledgement to the work and thus cautiously selected, and usually takes several months to obtain.

Stratification of altmetric indicators

The stratification is conducted in two main steps. The first step is to judge which level the indicator belongs to. The second step is to compare the cost of indicators in each level. The result is demonstrated in Figure 3, where each indicator finds its place in the triangle pyramid.

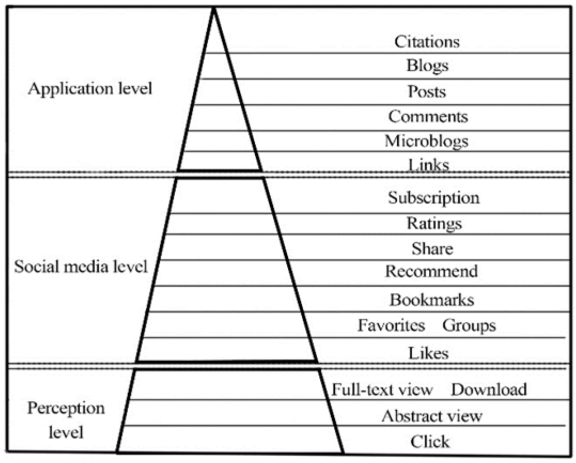


Figure 3. Stratification of altmetrics indicators in the pyramid form.

Merits of the stratification

The stratification has several important advantages compared with the previous classification systems. (1)It clarified the logical relationship between groups of altmetrics indicators. (2) It introduced the transformation relationship between specific indicators. (3) It integrates the previous classifications and helps unify the aggregators' standards in collecting data. (4) It is beneficial in understanding the meaning of impact and the contribution of altmetrics in shaping the current landscape. (5) It can be used to illustrate the relationship between altmetrics and traditional bibliometrics.

Acknowledgments

This paper is supported by the Fundamental Research Funds for the Central Universities NO. 2014104010201.

References

Adie, E., Francois, S., & Nixon, W. (2014). Altmetrics in practice: How are institutional repositories using altmetrics today? *Open Repositories*.

Mounce, R. (2013). Open access and altmetrics: Distinct but complementary. *JASIST, 39*(4), 14-17.

Piwowar, H. (2013). Altmetrics: Value all research products. *Nature, 493*(7431), 159-159.

Priem, J. (2013). Scholarship: Beyond the paper. *Nature*, 495(7442), 437-440.

Priem, J, & Costello, K L. (2010). How and why scholars cite on Twitter. *JASIST, 47*(1), 1-4.

Wang, X., Liu, C., Fang, Z C, & Mao, W L. (2014). From Attention to Citation, What and How Does Altmetrics Work? arXiv preprint arXiv:1409.4269.

ISSI 2015

CITATION AND CO-CITATION ANALYSIS

Citation Type Analysis for Social Science Literature in Taiwan

Ming-yueh Tsay

mytsay@nccu.edu.tw
Graduate Institute of Library, Information and Archival Studies, National ChengChi University (Taiwan)

Abstract
Through citation analysis, this study explored the distribution of document type, language and publication year for citations in social science journals. Samples were research articles published in 2010 from first-rank journals, as assessed by the Department of Humanities and Social Sciences, National Science Council and indexed in Taiwan Social Sciences Citation Index (TSSCI). The section in which citations appeared, namely introductions, methodologies, results, and conclusions, were also examined. Conclusions and suggestions are made based on the research results and interdisciplinary comparisons. For social science studies in Taiwan, the major findings are as follows: 1. Journals and books were the most cited materials, and English was the language of most citations. 2. Social scientists in Taiwan tended to cite materials published within 10 years with a citing half-life of approximately 11 years. 3. The ratio of articles following the IMRAD format was high in Taiwan social science journals. 4. Citations in these social science journals occurred most frequently in the introduction section, while they occurred least frequently in the conclusions. 5. Social scientists mostly cite to set the stage for their current studies. 6. The citation type is highly related to the citation location.

Conference Topic
Citation and co-citation analysis

Introduction
Since the social sciences are associated with human society, its patterns, where it goes and how it works, it can enrich the values and contents of our lives. In contrast, the "hard sciences" have been the focus of attention with the rapid growth of technology grew, and the social sciences have received less attention. This has led to a lack of balance between academic and technical research in many developing countries. To gain attention and support from governments and the public, social scientists need to promote their research outcomes and impacts much more effectively via the presentation and communication of their scholarly articles.

A research article may consist of body text and references; the former is the citing article, and the latter are cited articles. Relations between the citing and the citied may explain the interaction, development and communication among disciplines, and can reveal current research interests and future trends. Citations have multiple roles and unique functions in scholarly communication; for example, a cited article may present broader research contents, explain methods applied in a research or provide information and discussion that support a specific perspective.

The importance of journal articles for scholarly communication and academic assessment motivates the present study on Taiwanese social science journal articles to explore and compare their characteristics and types of citations via methods of bibliometric and citation analysis. The research outcomes may improve the knowledge of citation, and serve as reference for future empirical researches for the social science studies in Taiwan.

Other Citation Studies
Citations have been studied using context or content analysis, whereby the analysis determines the citation type based on the surrounding text. Frost (1979) mentioned the complexity of citation function and that the classification of citation function and proper schemes for classification received little attention in citation studies. To explore the nature of

citation use, some various schemes of classification for different disciplines have been developed to explain the functions of citations and the relations between body text and citations.

In Moravcsik and Murugesan's (1975) study physics citations fall into the "applied/used" category with 60% and 40% of the citations being general acknowledgements. In the study of Voos and Dagaev (1976), inspected the locations of each citation in sample articles and found that articles of biology and medicine were mostly cited within two to three years after their publication, and were cited the most in the introduction section, and next in the discussion section.

Peritz (1983) selected a variety of social science journals in which the basic methodologies of empirical social research were used and analyzed into the categories of a citation classification scheme. That study revealed that generally, "setting the stage for the present study" citations rank first. To carry out the reliability citation classification scheme, Peritz further investigated the association between classification and location and found that the marginal frequencies of the location introduction, methods and discussion were fairly close to the frequencies of the classification categories of setting the stage, methodology, and comparison and argument, respectively.

More recently, Harwood (2008) interviewed six informants who were computer scientists and six who were sociologists on the functions of citations in their writing. His findings reveal that position, supporting, and credit are relatively frequent across both disciplines, although the engaging function is far more frequent in the sociology texts.

Case and Miller (2011) investigated the citation practice of a group of citing authors with an interest in bibliometric or scientometric research, finding that the most popular reason was "this reference is a 'concept marker'," which distantly followed by "reviews prior work in the area" and other reasons.

The above literature survey shows there have been many studies investigating citation category and citation practices, which are likely to vary from discipline to discipline. This motivates the present study to further explore the citation type of articles cited in the social science journals published in Taiwan.

Research Method and Limitation

The journals selected in this research were six first-ranked journals indexed in the Taiwan Social Sciences Citation Index (TSSCI) in the disciplines of sociology, education, psychology, political science, economics and management. In this study, it is assumed that the first ranked journal of each discipline may represent the research characteristics of that discipline.

Articles published in 2010 and following the IMRAD format were selected as research samples, though articles published earlier than 2010 were also collected if there were insufficient samples. The titles of journals and number of articles selected for the six disciplines were: sociology, *Taiwanese Journal of Sociology*, 15 articles (2008-2010); education, *Bulletin of Educational Psychology*, 31 articles (2010); psychology, *Chinese Journal of Psychology*, 16 articles (2010); political science, *Taiwan Political Science Review*, 16 articles (2008-2010); economics, *Academia Economic Papers*, 13 articles (2010); management, *Journal of Management*, 25 articles (2010).

In the present study, if introductions and literature reviews were in two different sections, they were considered as an introduction in combination; if results and discussion were in one section, they would be categorized as result. Citations were categorized, on the basis of the classification scheme proposed by Peritz (1983), which requires little subjective judgment and is easy to carry out even without in-depth knowledge of the subject field.

Full texts and references of all 116 research articles were downloaded from online databases or photocopied from printed journal and processed with Excel (Microsoft, U.S.) into bibliographical files. Employing bibliometric techniques and citation analysis, this study explored article type of journal, language of citation, citation years, document types of citations, citation types and locations of citations, the relations between citation type and location, and comparison among six disciplines of social sciences in Taiwan.

This study conducted purposive sampling to acquire journal articles for citation analysis, whose results might thus be limited indeed and less representative for each or the whole of humanities disciplines. Nevertheless, the current study aims to distinguish the meaningful characteristics of article structures, citation locations, and citation types of the six social science disciplines of Taiwan; also the method of "citation content analysis" used in this study to explore the nature of citation types is qualitative and justified by the attempt to interpret the existing phenomena. In the above senses, purposive sampling and unequal sample size seemed to be acceptable limitations.

Results

In this study, citation characteristics and locations in body texts are discussed according to article type of journal, language of citation, year of the highest citation and citation half-life, document type of citation, citation location and citation type.

Article Type of Journal

Papers published in social science journals in Taiwan are mainly divided into research articles and review articles. In general, research articles comply with the IMRAD format. The ratio of articles following the IMRAD format was high in the social sciences. Table 1 demonstrates review and research articles, both appeared in the disciplines of political science and sociology, while journals in the fields of psychology, education, economics and management preferred research articles.

Table 1 shows that, among the six disciplines, education, economics and management composed completely (100%) of research articles that follow the format of IMRAD.

Table 1. Article types in social science journals of Taiwan.

Discipline (Journal name)	*Papers*	*English article*	*Chinese review article*	*Chinese research article*	*% of Chinese Research article*
Political Science (*Taiwan Political Science Review*)*	30	1	13	16	55.2%
Sociology (*Taiwanese Journal of Sociology*)*	25	0	10	15	60.0%
Education (*Bulletin of Educational Psychology*)	32	1	0	31	100.0%
Psychology (*Chinese Journal of Psychology*)	23	4	3	16	84.2%
Economics (*Academia Economic Papers*)	18	5	0	13	100.0%
Management (*Journal of Management*)	30	5	0	25	100.0%
Total	158	16	26	116	81.7%

*Semi-annual journal. Sample articles of these journals were dated back to 2008 from 2010; samples of other journals were articles published in 2010.

Language of Citation

Materials in Chinese and English were the major source of references cited in social science articles, with the former accounting for 21.5% and the latter 78% of the total references collected. Most of the references in economics (93.5%) and management (92.1%) were English papers, while Chinese articles were infrequently used in both disciplines. Domestic research articles and reference materials, however, were used quite often by scholars of sociology (30.7%) and political science (43.4%).

Year of the Highest Citation and Citing Half-Life

Table 2 reveals the year of highest citation, citation age and citing half-life of articles in sample journals. Citing half-life refers to the time span from the current year to the year whose accumulated number of citations accounts for 50% of total citations in the journal. For example, the citing half-life of *Chinese Journal of Psychology* shown in Table 2 was 11, indicating that half of its citations were younger than 11 years as the citing articles being published. The time span of citation half-life reflects the currency of cited materials: the longer the citing half-life, the older the cited materials, and vice versa.

Table 2. Distribution of year of the highest citation, citation age of the highest citation and citing half-life.

Discipline (Journal name)	*Year of the highest citation*	*Citation Age*	*Citing half-life*
Political Science (*Taiwan Political Science Review*)*	2007	4	10.6
Sociology (*Taiwanese Journal of Sociology*)*	2006	5	11.5
Education (*Bulletin of Educational Psychology*)	2005	6	11.2
Psychology (*Chinese Journal of Psychology*)	2006	5	11.0
Economics (*Academia Economic Papers*)	2007	4	11.4
Management (*Journal of Management*)	2004	7	11.2
Average	2006	5.2	11.2

Based on the year of highest citations, the number of citations earlier than 2004 is decreasing for earlier articles. In other words, the older the articles were, the fewer citations they received. In general, for articles published in 2010 the peak of citations fell between 2004 and 2007 that suggests citations that received from the sample journals reached a peak after four to seven years of its publication, five years in most cases. A large number of citations came from articles published in the recent several years, indicating that social scientists have a tendency to cite the most recent articles. In the social science fields, scholars tended to cite materials with a citing half-life of approximately 11 years. For social scientists in most disciplines, 50% of their research needs could be satisfied by articles published after 2000, and the tendency to cite the most recent articles indicates the social science research depends on more current literature.

Document Type of Citation

In the six top journals selected as samples in this study, there were 116 Chinese articles following the IMRAD format, citing 6,063 references to the bibliographic files built by this study. According to the bibliographic data collected, journals and books were the most frequently cited, accounting for 88% (journals 65% and books 23%) of all types of cited materials. The uses of journals and of books in economics were quite different, with the highest interval over 76%, in which journals accounted for 82% of cited materials while books accounted for 6%. The second-highest difference between citations of journals and of books was in management, where journals accounted for 80% of the citations, which was 67% higher than books. For other disciplines, such as social science (journals 53% vs. books 35%) and political science (journals 49% vs. books 34%), the differences between the use of journals and books were not as great, indicating that they have a closer value in both disciplines. On the average, over all types of documents, social scientists preferred to use journals in exploration and support of their own research.

Aside from the citations of journal articles and book materials, the number of theses and dissertations cited in the journal of education was higher than those in journals of other disciplines. Online resources such as websites or electronic files were cited more frequently in the political science journal, suggesting that political scientists use more digital literature as references in their research. Research reports were cited more in the journal of economics than in other disciplines, which indicates that economists tended to prove or support their own research by data or results provided by research reports. Furthermore, the fact that economists and scholars of management cited a few unpublished manuscripts and working papers showed the significance of informal and unpublished materials to these two disciplines.

Citation Location

The number and location of citations from the 116 articles complying with the IMRAD format were calculated to analyze the distribution of citations in structured research articles. There were 11,149 citations collected in the section of introduction (literature review included), methods and materials, results, and discussion.

The distribution of citations in different sections of an article may help to determine the status, research patterns and characteristics of a discipline. As Table 3 shows, citations appeared the most in the introduction section of articles in every discipline of social science. The Introduction may include literature reviews, and both sections need a few references for proving points or serving as motivations. In the six disciplines of social science, the highest number of citations in the introduction sections occurred in the journals of sociology and political science, while the lowest was in the journal of economics. For the method section, scholars of economics and management cite more frequently in the section of methods and materials. In contrast, the sociologists cite the least frequently.

Table 3. Distribution of citation location in social science journals of Taiwan.

Discipline	*Introduction*		*Methodology & Materials*		*Results*		*Discussion*		*Total*	
	No.	*%*	*No.*	*%*	*No.*	*%*	*No.*	*%*	*No.*	*%*
Management[1]	1,799	65.4	480	17.3	256	9.3	216	7.9	2,751	24.7
Economics[2]	499	54.6	164	17.9	189	20.7	62	6.8	914	8.2
Political Sci.[3]	931	70.5	171	13.0	165	12.5	53	4.0	1,320	11.8
Psychology[4]	1,048	58.6	198	11.1	222	12.4	320	17.9	1,788	16.0
Education[5]	1,888	64.1	282	9.6	295	10.0	481	16.3	2,946	26.4
Sociology[6]	993	69.4	63	4.4	254	17.8	120	8.4	1,430	12.8
Total	7,158	64.2	1,358	12.2	1,381	12.4	1,252	11.2	11,149	100

1. Journal of Management; 2. Academia Economic Papers; 3. Taiwan Political Science Review; 4. Chinese Journal of Psychology; 5. Bulletin of Educational Psychology; 6. Taiwanese Journal of Sociology

In the results section, economists tended to cite more articles for comparison and contrast. Aside from economics, the number of citations in the results section of the sociology journal also high. In the management journal, descriptive statistics and quantitative analysis may be the major causes of its lower number of citations in the results section.

In the discussion section, the number of citations may reflect scholars' degree of concern about deliberations and evaluation of research outcomes. The top two numbers of citations in discussion section occurred in the journals of psychology and education.

Citation Type

In addition to the distribution of citation location, Peritz's classification scheme of citation type is used to classify articles cited in the sample journals. Mapping was made to inspect the relations between citation type and citation location and to analyze the differences among the six disciplines. The eight categories of citation classification scheme proposed by Peritz (1983, pp.304-305) are: 1. Setting the stage for the present study; 2. Background information; 3. Methodological; 4. Comparative; 5. Argumentative speculative, hypothetical; 6. Documentary; 7. Historical and 8. Casual.

Table 4. Distribution of citation type.

Citation type	*Sociology*		*Education*		*Psychology*		*Political Science*		*Economics*		*Management*		*Total*	
	No.	*%*	*No.*	*%*	*No.*	*%*	*No.*	*%*	*No.*	*%*	*No.*	*%*	*No.*	*%*
Setting the stage for the present study	271	56.7	153	53.5	235	58.6	311	56.3	133	48.9	411	63.8	1,514	57.5
Background information	44	9.2	15	5.2	21	5.2	23	4.2	9	3.3	13	2.0	125	4.7
Methodological	33	6.9	34	11.9	54	13.5	85	15.4	85	31.3	109	16.9	400	15.2
Comparative	70	14.6	35	12.2	68	17.0	46	8.3	38	14.0	72	11.2	329	12.5
Argumentative, speculative, hypothetical	45	9.4	48	16.8	23	5.7	17	3.1	5	1.8	38	5.9	176	6.7
Documentary	15	3.1	0	0.0	0	0.0	66	12.0	2	0.7	1	0.2	84	3.2
Historical	0	0.0	1	0.3	0	0.0	0	0.0	0	0.0	0	0.0	1	0.0
Casual	0	0.0	0	0.0	0	0.0	4	0.7	0	0.0	0	0.0	4	0.2
Total	478	-	286	-	401	-	552	-	272	-	644	-	2,633	-

Based on Table 4, the highest percentage of citations classified as "setting the stage for the present research" appeared in the journal of management (64%), and the lowest in the journal of economics (49%). Compared to other types of citation, citations that set the stage for the present study were significantly high in all six disciplines. The citation type of "background information" was most frequently found in the journal of sociology, while it was least frequent in the journal of management. The journal of economics contained the most methodological citations, which accounted for 31% of total citations, while the journal of sociology the least, which accounted for 7%; the interval between was rather large. Comparative citations were most found in the journal of psychology (17%) and the least in the journal of political science (8%). The journal of education included the most citations (17%), which were used in the presentation of argument, speculation, and hypothesis while the journal of political science the least (merely 3%). Documentary citations accounted for 12% of total citations in the journal of political science, which was the top among the six disciplines; whereas there was no such type of citations found in the journals of education and psychology. The citation types of "historical" and "casual" were hardly found in the journals of six disciplines, with only one historical citation in the journal of education and four casual citations in the journal of political science.

The distribution of citation type may reveal the research characteristics of a certain disciplines. For example, scholars of management tend to cite a large amount of literature to support or motivate their own research, whereas economists cite more methodological materials in their works, which indicates that research methods are valued more in economics. Political scientists tended to cite more raw data to support their studies; whereas scholars of education cited more articles for argumentation, speculation, and hypothesis. Comparative citations appeared the most in the journal of psychology, suggesting that psychological researchers tend to introduce other research in their own studies for comparison, correction, or corroboration.

Citation Type and Citation Location

According to Peritz's study, citation type was highly relevant to citation location. In this study, therefore, the relation between citation type and citation location in the six discipline sample journals was analyzed as follows.

Sociology

As Table 5 shows, in the journal of sociology, the number of citations that set the stage for the present study was 271, accounting for 56.7% of the total citations. Comparative citations accounted for 14.6% of the total citations, suggesting that the materials being cited in the journal articles were used to describe or support the present research. The citation type of "setting the stage for the present study" appeared primarily in the introduction section, while methodological citations that introduced the process of other research were mostly in the methods and materials section. In the results section, comparative, argumentative, speculative, and hypothetical citations accounted for the greatest number of citations. In the discussion session, comparative citations comprised the major part of total citations.

Table 5. Citations in Taiwanese Journal of Sociology by category and location.

Location / Category	Introduction		Methodology & Materials		Results		Discussion		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Setting the stage	271	86.0	0	0.0	0	0.0	0	0.0	271	56.7
Background information	30	9.5	1	4.0	13	18.8	0	0.0	44	9.2
Methodology	7	2.2	19	76.0	7	10.1	0	0.0	33	6.9
Comparative	0	0.0	0	0.0	23	33.3	47	68.1	70	14.6
Argumentative, speculative, hypothetical	3	1.0	0	0.0	23	33.3	19	27.5	45	9.4
Documentary	4	1.3	5	20.0	3	4.3	3	4.3	15	3.1
Historical	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Casual	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Total	315	100.0	25	100.0	69	100.0	69	100.0	478	100.0

Education

In the journal of education, the citation type of "setting the stage for the present study" accounted for the largest percentage of the total citations, 53.5%, as shown in Table 6. The distribution of methodological citations, comparative citations, and argumentative, speculative and hypothetical citations was rather even. Similar to the distribution in the sociology journal, all of the citations that set the stage for the present study appeared in the introduction section, and the citations in methods and materials section were mostly methodological citations, while there were few citations in the results section. As for the discussion part, the numbers of comparative, argumentative, speculative and hypothetical citations, especially the last three types, greatly exceeded other types of citation, indicating that scholars of education often introduced other research for detailed exploration, or made further inference based on previous studies.

Table 6. Citations in Bulletin of Educational Psychology by category and location.

Category \ Location	Introduction		Methodology & Materials		Results		Discussion		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Setting the stage	153	90.5	0	0.0	0	0.0	0	0.0	153	53.5
Background information	12	7.1	0	0.0	0	0.0	3	3.8	15	5.2
Methodology	3	1.8	30	100.0	1	12.5	0	0.0	34	11.9
Comparative	0	0.0	0	0.0	1	12.5	34	43.0	35	12.2
Argumentative, speculative, hypothetical	0	0.0	0	0.0	6	75.0	42	53.2	48	16.8
Historical	1	0.6	0	0.0	0	0.0	0	0.0	1	0.3
Total	169	100.0	30	100.0	8	100.0	79	100.0	286	100.0

Psychology

In the journal of psychology, as Table 7 presented, over half of its citations were classified as the type of "setting the stage for the present studies" (58.6%). In the discussion section, comparative citations accounting for 71% of total citations appeared in the discussion section, which suggests that psychologists tend to cite other materials as comparisons to examine whether their research results were consistent with previous studies, or to correct previous research and hereafter propose their own unique results.

Table 7. Citations in Chinese Journal of Psychology by category and location.

Category \ Location	Introduction		Methodology & Materials		Results		Discussion		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Setting the stage	235	93.6	0	0.0	0	0.0	0	0.0	235	58.6
Background information	16	6.4	4	7.3	0	0.0	1	1.3	21	5.2
Methodology	0	0.0	48	87.3	6	30.0	0	0.0	54	13.5
Comparative	0	0.0	3	5.5	12	60.0	53	70.7	68	17.0
Argumentative, speculative, hypothetical	0	0.0	0	0.0	2	10.0	21	28.0	23	5.7
Total	251	100.0	55	100.0	20	100.0	75	100.0	401	100.0

Political Science

From Table 8, it is clear that "setting the stage for the present study" citations were the most numerous of the eight types of citation, accounting for 56% of the total citations in the journal of political science. The second most numerous were the methodological citations, though they comprised only 15% of total citations, while the percentage of other types of citations was even lower. Interestingly, political scientists cited much more statistical data in the introduction section, which indicates that they tended to use quantitative data or factual information to support their studies when writing introduction and literature review. As for the other locations, comparison was often made in the results section, while citations in the discussion section mostly served as bases for inference.

Table 8. Citations in Taiwan Political Science Review by category and location.

Category \ Location	Introduction		Methodology & Materials		Results		Discussion		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Setting the stage	305	79.0	6	6.3	0	0.0	0	0.0	311	56.3
Background information	16	4.1	6	6.3	1	1.9	0	0.0	23	4.2
Methodology	7	1.8	73	76.0	5	9.3	0	0.0	85	15.4
Comparative	0	0.0	3	3.1	40	74.1	3	18.8	46	8.3
Argumentative, speculative, hypothetical	0	0.0	0	0.0	4	7.4	13	81.3	17	3.1
Statistical data	58	15.0	4	4.2	4	7.4	0	0.0	66	12.0
Casual	0	0.0	4	4.2	0	0.0	0	0.0	4	0.7
Total	386	100.0	96	100.0	54	100.0	16	100.0	552	100.0

Economics

Though the "setting the stage for the present study" citations were more numerous than other types of citations in the journal of economics, its percentage was a bit lower than in other disciplines, accounting for only 49% of all the citations in the journal. Table 9 also shows that economists cited more methodological materials, accounting for 31% of all citations, indicating a preference for empirical study in the field of economics. Models or methods proposed by other research were frequently found in the studies of economics, and comparative citations were mostly made in the section of results, which is consistent with the inference that economists were used to comparing their research results with previous studies. However, few citations in the discussion section revealed little of the characteristics of citation types in the journal of economics.

Table 9. Citations in Academia Economic Papers by category and location.

Category \ Location	Introduction		Methodology & Materials		Results		Discussion		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Setting the stage	129	92.8	2	3.8	2	2.7	0	0.0	133	48.9
Background information	5	3.6	0	0.0	4	5.3	0	0.0	9	3.3
Methodology	5	3.6	49	94.2	29	38.7	2	33.3	85	31.3
Comparative	0	0.0	0	0.0	37	49.3	1	16.7	38	14.0
Argumentative, speculative, hypothetical	0	0.0	0	0.0	2	2.7	3	50.0	5	1.8
Statistical data	0	0.0	1	1.9	1	1.3	0	0.0	2	0.7
Total	139	100.0	52	100.0	75	100.0	6	100.0	272	100.0

Management

The relations between citation type and location in the journal of management can be seen in Table 10. The percentage of citations that set the stage for the present study was comparatively high (64%) in the journal of management, which was the only discipline whose percentage exceeded 60% among all six disciplines discussed in this study. Unlike economists, who were found to care more about methods and materials, scholars of management focused more on literature reviews, tending to project the importance of their research questions by contrasting them with previous studies. Yet they still valued the implementation of research methods from other studies, according to the second top percentage (17%) of methodological citations. Comparative, argumentative, speculative and hypothetical citations also appeared in the section of discussion, while comparative citations accounted for more percentage (11%) of total citations in the journal of management.

Table 10. Citations in Journal of Management by category and location.

Category \ Location	Introduction		Methodology & Materials		Results		Discussion		Total	
	No.	%	No.	%	No.	%	No.	%	No.	%
Setting the stage	400	97.3	10	8.3	0	0.0	1	1.6	411	63.8
Background information	6	1.5	7	5.8	0	0.0	0	0.0	13	2.0
Methodology	5	1.2	90	75.0	11	22.4	3	4.7	109	16.9
Comparative	0	0.0	12	10.0	22	44.9	38	59.4	72	11.2
Argumentative, speculative, hypothetical	0	0.0	0	0.0	16	32.7	22	34.4	38	5.9
Statistical data	0	0.0	1	0.8	0	0.0	0	0.0	1	0.2
Total	411	100.0	120	100.0	49	100.0	64	100.0	644	100.0

Table 11. Citations in social science journals in Taiwan by category and location.

Category \ Location	Introduction (%)		Methodology & Materials (%)		Results (%)		Discussion (%)		Total (%)	
Setting the stage	89.87	3.07	0.45	0.27	56.3	89.87	3.07	0.45	0.27	56.3
Background information	5.37	3.9	4.33	0.85	4.85	5.37	3.9	4.33	0.85	4.85
Methodology	1.77	84.75	20.5	6.33	15.98	1.77	84.75	20.5	6.33	15.98
Comparative	0	3.1	45.68	46.12	12.88	0	3.1	45.68	46.12	12.88
Argumentative, speculative, hypothetical	0.17	0	26.85	45.73	7.12	0.17	0	26.85	45.73	7.12
Documentary	0.22	3.33	0.72	0.72	0.52	0.22	3.33	0.72	0.72	0.52
Historical	0.1	0	0	0	0.05	0.1	0	0	0	0.05
Statistical data	2.5	1.15	1.45	0	2.15	2.5	1.15	1.45	0	2.15
Casual	0	0.7	0	0	0.12	0	0.7	0	0	0.12
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

In sum, the percentages of "setting the stage for the present study" citations ranked first in the journals of all six disciplines, with management accounting for 63.8%, psychology 58.6%, sociology 56.7%, political science 56.3%, education 53.5%, and economics 48.9%. The percentage of methodological citations to total citations was 15.2%, which made the second high among the six journals, with economics accounting for 13.3%, management 16.9%, and political science 15.4%. As for the comparative citations, psychology (17%) and sociology (14.6%) covered more than other disciplines, while education exceeded other disciplines in the argumentative, speculative and hypothetical citations, with a percentage of 16.8%.

In Peritz's study, the citation type was highly relevant to the citation location, as confirmed by the results of this research shown in Table 11. In the introduction section, most citations belonged to the category of "setting the stage for the present study"; in the section of methods and materials, methodological citations appeared the most; as for the section of results and discussion, although the distribution of citation types varied among the six disciplines, comparative, argumentative, speculative, and hypothetical citations were the most on the average. Overall, the research outcomes indicated that most social scientists of Taiwan complied with international writing format, and confirmed the hypothesis proposed by Peritz that the citation location was highly relevant to the citation type.

Summary and Discussions

This study explores and compares the distribution of article types of journals, languages of citation, citation years, document types of citations, citation types and locations of citations among citations in the top social science journals of six disciplines published in Taiwan and indexed in the Taiwan Social Sciences Citation Index (TSSCI). The following conclusions may be drawn from the results.

1. Journals and books were the most cited materials; English language articles were the most cited in social science studies in Taiwan.
2. Social scientists in Taiwan tended to cite materials published within the past 10 year, most citations in the sample journals were for articles with four to seven years of the journal publication, indicating that social scientists in Taiwan tend to cite the most recent articles.
3. The ratio of articles following IMRAD format was high in social science journal in Taiwan, suggesting that the top social science journals comply strictly with the IMRAD format of structured articles in Taiwan.
4. In Taiwan, citations in social science journals occurred the most in the introduction section, while the conclusions section had the least: The distribution of citations in different sections of an article may indicate the status and characteristics of a research domain. In this study, citations occurred most frequently in the introduction section for each of the social science disciplines. The introduction may include research background and literature review, and both sections need quite a few references for proving points or indicating motivation. For the methods section, economics and management had high percentage of citations, indicating that scholars in these two disciplines were used to adopting models, designations or operations from previously published research. In the results section, economists and psychologists tended to cite more articles for comparison and contrast. In general, citations appeared least frequently in the conclusions section, though the percentage rates were still a little higher in psychology and education, revealing their concern for further discussion and evaluation of research results.
5. Social scientists mostly cite to set the stage for their present studies: The “setting the stage for the present study” citations were the most frequently used in the sampled social science journals, accounting for 57.5% of all citations. From the distribution of citation type, it is clear that social scientists tended to cite in order to provide support or motivation for their own studies, which as shown by the large number of “setting the stage for the present study” type of citations. Scholars of economics, management and political science used to introduce methods and materials to compare or verify their findings. Psychologists and sociologists tended to compare their research results with previous studies, whereas scholars of education emphasized discussion greater than other sections.
6. Citation type is highly relevant to the citation location, which is consistent with the findings of Peritz’s study.

In this study, citation characteristics of social scientists in Taiwan were analyzed via bibliographic data such as types of cited materials and languages of citations. The results revealed the citation characteristics and information need of Taiwan’s social scientists, which could be valuable in collection development of libraries or refinement of information services. Under the assumption that citations indicate the actual use of materials, the distribution of publication years and citing half-life may serve as evidence for libraries to order or suspend information resources (electronic journals, for instance), which could help to achieve similar goals on better budget allocation. Providing further exploration and examination of citations, this study is also expected to provide a better understanding of citation nature, and is anticipated to serve as a basis for future empirical studies.

There are limitations for the citation type determination by the textual analyst on the basis of the surrounding text. This is because, first, citation types may not be apparent simply by studying the text and, second, effective analysis sometimes requires specialist knowledge in the discipline of the texts being studied. Therefore, conducting an interview study with authors of the text to obtain their own views of citation types is suggested for further study. The small number of samples involved in this study preclude from making confident generalizations regarding the frequency of the citation types across these social science disciplines as a whole. Thus, the collection and analysis of a larger sample size is also suggested for further study.

Conclusion and Suggestion

The study is still to be improved owing to its restrictions and limitations. For better interpretation of the research trend, paradigm shifts and citation distribution of social sciences, it is suggested that the time frame, scope and quantity of sample collection be extended, including citations from both domestic and foreign articles. Co-research with experts and scholars in concerning disciplines are recommended as well. Even more, to reach a fuller apprehension of research features in academia by means of citation characteristics, samples in humanities and sciences may be examined in the future studies. Though Periz's classification scheme is known for its simplicity and directness, it is not quite suitable for those non-empirical studies. However, the Citation Content Analysis (CCA) framework proposed by Zhang, Ding and Milojevic (2013) may serve as solution to the problem, since it adopts both syntactic and semantic measurement of citation, which thus makes cross-field comparison possible. As for the essence of citation, the purposes and motives of citation are also valuable topics for further studying.

Acknowledgments

This work was supported by grant NSC1007-2410-H-004-153-MY2 from the National Science Council, Taiwan, R.O.C. Bibliometric data collected by Min-yee Lee, Graduate Institute of Library, Information and Archival Studies, National Chengchi University, Taiwan is very much appreciated.

References

Case, D.O. & Miller, J.B. (2011). Do bibliometricians cite differently from other scholars? *Journal of the American Society for Information Science and Technology*, *62*(3), 421-432.

Frost, C.O. (1979). The use of citations in library research: a preliminary classification of citation functions. *Library Quarterly*, *49*(4), 399.

Harwood, N. (2008). An interview-based study of the functions of citations in academic writing across two disciplines. *Journal of Pragmatics*, *41*(3), 497-518.

Moravcsik, M.J. & Murugesan, P. (1975). Some results on the function and quality of citations. *Social Studies of Science*, *5*, 86-92.

Peritz, B C. (1983). A classification of citation roles for the social sciences and related fields. *Scientometrics*, 5, 303-312.

Voos, H. and Dagaev, K.S. (1976). Are all citation equal? Or, did we op cit your idem? *The Journal of Academic Librarianship*, *1*(6), 19-21.

Zhang, G, Ding, Y. & Milojevic, S. Citation Content Analysis (CCA): A Framework for Syntactic and Semantic Analysis of Citation Content. *Journal of the American Society for Information Science and Technology*, *64*(7), 1490-1503.

University Citation Distributions

Antonio Perianes-Rodriguez[1] and Javier Ruiz-Castillo[2]

[1] *antonio.perianes@uc3m.es*
Universidad Carlos III, Department of Library and Information Science, SCImago Research Group, C/ Madrid, 128, 28903 Getafe, Madrid (Spain)

[2] *jrc@eco.uc3m.es*
Universidad Carlos III, Departamento de Economía, C/ Madrid, 126, 28903 Getafe, Madrid (Spain)

Abstract
In this paper we investigate the characteristics of the citation distributions of the 500 universities in the 2013 edition of the CWTS Leiden Ranking. We use a WoS dataset consisting of 3.6 million articles published in 2003-2008 with a five-year citation window, and classified into 5,119 clusters. The main findings are the following four. Firstly, The universality claim, according to which all university citation distributions, appropriately normalized, follow a single functional form, is not supported by the data. Secondly, nevertheless, the 500 university citation distributions are all highly skewed and very similar. Broadly speaking, university citation distributions appear to behave as if they differ by a relatively constant scale factor over a large, intermediate part of their support. Thirdly, citation impact differences between universities account for 3.85% of overall citation inequality. However, these differences are greatly reduced when university citation distributions are normalized using their MNCS values as normalization factors. Finally, the above results have important practical consequences. On one hand, we only need a single explanatory model for the single type of high skewness characterizing all university citation distributions. On the other hand, the similarity of university citation distributions goes a long way in explaining the similarity of the university rankings obtained with the MNCS and the top 10% indicator.

Conference Topic
Citation and co-citation analysis

Introduction
Universities constitute a key vehicle in the production of knowledge in contemporary societies. However, the evaluation of the quality, or the relevance of the research done by universities in a myriad of scientific fields is a very difficult problem. For the assessment of the performance of research units of all types during the last decades, academic bodies, public officials in charge of science policy, and specialists in the field of Scientometrics have been paying increasing attention to one observable aspect of research in all fields: the citation impact of publications in the periodical literature.

In this paper, we focus on this aspect of research for the 500 universities included in the 2013 edition of the CWTS Leiden Ranking (LR universities) (Waltman et al., 2012a). We use a Web of Science (WoS) dataset consisting of 3.6 million publications in the 2005-2008 period, the citations they receive during a five-year citation window for each year in that period, and a classification system consisting of 5,119 clusters (Ruiz-Castillo & Waltman, 2015).

The construction of university citation distributions in the all-sciences case requires the prior solution of two methodological problems: the assignment of responsibility for publications with two or more co-authors belonging to different institutions, and the aggregation of the citation impact achieved by research units working in different scientific clusters. We solve these problems using a fractional counting approach in the presence of co-authorship, and the standard field-normalization procedure where cluster mean citations are used as normalization factors.

Once these two problems have been solved, specialists typically debate the properties of alternative citation impact indicators. In this paper, we study a basic aspect of the research evaluation problem that comes *before* the comparison of the advantages and shortcomings of specific indicators, namely, the characteristics of the university citation distributions themselves. These distributions arise from the interplay of a complex set of economic, sociological, and intellectual factors that influence in a way hard to summarize the research performance of each university in every field. In this scenario, it is well known that some universities are more productive or successful than others in terms of the number of publications and/or the mean citation that these publications receive. However, little is known concerning the shape of university citation distributions abstracting from size and mean citation differences. In order to contribute to this knowledge, in this paper we investigate the following four issues.

Firstly, we inquire whether university citation distributions are universally distributed. The universality condition, borrowed from statistical physics, means that, appropriately normalized, citation distributions follow a unique functional form within the bounds set by random variation. Radichhi *et al.* (2008) suggest a statistical test of this condition in their study of 14 WoS journal subject categories. According to this test, the universality condition is not satisfied for our 500 university citation distributions. This is consistent with previous results for large classification systems in WoS datasets consisting of complete field citation distributions that include publications with zero citations (Albarrán & Ruiz-Castillo, 2011, Albarrán *et al.*, 2011a, Waltman *et al.*, 2012a, Perianes-Rodriguez & Ruiz-Castillo, 2014).

Secondly, in view of the above finding, we ask: are at least university citation distributions as highly skewed and as similar among each other as previous results indicate for field citation distributions? Using the same size- and scale-independent techniques that have been used in previous research, we confirm that this is the case in our dataset. This result has been established at different aggregation levels, publication years, and citation window lengths, and independently of whether the problem of the multiple assignment of publications to sub-fields in WoS datasets is solved by following a multiplicative or a fractional approach (Glänzel, 2007, Radicchi *et al.*, 2008, Albarrán & Ruiz-Castillo, 2011, Albarrán *et al.*, 2012, Herranz & Ruiz-Castillo, 2012, Waltman *et al.*, 2012a, Radicci & Castellano, 2012, Li *et al.*, 2013, Ruiz-Castillo & Waltman, 2015, Perianes-Rodriguez & Ruiz-Castillo, 2014). Similar conclusions concerning the skewness and similarity of individual productivity distributions are found when authors are classified into 30 broad scientific fields (Ruiz-Castillo & Costas, 2014).

Thirdly, using the measuring framework introduced in Crespo *et al.* (2013), we investigate how important is the effect of differences in citation impact between LR universities in the overall citation inequality in the union of the 500 LR university citation distributions. Furthermore, we inquire up to what point this effect can be accounted for by scale factors captured by the universities' Mean Normalized Citation Score (*MNCS* hereafter). The answer is that citation impact differences between universities account for 3.85% of overall citation inequality –a much smaller percentage than what is found in the context of production and citation practice differences between scientific fields (Crespo *et al.*, 2013, 2014, Ruiz-Castillo & Waltman, 2015, Perianes-Rodriguez & Ruiz-Castillo, 2014). These differences are greatly reduced when university citation distributions are normalized using their *MNCS* values as normalization factors.

Finally, we discuss the implications of these results for the understanding of the high correlation between the university rankings according to two citation impact indicators: the *MNCS*, and the Top 10% indicator of scientific excellence (the $PP_{top\ 10\%}$ indicator hereafter), defined as the percentage of an institution's output included into the set formed by 10% of the world most cited papers in the different scientific fields. The latter indicator has been recently adopted by well-established institutions, such as the CWTS in the Netherlands, and SCImago in Spain.

The rest of the paper is organized into two Sections. The first section presents the empirical results, while the next section discusses further research.

Empirical results

The universality of university citation distributions

Let $\boldsymbol{c_i}$ be the LR university i field-normalized citation distribution. Note that, for each university, the mean citation of $\boldsymbol{c_i}$ is precisely the Mean Normalized Citation Score (MNCS hereafter). Let $\boldsymbol{c^*_i}$ be the normalized citation distribution of university i using the university MNCS as the normalization factor. Let $\boldsymbol{C^*}$ be the union of the universities' normalized citation distributions, $\boldsymbol{C^*} = \cup_i \{\boldsymbol{c^*_i}\}$, where publications are ranked in increasing order of the number of normalized citations. Let $\boldsymbol{X_z}$ be the set of publications in the top z% of distribution $\boldsymbol{C^*}$, and let $\boldsymbol{x_{zi}}$ be the publications in X_z that belongs to the i-th university, so that $\boldsymbol{X_z} = \cup_i \{\boldsymbol{x_{zi}}\}$. In the terminology of Radicchi *et al.* (2008), if the ranking is fair, or unbiased, the percentage of publications that the set $\boldsymbol{x_{zi}}$ represents within each university should be near z% with small fluctuations. Let N_c and N_i be, respectively, the number of universities and the number of publications in the i-th university. Assuming that publications of the various universities are scattered uniformly along the rank axis, for any value z% one would expect the average relative frequency of the number of articles in any university to be z% with a standard deviation $\sigma_z = \{[z(100 - z)\Sigma_i (1/N_i)]/N_c\}^{1/2}$, which is equation (2) in Radicchi *et al.* (2008).

Table 1. Percentage of publications in each sub-field that appear in the top z% of the global rank, together with the standard deviation, σ_z, and the coefficient of variation, σ_z/z.

Theoretical values			Normalised distribution		
z%	σ_z	σ_z/z	**z%**	σ_z	σ_z/z
(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
1	0.20	0.20	0.96	0.29	0.30
5	0.43	0.09	4.95	0.90	0.18
10	0.59	0.06	10.00	1.46	0.15
20	0.79	0.04	20.03	2.41	0.12
30	0.91	0.03	30.04	3.11	0.10
40	0.97	0.02	40.00	3.49	0.09
50	0.99	0.02	49.88	3.76	0.08
75	0.86	0.01	74.73	4.08	0.05
90	0.59	0.01	88.94	4.08	0.05

For each z value in a certain sequence, column 2 in Table 1 presents the standard deviations σ_z, while column 3 is the theoretical coefficient of variation, namely, σ_z/z. Columns 4 to 6 contain the values for the average z, the standard deviation σ_z, and the coefficient of variation σ_z/z obtained empirically in distribution $\boldsymbol{C^*}$.

Although σ_z varies non-linearly with z, the theoretical coefficient of variation in column 3 raises from 0.01 to 0.20 when we proceed from z = 90% towards z = 1%. In the normalized case, the considerable differences with the theoretical values in column 6, above all for lower values of z, indicate the lack of universality for this set of 500 university citation distributions. This conclusion contrasts with the universality claim in Chatterjee *et al.* (2014), who study 42 academic institutions across the world, their publications in four years, 1980, 1990, 2000, and 2010, and the citations they receive according to the WoS until July 2014. We should emphasize that this paper has a number of technical problems. The criterion for selecting their

42 academic institutions is not given, and there is no information on how the following three problems have been solved: the assignment of publications in WoS datasets to multiple journal subject categories, the assignment of responsibility for co-authored publications, and the all-sciences aggregation problem. Nevertheless, we will proceed discussing their results. Chatterjee *et al.* (2014) explain that, for each publication year, the university normalized citation distributions fit well to a lognormal for most of the range, although the poorly cited publications seem to follow another distribution, while the upper tail is better described by a power law. This is quite different from the claim that there is a single functional form for the entire domain of definition of the 42 institutions in their sample. Our statistical approach tests whether the universality claim is supported by the data over the entire domain of the 500 LR universities. In this sense, our results do not contradict each other. We both agree that the universality claim over the entire domain is not the case in our respective samples.

On the other hand, the main problem with the still unpublished version of Chatterjee *et al.* (2014) is that, in our opinion, their statistical methods are not clearly explained. Unfortunately, the authors do not explain the following three aspects: (i) how the partition of the domain into three segments is estimated for each university, and whether this partition is universal; (ii) which tests have been used to determine the functional form chosen in each segment versus possible alternatives; (iii) how the confidence interval for the power law parameter has been estimated, and which is the confidence interval for the lognormal parameters. As a matter of fact, the only clear evidence for the distributions collapse into a universal curve is the graphical illustration provided for a sample –whose selection is unexplained– of 24 of the original 42 academic institutions.

The skewness and similarity of university citation distributions

The skewness of citation distributions is assessed by simply partitioning citation distributions into three classes of articles with low, fair, and very high number of citations. For this purpose, we follow the Characteristic Scores and Scale (CSS hereafter) approach, first introduced in Scientometrics by Schubert *et al.* (1987). In our application of the CSS technique, the following two *characteristic scores* are determined for every university: μ_1 = mean citation, which in our context is equal to the *MNCS*, and μ_2 = mean citation for articles with citations greater than μ_1. We consider the partition of the distribution into three broad categories: (i) articles with a low number of citations, smaller than or equal to μ_1; (iii) fairly cited articles, with a number of citations greater than μ_1 and smaller than or equal to μ_2, and (iii) articles with a remarkable or outstanding number of citations greater than μ_2. For each citation distribution, we measure the percentages of publications in the three categories, as well as the percentages of the total citations accounted for by the three categories. The average, standard deviation, and coefficient of variation for the 500 university values of the percentages of publications, the percentages of the total citations in the three categories are included in Table 2.

The results are remarkable. In principle, differences in resources, intellectual traditions, organization, the structure of incentives, and other factors lead us to expect large differences between the 500 LR university citation distributions in different parts of the world. However, judging from the size of the standard deviations and the coefficient of variations for the 500 universities, we find that university citation distributions are extremely similar. At the same time, the distributions are highly skewed: on average, the MNCS values of the 500 universities is 12.9 percentage points above the median, while the 12.5 of outstanding articles account for 44.4% of all normalized citations.

Table 2. The skewness of citation distributions according to the CSS approach. Percentages of articles, and percentages of citations by category. Average, standard deviation, and coefficient of variation over the 500 LR universities, and results for the overall citation distribution.

	Percentage of articles in category:			Percentage of citations in category:		
	1	**2**	**3**	**1**	**2**	**3**
Average (Std. deviation)	62.9 (1.9)	24.6 (1.2)	12.5 (1.2)	22.9 (1.7)	32.7 (0.8)	44.4 (1.5)
Coefficient of variation	0.03	0.05	0.10	0.08	0.02	0.03

For the sake of robustness, we have conducted two more sets of computations. In the first place, in the presence of co-authorship we have assigned publications to universities in a multiplicative way. In the second place, we have studied the raw citation distributions without the benefit of any field-normalization procedure. Interestingly enough, the results are very similar to those obtained for field-normalized university citation distributions in the fractional case. Thus, we conclude that the characteristics of university citation distributions are robust to the way the assignment of publications to universities in the presence of co-authorship and the all-sciences aggregation problem are solved.

Finally, we should mention the results of two contributions closer to our own in which research publications are aggregated into the type of organization unit to which the authors belong. Firstly, Albarrán *et al.* (2015) study the partition of world citation distributions into 36 countries and two residual geographical areas using a dataset, comparable to ours, consisting of 4.4 million articles published in 1998-2003 with a five-year citation window for each year. They find that, at least in some broad fields and in the all-sciences case, the country citation distributions are not only highly skewed, but also very similar across countries –a result parallel to our own for the 500 LR universities. Secondly, Perianes-Rodriguez & Ruiz-Castillo (2015) study a set of 2,530 highly productive economists who work in 2007 in a selection of the top 81 economics departments in the world. Contrary to previous results for field or country citation distributions, we find that productivity distributions are very different across the 81 economics departments. However, the data in Perianes-Rodriguez & Ruiz-Castillo (2015) does not consist of department citation distributions of articles published in a certain period of time with a citation window of common length, but of the individual productivity of faculty members in each department, where individual productivity is measured as a quality index that weights differently the articles published up to 2007 by each researcher in four journal equivalent classes. Nevertheless, we cannot rule out that the similarity of citation distributions is a phenomenon present at certain aggregate levels. To settle this issue, we need more work at the department level with citation distributions articles published in a certain period of time with a common citation window.

The importance of citation impact differences between universities

Together with the assessment of the between-group variability concerning the shape of university citation distributions, we are interested in measuring how important are the citation impact differences between universities. Formally, this problem is analogous to the measurement of the importance of differences in production and citation practices between scientific fields. For the latter, Crespo *et al.* (2013) suggested to measure the impact of such differences on the overall citation inequality for the entire set of field citation distributions applying an additively decomposable citation inequality index to a double partition into scientific fields and quantiles. Similarly, in our case we measure how much of the overall citation inequality exhibited by the union of the 500 LR university citation distributions can be attributed to the citation impact differences between universities (this is also the approach adopted in Perianes-Rodriguez & Ruiz-Castillo, 2014a, to assess the effect of citation impact between countries).

For that purpose, we begin with the partition of, say, each university citation distribution into Π quantiles, indexed by π = 1,.., Π. In practice, in this paper we use the partition into percentiles, that is, we choose Π= 100. Assume for a moment that, in any university u, we disregard the citation inequality within every percentile by assigning to every article in that percentile the mean citation of the percentile itself, μ_u^{π}. The interpretation of the fact that, for example, $\mu_u^{\pi} = 2\,\mu_v^{\pi}$ is that, on average, the citation impact of university u is twice as large as the citation impact of university v in spite of the fact that both quantities represent a common underlying phenomenon, namely, the same *degree of citation impact* in both universities. In other words, for any π, the distance between μ_u^{π} and μ_v^{π} is entirely attributable to the difference in the citation impact that prevails in the two universities for publications with the same degree of excellence in each of them. Thus, the citation inequality between universities at each percentile, denoted by $I(\pi)$, is entirely attributable to the citation impact differences between the 500 LR universities holding constant the degree of excellence in all universities at quantile π. Hence, any weighted average of these quantities, denoted by *IDCU* (*I*nequality due to *D*ifferences in *C*itation impact between *U*niversities), provides a good measure of the total impact on overall citation inequality that can be attributed to such differences. Let $\boldsymbol{c_i}$ be university i citation distribution, and let $\boldsymbol{C}$ be the union of the universities citation distributions, $\boldsymbol{C} = \cup \{\boldsymbol{c_i}\}$. We use the ratio

$$IDCU/I(\boldsymbol{C}) \tag{1}$$

to assess the relative effect on overall citation inequality, $I(\boldsymbol{C})$, attributed to citation impact differences between universities (for details, see Crespo *et al.*, 2013).

Finally, we are interested in estimating how important scale differences between university citation distributions are in accounting for the effect measured by expression (1). Following the experience in other contexts, we choose the university mean citations as normalization factors. To assess the importance of such scale factors, we use the relative change in the *IDPD* term, that is, the ratio

$$[IDCU - IDCU^*]/IDCU, \tag{2}$$

where *IDCU** is the term that measures the effect on overall citation inequality attributed to the differences in university distributions after the normalization of university citation distributions using university mean citations as normalization factors (for details, see again Crespo *et al.*, 2013). The estimates for expressions (1) and (2) in our dataset are included in table 3:

Table 3. The effect on overall citation inequality, I(C), of the differences in citation impact between universities before and after MNCS normalization, and the impact of normalization on this effect.

		Normalization impact = 100 [*IDPD – IDCP*/IDCP*]
Before MNCS normalization, 100 [*IDPU/I(C)*]	3.85 %	-
After MNCS normalization, 100 [*IDPU*/I(C)*]	0.72 %	81.9 %

It is interesting to compare these figures with what was obtained in two instances in the previous literature. The first case concerns the partition into 36 countries and two residual geographical areas in the all-sciences case (Albarrán *et al.*, 2014), while the second case refers to 219 WoS sub-fields (Crespo *et al.*, 2014). Two comments are in order. Firstly, the effect on overall citation inequality due to citation impact differences between the 500 LR universities (3.85%) is comparable to the effect due to citation impact differences between countries (5.4%). However, both of them are considerably smaller than the corresponding effect on overall citation inequality attributable to differences in production and citation practices across the 219 sub-fields (approximately 18%). Secondly, the reduction of the total effect generated by *MNCS* normalization in our dataset (81.9% of the total effect) is of a comparable order of magnitude to the same phenomenon in the context of country (85.2%) or sub-field citation distributions (83.2%).

It should be noted that these results summarize in a pair of scalars a complex phenomenon that takes place along the entire support of our university citation distributions. As a matter of fact, the term *IDCU* is simply a weighted average of the $I(\pi)$ terms, $\pi = 1,\ldots, 100$, that capture the effect on overall inequality of the citation impact differences between the 500 LR universities holding constant the degree of excellence in all universities at percentile π. Therefore, it is instructive to study how $I(\pi)$ changes with π both before and after the MNCS normalization. The results appear in Figure 1 (since $I(\pi)$ is very high for $\pi < 27$, for clarity these percentiles are omitted from Figure 1), which deserves the following two comments. Firstly, the strong impact of *MNCS* normalization is readily apparent. Secondly, it is useful to informally partition the support of our citation distributions into the following three intervals: [0, 57], [58, 96], and [98, 100]. In the first and the third one, $I(\pi)$ values are very high. This means that, since in these two intervals university citation distributions differ by more than a scale factor, the universality condition can hardly be satisfied in them. However, $I(\pi)$ is approximately constant for a wide range of intermediate values in the second interval. Thus, this is the range of values where the search for a single functional form in Chatterjee *et al.* (2014) may give good results in our dataset.

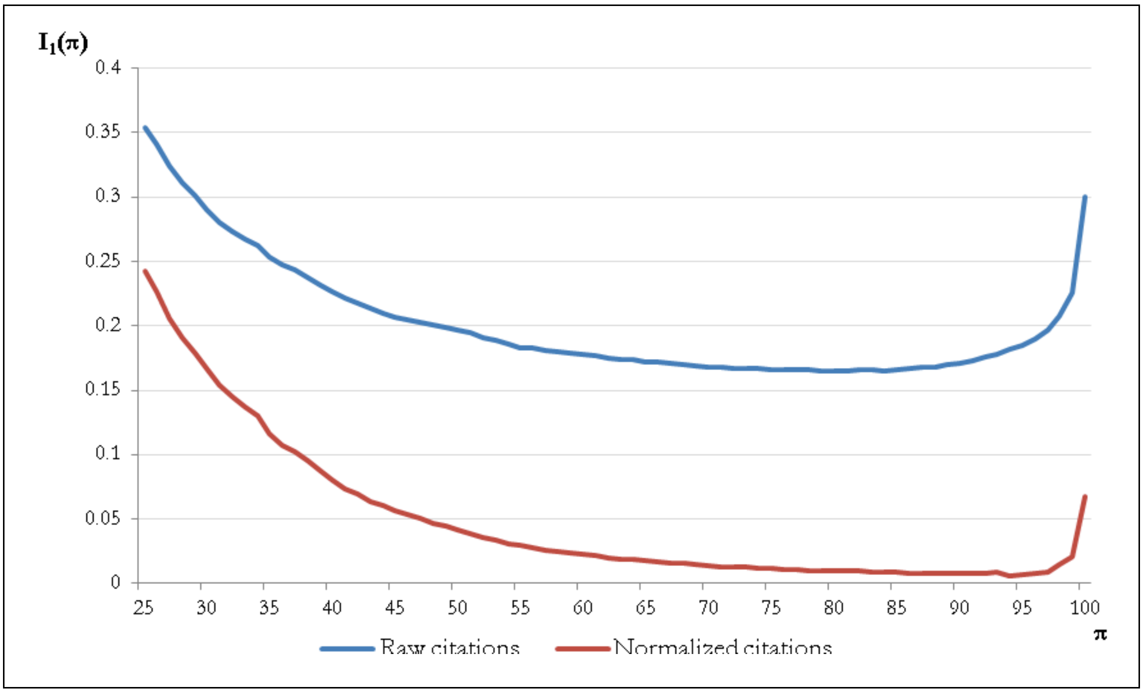


Figure 1. Citation Inequality Due to Differences in Citation Practices, I(π), as a function of π. Results for the [27, 100] quantile interval.

Implications of the results

Our results have two types of practical implications. In the first place, assume that the top, intermediate, and worse universities have different types of citation distributions. In this case, we would need to build different models to explain the citation impact variability within the universities of the three types. On the contrary, since we have found that, although not universal, university citation distributions are rather similar, we need a single model to explain the high within-universities variability.

In the second place, recall that the move in the CWTS and SCImago rankings from an average-based citation impact indicator –such as the *MNCS*– towards a rank percentile approach that throws all the weight on the top x% of most cited papers –such as the $PP_{top\ 10\%}$ indicator– is surely due to the idea that, for highly skewed citation distributions, average-based indicators might not represent well the excellence in citation impact. However, the two rankings are rather similar: the Pearson correlation coefficient between university values is 0.981, while the Spearman correlation coefficient between ranks is 0.986. The situation is illustrated in Figure 2, where the positive slope indicates that to low (high) *MNCS* values there correspond lower (higher) $PP_{top\ 10\%}$ values.

We conclude that ordinal differences between the university rankings according to the MNCS and the $PP_{top\ 10\%}$ indicators are of a small order of magnitude. As a matter of fact, we find a strong, more or less linear relationship between the $PP_{top\ 10\%}$ and the MNCS in two other instances: for the 500 universities in the 2011/2012 edition of the Leiden Ranking (see Figure 2 in Waltman *et al.*, 2012b), and for the partition of the world into 39 countries and eight geographical areas studied in Albarrán and Ruiz-Castillo (2012). How can we explain these results? We have seen already that, university citation distributions behave as if they differ by a relatively constant scale factor over the [58, 96] percentile interval in their support. In this empirical scenario, it is not surprising that the MNCS values, which are reached at approximately the 63th percentile of citation distributions, and the $PP_{top\ 10\%}$ indicator that focus on the last 10 percentiles, provide very similar rankings. A convenient practical consequence is that the citation impact university ranking provided by the MNCS indicator is an adequate one. The $PP_{top\ 10\%}$ indicator would only add greater cardinal differences between the best and worse universities with relatively few re-rankings.

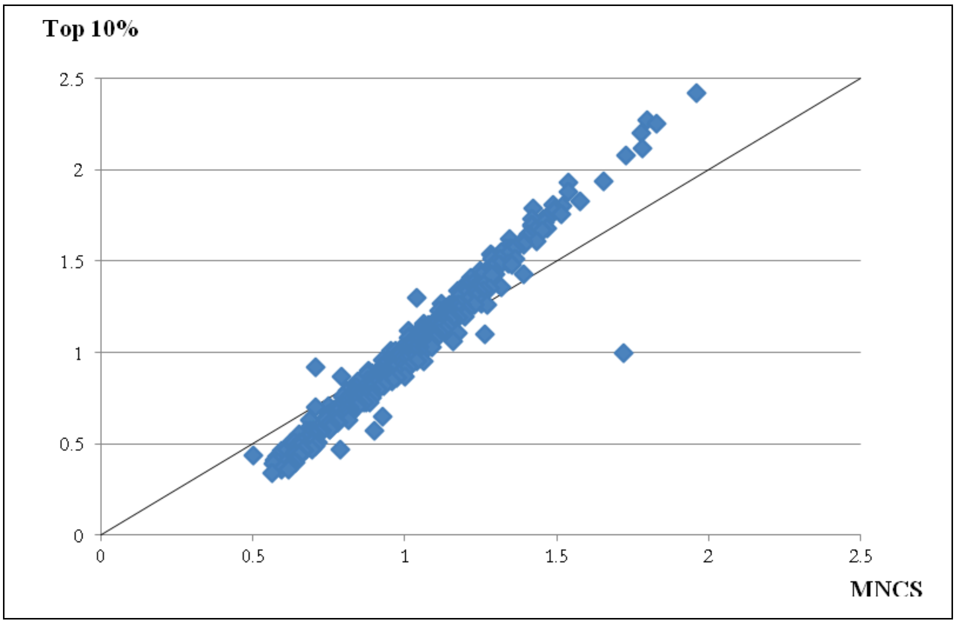


Figure 2. Scatterplot of the relation between the MNCS indicator and the PPtop 10% indicator for the 500 Leiden Ranking universities

It should be noted that further details concerning the following topics can be found in the Working Paper version of this paper, Perianes-Rodriguez & Ruiz-Castillo, 2014b): (i) the distribution of the total number of publications by universities; (2) the means μ_1 and μ_2, as well as the results of the CSSS approach for individual universities; (3) the graphical illustration of these results; (4) the measurement of the skewness of university citation distributions by means of a skewness index robust to extreme observations; (5) the robustness of all skewness results for the assignment of publications to universities in a multiplicative way, as well as the treatment of raw citation distributions without the benefit of any field-normalization procedure; (6) the re-rankings involved in the move from the MNCS towards the $PP_{top\ 10\%}$ indicator, as well as the cardinal differences between their values. In any case, the robustness of all of our results must be investigated with other datasets characterized by other publication years, and other citation windows, as well as other data sources different from the WoS.

Further research

Here are the possibilities for further research:

1. The effect on overall citation inequality attributable to the differences in citation impact between universities shows a characteristic pattern: broadly speaking, university citation distributions appear to behave as if they differ by a relatively constant scale factor over a large, intermediate part of their support. Consequently, it might be interesting to compute the exchange rates introduced in Crespo *et al.* (2013, 2014) to exploit this feature, and to use them as normalization factors. More generally, one could experiment with other normalization approaches that have been found useful in other contexts, notably the two parameter scheme introduced by Radicci & Castellano (2012).
2. Chatterjee *et al.*'s (2014) idea of fitting specific functional forms to university citation distributions in different intervals of their support is worth pursuing. The threshold determining the upper tail where a power law might be the best alternative could be estimated following the methods advocated in Clauset *et al.* (2009). Similar grid techniques could be applied to determine the lower bound of the interval where a lognormal might be the best alternative. In any case, standard methods should be used to test which specific functional form is best in each interval, as well as to estimate the parameters' confidence intervals (Thelwall & Wilson, 2014, and Brzezinski, 2015).
3. As we have seen in Section III.4, differences in citation impact between universities after MNCS normalization tend to rise when we reach the last few percentiles including the most highly cited articles. The question left for further research is how to complement average-based or $PP_{top\ 10\%}$ indicators with other measurement instruments that highlight the behavior of citation distributions over the last few percentiles. Given the important role of extreme observations in citation distributions, robustness of alternative high-impact indicators to these extreme situations will be an important element in the discussion.
4. Consider an array of citation distributions with a smaller number of scientific fields than in this paper in the columns, and the 500 LR universities in the rows. We already know much concerning field citation distributions and university citation distributions in the all-sciences case. A possible next step is to study the characteristics of university citation distributions column by column, that is, restricted to each field. The results will determine to what extent the similarities between citation distributions is a question depending on the aggregation level at which the study is conducted.

References

Albarrán, P., & Ruiz-Castillo, J. (2011). References made and citations received by scientific articles. *Journal of the American Society for Information Science and Technology*, *62*, 40–49.

Albarrán, P. and Ruiz-Castillo, J. (2012). *The Measurement of Scientific Excellence Around the World*. Working Paper, Economic Series 12-08, Universidad Carlos III (http://hdl.handle.net/10016/13896).

Albarrán, P., Ortuño, I., & Ruiz-Castillo, J. (2011a). The measurement of low- and high-impact in citation distributions: technical results. *Journal of Informetrics*, 5, 48–63.

Albarrán, P., Crespo, J., Ortuño, I., & Ruiz-Castillo, J. (2011b). The skewness of science in 219 sub-fields and a number of aggregates. *Scientometrics*, *88,* 385–397.

Albarrán, P., Perianes-Rodriguez, A., & Ruiz-Castillo, J. (2014). Differences in citation impact across countries. *Journal of the American Society for Information Science and Technology*. *66*, 512-525.

Brzezinski, M. (2015). Power laws in citation distributions: Evidence from Scopus. *Scientometrics*, *103*, 213-228.

Chatterjee, A., Ghosh, A., and Chakrabarty, B. K. (2014). *Universality of citation distributions for academic institutions and journals*, 29 September, arViv:1409.8029 [physics.soc-ph].

Clauset, A., C. R. Shalizi, and M. E. J. Newman (2009). Power-law Distributions In Empirical Data. *SIAM Review*, *51*, 661-703.

Crespo, J. A., Li, Y., & Ruiz-Castillo, J. (2013). The measurement of the effect on citation inequality of differences in citation practices across scientific fields. *PLoS ONE*, *8,* e58727.

Crespo, J. A., Herranz, N., Li, Y., & Ruiz-Castillo, J. (2014). The effect on citation inequality of differences in citation practices at the Web of Science subject category level. *Journal of the Association for Information Science and Technology*, *65*, 1244–1256.

Glänzel, W. (2007). Characteristic scores and scales: A bibliometric analysis of subject characteristics based on long-term citation observation. *Journal of Informetrics*, *1*, 92–102.

Herranz, N., & Ruiz-Castillo, J. (2012). Multiplicative and fractional strategies when journals are assigned to several sub-fields. *Journal of the American Society for Information Science and Technology*, *63*, 2195–2205.

Leydesdorff, L., Radicchi, F., Bornmann, L., Castellano, C., and de Nooye, W. (2012). Field-normalized Impact Factors: A Comparison of Rescaling versus Fractionally Counted Ifs. *Journal of the American Society for Information Science and Technology*, *27*, 292-306.

Li, Y., Castellano, C., Radicchi, F., & Ruiz-Castillo, J. (2013). Quantitative evaluation of alternative field normalization procedures. *Journal of Informetrics*, *7*, 746–755.

Perianes-Rodriguez, A. & Ruiz-Castillo, J. (2015). Within- and between-department variability in individual productivity. The case of economics. *Scientometrics*, *102*: 1497-1520.

Perianes-Rodriguez, A. & Ruiz-Castillo, J. (2014). *University citation distributions*. Working Paper, Economic Series 14-26, Universidad Carlos III (http://hdl.handle.net/10016/19811).

Radicchi, F., & Castellano, C. (2012). A reverse engineering approach to the suppression of citation biases reveals universal properties of citation distributions. *PLoS ONE*, *7*, e33833.

Radicchi, F., Fortunato, S., and Castellano, C. (2008), "Universality of Citation Distributions: Toward An Objective Measure of Scientific Impact", *PNAS*, 105: 17268-17272.

Ruiz-Castillo, J. (2014). The comparison of classification-system-based normalization procedures with source normalization alternatives in Waltman and Van Eck. *Journal of Informetrics*, *8*, 25–28.

Ruiz-Castillo, J. & Costas, R. (2014). The Skewness of Scientific Productivity. *Journal of Informetrics*, *8*, 917-934.

Ruiz-Castillo, J. & Waltman, L. (2015). Field-normalized citation impact indicators using algorithmically constructed classification systems of science. *Journal of Informetrics*, *9*, 102-117.

Schubert, A., Glänzel, W., & Braun, T. (1987). A New Methodology for Ranking Scientific Institutions". *Scientometrics*, *12*, 267-292.

Thelwall, M., & Wilson, P. (2014). Distributions for cited articles from individual subjects and years. *Journal of Informetrics*, *8*, 824-839.

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

Waltman, L., Van Eck, N. J., & Van Raan, A. F. J. (2012a). Universality of citation distributions revisited. *Journal of the American Society for Information Science and Technology*, *63*, 72–77.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., Van Eck, N. J., Van Leeuwen, T. N., Van Raan, A. F. J., Visser, M. S., & Wouters, P. (2012b). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, *63*, 2419–2432.

Exploration of the Bibliometric Coordinates for the Field of 'Geography'

Juan Gorraiz and Christian Gumpenberger

christian.gumpenberger, juan.gorraiz@univie.ac.at
University of Vienna, Library and Archive Services, Bibliometrics and Publication Strategies, Boltzmanngasse 5, A-1090 Vienna (Austria)

Abstract
This study is a bibliometric analysis of a highly complex research discipline, namely geography, in order to identify the most used and cited publication channels, to reveal publication strategies, and to analyse the discipline's coverage in the three main data sources for citation analyses: Web of Science, Scopus and Google Scholar. The results show very heterogeneous and individual publication strategies when considering the selection of adequate publication channels even in the same research fields. Monographs, journal articles (including proceedings papers) and book chapters are the most cited document types. Differences between research fields more related to the natural sciences than to the social sciences are clearly visible but not so considerable when taking into account the higher number of co-authors. General publication strategies are more established in the fields related to the natural sciences. Although an "iceberg citation model" is suggested, citation analyses for monographs, book chapters and reports (working papers) should be conducted separately and include complementary data sources, such as Google Scholar, in order to enhance the coverage and improve the quality of the citation analysis.

Conference Topics
Citation and co-citation analysis – Social Sciences

Introduction and background

From a bibliometric point of view, geography is a very challenging discipline, because it belongs to the natural sciences (geography, physical) as well as to the social sciences (geography), as it is clearly depicted in each edition of Journal Citation Reports (see Table 1).

Table 1. Category data of geography in both Editions of JCR (2013)

JCR EDITION 2013	*Category*	*Total Cites*	*Median IF*	*Aggregate IF*	*Aggregate Immediacy Index*	*Aggregate Cited Half-Life*	*# Journals*	*# Articles*
Sciences	GEOGRAPHY, PHYSICAL	159297	2.152	2.574	0.72	7.5	46	4972
Social Sciences	GEOGRAPHY	79207	1.059	1.612	0.343	7.4	76	3762

Table 1 shows very different citation characteristics according to the corresponding JCR edition. Furthermore, geography is a highly interdisciplinary field, very strongly related to geosciences, environmental sciences, ecology and remote sensing (natural sciences), or to economics, urban studies and political sciences (social science), as a quick search and refine analysis in WoS (Web of Sciences - core collection) illustrates.

Although there are many studies illustrating the differences between natural and social sciences and the different publication cultures depending on the discipline (e.g. Nederhof, 2006; Australian Research Council, 2012; Ossenblok et al., 2012; van Leeuwen, 2013; Moksony, 2014), no literature focusing on this specific could be retrieved by the authors.

The main research questions of this study are:

- What are the publication characteristics depending on the different research field?
- Can differences be observed concerning research fields? What is their time evolution?

- Which are the most used publication channels? Which document types are the most cited ones? Is it possible to identify publication strategies?
- What is the coverage in the three main citation data sources, Web of Science, Scopus and Google Scholar? Could Google Scholar be used as a complementary data source?

Data sources and methodology

This study is primarily based on publication data collected for three professorial appointments at the University of Vienna (Department for Geography): the first one, related to Geosciences and comprising of twelve candidates, and the second one, related to Social and Economic Geography and comprising of ten candidates, were performed during 2013. The third one, related to Demography and comprising of nine candidates, was performed in August 2014.

All the publication data were delivered directly by the applicants, whose identity has to remain anonymous. All bibliometric indicators added to the list of publications by the authors themselves, such as citation counts, impact factor or the h-index, were controlled or recalculated in order to enable a correct and comparable analysis (Gorraiz, J. & Gumpenberger, C., 2015). Document types used by the authors in their list of publications were manually reassigned to the following standard groups: Monographs (Books), Book chapters, Journal articles, Proceedings Papers, Conferences (including meeting abstracts and talks), Reports (Working Papers), Book Reviews, Edited Books and Journals Issues, and other publications (or Miscellaneous). A clear distinction between "Proceedings Papers" and "Conferences" was not always possible when relying on the lists of publications.

The main data source for coverage and citation analyses was Web of Science - Core Collection (WoS) including the Conference Proceedings and Book Citation Index. Since coverage in the usual multidisciplinary bibliographic and citation databases (Web of Science, Scopus) is very low and unsatisfactory for citation analyses, we have included Google Scholar (GS) as an additional data source in a first explorative attempt (Jacso, 2005; Kousha & Thelwall, 2007; Meho, & Yang, 2007; Gorraiz et al., 2013).

The analysis in GS was performed by using the Google Scholar Citation Profiles (applicants for the third appointment were invited to create their individual profiles and make them publicly available for a couple of weeks) as well as by applying the tool 'Publish or Perish' particularly designed for this purpose.

In spite of the fact that citations were checked and the percentage of self-citations was determined, citation analyses in GS should be taken with a pinch of salt. Google Scholar is not a database but a search engine, and therefore indexing remains non-transparent and documentation is lacking. That is why the analyses were also performed in Web of Science, including the Cited Reference Search (which means considering citations originating from Web of Science (WoS) 'core journals' to all document types without any restrictions), and in Scopus.

Publication windows were the last ten years (general for all authors, appointments no.1 and 2) and the career length of each applicant (for all appointments). In order to distinguish individual scientific career lengths, the year of the first publication activity is always included.

The observed citations window was identical for all applicants per professorial appointment procedure. It covers the date from publication until April - May 2013 for appointments no. 1 and 2, and until July - August 2014 for appointment procedure no.3.

Visibility analyses were performed according to the data in the Journal Citation Reports (JCR), Science Edition 2012 (appointments no. 1&2).

The quartiles (Q1= top 25%; Q2= top 25-50%; Q3= top 50-75% and Q4= top 75-100%) were calculated according to the 2-years impact factor (IF) in the corresponding WoS category.

Results

Comparison between appointments no.1 and no.2

Table 2 and 3 show the most important publication document types used by the candidates for both appointments. The spectrum is much more heterogeneous in the social sciences, where journal articles are not always the most common publication channel.

Table 2. Publication spectrum and WoS coverage according to provided publication list for appointment no.1 – Geosciences - 12 candidates. (In parenthesis, the number of document types indexed in WoS; PY=all years; *no distinction).

Candidate no.	1st Pub Year	Books	Edited Books/ Issues	Book Chapters	Proceedings & Conference Papers*	Book Reviews	Miscellaneous	Journal Articles (JA)
1	2004	1	0	5 (1)	14 (1)	0	3	28 (24)
2	2002	0	0	6 (1)	35 (3)	0	2	33 (30)
3	1996	13	7	12 (4)	26 (1)	0	0	38 (28)
4	1990	2	4 (2)	25 (6)	17	0	29	17 (11)
5	1998	4	2	1	6 (2)	0	65	75 (61)
6	1998	2	0	8 (2)	55 (2)	0	3	31 (21)
7	2007	4	0	1	41	0	1	35 (33)
8	1994	9	0	16	192	0	0	66 (53)
9	1999	0	0	7	13 (3)	0	5	28 (28)
10	2005	3	0	12	12(2)	10 (5)	10	18 (11)
11	2002	0	0	5 (1)	70	0	0	28 (18)
12	1994	1	0	2 (1)	8	0	1	51 (51)

Table 3. Publication spectrum and WoS coverage according to provided publication list for appointment no. 2 - Social & Economic Geography - 10 candidates. (In parenthesis, the number of document types indexed in WoS; PY=all years; *no distinction).

Candidate no.	1st Pub Year	Books	Edited Books/ Issues	Book Chapters	Proceedings & Conference Papers*	Book Reviews	Miscellaneous	Journal Articles (JA)
1	1999	3	2	8 (1)	2	8	50	72 (35)
2	2002	3	11	21	5 +*56	0	0	16 (8)
3	1991	7	0	19(1)	*87	0	13	37 (18)
4	1993	3	0	17 (2)	*67	19(9)	44	46 (24)
5	1994	7	2	16	2 + *34	0	9	31 (17)
6	2005	3	5	15	*42	0	5	15 (4)
7	1990	3	11	58	*4*	10	14	35 (22)
8	2005	1	1	5	**40*	0	9	20 (7)
9	2004	3 (1)	0	21 (7)	**10*	2	10	16 (11)
10	2000	3	1	17	*72	0	49	22 (11)

Miscellaneous were principally Reports and Working Papers in both appointments. Therefore this document type was considered separately in the second part of the study.

In appointment no. 2, other document types such as Films, Policy Briefs, Newspapers and Special Issues were mentioned but only individually. For two candidates (one in appointment no.1 and one in no.2), articles in other (non-scientific or non-peer-reviewed) journals were also assigned to the group Miscellaneous.

Concerning the coverage in WoS both tables corroborate the low coverage of books and book chapters in both editions of the Book Citation Index. For articles in peer-reviewed journals, the WoS coverage in appointment no.1 varies between 60 and 100% and the trend in the last 10 years was constantly increasing until it reached a quota of almost 90% for all candidates. In appointment no. 2, the coverage was lower, varying between about 30 and 60%, but a similar trend was also observed even if not as steep.

Tables 4 and 5 show the results of the visibility (publication strategies) and citation analyses performed for both appointments. Only publications indexed in WoS in the last ten complete years (2003-2012) were considered.

Table 4. Visibility (Q1 and %Q1) and citation analysis in WoS for appointment no. 1 – Geosciences - 12 candidates. (PY=2003 -2012, ARPP= Articles, Reviews & Proceedings Papers).

Candidate no.	*1st Pub Year*	*Publications*			*# Authors per Paper*	*Citations ARPP*			*h-Index*	*% Self-citations*	*Q1*	*% Q1*
		Total	*ARPP*	*per Y*		*Sum*	*per P*	*Max*				
1	2004	25	25	2.78	6.36	147	5.88	28	7	16.22%	16	69.57%
2	2002	28	28	2.80	4.93	181	6.46	36	7	24.31%	14	87.50%
3	1996	29	26	2.60	4.83	249	9.58	31	10	19.05%	14	53.85%
4	1990	11	7	0.70	2.73	29	4.14	21	3	12.50%	5	100.00%
5	1998	49	48	4.80	5.57	458	9.54	42	12	30.07%	34	72.34%
6	1998	18	18	1.80	3.72	180	10.00	44	7	7.78%	8	53.33%
7	2007	32	32	5.33	5.53	428	13.38	155	12	21.26%	20	62.50%
8	1994	31	29	2.90	5.06	598	21.36	110	15	7.18%	29	93.55%
9	1999	17	17	1.70	4.94	317	18.65	102	7	4.73%	6	42.86%
10	2005	16	11	1.38	2.94	40	3.64	24	3	10.00%	2	14.29%
11	2002	16	16	1.60	4.38	129	8.06	21	8	15.50%	9	60.00%
12	1994	36	26	2.60	4.69	294	11.31	44	12	17.06%	32	91.43%
	Mean	**25.67**	**23.583**	**2.582**	**4.64**	**254.2**	**10.166**	**54.8**	**8.583**	**15.47%**	**16**	**66.77%**

Table 5. Visibility (Q1 and %Q1) and citation analysis in WoS for appointment no. 2 - Social & Economic Geography - 10 candidates. (PY=2003-2012; ARPP= Articles, Reviews & Proceedings Papers).

Candidate no.	*1st Pub Year*	*Publications*			*# Authors per Paper*	*Citations ARPP*			*h-Index*	*% Self-citations*	*Q1*	*% Q1*
		Total	*ARPP*	*per Y*		*Sum*	*per P*	*Max*				
1	1999	22	15	1.50	1.14	122	8.13	53	6	11.02%	12	60.00%
2	2002	7	4	0.40	2.00	22	5.50	10	3	9.09%	0	0.00%
3	1991	12	9	0.90	1.75	352	39.11	94	7	3.13%	9	81.82%
4	1993	**23**	12	1.20	2.61	134	11.167	76	6	13.41%	7	31.82%
5	1994	13	9	0.90	2.23	76	8.44	34	4	3.13%	3	23.08%
6	**2005**	4	3	0.38	**1.00**	3	1.00	2	1	0.00%	0	0.00%
7	**1990**	18	13	1.30	2	36	2.77	11	3	24.32%	3	18.75%
8	**2005**	7	6	0.75	2.57	48	8.00	17	4	8.33%	1	14.29%
9	**2004**	17	14	1.56	1.82	259	18.50	149	5	8.33%	7	70.00%
10	2000	8	7	0.70	1.13	53	7.57	40	3	9.26%	1	12.50%
	Mean	**13.1**	**9.2**	**0.958**	**1.82**	**110.5**	**11.02**	**48.6**	**4.2**	**9.00%**	**4.3**	**31.22%**

These results corroborate the higher number of publications and citations in the discipline related to the natural sciences (about twice as many). But taking into account the number of co-authors and the percentage of self-citations, which is almost twice as high in the natural sciences, there is not really a considerable difference.

The visibility analysis (number of Q1- journal articles) shows that publishing in top journals with impact factor, result in a much higher visibility in the appointment related to natural sciences than in the one related to the social sciences.
Finally, tables 6 and 7 show that the citation differences, according to the aggregate impact factor of the main WoS category, are higher in appointment no.1 than in no.2.

Table 6. First and second research field according to WoS categories for appointment no. 1 - Geosciences – 12 candidates.

Candidate no.	First Research Field (2003-2012)		Second Research Field (2003-2012)
	WoS Category	IF aggregate 2012	WoS Category
1	Ecology	3.095	Environmental Sciences
2	Remote Sensing	1.845	Geosciences, Multidisciplinary
3	Water Resources	1.803	Geosciences, Multidisciplinary
4	Water Resources	1.803	Geosciences, Multidisciplinary
5	Soil Science	1.780	Geosciences, Multidisciplinary
6	Ecology	3.095	Forestry / Soil Science/ Environm. Sci.
7	Ecology	3.095	Forestry / Plant Sciences
8	Geosciences, Multidisciplinary	2.176	Geography, Physical
9	Geosciences, Multidisciplinary	2.176	Geography/ Water Resources
10	Geography, Physical	2.206	Geography / Remote Sensing
11	Water Resources	1.803	Soil Sciences /Environmental Sci.
12	Geochemistry & Geophysics	1.474	Oceanography/Geosciences, Multi.

Table 7. First and second research field according to WoS categories for appointment no. 2 - Social & Political Geography – 10 candidates.

Candidate no.	First Research Field (2003-2012)		Second Research Field (2003-2012)
	WoS Category	IF aggregate 2012	WoS Category
1	Geography	1.469	Industrial Relations & Labor
2	Geography	1.469	Environmental Sciences
3	Geography	1.469	Economics; Management
4	Geography	1.469	Environmental Studies; Economics
5	Geography	1.469	Economics
6	Geography	1.469	Geography, Physical
7	Geography	1.469	Urban Studies
8	Geography	1.469	Environmental Studies & Sciences
9	Economics	1.148	Geography; Planning & Development
10	Geography	1.469	Economics

Results obtained in appointment no. 3 (Demography &Population Geography)

Applicants were invited to create their individual Google Scholar Citations profiles and make them publicly available for a couple of weeks.
From the nine applicants:

- six created a GS Citation Profile

- two refused to create one
- one followed the invitation, but the profile was incomplete

The tool 'Publish or Perish', particularly designed for this purpose, was then used for collecting and checking the data.

First of all, two key aspects (Focus 1 and 2) of each candidate's publications were determined in GS (free keywords) and in Web of Science according to the assigned Subject Categories (WoS categories) in the database. The results are shown in Table 8.

Table 8. First and second research field in WoS categories and GS for appointment no. 3– 9 candidates.

Candidate no.	Google Scholar		Web of Science	
	Focus 1	**Focus 2**	**WoS Category 1**	**WoS Category 2**
1	Human Geography -	Area Studies - East Asia - Japan	Urban Studies	Area Studies
2	Human Geography	Population Geography	Geography	Geography
3	Migration Studies	Demographic Change	Geography	Geography, Physical
4	Migration	Urban Studies	Geography; Planning & Development	Urban Studies
5	Urbanization	Cross-border Mobility	Geography	Geography, Physical
6	Demography	Fertility	Demography	Geography
7	Demography	Population	Demography	Public, Environmental & Occupational Health
8	Population Geography	Migration and Labour Markets	Geography	Political Science
9	Resilience	Livelihood	Public, Environmental & Occupational Health	Geography, Physical

Table 9 represents the publication activity for each scientist according to the most relevant publication types. The data are based on the list of publications submitted by the candidates. In order to distinguish individual scientific career lengths, the year of the first publication activity has been included.

The results hint at very heterogeneous and individual publication strategies taking into account publication types. The three next sections contain coverage and citation analyses performed in the three considered data sources. Table 10 shows the percentage of coverage in Google Scholar for each publication type. Monographs (Books) and Edited Books or Issues are very well covered, probably due to the inclusion of Google Books (Kousha & Thelwall, 2009).

The coverage of Journal Articles is also much higher than in WoS or Scopus (see Table 11). Also of interest is the high coverage of Reports (Working Papers). Chapters in Books are not so well covered, but this is probably due to incidental incorrect citations.

Table 9. Publication spectrum (publication types) for appointment no. 3. (*no distinction).

Candidate no.	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Total (excl. Conferences)	**58**	**73**	**36**	**121**	**73**	**80**	**75**	**60**	**42**
Monographs	5	1	4	4	3	3	3	2	1
Book Chapters	13	32	15	48	17	11	11	21	7
Journal Articles	20	20	5	21	17	44	28	27	20
Proceedings Papers*	2	0	2	1	8	0	8	0	0
Reports (Working Papers)	3	0	7	11	7	13	10	3	11
Book Reviews	8	0	2	8	2	0	0	0	0
Edited Books/Journals	5	20	1	11	5	6	3	3	2
Other Publications	*1*	*0*	*0*	*17*	*14*	*3*	*12*	*4*	*1*
Conferences*	64	94	33	94	4	38	109	90	34
1st Year Publication	**1998**	**1994**	**2000**	**1993**	**1999**	**1992**	**1999**	**1989**	**2000**

Table 10. Coverage (%) in Google Scholar for each publication type (Appointment no. 3) (*no distinction).

Candidate no.	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Total (excl. Conferences)	**58**	**73**	**36**	**121**	**73**	**80**	**75**	**60**	**42**
GS Profile	Yes	Incom-plete	Yes	No	No	Yes	Yes	Yes	Yes
Total Pub (excl. Conf)	44.83%	52.05%	44.44%	57.02%	35.62%	72.50%	77.33%	68.33%	97.62%
Monographs	60.00%	100.00%	50.00%	75.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Book Chapters	16.67%	12.50%	40.00%	56.25%	35.29%	45.45%	90.91%	42.86%	100.00%
Journal Articles	85.00%	50.00%	60.00%	71.43%	41.18%	81.82%	82.14%	100.00%	100.00%
Proceedings Papers*			50.00%		25.00%		100.00%		
Reports	66.67%		28.57%	54.55%	28.57%	46.15%	60.00%	33.33%	90.91%
Book Reviews			50.00%	25.00%	100.00%				
Edited Books/Journals	20.00%	70.00%	100.00%	81.82%	80.00%	83.33%	100.00%	66.67%	100.00%
Other Publications				41.18%		33.33%	41.67%		100.00%
1st Year	**1998**	**1994**	**1998**	**1995**	**1999**	**1992**	**1999**	**1995**	**2002**

Table 11 shows the results of the coverage and citation analyses performed in WoS, including the Cited Reference Search, in Scopus and in Google Scholar. The higher coverage scores in WoS over those in Scopus are due to the inclusion of the Cited Reference Search. This enabled citations not only of journal articles and book indexed in WoS to be retrieved, but also of other books, reports and other document types cited by the core journals in WoS.

All sections include the same indicators for each data source: 1) number of indexed publications; 2) percentage of publications covered according to the provided publication list; 3) number of cited documents; 4) total number of citations; 5) number of citations per cited publication; 6) maximum number of citations attracted by a publication; 7) total h-index and 8) i-index (number of publications with more than 10 citations).

The percentage of self-citations was only calculated for GS, where the number of citations was of sufficient significance.

Table 11 confirms that the values of the main citation indicators (number of citations, citations per cited publication and h-index) are different in absolute values in GS, WoS and Scopus, but are comparable in terms of relative values. Spearman correlations performed for these indicators (number of citations, citations per cited publication and h-index) in the three data sources (WoS, Scopus and Google Scholar) were very strong (varying from 0.8 to 0.95). A detailed coverage and citation analysis for the three most cited document types in Google Scholar, Monographs, Book Chapters and Journal Articles (see Table 12) is shown in Table 13.

Table 11. Coverage and citation analysis in the three data sources for each candidate (Appointment no. 3)

	Candidate no	1	2	3	4	5	6	7	8	9
Google Scholar	GS Profile available	Yes	Incom-plete	Yes	No	No	Yes	Yes	Yes	Yes
	Total Pub (excl. Conf)	26	38	22	74	26	60	60	55	44
	% covered in GS	44.83%	52.05%	44.44%	57.02%	35.62%	72.50%	77.33%	68.33%	97.62%
	# cited documents	20	15	16	60	14	53	43	33	23
	Total Citations	123	36	106	667	80	1026	699	320	142
	% Self-citations	5.69%	13.89%	15.09%	7.65%	7.50%	14.52%	16.45%	20.94%	21.13%
	Citations/Cited Pub	6.15	2.40	6.63	11.12	5.71	19.36	16.26	9.70	6.17
	Maximum Citations	20	6	49	86	16	144	165	128	14
	h-index	7	3	5	14	5	19	13	9	8
	i-index (more than 10 cit)	5	0	2	21	3	25	18	8	5
WoS + Cited Ref Search	Total Pub (excl. Conf)	13	11	7	31	10	47	35	15	26
	% covered in WoS + CRS	17.24%	8.22%	16.67%	22.31%	9.59%	53.75%	38.67%	13.33%	52.38%
	# cited documents	11	6	6	29	9	44	31	12	24
	Total Citations	30	6	16	86	17	435	102	39	60
	Citations/Cited Pub	2.73	1.00	2.67	2.97	1.89	9.89	3.29	3.25	2.50
	Maximum Citations	9	1	10	16	4	55	21	24	7
	h-index	4	1	2	6	3	12	5	2	4
	i-index (more than 10 cit)	0	0	1	2	0	14	2	1	0
Scopus	Total Pub (excl. Conf)	9	10	2	11	6	30	16	11	10
	% covered in Scopus	15.52%	13.70%	5.56%	9.09%	8.22%	36.25%	21.33%	18.33%	23.81%
	# cited documents	5	5	1	7	2	24	10	8	9
	Total Citations	22	6	2	35	3	384	58	50	27
	Citations/Cited Pub	4.40	1.20	2.00	5.00	1.50	16.00	5.80	6.25	3.00
	Maximum Citations	11	2	2	22	2	57	23	31	8
	h-index	2	1	1	2	1	11	4	4	3
	i-index (more than 10)	1	0	0	1	0	13	2	1	0
1st Year Publication		1998	1994	2000	1993	1999	1992	1999	1989	2000

Table 12. Summary of the three most cited publication types in Google Scholar (Appointment no. 3).

Document Type	% Coverage	% Cited	Citations/Cited P	Maximum Citations	% Self-citations
Book Chapters	48.74%	68.77%	6.21	86	23.04%
Journal Articles	74.62%	74.20%	10.06	144	11.22%
Monographs	87.22%	92.59%	21.17	165	9.76%

The results show that not always the same publication types are the most cited for each candidate. There are individual differences. A separate citation analysis of these publication types is then recommended for evaluation purposes.

Table 13. Detailed Citation analysis in Google Scholar for each candidate and the three most cited publication types (Appointment no. 3). (the three highest values for each document type are highlighted in different colours).

Candi date no.	Publication Types	Liste	Google Scholar										
			Publications						Citations				
		# P	1st year	# Total	# Not list	# Cited	% cited	% Coverage	# Total	Mean	# Max	# Self	% Self
1	Monographs	5	1998	3	0	3	100.00%	60.00%	19	6.33	7	2	10.53%
	Book chapters	13	2001	3	1	3	100.00%	15.38%	44	14.67	19	4	9.09%
	Journal articles	20	1998	17	0	12	70.59%	85.00%	58	4.83	20	1	1.72%
2	Monographs	1	1994	2	1	2	100.00%	100.00%	7	3.50	6		0.00%
	Book chapters	32	1996	4	0	1	25.00%	12.50%	2	2.00	2		0.00%
	Journal articles	20	1998	10	0	9	90.00%	50.00%	21	2.33	4	3	14.29%
3	Monographs	4	2002	2	0	2	100.00%	50.00%	55	27.50	49	2	3.64%
	Book chapters	15	2003	6	0	4	66.67%	40.00%	8	2.00	4	2	25.00%
	Journal articles	5	2009	3	0	2	66.67%	60.00%	10	5.00	6	0	0.00%
4	Monographs	4	1996	3	0	2	66.67%	75.00%	20	10.00	18	0	0.00%
	Book chapters	48	1996	27	0	25	92.59%	56.25%	313	12.52	86	20	6.39%
	Journal articles	21	1996	15	0	14	93.33%	71.43%	151	10.79	48	7	4.64%
5	Monographs	3	1999	3	0	2	66.67%	100.00%	25	12.50	16	4	16.00%
	Book chapters	17	2001	6	0	4	66.67%	35.29%	12	3.00	5	0	0.00%
	Journal articles	17	2000	7	0	4	57.14%	41.18%	25	6.25	12	1	4.00%
6	Monographs	3	1992	3	0	3	100.00%	100.00%	74	24.67	27	8	10.81%
	Book chapters	11	1997	5	0	5	100.00%	45.45%	11	2.20	4	5	45.45%
	Journal articles	44	1996	36	0	34	94.44%	81.82%	892	26.24	144	126	14.13%
7	Monographs	3	2002	3	0	3	100.00%	100.00%	249	83.00	165	10	4.02%
	Book chapters	11	2005	11	1	8	72.73%	90.91%	64	8.00	16	25	39.06%
	Journal articles	28	1999	23	0	17	73.91%	82.14%	278	16.35	66	68	24.46%
8	Monographs	2	2003	2	0	2	100.00%	100.00%	18	9.00	17	0	0.00%
	Book chapters	21	1995	9	0	6	66.67%	42.86%	36	6.00	15	10	27.78%
	Journal articles	27	1999	27	0	18	66.67%	100.00%	227	12.61	83	39	17.18%
9	Monographs	1	2010	1	0	1	100.00%	100.00%	14	14.00	14	6	42.86%
	Book chapters	7	2005	7	0	2	28.57%	100.00%	11	5.50	8	6	54.55%
	Journal articles	20	2005	20	0	11	55.00%	100.00%	68	6.18	13	14	20.59%

Conclusions and discussion

The main conclusions of this case study for the field geography can be summarized in the following points:

Differences between research fields more related to the natural sciences than to the social sciences are clearly visible. However, the higher productivity (number of publications per year) and citation counts, are relativized when also considering the higher number of co-authors and percentage of self-citations

- General publication strategies, especially these based on the impact factor, are still more evident in the fields related to the natural sciences
- The results hint at very heterogeneous and individual publication strategies considering the selection of adequate publication channels even in the same research fields
- Journal Articles and Book Chapters are the most used publication channels
- Monographs, Journal Articles (including Proceedings Papers) and Book Chapters are the most cited document types
- The coverage, especially books, is much higher in Google Scholar and suggests the recommendation of this data source as complementary one, although this data source is still a black box (no transparency, missing content information, etc.). In this study the accuracy of the citations in GS was very high (~95%). Nevertheless further

measures are needed to reduce the noise of Google Scholar data in order to increase the significance of this alternative data source for evaluative purposes.

- The values of the main citation indicators might differ in absolute values in GS, WoS and Scopus, but are comparable in terms of relative values.
- This fact suggests a "citation iceberg model" (see Figure 1). The citation analysis in WoS or Scopus shows only the 'visible part' but this is generally still related to and indicates the 'invisible part'.
- Therefore, citation analyses for monographs, book chapters and reports (working papers) should be conducted separately and require the inclusion of complementary data sources. Otherwise relevant publications can be easily missed, resulting in wrong interpretations.
- Peers still have to be aware of blind spots in 'citation analyses' (e.g. 'non cited' document types and publications) with potentially harmful consequences in evaluation exercises

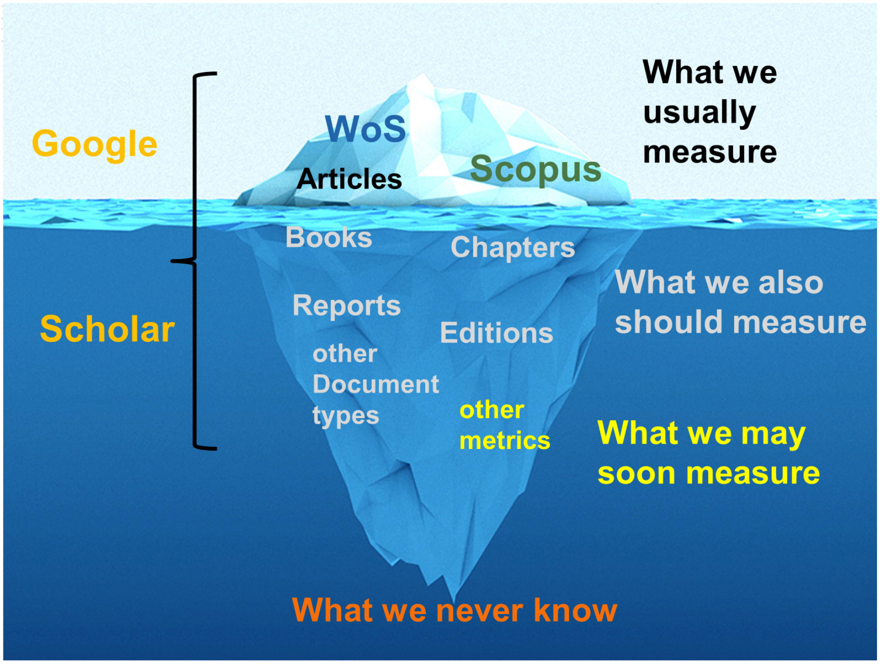


Figure 1. Citation "iceberg" model.

Finally, it should be stressed that citations can only used as a proxy for impact (and not for the quality) of publications produced in the 'publish or perish' community (i.e. the scientists who are committed to publishing their results). However, the scientific community is much broader and also comprises teaching academics as well as representatives from government or industry, who rather use than cite scientific output. Furthermore, our society has become progressively informed ('societal impact'). Unfortunately alternative metrics (like usage metrics and altmetrics) are still in their infancy (Kurtz M.J. & Bollen. J., 2010; Priem, J. et al., 2012; Gorraiz, J. et al., 2014; Hammarfelt, B., 2014) to measure the impact beyond citations and could not yet be applied to the described appointment procedures due to the current lack of available and reliable data.

Acknowledgments

We wish to thank Christian Buchmayer for his help collecting and disambiguating the data, and Prof. Thomas Glade for his support and suggestions.

References

Australian Research Council (2012). The Excellence in Research for Australia (ERA) Initiative, Retrieved January 5, 2015 from http://www.arc.gov.au/era/.

Gorraiz, J., Purnell, P.J. & Glänzel, W. (2013). Opportunities for and limitations of the book citation index. *Journal of the American Society for Information Science and Technology*, *64*(7), 1388-1398.

Gorraiz, J., Gumpenberger, C. & Schlögl, C. (2014). Usage Versus Citation Behaviours in Four Subject Areas. Scientometrics, *101*(2), 1077-1095.

Gorraiz, J. & Gumpenberger, C. (2015). A bibliometric model for the enhancement of professorial appointments, in press.

Hammarfelt, B. (2014). Using altmetrics for assessing research impact in the humanities. *Scientometrics*, *101*(2), 1419-1430.

Jacso, P. (2005). Google Scholar: the pros and the cons. *Online Information Review*, *29*(2), 208-214.

Kousha, K., & Thelwall, M. (2009). Google book search: Citation analysis for social science and the humanities. *Journal of the American Society for Information Science and Technology*, *60*(8), 1537-1549.

Kousha, K. & Thelwall, M. (2007). Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology*, *58*(7), 1055-1065.

Kurtz, M. J. & Bollen, J. (2010). Usage bibliometrics. *Annual Review of Information Science and Technology*, *44*(1), 1-64.

Meho, L. I. & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar. *Journal of the American Society for Information Science and Technology*, *58*(13), 2105-2125.

Moksony, F., Hegedus, R. & Csaszar, M. (2014). Rankings, research styles, and publication cultures: a study of American sociology departments. *Scientometrics*, *101*(3), 1715-1729.

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, *66*(1), 81-100.

Ossenblok, T.L.B., Engels, T.C.E. & Sivertsen, G. (2012). The representation of the social sciences and humanities in the Web of Science - a comparison of publication patterns and incentive structures in Flanders and Norway (2005-9). *Research Evaluation*, *21*(4), 280-290.

Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. Retrieved January 5, 2015 from arXiv preprint:1203.4745, http://arxiv.org/abs/1203.4745.

van Leeuwen, T.N. (2013). Bibliometric research evaluations, Web of Science and the Social Sciences and Humanities: a problematic relationship? *Bibliometrie-Praxis & Forschung*, *2*, Retrieved January 5, 2015 from http://www.bibliometrie-pf.de/article/viewFile/173/218.

The Most-Cited Articles of the 21st Century

Elias Sanz-Casado[1,2], Carlos García-Zorita[1,2] and Ronald Rousseau[3,4]

[1,2] *elias.sanz@uc3m.es; czorita@bib.uc3m.es*
[1] Department of Library and Information Science. Laboratory of Metric Studies on Information (LEMI). Carlos III University of Madrid. C/Madrid 126, Getafe, 28903 Madrid, Spain

[2] *Research Institute for Higher Education and Science (INAECU).*
Carlos III University of Madrid-Autonomous University of Madrid. C/Madrid 126, Getafe, 28903 Madrid, Spain

[3,4] *ronald.rousseau@kuleuven.be*
[3] KU Leuven, Department of Mathematics, Celestijnenlaan 200B, B-3000 Leuven (Heverlee), Belgium
[4] Universiteit Antwerpen, IBW, Venusstraat 35, B-2000 Antwerpen, Belgium

Abstract
The aim of this paper is to collect the most-cited articles of the 21st century and to study how this group changed over time. Here the term "most-cited" is operationalized by considering yearly h-cores in the Web of Science. These h-cores are analysed in terms of authors, research areas, countries, institutions, journals and average number of authors per paper. We only consider publications of article or proceedings type. The research of some of the more prolific authors is on genetics and genomes publishing in multidisciplinary journals, such as *Nature* and *Science*, while the results show that writing a software tool for crystallography or molecular biology may help collecting large numbers of citations. English is the language of all articles in any h-core. The core institutions are largely those best placed in most rankings of world universities. Some attention is given on the relation between h-core articles and the information sciences. We conclude by stating that the notion of an h-core provides a new perspective on leading countries, articles and scientists.

Conference Topic
Citation and co-citation analysis

Introduction
The objective of this paper is to collect the most-cited articles of the 21st century and to study how this group changed over time. The term "most-cited" is operationalized by considering the h-core (Hirsch, 2005; Rousseau, 2006) in the Web of Science (WoS) for each period of time, starting with the period 2001-2005, continuing with 2001-2006 and ending with 2001-2013. These periods refer to the publication and the citation window. We recall that the h-core at a given moment in time, for instance on January 1, 2009, consists of the set of articles which at that time received a number of citations at least equal to their rank among all articles published during the period 2001-2008. This approach is different from the one taken in (Van Noorden et al., 2014) where a fixed number, concretely 100, of articles is considered. Furthermore, we study the papers making up the corresponding h-cores in terms of authors, research areas, countries, institutions, journals and average number of authors per paper.

Methodology
We have to point out that the 21st century starts on January 1, 2001. This implies that we only consider publications from 2001 on. Moreover, we only consider publications in Thomson Reuters' Web of Science (WoS) and we restrict ourselves to publications of article or proceedings type.
Although finding today's h-core for a set of articles in the Web of Science is easy, finding an h-core in the past needs some specific knowledge of the tools available in the WoS. First one retrieves the set for which one wants to determine the h-core (ending in the year Y). Its

articles are ranked from most cited to least cited. These are collected as a marked list. This is possible for at most 5,000 items. Clicking on Marked List shows this list and now, on this page, the system can provide a Citation Report, which is downloaded as an Excel file showing yearly citations for each of these records. Now we add the same data for the next 5,000 items (more was not necessary for our investigation). In this Excel file, we remove the columns corresponding to the year Y+1 and all later ones. In a next step we sum all remaining citations of each article. Sorting these sums from highest to lowest and comparing with a column of natural numbers leads to the h-index and the h-core. More details of this procedure are provided in (Rousseau & Zhang, 2014).

Results

The most-cited papers

The most-cited articles over the period 2001-2013 (the latest h-core) are shown in Table 1. It is clear that writing a software tool for crystallography or molecular biology may give one's paper a huge boost. The article by the National Cholesterol Education Program Expert Panel (2001) was the most-cited one from 2005 till 2008. From the year 2009 on Sheldrick's became the most-cited one.

Table 1. Most-cited articles over the period 2001-2013.

Rank	Article cited	Times cited
1	Sheldrick, G.M. (2008). A short history of SHELX. *Acta Crystallographica Section A*, 64, 112-122.	34,533
2	Livak, K.J. & Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods*, 25(4), 402-408.	24,796
3	Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, 24(8), 1596-1599.	17,049
4	Novoselov, K.S., Geim, A.K., Morozov, S.V., Jiang, D., Zhang, Y., Dubonos, S.V., Grigorieva, I.V. & Firsov, A.A. (2004). Electric field effect in atomically thin carbon films. *Science*, 306(5696), 666-669.	12,512
5	Ronquist, F. & Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics,* 19(12), 1572-1574.	11,185
6	National Cholesterol Education Program Expert Panel (Group author; includes 28 members). (2001). Executive summary of the Third Report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III). *JAMA-Journal of the American Medical Association*, 285(19), 2486-2497.	11,160
7	Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D – Biological Crystallography*, 60(special issue 1), 2126-2132.	10,392
8	Huelsenbeck, J.P. & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8), 754-755.	10,317
9	Spek, A.L. (2003). Single-crystal structure validation with the program PLATON. *Journal of Applied Crystallography*, 36, 7-13.	9,920
10	Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics*, 5(2), 150-163.	9,175

Time evolution of h-index and h-cores

The difference between the h-index and the number of items in the h-core is due to the possible existence of more than one document with the same number of citations as the h-index, as illustrated in Table 2. For the year 2005, for example, there were five articles with 359 citations.

Table 2. H-indices and h-cores for the periods 2001-2005 till 2001-2013.

End year	h-index	# articles in the h-core
2005	359	363
2006	441	442
2007	526	527
2008	614	616
2009	704	704
2010	800	800
2011	902	902
2012	1014	1014
2013	1122	1122

It is obvious that only a small percentage of articles included in the WoS belongs to the h-core of a specific period. In order to show the evolution of the ratio of the h-core with respect to all articles we put their values for the period 2001-2004 equal to 100. Figure 1 shows the total number of papers in each period and the number of papers in each h-core when this rescaling has been performed. Linear regression is almost perfect for the two lines: all publications (R^2= 0,9982) and h-core (R^2= 0,9967). For this reason we can forecast the 21[st] century h-index for, at least, the next years to come. This would lead to an h-core of 1195 documents in 2014 and 1290 in the year 2015.

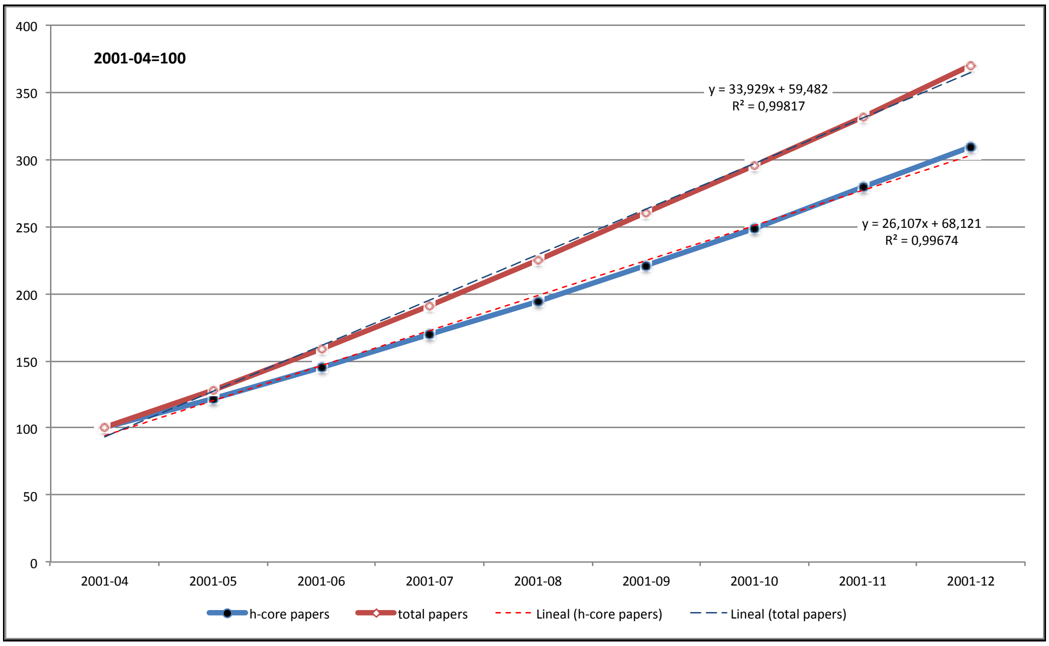


Figure 1. Evolution of the h-core.

In Table 3, we show the number of articles published in the years 2001 to 2011 included in each of the h-cores. For each h-core these numbers follow the order of publication, i.e. most articles are published in the year 2001 and least in the latest year included in the core. Core13 has exactly the same number of articles published in 2001 as in 2002 (209 articles), while it does not contain articles published in 2013.

Table 3. Evolution of h-cores.

Year of Publication	Core-05	Core-06	Core-07	Core-08	Core-09	Core-10	Core-11	Core-12	Core-13
2001	196	210	218	217	217	213	213	209	209
2002	116	137	158	173	187	197	201	205	209
2003	43	72	96	120	138	151	159	163	169
2004	7	21	41	62	82	99	117	138	146
2005	1	2	11	31	49	74	93	110	121
2006			3	9	17	35	56	70	95
2007				3	10	23	36	47	58
2008				1	3	6	19	39	54
2009					1	2	6	21	32
2010							2	9	19
2011								3	8
2012									2
Total	363	442	527	616	704	800	902	1014	1122

Table 4 shows the number of articles in the h-core (on the diagonal) and on the last line the number of unique articles in the union of all h-cores until the year indicated on top of the column. The other numbers refer to the number of articles originally belonging to the core referred to on the left, but which do not anymore belong to the h-core. We note that there is one article that left the core (in 2007) but re-entered (in 2008) and from then on stayed in the core. This paper is:

> Minokoshi, Y., Kim, Y., Peroni, O., Fryer, L., Muller, C., Carling, D., & Kahn, B. (2002). Leptin stimulates fatty-acid oxidation by activating AMP-activated protein kinase. *NATURE*, *415* (6869), 339–343. doi:10.1038/415339a

Table 4. H-cores and h-core losses

	Core-05	Core-06	Core-07	Core-08	Core-09	Core-10	Core-11	Core-12	Core-13
Core-05	363	9	9	9	9	9	9	9	9
Core-06		442	13	12	12	12	12	12	12
Core-07			527	17	17	17	17	17	17
Core-08				616	15	15	15	15	15
Core-09					704	26	26	26	26
Core-10						800	27	27	27
Core-11							902	24	24
Core-12								1014	22
Core-13									1122
Total	363	451	549	654	757	879	1008	1144	1274

H-cores characteristics

All articles in any h-core are written in English. We note that the 2001-2005 h-core contains one article that was later retracted (Chang and Roth, published in *Science*, which has now 533 citations and had 359 citations by the end of 2005, being the last one in the 2005 core). Some of the more prolific authors (E.S. Lander, M.J. Daly, R.A. Gibbs, J. Wang) perform research on genetics and genomes publishing in multidisciplinary journals, such as *Nature* and *Science*, often in hyper co-authored papers (with dozens and even hundreds of authors). A. Jemal and E. Ward publish yearly statistics on cancer, which all enter the h-core. R. Collins and R. Peto work on internal medicine and publish almost exclusively in *Lancet*. The fields of nanotechnology and grapheme research are represented by C.M. Lieber and Nobel Prize winners A.K. Geim and K.S. Novoselov (Table 5).

Table 5. Authors with highest number of papers in the h-core (Authors with more than 7 papers in the latest core).

Author	Core-05	Core-06	Core-07	Core-08	Core-09	Core-10	Core-11	Core-12	Core-13
Lander, ES	11	13	14	15	16	17	17	19	18
Wang, J	7	8	8	9	9	10	10	14	14
Jemal, A	4	4	5	6	7	8	9	10	12
Collins, R	5	6	7	8	9	11	11	11	10
Daly, MJ	4	5	6	6	7	10	10	12	10
Peto, R	4	5	7	8	8	9	9	9	10
Lieber, CM	5	6	7	7	7	8	8	9	10
Ward, E	3	3	4	5	6	7	8	9	10
Gibbs, RA	2	3	3	3	3	4	4	11	10
Geim, AK				3	3	5	6	8	10
Novoselov, KS				3	3	5	6	8	10
Thun, MJ	5	5	6	7	8	9	9	9	9
Altshuler, D	4	4	5	5	6	8	8	10	9
Abecasis, GR	2	2	2	2	2	4	4	9	9
Golub, TR	4	5	6	8	8	9	9	8	8
Murray, T	5	5	6	7	8	8	8	8	8
Gabriel, SB	3	3	3	3	4	5	5	9	8
Li, Y	1	2	3	3	4	7	7	8	8
Bartel, DP	1	1	1	3	3	4	4	7	8

The multidisciplinary areas (which include journals such as *Nature*, *Science* and *PNAS*), and the ones related to general and internal Medicine (such as *Lancet* or the *New England Journal of Medicine*) occur the most in each of the cores, as illustrated in Table 6.

Table 6. H-cores in different research areas (Areas with more than 10 papers in the last core).

Research area	Core-05	Core-06	Core-07	Core-08	Core-09	Core-10	Core-11	Core-12	Core-13
Science & Technology - Other Topics	39,1%	38,0%	35,3%	34,9%	32,8%	33,4%	32,7%	32,0%	31,9%
General & Internal Medicine	27,8%	26,2%	26,4%	25,0%	24,6%	23,1%	21,6%	20,4%	20,0%
Biochemistry & Molecular Biology	8,3%	9,0%	8,3%	9,7%	10,1%	10,6%	11,4%	12,8%	13,3%
Physics	5,5%	5,0%	4,9%	4,5%	5,0%	5,5%	6,4%	6,9%	7,0%
Chemistry	0,8%	1,4%	2,1%	1,9%	3,0%	3,9%	5,3%	6,0%	6,1%
Computer Science	2,5%	3,6%	4,7%	4,2%	4,5%	4,5%	5,1%	5,3%	5,5%
Cell Biology	4,1%	4,3%	4,0%	4,5%	4,5%	5,0%	5,2%	5,3%	5,1%
Engineering	1,4%	1,6%	3,0%	3,4%	3,6%	3,5%	3,8%	3,6%	3,9%
Biotechnology & Applied Microbiology	2,2%	3,4%	2,8%	3,1%	3,4%	3,1%	3,3%	3,8%	3,8%
Materials Science	0,6%	0,7%	0,8%	0,8%	1,7%	2,1%	3,0%	3,4%	3,8%
Oncology	2,8%	2,3%	2,3%	2,4%	2,7%	2,9%	2,5%	2,6%	2,9%
Genetics & Heredity	3,6%	3,4%	3,4%	3,2%	3,7%	3,4%	3,2%	3,3%	2,8%
Mathematics	0,8%	1,8%	1,7%	1,8%	1,7%	1,8%	2,0%	2,5%	2,7%
Mathematical & Computational Biology	0,8%	2,0%	1,7%	1,9%	1,8%	1,8%	1,8%	2,4%	2,4%
Research & Experimental Medicine	3,0%	3,2%	3,4%	3,2%	3,1%	2,9%	2,5%	2,5%	2,2%
Crystallography	0,8%	0,7%	0,9%	1,1%	1,3%	1,5%	1,6%	1,8%	2,0%
Neurosciences & Neurology	0,3%	0,7%	0,8%	0,8%	0,9%	1,4%	1,4%	1,9%	2,0%
Astronomy & Astrophysics	2,5%	2,9%	2,5%	2,3%	2,1%	2,1%	1,9%	1,9%	1,6%
Cardiovascular System & Cardiology	1,4%	1,8%	1,9%	1,8%	1,6%	1,4%	1,4%	1,5%	1,5%
Evolutionary Biology	0,0%	0,2%	0,2%	0,5%	0,7%	0,9%	1,2%	1,4%	1,5%
Immunology	2,8%	3,2%	3,2%	2,4%	2,7%	2,1%	1,8%	1,6%	1,3%
Biophysics	0,0%	0,0%	0,4%	0,5%	0,9%	1,0%	1,0%	1,1%	1,3%
Environmental Sciences & Ecology	0,3%	0,5%	0,4%	0,2%	0,4%	0,9%	0,9%	1,1%	1,3%
Radiology, Nuclear Medicine & Medical Imaging	0,0%	0,2%	0,4%	0,5%	0,6%	0,8%	1,0%	1,2%	1,2%
Endocrinology & Metabolism	1,4%	1,1%	1,1%	1,5%	1,4%	1,3%	1,1%	1,0%	1,1%

Table 7 shows a list of most used sources, where we observe, together with the mentioned multidisciplinary journals, the presence of medicine-related journals, including the specialized journal, *CA-A Cancer Journal for Clinicians,* whose presence is due to the systematic publication of the highly-cited annual statistics on cancer (all of them are in core 13). Other journal in the top positions, such as *Physical Review Letters* or *Nature Materials* occur less frequently.

Table 7. Journals of h-core publications (sources with 10 or more papers).

Source Titles	Core-05	Core-06	Core-07	Core-08	Core-09	Core-10	Core-11	Core-12	Core-13
NATURE	19,6%	17,4%	15,6%	15,9%	14,6%	14,9%	14,6%	14,4%	13,9%
SCIENCE	15,2%	16,1%	15,6%	15,1%	14,1%	14,0%	13,4%	12,9%	12,7%
NEW ENGLAND JOURNAL OF MEDICINE	16,5%	15,4%	15,2%	14,9%	14,8%	14,0%	13,1%	12,1%	11,9%
LANCET	5,2%	5,0%	5,1%	4,5%	4,4%	4,4%	4,1%	3,7%	3,6%
JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION	5,5%	5,2%	5,1%	4,7%	4,4%	3,9%	3,5%	3,3%	3,1%
PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA	3,6%	3,8%	3,6%	3,4%	3,3%	3,5%	3,4%	3,2%	3,1%
CELL	0,8%	0,7%	0,9%	1,5%	1,8%	2,4%	2,7%	2,9%	2,9%
NUCLEIC ACIDS RESEARCH	3,3%	2,9%	2,5%	2,8%	2,4%	2,1%	2,2%	2,5%	2,6%
BIOINFORMATICS	0,8%	1,6%	1,3%	1,1%	1,1%	1,1%	1,1%	1,5%	1,6%
PHYSICAL REVIEW LETTERS	3,6%	2,5%	2,3%	1,8%	1,6%	1,1%	1,4%	1,5%	1,4%
CA-A CANCER JOURNAL FOR CLINICIANS	1,4%	1,1%	1,3%	1,3%	1,3%	1,3%	1,2%	1,2%	1,2%
NATURE MATERIALS	0,0%	0,0%	0,0%	0,2%	0,6%	0,9%	1,2%	1,4%	1,2%
ACTA CRYSTALLOGRAPHICA SECTION D-BIOLOGICAL CRYSTALLOGRAPHY	0,0%	0,0%	0,4%	0,5%	0,7%	0,9%	0,9%	1,0%	1,2%
NATURE MEDICINE	1,7%	1,8%	1,5%	1,6%	1,6%	1,5%	1,4%	1,3%	1,2%
CIRCULATION	1,1%	1,4%	1,5%	1,3%	1,1%	1,0%	0,9%	1,0%	1,1%
IEEE TRANSACTIONS ON INFORMATION THEORY	0,8%	0,7%	1,3%	1,1%	1,1%	1,1%	1,0%	0,9%	1,0%
JOURNAL OF CLINICAL ONCOLOGY	0,8%	0,7%	0,6%	0,8%	0,9%	1,0%	0,8%	0,7%	0,9%
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY	0,0%	0,2%	0,6%	0,5%	0,6%	0,6%	0,9%	1,0%	0,9%
NATURE GENETICS	2,5%	2,0%	1,9%	1,6%	1,8%	1,6%	1,4%	1,3%	0,9%

We observe that the shares of the top journals such as *Nature*, *Science* and the *NEJM* are slowly declining over the years, while the share of *Cell* is increasing. This corresponds with recent findings (Lozano et al., 2012; Larivière et al., 2014; Acharya et al., 2014) that more and more highly-cited publications are published in journals that do not have the highest impact factors, say "non-elite journals". Of course, this is as such not surprising as the number of publications world-wide increases faster than the publication opportunities provided by so-called elite journals.

In Table 8 we show the distribution of countries in the h-cores, where an article is classified as belonging to a country if at least one author has an address in this country. The first place goes to the USA. If, however, we consider the European Union (EU-28) as one entity then it leads the rankings in all except one year. Our results correspond to those obtained by King (2004) for the percentage of documents published by USA in the 1% most cited papers. Our results are also similar to those found by Leydesdorff et al. (2014). In their work the EU-28 gains gradually in the top-10% segment at the expense of the USA, and one can expect a cross-over between the EU28 and the USA in the near future within the top-10% segment. However, the distance between the U.S. and the EU is much larger in the top-1% segment.

Also here we see that the top performers (USA, EU-28 and Germany) lose in the share of h-core articles. This observation also holds for the Netherlands and most Scandinavian countries. England and Scotland consolidate their share, while Brazil and New Zealand show an increase. Although China's share in publications shows an exponential growth (Jin & Rousseau, 2005; Zhou & Leydesdorff, 2006, 2008) its share in h-core papers is much lower and shows at best a small increase in the latest years, after a decrease in the period 2008-2009.

Core institutions are shown in Table 9. Leading institutions are those that one can find in most rankings of world universities, although The University of Texas (Austin) is only 39th in the latest ARWU ranking.

Table 8. Countries of publication (with 10 or more papers in the latest core).

Countries	Core-05	Core-06	Core-07	Core-08	Core-09	Core-10	Core-11	Core-12	Core-13
European Union	78,8%	76,9%	76,5%	76,8%	73,6%	75,3%	73,8%	75,8%	76,0%
USA	75,2%	75,1%	75,5%	74,8%	75,1%	74,5%	73,1%	72,0%	71,7%
England	18,2%	19,0%	17,5%	17,9%	17,3%	17,6%	17,1%	17,9%	17,8%
Germany	14,0%	13,6%	13,5%	12,5%	11,9%	12,0%	12,2%	12,2%	11,7%
France	8,5%	8,8%	9,1%	9,3%	8,9%	8,8%	8,2%	8,3%	8,5%
Canada	9,9%	9,0%	8,7%	8,6%	8,1%	8,6%	8,0%	8,4%	8,3%
Japan	7,4%	8,8%	8,3%	8,3%	7,7%	7,9%	7,3%	7,8%	7,7%
Italy	5,8%	5,7%	6,1%	5,8%	5,5%	6,4%	6,3%	6,4%	6,1%
Switzerland	5,5%	4,8%	5,1%	5,2%	4,8%	5,1%	5,2%	5,6%	6,0%
Netherlands	6,9%	6,3%	5,7%	5,7%	5,5%	5,5%	5,1%	5,7%	5,8%
Australia	5,0%	5,2%	5,1%	5,4%	5,3%	5,4%	5,5%	5,3%	5,7%
Sweden	5,2%	5,4%	5,1%	5,4%	5,3%	5,3%	5,4%	5,5%	5,3%
Spain	3,6%	3,4%	3,6%	3,4%	3,0%	3,1%	3,2%	3,5%	3,8%
Belgium	4,1%	3,8%	3,6%	4,1%	4,0%	4,0%	3,7%	3,7%	3,7%
Scotland	2,8%	2,7%	3,2%	3,4%	3,3%	3,5%	3,3%	3,3%	3,1%
Denmark	3,6%	3,2%	3,0%	3,1%	2,8%	3,1%	3,0%	2,7%	2,8%
Finland	3,3%	2,7%	2,8%	2,3%	2,1%	2,3%	2,3%	2,6%	2,6%
Peoples R China	2,2%	1,8%	1,9%	1,5%	1,4%	1,8%	1,8%	2,5%	2,4%
Austria	2,2%	1,8%	1,9%	1,9%	1,8%	2,0%	2,2%	2,1%	2,1%
Israel	1,4%	1,6%	1,9%	1,6%	1,6%	1,6%	1,6%	1,7%	1,6%
Norway	1,7%	1,6%	2,3%	2,1%	1,8%	1,9%	1,8%	1,6%	1,5%
Russia	1,4%	0,7%	0,9%	1,1%	1,1%	1,3%	1,4%	1,5%	1,5%
South Korea	1,1%	0,9%	0,8%	1,0%	0,9%	0,9%	1,1%	1,4%	1,5%
Poland	1,1%	0,9%	0,8%	1,1%	1,4%	1,3%	1,3%	1,3%	1,4%
Ireland	1,4%	1,4%	1,5%	1,3%	1,3%	1,6%	1,3%	1,3%	1,3%
Brazil	0,8%	0,7%	0,8%	1,0%	1,0%	1,0%	1,1%	1,1%	1,2%
New Zealand	0,3%	0,5%	0,6%	0,6%	0,9%	1,0%	1,0%	1,0%	1,2%
Taiwan	1,1%	0,9%	0,9%	1,0%	1,0%	0,9%	0,7%	0,7%	0,9%

Table 9. Core institutions restricted to those with 25 or more papers in the latest core.

Institution	Core-05	Core-06	Core-07	Core-08	Core-09	Core-10	Core-11	Core-12	Core-13
Harvard Univ	37	47	52	63	69	80	86	97	106
MIT	16	18	23	29	33	41	43	53	56
Univ Calif Berkeley	17	22	28	34	39	39	49	54	54
Univ Texas	11	16	20	25	30	35	39	41	45
Johns Hopkins Univ	12	17	19	26	29	34	33	40	43
Univ Washington	21	25	30	36	38	38	38	39	42
Univ Michigan	10	12	18	20	20	27	27	35	41
Univ Cambridge	11	13	16	20	22	26	29	34	39
Univ Oxford	15	14	16	18	19	24	27	34	39
Stanford Univ	15	21	24	24	26	26	33	37	38
Brigham & Womens Hosp	13	18	24	29	32	32	31	34	35
Univ Calif Los Angeles	13	19	19	20	21	24	26	28	35
Univ Calif San Diego	9	12	13	15	18	23	25	29	32
Columbia Univ	3	4	8	13	15	19	22	28	31
Massachusetts Gen Hosp	9	11	13	15	18	24	25	27	31
Univ Calif San Francisco	13	14	18	21	22	23	25	28	29
Univ Penn	13	13	14	15	17	19	19	25	26
Duke Univ	8	9	11	12	17	18	18	23	25
NCI	12	14	16	20	21	24	25	27	25
Univ Pittsburgh	7	9	11	16	16	18	19	22	25

In table 10 we have calculated average co-authorship values of articles in h-cores by research areas. For several research areas these values are higher than the co-authorship values of all publications: for example, in Clinical Medicine the co-authorship value for all publications was 4.5 authors per document and 5 in Bioscience and Biomedical Research (Bordons & Gómez 2000; Glänzel & Schubert, 2005). For several research areas these values are higher than the co-authorship values expected from previous research. For example, in Clinical Medicine the co-authorship value for all publications was 4.5 authors per document and 5 in Bioscience and Biomedical Research (Bordons & Gómez 2000; Glänzel & Schubert, 2005).

Table 10. Average numbers of authors for papers in the h-cores by research areas (areas with more than 10 papers in 2013).

Research Area	Core-05	Core-06	Core-07	Core-08	Core-09	Core-10	Core-11	Core-12	Core-13	Average
Science & Technology - Other Topics	15,5	16,1	14,6	13,9	14,7	14,5	14,5	17,0	15,9	15,3
General & Internal Medicine	19,8	20,4	23,4	25,6	24,2	25,9	22,7	22,1	22,1	23,1
Biochemistry & Molecular Biology	8,2	8,6	8,3	8,5	8,4	7,9	7,3	7,5	7,4	7,8
Physics	52,2	45,0	40,4	37,9	31,3	19,4	15,3	13,6	49,6	31,0
Chemistry	4,0	3,8	4,5	4,4	4,8	5,4	5,2	5,2	5,2	5,1
Computer Science	3,6	3,3	3,0	3,0	3,2	3,1	3,0	3,1	3,0	3,1
Cell Biology	11,4	11,8	11,7	10,9	10,8	10,7	10,2	11,1	11,1	10,9
Engineering	3,8	3,6	3,1	2,9	2,8	2,9	2,8	2,8	2,8	2,9
Biotechnology & Applied Microbiolog	6,8	5,9	7,0	7,4	7,4	6,5	6,0	5,6	5,4	6,2
Materials Science	4,5	3,3	6,5	5,6	5,0	5,2	5,6	5,8	6,3	5,7
Oncology	10,6	10,6	9,8	10,1	10,8	11,2	11,1	11,2	11,1	10,8
Genetics & Heredity	7,1	6,7	8,4	8,0	7,5	7,0	6,5	6,2	5,9	6,9
Mathematics	3,3	3,9	3,9	3,5	4,3	3,9	3,8	3,7	3,6	3,8
Mathematical & Computational Biolo	3,3	3,8	3,8	4,7	5,3	5,0	4,7	4,3	4,3	4,5
Research & Experimental Medicine	11,5	12,1	11,6	11,6	11,0	11,5	11,8	11,4	11,4	11,5
Crystallography	3,3	3,3	3,0	2,6	2,6	3,4	3,1	4,2	5,1	3,7
Neurosciences & Neurology	16,0	10,7	8,8	8,6	8,7	8,5	8,3	7,6	7,8	8,3
Astronomy & Astrophysics	41,8	30,7	30,7	37,3	35,9	37,5	38,8	46,5	45,8	38,8
Cardiovascular System & Cardiology	12,6	10,5	8,8	10,1	10,1	9,7	9,8	11,9	13,5	10,9
Evolutionary Biology		2,0	2,0	3,0	3,4	3,1	2,6	2,7	2,9	2,8
Immunology	8,1	7,3	7,2	7,4	7,5	7,7	7,6	7,6	7,8	7,6
Biophysics			2,5	2,3	3,3	4,0	3,8	5,1	5,9	4,5
Environmental Sciences & Ecology	7,0	4,0	4,0	7,0	5,0	3,1	2,9	2,6	2,8	3,2
Radiology, Nuclear Medicine & Medical Imagin		6,0	4,5	5,7	6,3	5,0	4,3	5,0	5,5	5,1
Endocrinology & Metabolism	7,2	7,2	6,8	8,4	6,9	6,9	6,9	6,6	5,9	6,9

Areas with an average of less than 5 authors (in 2013) are: computer science, engineering, mathematics, mathematical and computational biology, crystallography, evolutionary biology, biophysics and environmental sciences & ecology. Areas with an average larger than 15 are: science & technology – other topics, general & internal medicine, physics and astronomy & astrophysics.

The 21st century h-core (2001-2013) and the information sciences

Only one article classified by Thomson Reuters as *Information science and library science* belongs to this h-core, namely Venkatesh, V., Morris, M.G., Davis, G.B. et al. (2003). User acceptance of information technology: toward a unified view. *MIS Quarterly*, 27(3), 425-478 (cited 2261 times in total).

Yet, other ones were used and cited in *Information science and library science* articles. We list those that were cited at least 30 times by ILS researchers (on December 25, 2014).

1. Hirsch, J.E. (2005). An index to quantify an individual's research output. *Proceedings of the National Academy of Sciences of the USA*, 102(46), 16569-16572. Cited 682 times by ILS researchers.

2. Venkatesh, V., Morris, M.G., Davis, G.B. et al. (2003). User acceptance of information technology: toward a unified view. *MIS Quarterly*, 27(3), 425-478. Cited 595 times.
3. Newman, M.E.J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the USA*, 98(2), 404-409. Cited 118 times.
4. Blei, D.M., Ng, A.Y. & Jordan, M/I. (2003). Latent Dirichlet allocation. *Journal of Machine-Learning Research*, 3(4-5), 993-1022. Cited 93 times.
5. Zhara, S.A. & George, G. (2002). Absorptive capacity: a review, reconceptualization, and extension. *Academy of Management Review*, 27(2), 185-203. Cited 91 times.
6. Berners-Lee, T., Hendler, L. & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 28-37. Cited 64 times.
7. Newman, M.E.J., Strogatz, S.H. & Watts, D.J. (2001). Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 62(2), article number 026118. Cited 60 times
8. Girvan, M. & Newman, M.E.J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the USA*, 99(12), 7821-7826. Cited 50 times.
9. Newmann. M.E.J. & Girvan, M. (2004). Finding and evaluating community structure in networls. *Physical Review E*, 69(2), article number 026113. Cited 36 times
10. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. Cited 30 times.

Besides Hirsch's famous article on the h-index (Hirsch, 2005), we see also Berners-Lee's article on the semantic web (Berners-Lee et al., 2001) and note the fact that Mark Newman occurs four times in this ILS h-core.

Conclusions

-Using the notion of an h-core provides a new perspective on leading countries, articles and scientists.
-The scientific contribution to the h-cores by the EU-28 is slightly higher than the USA's.
-The trend of annual h-cores since 2001 can predict future values of this indicator.
Of course, the view provided in this contribution is highly biased in favor of certain research areas such as General & Internal Medicine, or Biochemistry & Molecular Biology, and certain methodologies (writing heavily used software programs). Yet, it is a fact of life that these areas provide today's leading research. One should clearly realize that publishing highly cited research is different from realizing outstanding intellectual achievements.

References

Acharya, A., Verstak, A., Suzuki, H., Henderson, S., Iakhiaev, M., Chiung, C., Lin, Y. & Shetty, N. (2014). Rise of the rest: the growing impact of non-elite journals. Retrieved October 8, 2014 from arXiv:1410:2217v1.

Bordons, M., & Gómez, I. (2000). Collaboration networks in science. In: B. Cronin & H.B. Atkins (Eds.), *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*. Information Today Inc, Medford, NJ, 197–213.

Glänzel, W., & Schubert, A. (2005). Analysing scientific networks through co-Authorship. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research* (pp. 257–276). Dordrecht: Kluwer Academic Publishers.

Hirsch, J.E. (2005). An index to quantify an individual's research output. *Proceedings of the National Academy of Sciences of the USA*, *102*(46), 16569-16572.

Jin, BH. & Rousseau, R. (2005). China's quantitative expansion phase: exponential growth but low impact. In: P. Ingwersen & B. Larsen (Eds.) *Proceedings of ISSI 2005* (pp. 362-370). Stockholm: Karolinska University Press,.

King, D. A. (2004). The scientific impact of nations. *Nature*, *430*(6997), 311–316.

Larivière, V., Lozano, G.A., & Gingras, Y. (2014). Are elite journals declining? J*ournal of the Association for Information Science and Technology, 65*(4), 649-655.

Leydesdorff, L., Wagner, C.S. & Bornmann, L. (2014). The European Union, China, and the United States in the top-1% and top-10% layers of most-frequently cited publications: Competition and collaborations. *Journal of Informetrics*, *8*(3), 606-617.

Lozano, G.A., Larivière, V. & Gingras, Y. (2012). The weakening relationship between the impact factor and papers' citations in the digital age. *Journal of the American Society for Information Science and Technology*, *63*(11), 2140-2145.

Rousseau, R. & Zhang, L. (2014). How to determine the h-index of a set of publications in the WoS? *ISSI Newsletter*, *10*(3), 63-65.

Van Noorden, R., Maher, B. & Nuzzo, R. (2014). The Top-100 papers. *Nature*, *514*(7524), 550-553.

Rousseau, R. (2006). New developments related to the Hirsch index. *Science Focus 1*(4), 23-25 (in Chinese). An English version is available at: http://eprints.rclis.org/7616/

Zhou, P. & Leydesdorff, L. (2006). The emergence of China as a leading nation in science. *Research Policy*, *35*(1), 83-104.

Zhou, P. & Leydesdorff, L. (2008). China ranks second in scientific publications since 2006. *ISSI Newsletter*, *4*(1), 7–9.

An International Comparison of the Citation Impact of Chinese Journals with Priority Funding

Ping Zhou[1*] and Loet Leydesdorff[2]

[1] *pingzhou@zju.edu.cn*
Department of Information Resources Management, School of Public Affairs, Zhejiang University, No. 866 Yuhangtang Road, Hangzhou, 310058 (China)

[2] *loet@leydesdorff.net*
Amsterdam School of Communication Research (ASCoR), University of Amsterdam, PO Box 15793, 1001 NG Amsterdam (The Netherlands); http://www.leydesdorff.net

Abstract
We have investigated the citation impact of four pairs of journals in four subject categories including the category of multidisciplinary journals, journals in environmental sciences, applied mathematics, as well as metallurgy and metallurgical engineering. Each pair is composed of one Chinese journal and one leading international journal in the same subject category. Comparison is done between the selected Chinese and international journals in each pair. The four Chinese journals are selected because of priority funding by the Chinese CIU Plan in categories A and B. Compared with leading international journals in the same subject category, citation impacts of the four Chinese journals in their relevant environments are low, although they have been improving from 2004 to 2013. Leading international journals are more intensively and systematically cited than Chinese ones in the same subject category of the JCR. Regarding the CIU Plan, the level of funding seems not to follow exactly the citation impacts: Journals receiving larger amounts of funding do not necessarily perform better in citation impact, and journals receiving the same amount of subsidy may have different citation performances.

Keywords:
Citation and co-citation analysis

Introduction
Right after the United States, China has been the second largest producer of scientific publications since 2006 (Zhou & Leydesdorff, 2008; ISTIC, 2013). With citation impact rising continuously China jumped to the fifth position in 2013 in terms of national total citation impact from the eighth in 2010 (ISTIC, 2013), two years earlier in reaching the target set by the Ministry of Science and Technology (MOST) of China in the 12th National Plan for the Development of Science and Technology (NPDST). In terms of total citations received by disciplines, however, China's performance was not evenly distributed: chemistry, materials science, engineering technology, mathematics, computer science, and physics performed best by taking the second position in the world total (ISTIC, 2013).

In addition to being a second largest producer of academic papers, China is also the second largest publishing nation of academic journals. Of the 9,884 journals, approximately 5,300 are in science and technology (Liu, 2012; Yao et al., 2014). Nevertheless, international visibility of Chinese journals is still low (Jin & Rousseau, 2004; Leydesdorff & Jin, 2005; Zhou & Leydesdorff, 2007a, 2007b; ISTIC, 2014). In 2013, only 162 Chinese journals (i.e., about 3% of China's total S&T journals) were indexed in the Science Citation Index (SCI) of Thomson Reuters. Journals to be indexed in the SCI are required to satisfy basic criteria, and thus one can expect these 162 Chinese journals to be of relatively higher quality among the 5,300 Chinese S&T journals. Nevertheless, most of the SCI indexed Chinese journals do not perform well in terms of citation impact as measured by the Impact Factor. Take the data of 2011 for example, of the 114 Chinese journals indexed in the SCI, only four were in the first quartile and 23 in the second of the corresponding subject categories of JCR 2011 (Liu, 2012).

The administrative structure of Chinese journals is special, and sometimes, confusing because of the involvement of both government agencies and the practical management by editorial boards. Administration at the national level is carried out by the General Administration of Press and Publication (GAPP) that is directly led by the State Council of China. At the provincial/regional level, the Administration of Press and Publication (APP) is responsible in each province or municipality. In addition to making regulations and policies relevant to journal publication and development, the GAPP is responsible for the approval of new journals and regular censorship; provincial APPs are responsible for administration and controls (including censorship) of local journals.

Practical management of Chinese academic journals is carried out by the editorial boards affiliated to research institutes, universities, and academic associations/societies. These institutions are affiliated to respective government agencies. Different governmental agencies are responsible for different sets of journals with different policies aiming at quality improvement with a special focus on international visibility. For example, at the national level are projects such as 'Journal Phalanx of China' of the GAPP, the 'Development Strategy Research for Competitive S&T Journals' of the Ministry of Science and Technology (MOST), and the 'Key Academic Specific Foundation' of the National Natural Science Foundation of China (NSFC). Years have passed since these projects were adopted, but the original targets of raising journal quality and international visibility have remained too far to reach.

In November 2013, in order to fasten the process towards international visibility of Chinese journals, six government agencies including the China Association for Science and Technology (CASST), the Ministry of Finance, the Ministry of Education (MOE), The State Press and Publication Administration (SPPA), the Chinese Academy of Sciences (CAS), and Chinese Academy of Engineering (CAE) jointly issued a unified standard of journal selection and funding: the International Impact Upgrading Plan for Chinese S&T Journals (abbreviated as CIU Plan). The CIU Plan is carried out in two steps. The objective of the first step is to raise the Journal Impact Factor (JIF) of a selected set of Chinese journals published in English to Quartile 1 and 2 of the Impact Factor in the Journal Citation Reports (JCR), by the end of the 12th Five-Year Plan (2011-2015), and to establish a journal set in the English language that can represent research frontiers or dominant fields of China, or in fields in which China does not yet have its own journals. The second step is to form a world top-journal set to which China has independent intellectual property rights by the year 2020.

Candidate journals must be in English and under the management of the above listed six government agencies. To ensure high-quality journals to be funded, the selection scheme combines bibliometric indicators, expert reviews, and a response by editorial boards. Journals being funded are classified into four categories, namely A, B, C and D. Those in categories A, B and C already have English version and are funded for three years. The funding amount in categories A, B, and C are respectively 2 million RMB or 322,092 US$, 1 million RMB (US$ 161,046), and 0.5 million RMB (US$ 85,230), respectively. Journals in category D are those that do not but will have an English edition; they receive 0.5 million RMB each. Of the nearly 5,300 scholarly journals in science and technology, only 76 are covered by the CIU Plan, among which 66 are in the categories of A, B, and C (Yao et al., 2014).

Journals receiving the largest funding are distributed among different Subject Categories and with different performances as measured by Journal Impact Factor (JIF) in the *Journal Citation Reports*. The rank of *Nano Research* is the highest whereas that of the *Journal of Zhejiang University-Science A* is the lowest. Questions arise such as: Are these journals selected because they outperform the rest of Chinese journals in the same subject category based on the selection scheme mentioned above? How do they perform in comparison with their past, and their international counterparts?

Comparative studies between Chinese and international journals have been done before (Li, 2006; Zhou, et al., 2010; Jin & Leydesdorff, 2005; Zhou & Leydesdorff, 2007a, 2007b). Based on data of the *Journal Citation Reports* (*JCR*) of Thomson Reuters and the China Scientific and Technical Papers and Citations Database (CSTPCD) of the Institute of Scientific and Technological Information of China (ISTIC), Zhou and Leydesdorff (2007a, 2007b), for example, compared journal-journal citation relations from different perspectives, and found that international visibility of high-quality Chinese journals was low. These studies were based on data of ten or more years ago (i.e., *JCR* 2003 and 2004). The situation has changed given China's rapid development in science and technology and its increasing R&D investment during the last ten years (MOST, 2012; NBS, 2013). The CIU Plan further stimulated our interests in mapping an updated picture of the citation performance of Chinese journals in the international scholarly community. To highlight scholarly impact the current study mainly focuses on the citation impact environments of Chinese journals supported by the CIU Plan.

Methods and materials

We use routines developed by Leydesdorff & Cozzens (1992): aggregated journal-journal citation matrices of the environment of a seed journal can be harvested from *JCR* data. A seed journal is the one under investigation and acts as a starter to run the routines. Any journal indexed in the Science Citation Index (*SCI*) or Social Science Citation (*SSCI*) can be used as a seed. The relevant citation networks of the seed journal is determined by including all journals which cite or are cited by the seed journal to the extent of a contribution of (e.g.) 1% of its citation rate (He & Pao, 1986; Leydesdorff, 1986). By default the threshold is 1%, but this can be changed so as to include an appropriate number of journals in a local citation environment. For a network with too many journals, one may raise the threshold to reduce the size of the network, and vice versa.

Each journal in a network is represented by a node, which can be a circle or an ellipse in a Pajek map. The size of an ellipse is determined by the corresponding journal's contribution to the citing or citation impact environment in the year under investigation. The distinction of the vertical and horizontal size of the ellipse, informs the reader about the extent to which within-journal (self-) citations participate in the citation impact (Leydesdorff, 2007; Zhou & Leydesdorff, 2007). Note that within-journal citations can be author self-citations or citations among authors publishing in the same journal. Citation excluding journal self-citations can be considered as a measure of inter-journal communication.

In a citation impact environment, a journal's node size in the representation is determined by the logarithm of its contribution to the total number of citations in a local environment during the year under investigation. Citation counts are total of a journal during the current year; citation counts are combined for both the *SCI* and *SSCI*.

Many programs such as VOSviewer, Pajek, or Gephi can be used to visualize journal citation networks. In this study, we use Pajek because it serves the purpose of illustrating relative cited size of individual journals in local environments. Data of a citation impact environment can be imported into Pajek after being generated by the routines. The cosine between two vectors (Salton & McGill, 1983) is used to measure the similarity between the distributions for the various journals included in a citation environment (Leydesdorff, 2007). A visualized citation network showing strength of citation relations between journals in a local environment can thus be obtained.

Table 1. Journals to be investigated.

Journal Pair	*Journal Title*	*Country*	*Items in 2012*	*CIU Plan Category*	*JIF 2013*	*Rank in JIF*	*Quartile in Category*	*Category Name*
1	Chinese Science Bulletin	China	631	A	1.365	14/55	Q2	Multidisciplinary Sciences
	Science	USA	832		31.47	2/55	Q1	
2	Journal of Environmental Sciences-China	China	281	A	1.922	95/216	Q2	Environmental Sciences
	Environment International	USA	199		5.664	7/216	Q1	
3	Journal of Computational Mathematics	China	42	B	1.049	73/251	Q2	Mathematics, applied
	Foundations of Computational Mathematics	USA	23		2.152	13/251	Q1	
4	Acta Metallurgica Sinica	China	215	B	0.548	42/75	Q3	Metallurgy & Metallurgical Engineering
	Acta Materialia	USA	681		3.940	1/75	Q1	

In 2004, 71 Chinese journals were indexed in the *JCR*. Only a few journals satisfied the above three conditions; four journals were selected for the current study. For horizontal comparison, both Chinese and foreign journals must be in the same subject category of the *JCR*. Furthermore, the foreign journals do not have to be ranked first in the corresponding subject categories, but they should be in the first Quartile of Impact Factors and in the same subject category of the JCR as the selected Chinese journals. Table 1 lists journals satisfying the above conditions and will be used to study.

Results

Cited patterns of the selected journals will be investigated. The threshold is set at 1%, which means in a seed journal's citation environment, only journals contributing to 1% or more of the seed journal's total citations will be included. Due to the page limit of the ISSI 2015, only the results of the first two pairs of journals listed in Table 1 will be presented in detail. Conclusions and discussion, however, are based on the results of the four pairs of journals.

Chinese Science Bulletin versus Science

Chinese Science Bulletin. Only 10 journals contributed at least 1% of the total citation counts of *Chinese Science Bulletin* (*CSB*) in 2004, and these journals were all from China. In other words, visibility of *CSB* among foreign journals that were indexed in the *SCI*/*SSCI* was very low. As a multidisciplinary journal, citation impact of *CSB* was multidisciplinary with specific impacts in the geosciences, geology, and chemistry (Fig. 1a). In the citation impact environment of *CSB*, citation to *CSB* was highest even if within-journal citations were excluded. Within-journal citations of some Chinese journals took high proportions in their total citations, among which journals like *Acta Physica Sinca* and *Advances in Atmospheric*

Sciences were most obvious. In terms of Impact Factor, however, *Acta Geologica Sinica-English Edition* (2.150), *Science in China Series D – Earth Sciences* (0.909), *Acta Chimica Sinica* (0.895), and *Acta Petrologica Sinica* (0.805) performed relatively better than *CSB* (0.683) (Fig. 1a).

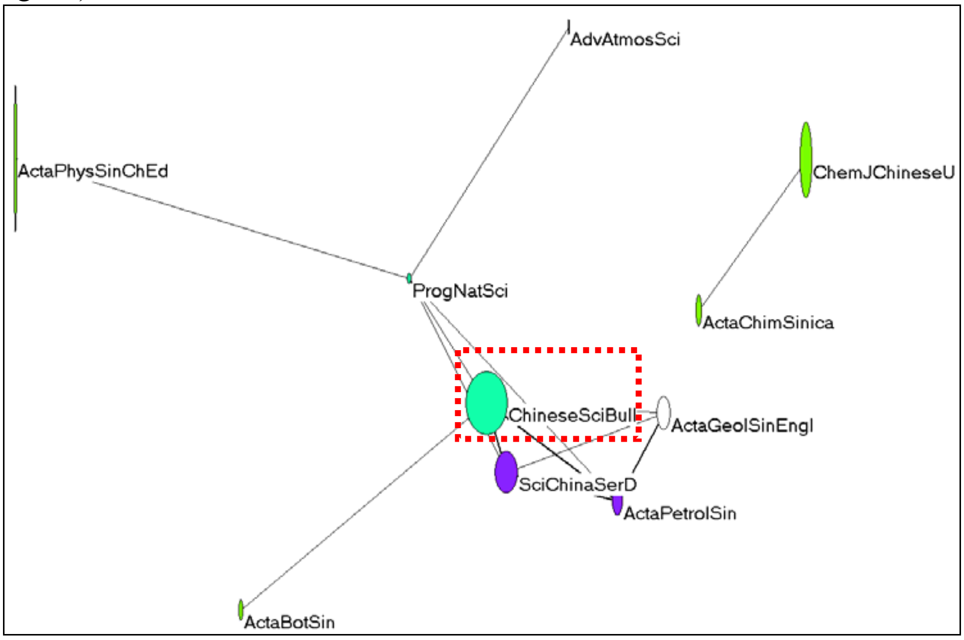


Figure 1a. Citation impact environment of *Chinese Science Bulletin* in 2004 (threshold = 1%, cosine ≥ 0.2).

Citation impact of *CSB* was enlarged to 13 journals in 2013 in terms of number of journals contributing at least 1% to the total citations of *CSB*. Most importantly, of these 13 journals eight were from other countries, which is a significant progress for Chinese journals in terms of citation impact on foreign journals compared to the year 2004. Within-journal citations contributed the most to the total citations of *CSB*. Citation impact of *CSB* on disciplines was similar to that in 2004 – involving multidisciplinary areas, geosciences, geology, and chemistry (Fig. 1b).

Impact Factor value of *CSB* were increased from 0.683 in 2004 to 1.365 in 2013.With the addition of foreign journals in the citation impact environment of *CSB*, journals with the highest citation impact is no longer *CSB* itself as in the year 2004; but instead, foreign journals such as the *Journal of Geophysical Research*, *Lithos*, and *Precambrian Research*, take the lead. In other words, in the citation impact environment of the Chinese journal *CSB*, citation impact of foreign journals was higher than that of Chinese journals. In terms of within-journal citations, *Journal of Geophysical Research* and *PLoS ONE* are most pronouncedly present. The heavy within-journal citations made the node of *PLoS ONE* a vertical line - citations from other journals in this environment were almost negligible (Fig. 1b).

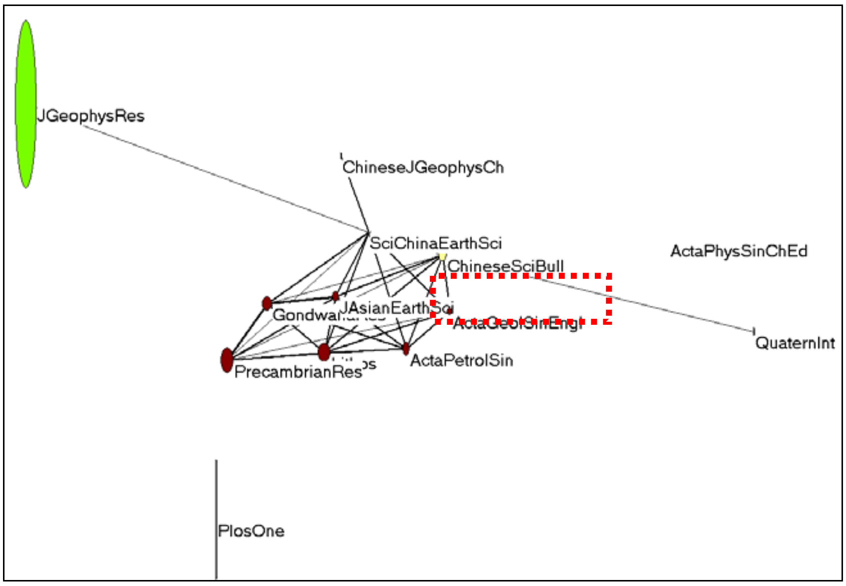


Figure 1b. Citation impact environment of *Chinese Science Bulletin* in 2013 (threshold = 1%, cosine ≥ 0.2).

***Science*.** The citation impact network of *Science* was very much focused in 2004: Three journals contributed mostly to the citations of *Science*, and none of these three was from China. Except within journal citations of *Science*, the other two top contributors were *Journal of Biological Chemistry* (*JBC*) and *Proceedings of the National Academy of Sciences of the United States of America* (*PNAS*) (Fig. 2a). Unlike the multidisciplinary journal *Chinese Science Bulletin* with distinct impact on geosciences and geology, citation impact of *Science* was more in biochemistry, in addition to impact in multiple disciplines. In terms of citation impact in the citation environment of *Science*, all the three journals are high with *JBC* having the highest impact. When within-journal citations are excluded, however, *PNAS* performed the best, and *Science* came next. In other words, compared with *JBC, PNAS* and *Science* had higher visibility in other journals. The distinct performance of citation impact of *JBC* and *PNAS* might largely be attributed to their high volumes of publications. In 2003, publications of *JBC*, *PNAS*, and *Science* were 6,585, 3084, and 845, respectively. In terms of average citation impact measured by the Impact Factor, however, *Science* performed the best (IF = 31.85), and followed by *PNAS* (IF = 10.452) and *JBC* (IF = 6.355).

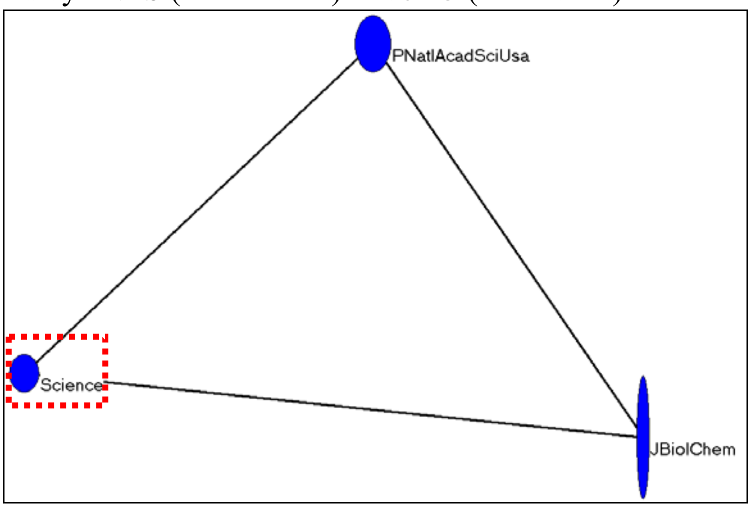


Figure 2a. Citation impact environment of *Science* in 2004 (threshold = 1%, cosine ≥ 0.2).

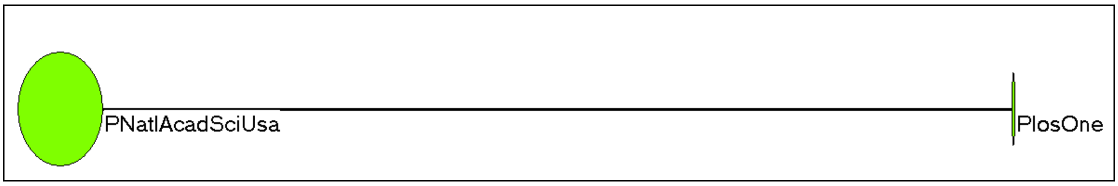


Figure 2b. Citation impact environment of *Science* in 2013 (threshold = 1%, cosine ≥ 0.2).

In the citation environment of *Science* in 2013, the percentage of within-journal citations of *Science* declined to less than 1% of its total citations. As a result, *Science* did not appear in its citation impact environment. In other words, the citation impact of *Science* was even more concentrated than in 2004. Impact Factor value of *Science* had increased from 31.853 in 2004 to 34.463 in 2013. Again, no Chinese journals appeared in this environment. *Science* was mostly cited by two multidisciplinary journals – *PNAS* and *PLoS ONE*, implying the multidisciplinary citation impact of *Science* with no distinct field emphasis like the situation in 2004. The high total citation impact of *PNAS* and *PLoS ONE* can be partially attributed to their high volume of publications: In 2013 *PLoS ONE* published 31,496 papers, which was eight times of that of the *PNAS* (3,901) and 37 times of that of *Science* (841). In terms of average citation impact (i.e., JIF), however, *Science* performed the best (31.477), and *PNAS* (9.809) came next. Average citation impact of *PLoS ONE* was the lowest (3.534), and furthermore, with heavy within-journal citations (Fig. 2b).

In summary, *Science* is widely cited in many journals in a range of different disciplines. When the threshold is set at 1%, however, only two or three journals are left in the citation impact environment of *Science*. In other words, these journals cited *Science* more intensively than other journals.

Journal of Environmental Sciences-China versus Environment International

Journal of Environmental Sciences-China. By 2004, the *Journal of Environmental Sciences-China* (*JES*) only received in total 193 citations of which 27 within-journal citations contributed the most; the other citations were scattered among journals in the environmental sciences, geosciences, chemistry, and biosciences. Although journals contributing 1% or more to *JES*'s total citation were mostly foreign and were as many as 26, these journals cited *JES* for only two or three times. In other words, except within-journal citations, there were no other journals citing *JES* systematically. Impact Factors of journals citing the *JES* were also low, between the highest of *Applied Catalysis B- Environmental* (4.042) citing *JES* six times in total and the lowest (0.172) of *Journal of the Chemical Society of Pakistan* citing *JES* four times (Fig. 3a). In other words, the *JES* had very low impact on other journals, citation impact in terms of Impact Factors of those citing *JES* occasionally was also very low.

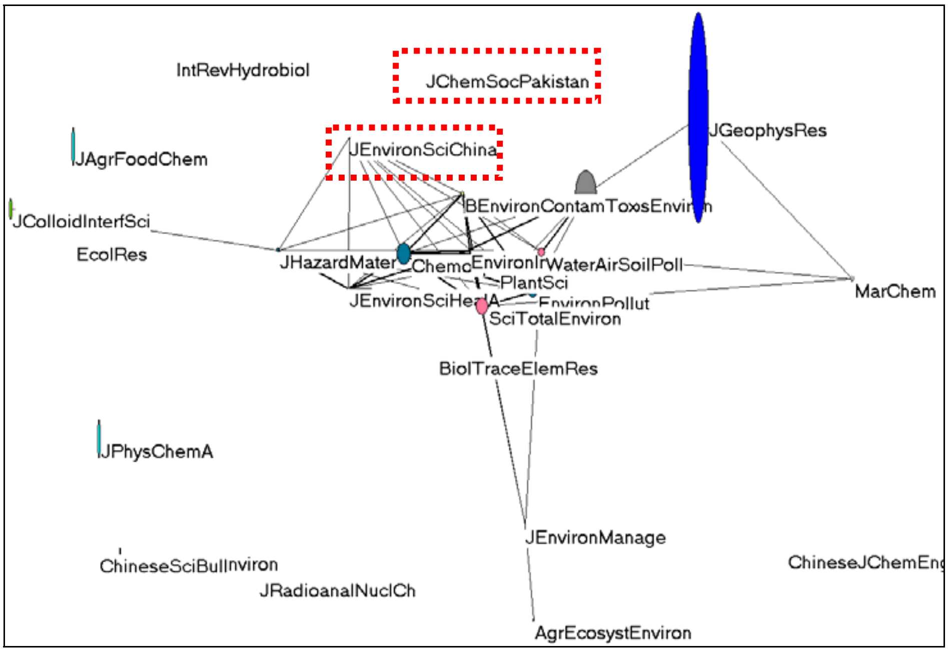


Figure 3a. Citation impact environment of *Journal of Environmental Sciences-China* in 2004 (threshold = 1%, cosine ≥ 0.2).

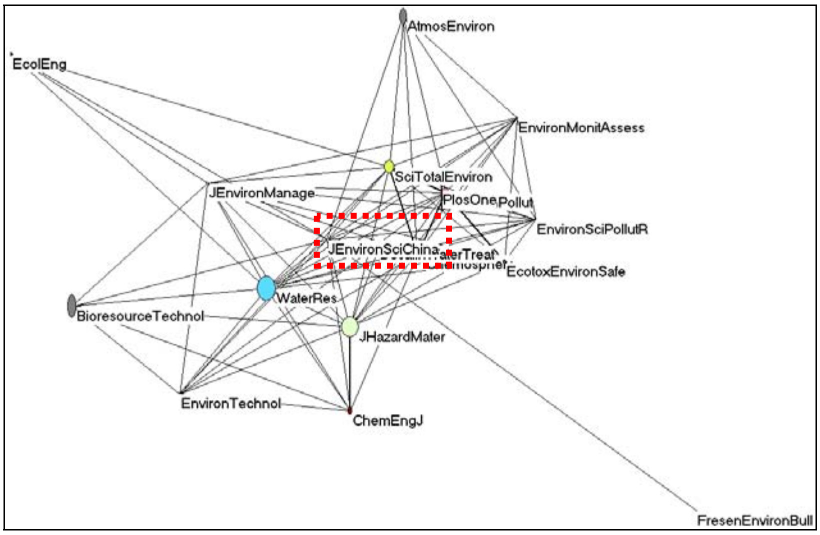


Figure 3b. Citation impact environment of *Journal of Environmental Sciences-China* in 2013 (threshold = 1%, cosine ≥ 0.2).

Performance of *JES* had been improved significantly in 2013, in addition to a large increase of the Impact Factor value from 0.254 in 2004 to 1.922 in 2013. Compared with the citation impact environment in 2004, the number of journals citing *JES* was less (i.e., 18 journals) but each contributed more citations. Journals citing *JES* were mostly foreign, although within-journal citations were still the first contributor. Instead of being cited occasionally like it was ten years ago, *JES* received more focused citation from other journals, and citation impact was more focused instead of scattering among different disciplines. For example, the foreign journal *Environmental Science and Pollution Research* contributed 28% of *JES*'s total citation by 2013, but did not appear in the citation environment of *JES* in 2004. Furthermore, journals citing *JES* had higher citation impact than those in 2004 ranging from 0.527 to 5.323. Citation relations among journals in the citation impact environment of *JES* formed closer relationship and thus interlinked with one another (Fig. 3b).

Environment International. In 2004, the citation impact of the *Environment International* was concentrated on environmental science. The journal contributing most to the total citations of *Environment International* was *Environmental Science & Technology*. Within-journal citations played much less a role than that of the *Journal of Environmental Sciences-China*. Impact Factors of journals citing the *Environment International* were much higher than those of the *Environment International*. For example, the Impact Factor of *Environmental Science & Technology,* the largest citation contributor to the citation impact of *Environment International*, was 3.557, which was even higher than that of *Environment International* (2.335). In other words, the *Environment International* had significant citation impact on high-quality journals. In the citation environment of *Environment International*, its citation impact was negligible whereas that of *Environmental Science & Technology* was highest (Fig. 4a).

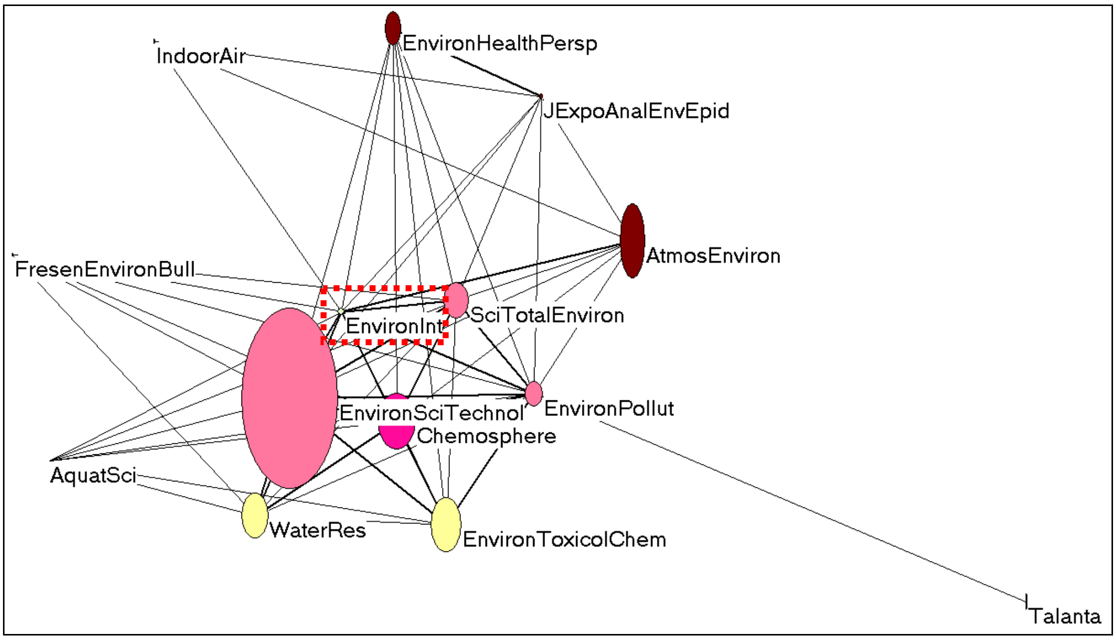


Figure 4a. Citation impact environment of *Environment International* in 2004 (threshold = 1%, cosine ≥ 0.2).

From 2004 to 2013, the Impact Factor value of *Environment International* increased from 2.335 to 5.664. Citation impact on number of journals extended from 14 to 18. Journals citing *Environment International* most frequently were *Chemosphere* (IF = 3.499) and *Science of the Total Environment* (IF = 3.163). Impact Factors of journals contributing at least 1% to the citation of *Environment International* were ranging from 1.679 to 5.664. In the citation impact environment of *Environment International*, the citation impact of *Environment International* itself became visible whereas that of *Environmental Science & Technology* was still the highest (Fig. 4b).

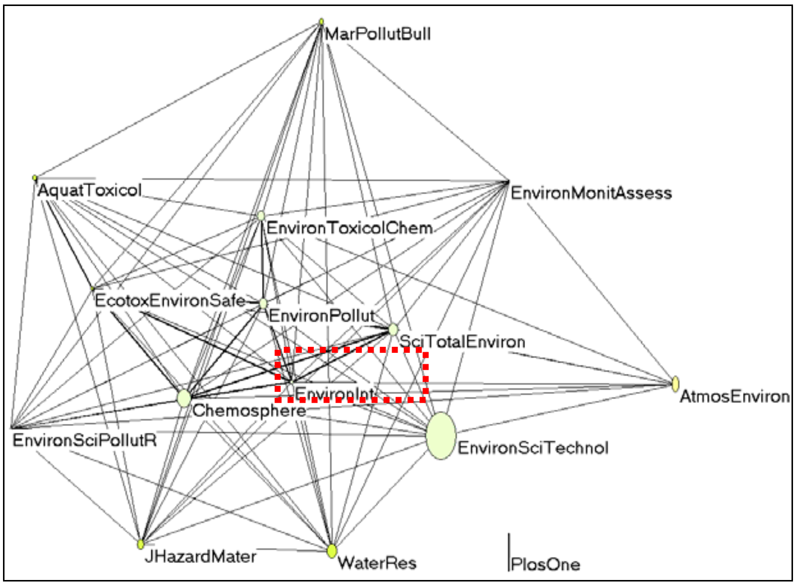


Figure 4b. Citation impact environment of *Environment International* in 2013 (threshold = 1%, cosine ≥ 0.2).

Conclusions and discussion

We have carried out a comparative study on journal citation impact between four pairs of journals in multiple disciplines, environmental sciences, applied mathematics, as well as metallurgy and metallurgical engineering. The four Chinese journals are selected because of additional funding by the Chinese CIU Plan in categories A and B. In Category A are *Chinese Science Bulletin* (*CSB*) and *Journal of Environmental Sciences-China* (*JES*), and in Category B are *Journal of Computational Mathematics* (*JCM*) and *Acta Metallurgica Sinica* (*AMS*). Leading foreign journals were used as matched pairs with the four Chinese journals. These are *Science*, *Environment International*, *Foundations of Computational Mathematics*, and *Acta Materialia* respectively.

International visibility of *CSB* was very low in 2004 although being indexed in the SCI and with a citation impact only on Chinese journals. The situation has been improved ten years later in 2013. More foreign journals cited *CSB*, but this may be by Chinese authors. Citation impact measured by Impact Factor of *CSB* has also been increased, but is still a long distance away from the best. Compared with *CSB*, *Science* has citation impact on higher quality journals measured by Impact Factor, and was cited more intensively with just two or three multidisciplinary journals contributing most to the citation counts of *Science*. By the year 2013, most citations to *Science* were from two multidisciplinary journals - *Proceedings of the National Academy of Sciences of the United States of America* (*PNAS*) and *PLoS ONE*.

Within-journal citations were the first contributor of *CSB*, whereas this is not the case for *Science*. As a multidisciplinary journal, *CSB* did not appear in the citation impact environment of *Science*, implying a weak contribution of references in *CSB* to *Science*. On the other hand, the absence of *Science* in the citation environment of *CSB* implies that *CSB* has a long way to go before coming into the sight of authors publishing in *Science*.

Although being cited by foreign journals in 2004, citations received by the *Journal of Environmental Sciences-China* (*JES*) remained occasional. The situation has improved ten years later in 2013. Citation impact of *JES* has been increased significantly, but is still far behind that of the leading foreign journals in the same subject category. Compared with the *JES*, *Environment International* has citation impact on journals with higher quality measured by Impact Factor. The citation impact of the *Environment International* was more focused: Fewer journals contributing to 1% of the total citations of *Environment International* but each journal contributed more; within-journal citations of *Environment International* were less significant to total citation counts than that of the *JES*.

Similar to the *Journal of Environmental Sciences-China*, the citation impact of the *Journal of Computational Mathematics* (*JCM*) was very low and was distributed among many journals in 2004. The situation was improved in 2013 with citation impact of the *JCM* being increased significantly, but still far behind that of leading foreign journals in the same subject category. The starting point of *Foundations of Computational Mathematics* was not high in 2004 because of a short history of being indexed in the *SCI*. Compared with the *JCM*, *Foundations of Computational Mathematics* (*FCM*) has citation impact on journals with higher quality measured by Impact Factor. Citation impact of *FCM* is also more focused: Fewer journals contributing to 1% of the total citations. Within-journal citations of *Foundations of Computational Mathematics* contributed less to its total citation than that of the *JCM*.

In 2004 the citation impact of *Acta Metallurgica Sinica* (*AMS*) was low and scattered among many journals, most of which were from China. Within-journal citation was rather heavy and became even heavier in 2013. Citation impact had been improved slightly in 2013 but was still very low. Furthermore, journal quality measured by Impact Factors of journals citing *AMS* had not been improved during 2004-2013. In contrast to *AMS*, *Acta Materialia* was able to generate citation impact in journals with higher quality measured by Impact Factors. Similar to *Acta Metallurgica Sinica*, within-journal citations of *Acta Materialia* also contributed first to its own total citation.

In general, the citation impact of leading Chinese journals has improved during the period 2004-2013, but there is still a long distance to catch up with leading foreign journals. Although being funded under Category B in the CIU Plan, *Journal of Computational Mathematics* performed as well as the other two in a higher rank of category – Category A of the CIU Plan. Being funded at the same level under Category B, the *Journal of Computational Mathematics* performed better than *Acta Metallurgica Sinica*. Foreign journals of higher Impact Factor are more intensively cited than Chinese journals at a given threshold (e.g., 1%) in the same subject category of the JCR, which may imply a positive correlation between journal quality and citation intensity in a specialist citation environment. In other words, journals with higher Impact Factor in the same subject category may be cited more intensively, or by a relatively stable number of journals in their citation impact environment across different years.

Acknowledgement

The study was supported by National Natural Science Foundation of China (NSFC) with Grant Number 71473219 and the Planning Office of Philosophy and Social Sciences of Hangzhou City, Zhejiang Province, with Grant Number B14TD02. The authors thank Thomson Reuters for access to the JCR data.

References

He, C., & Pao, M.L. (1986). A discipline-specific journal selection algorithm. *Information Processing & Management*, *22*(5), 405–416.

ISTIC (2013). 中国科技论文统计结果 2013 (Statistical Data of Chinese S&T Papers). Beijing: Institute of Scientific and Technical Information of China.

ISTIC (2014). 中国科技论文统计结果 2014 (Statistical Data of Chinese S&T Papers). Beijing: Institute of Scientific and Technical Information of China.

Jin, B.H.& Leydesdorff, L (2005), 中国科技期刊引文网络：国际影响和国内影响分析(Citation networks of Chinese S&T journals: analysis on international and domestic impact), 中国科技期刊研究*(Chinese Journal of Scientific and Technical Periodicals)*, *16*(2): 141–146.

Leydesdorff, L. (1986). The development of frames of references. *Scientometrics, 9*, 103–125.

Leydesdorff, L., & Cozzens, S.E. (1993). The delineation of specialties in terms of journals using the dynamic journal set of the Science Citation Index. *Scientometrics, 26*, 133–154.
Leydesdorff, L., & Jin, B.H. (2005). Mapping the Chinese Science Citation Database in terms of aggregated journal–journal citation relations. J*ournal of the American Society of Information Science & Technology, 56*(14), 1469–1479.
Leydesdorff, L. (2007). Visualization of the Citation Impact Environments of Scientific Journals: An Online Mapping Exercise, *Journal of the American Society for Information Science and Technology, 58*(1):25-38.
Li, L. (2012). 解读"科技期刊国际影响力提升计划" (Understanding the International Impact Upgrading Plan for Journals in Science and Technology). 科技日报 (Science and Technology Daily), Page 3, December 27, 2012.
Li, ZX (2006). How to establish a first-class international scientific journal in China? *World Journal of Gastroenterology, 12* (43): 6905-6908.
MOST (2012). China Science & Technology Statistics Database. http://www.sts.org.cn/sjkl/kjtjdt/index.htm
NBS (2013). Expenditure on R&D (100 million yuan) in: China Statistical Year Book. Available at: http://data.stats.gov.cn/english/easyquery.htm?cn=C01Ren, S.L. (2005). Editing scientific journals in Mainland China. European Science Editing, 31(1), 8–9.
Yao, Z.C., Luo, Z.F., Jin, X.Y. & Duan, R.Y. (2014). 《中国科技期刊国际影响力提升计划》资助期刊的分析与展望(New start, new task, new development: Analysis and expectation on journals funded by the International Impact Upgrading Plan for China's S&T journals) (in Chinese). 编辑学报 (*Acta Editologica*), *26*(4): 342-346.
Zhou, P., Leydesdorff, L. & Wu, Y.S. (2005).中国科技期刊引文环境的可视化(Visualization of the Citation Impact Environments in the CSTPC Journal Set), 中国科技期刊研究 (Chinese Journal of Scientific and Technical Periodicals), *16*(6), 773-780.
Zhou, Ping & Loet Leydesdorff (2007a). A Comparison Between the China Scientific and Technical Papers and Citations Database and the Science Citation Index in Terms of Journal Hierarchies and Inter-journal Citation Relations, *Journal of the American society for Information science and Technology*, *58*(2):223–236.
Zhou, Ping & Loet Leydesdorff (2007b). The citation impacts and citation environments of Chinese journals in mathematics, *Scientometrics, 72*(2).
Zhou, P. & Leydesdorff, L. (2008). China ranks second in scientific publications since 2006. *ISSI Newsletter*, *4*(1): 7-9.
Zhou, P. (2009). Mapping knowledge production and scholarly communication in China. PhD Thesis. University of Amsterdam. http://dare.uva.nl/document/2/64587
Zhou, Ping, Xinning Su & Loet Leydesdorff (2010), A comparative study on communication structures of Chinese journals in the social sciences. *Journal of the American society for Information science and Technology, 61*(7):1360–1376.

Research Data Explored: Citations versus Altmetrics

Isabella Peters[1], Peter Kraker[2], Elisabeth Lex[3], Christian Gumpenberger[4], and Juan Gorraiz[4]

[1] *i.peters@zbw.eu*
ZBW Leibniz Information Centre for Economics, Düsternbrooker Weg 120, D-24105 Kiel (Germany) & Kiel University, Christian-Albrechts-Platz 4, D-24118 Kiel (Germany)

[2] *pkraker@know-center.at*
Know-Center, Inffeldgasse 13, A-8010 Graz (Austria)

[3] *elex@know-center.at*
Graz University of Technology, Knowledge Technologies Institute, Inffeldgasse 13, A-8010 Graz (Austria)

[4] *christian.gumpenberger, juan.gorraiz@univie.ac.at*
University of Vienna, Vienna University Library, Dept of Bibliometrics, Boltzmanngasse 5, A-1090 Vienna (Austria)

Abstract
The study explores the citedness of research data, its distribution over time and how it is related to the availability of a DOI (Digital Object Identifier) in Thomson Reuters' DCI (Data Citation Index). We investigate if cited research data "impact" the (social) web, reflected by altmetrics scores, and if there is any relationship between the number of citations and the sum of altmetrics scores from various social media-platforms. Three tools are used to collect and compare altmetrics scores, i.e. PlumX, ImpactStory, and Altmetric.com. In terms of coverage, PlumX is the most helpful altmetrics tool. While research data remain mostly uncited (about 85%), there has been a growing trend in citing data sets published since 2007. Surprisingly, the percentage of the number of cited research data with a DOI in DCI has decreased in the last years. Only nine repositories account for research data with DOIs and two or more citations. The number of cited research data with altmetrics scores is even lower (4 to 9%) but shows a higher coverage of research data from the last decade. However, no correlation between the number of citations and the total number of altmetrics scores is observable. Certain data types (i.e. survey, aggregate data, and sequence data) are more often cited and receive higher altmetrics scores.

Conference Topic
Altmetrics, Citation and co-citation analysis

Introduction
Recently, data citations have gained momentum (Piwowar & Chapman, 2010; Borgman, 2012; Torres-Salinas, Martín-Martín, & Fuente-Gutiérrez, 2013). This is reflected, among others, in the development of data-level metrics (DLM), an initiative driven by PLOS, UC3 and DataONE[1], to track and measure activity on research data, and the recent announcement of CERN to provide DOIs for each dataset they share through their novel Open Data portal[2]. Data citations are citations included in the reference list of a publication that formally cite either the data that led to a research result or a data paper[3]. Thereby, data citations indicate the influence and reuse of data in scientific publications.
First studies on data citations showed that certain well-curated data sets receive far more citations or mentions in other articles than many traditional articles (Belter, 2014). Citations, however, are used as a proxy for the assessment of impact primarily in the "publish or perish" community; to consider other disciplines and stakeholders of research, such as industry, government and academia, and in a much broader sense, the society as a whole, altmetrics

[1] http://escholarship.org/uc/item/9kf081vf
[2] https://www.datacite.org/news/cern-launches-data-sharing-portal.html
[3] http://www.asis.org/Bulletin/Jun-12/JunJul12_MayernikDataCitation.html

(i.e. social media-based indicators) are emerging as a useful instrument to assess the "societal" impact of research data or at least to provide a more complete picture of research uptake, besides more traditional usage and citation metrics (Bornman, 2014; Konkiel, 2013). Previous work on altmetrics for research data has mainly focused on motivations for data sharing, creating reliable data metrics and effective reward systems (Costas et al., 2012).

This study contributes to the research on data citations in describing their characteristics as well as their impact in terms of citations and altmetrics scores. Specifically, we tackle the following research questions:

- How often are research data cited? Which and how many of these have a DOI? From which repositories do research data originate?
- What are the characteristics of the most cited research data? Which data types and disciplines are the most cited? How does citedness evolve over time?
- To what extent are cited research data visible on various altmetrics channels? Are there any differences between the tools used for altmetrics scores aggregation?

Data sources

On the Web, a large number of data repositories are available to store and disseminate research data. The Thomson Reuters Data Citation Index (DCI), launched in 2012, provides an index of high-quality research data from various data repositories across disciplines and around the world. It enables search, exploration and bibliometric analysis of research data through a single point of access, i.e. the Web of Science (Torres-Salinas, Martín-Martín & Fuente- Gutiérrez, 2013). The selection criteria are mainly based on the reputation and characteristics of the repositories[4]. Three document types are available in the DCI: data set, data study, and repository. The document type "repository" can distort bibliometric analyses, because repositories are mainly considered as a source, but not as a document type.

First coverage and citation analyses of the DCI have been performed April-June 2013 by the EC3 bibliometrics group of Granada (Torres-Salinas, Jimenez-Contreras & Robinson-Garcia, 2014; Torres-Salinas, Robinson-Garcia & Cabezas-Clavijo, 2013). They found that data is highly skewed: Science areas accounted for almost 80% of records in the database and four repositories contained 75% of all the records in the database; 88% of all records remained uncited. In Science, Engineering and Technology citations are concentrated among datasets, whereas in the Social Sciences and Arts & Humanities, citations often refer to data studies.

Since these first analyses, DCI has been constantly growing, now indexing nearly two million records from high-quality repositories around the world. One of the most important enhancements of the DCI has undoubtedly been the inclusion of "figshare[5]" as new data source which led to an increase of almost a half million of data sets and 40.000 data studies (i.e. about one fourth of the total coverage in the database).

Gathering altmetrics data is quite laborious since they are spread over a variety of social media platforms which each offer different applications programming interfaces (APIs). Tools, which collect and aggregate these altmetrics data come in handy and are now fighting for market shares since also large publishers increasingly display altmetrics for articles (e.g., Wiley[6]). There are currently three big altmetrics data providers: ImpactStory[7], Altmetric.com, and PlumX[8]. Whereas Altmetrics.com and PlumX focus more on gathering and providing

---

[4] http://thomsonreuters.com/data-citation-index, http://thomsonreuters.com/products/ip-science/04_037/dci-selection-essay.pdf

[5] http://figshare.com

[6] http://eu.wiley.com/WileyCDA/PressRelease/pressReleaseId-108763.html?campaign=wlytk-41414.4780439815

[7] https://impactstory.org

[8] https://plu.mx

data for institutions (e.g., publishers, libraries, or universities), ImpactStory's target group is the individual researcher who wants to include altmetrics information in her CV.

ImpactStory is a web-based tool, which works with individually assigned permanent identifiers (such as DOIs, URLs, PubMed IDs) or links to ORCID, Figshare, Publons, Slideshare, or Github to auto-import new research outputs like e.g. papers, data sets, slides. Altmetric scores from a large range of social media-platforms, including Twitter, Facebook, Mendeley, Figshare, Google+, and Wikipedia[9], can be downloaded as .json or .csv (as far as original data providers allow data sharing). With Altmetric.com, users can search within a variety of social media-platforms (e.g., Twitter, Facebook, Google+, or 8,000 blogs[10]) for keywords as well as for permanent identifiers (e.g., DOIs, arXiv IDs, RePEc identifiers, handles, or PubMed IDs). Queries can be restricted to certain dates, journals, publishers, social media-platforms, and Medline Subject Headings. The search results can be downloaded as .csv from the Altmetric Explorer (web-based application) or via the API. Plum Analytics or Plum X (the fee-based altmetrics dashboard) offers article-level metrics for so-called artifacts, which include articles, audios, videos, book chapters, or clinical trials[11]. Plum Analytics works with ORCID and other user IDs (e.g., from YouTube, Slideshare) as well as with DOIs, ISBNs, PubMed-IDs, patent numbers, and URLs. Because of its collaboration with EBSCO, Plum Analytics can provide statistics on the usage of articles and other artifacts (e.g., views to or downloads of html pages or pdfs), but also on, amongst others, Mendeley readers, GitHub forks, Facebook comments, and YouTube subscribers.

Methodology

We used DCI to retrieve the records of cited research data. All items published in the last 5.5 decades (1960-9, 1970-9, 1980-9, 1990-9, 2000-9, and 2010-4) with two or more citations (Sample 1, n=10,934 records) were downloaded and analysed. The criterion of having at least two citations is based on an operational reason (reduction of the number of items) as well as on a conceptual reason (to avoid self-citations). The following metadata fields were used in the analysis: available DOI or URL, document type, source, research area, publication year, data type, number of citations and ORCID availability[12]. The citedness in the database was computed for each decade considered in this study and investigated in detail for each year since 2000. We then analysed the distribution of document types, data types, sources and research areas with respect to the availability or non-availability of DOIs reported by DCI.

All research data with two or more citations and with an available DOI (n=2,907 items) were analysed with PlumX, ImpactStory, and Altmetric.com and their coverage on social media platforms and the altmetric scores was compared. All other items with 2 or more citations and an available URL (n=8,027) were also analysed in PlumX, the only tool enabling analyses based on URLs, and the results were compared with the ones obtained for items with a DOI. We also analysed the distribution of document types, data types, sources and research areas for all research data with 2 or more citations and at least one altmetric score (sample 2; n=301 items) with respect to the availability or non-availability of the permanent identifier DOI reported by DCI (items with DOI and URL or items with URL only).

[9] http://feedback.impactstory.org/knowledgebase/articles/367139-what-data-do-you-include-on-profiles
[10] http://support.altmetric.com/knowledgebase/articles/83335-which-data-sources-does-altmetric-track
[11] http://www.plumanalytics.com/metrics.html
[12] The DCI field "data type" was manually merged to more general categories; e.g. "survey data in social sciences" was merged with the category "survey data".

Table 1. Results of DCI-based citation and altmetrics analyses for the last 5.5 decades.

DCI	*1960-69*	*1970-79*	*1980-89*	*1990-99*	*2000-09*	*2010-14*
total # items	6 040	23 712	43 620	186 965	2 096 023	1 627 668
# items with > 2 citations	5	110	360	956	4 727	4 777
# items with at least 1 citation	5	4207	7519	43749	239867	218440
uncited %	99.9%	82.3%	82.8%	76.6%	88.6%	86.6%
items with DOI and >= 2 cit	4	107	343	846	1381	226
% with DOI and >=2 cit	0.8	97.27%	95.28%	88.49%	29.22%	4.73%
with Altmetrics Data (PlumX)	1	5	14	40	114	20
%	25.0%	4.7%	4.1%	4.7%	8.3%	8.8%
items with URL only and >= 2 cit	1	3	17	110	3 346	4551
% with URL only and >=2 cit	0.2	2.73%	4.72%	11.51%	70.78%	95.27%
with Altmetrics Data (PlumX)	1	1	8	11	54	33
%	100.0%	33.3%	47.1%	10.0%	1.6%	0.7%

Results and discussion

General Results

Table 1 gives an overview of the general results obtained in this study. The analysis revealed a high uncitedness of research data, which corresponds to the findings of Torres-Salinas, Martin-Martin and Fuente-Gutiérrez (2013). A more detailed analysis for each year (see Table 2) shows, however, that the citedness is comparatively higher for research data published in recent years (published after 2007) although the citation window is shorter.

Table 2. Evolution of uncitedness in DCI in the last 14 years.

PY	*Items*	*uncited*	*% uncited*
2000	28282	18152	64.18%
2001	36397	25367	69.70%
2002	64781	51464	79.44%
2003	115997	93538	80.64%
2004	141065	122802	87.05%
2005	212781	178146	83.72%
2006	299443	275216	91.91%
2007	362405	333136	91.92%
2008	398931	364236	91.30%
2009	435941	394099	90.40%
2010	390957	349623	89.43%
2011	270932	224790	82.97%
2012	492534	428752	87.05%
2013	448489	386507	86.18%
2014	24756	19556	78.99%

The results also show a very low percentage of altmetrics scores available for research data with two or more citations (see Table 1). But, two different trends can be observed: the percentage of data with DOI referred to on social media-platforms is steadily increasing while the percentage of data with URL only is steadily decreasing in the same time frame.

The percentage of research data with altmetrics scores in PlumX, the tool with the highest average in this study, is lower than expected (ranging between 4 and 9%) but actually has doubled for data published in the last decades, which confirms the interest in younger research data and an increase in social media activity of the scientific community in recent years.

Table 3. Overview on citation distribution of Sample 1 (n=10,934 items).

items with at least 2 citations	*Document Type*	*# items*	*Total Citations*	*Mean Citations*	*Maximum Citations*	*Standard Deviation*	*Variance*
all	Data set	5641	17984	3.19	121	3.38	11.46
	Data study	5242	91623	17.48	1236	50.22	2521.67
	Repository	51	10076	197.57	3193	618.73	382824.45
	Total	10934	119683	10.95	3193	56.39	3179.49
with DOI	Data set	342	977	2.86	52	3.86	14.93
	Data study	2565	53293	20.78	1236	63.44	4024.45
	Total	2907	54270	18.67	1236	59.88	3585.92
with URL only	Data set	5299	17007	3.21	121	3.35	11.23
	Data study	2677	38330	14.32	272	32.59	1062.31
	Repository	51	10076	197.57	3193	618.73	382824.45
	Total	8027	65413	8.15	3193	54.80	3003.30

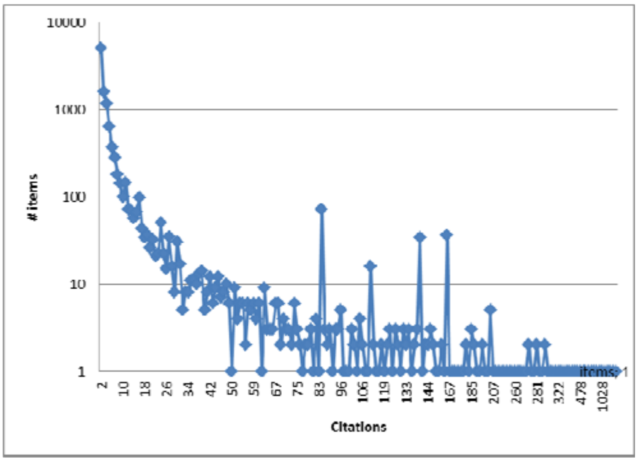


Figure 1. Citation distribution of Sample 1 (logarithmic scale).

Results for Sample 1

Table 3 shows the citation distribution of Sample 1 (10,934 items with at least two citations in DCI) for items with DOI or URL only separated according to the three main DCI document types (data set, data study, and repository[13]). The results reveal that almost half of the data studies have a DOI (48.9%) but only few data sets do so. Data studies are on average more often cited than data sets (17.5 vs. 3.2 citations per item), and data studies with a DOI attract more citations (mean values) than those with a URL (20 vs. 14 citations per item).

There were only few repositories (51) in the data set; it is the document type, which attracts the most citations per item. This finding is in line with the results of Belter (2014) who also found aggregated data sets – Belter calls them "global-level data sets" – to be more cited. However, such citing behaviour has a negative side effect on repository content (i.e., the single data sets) since it is not properly attributed in favour of citing the repository as a whole.

The high values of the standard deviation and variance illustrate the skewness of the citation distribution (see Figure 1). Almost half of the research data (4,974 items; 45.5%) have only two citations. Six items, two repositories and four data studies, from different decades (PY=1981, 1984, 1995, 2002, 2011, and 1998, sorted by descending number of citations) had more than 1,000 citations and account for almost 30% of the total number of citations.

[13] Table 3 includes repositories as document type to illustrate the citation volume in DCI.

Table 4 shows the top 10 repositories by the number of items. Considering the number of citations, there are three other repositories which account for more than 1,000 citations each: Manitoba Centre for Health Policy Population Health Research Data Repository (29 items; 1,631 citations), CHILDES - Child Language Data Exchange System (1 item; 3,082 citations), and World Values Survey (1 item; 3,193 citations). Interestingly, although "figshare" accounts for almost 25% of the DCI, no item from "figshare" was cited at least twice in DCI. We also noted that the categorization of "figshare" items is missing. All items are assigned to the Web of Science category (WC) "Multidisciplinary Sciences" or the Research Area (SU) "Science & Technology/Other Topics" preventing detailed topic-based citation analyses. Furthermore, only nine items from Sample 1 were related to an ORCID, three data sets with a DOI, and three data sets and data studies with a URL.

Table 4. Analysis of Sample 1 by sources (repositories) (n=10,934 items).

Sources (with DOI)	*# items*	*# citations*	*Sources (with URL)*	*# items*	*# citations*
Inter-university Consortium for Political and Social Research	2530	53041	miRBase	3456	10209
Worldwide Protein Data Bank	229	458	Cancer Models Database	864	2698
Oak Ridge National Laboratory Distributed Active Archive Center for Biogeochemical Dynamics	108	508	UK Data Archive	836	25479
Archaeology Data Service	21	75	European Nucleotide Archive	361	1346
3TU.Datacentrum	8	22	Gene Expression Omnibus	353	754
SHARE - Survey of Health, Ageing and Retirement in Europe	4	151	National Snow & Ice Data Center	298	2796
World Agroforestry Centre	3	6	Australian Data Archive	264	2469
Dryad	2	4	Australian Antarctic Data Centre	249	1621
GigaDB	2	5	nmrshiftdb2	219	445
			Finnish Social Science Data Archive	183	913

Considering their origin, considerable differences were reported in Sample 1 for items with or without a DOI (see Table 4). All twice or more frequently cited research data with a DOI are archived in nine repositories, while 92 repositories account for research data without a DOI.

Table 5 shows that there are big differences between the most cited data types when considering research data with a DOI or with a URL. Survey data, aggregate data, and clinical data are the most cited ones of the first group (with a DOI), while sequence data and numerical and individual level data are the most cited data types of the second group (with a URL). Apart from survey data, there is no overlap in the top 10 data types indexed in DCI. Similar results were obtained when considering data sets and data studies separately.

Disciplinary differences become apparent in the citations of DOIs and URLs as well as in the use of certain document types. As shown in Table 6 it is more common to refer to data studies via DOIs in the Social Sciences than in the Natural and Life Sciences, where the use of URLs for both data studies and data sets is more popular. Torres-Salinas, Jimenez-Contreras and Robinson-Garcia (2014) also report that citations in Science, Engineering and Technology citations are concentrated on data sets, whereas the majority of citations in the Social Sciences and Arts & Humanities refer to data studies. Table 6 suggests that these differences could be related to the availability of a DOI.

Table 5. Analysis of Sample 1 by data types (manually merged), top 10 types (n=10,934 items).

Data Types (with DOI)	*# items*	*# citations*	*Data Types (with URL only)*	*# items*	*# citations*
survey data	1734	43686	sequence data	3408	10458
administrative records data	302	3326	profiling by array, gen, etc	352	752
aggregate data	274	9440	Individual (micro) level	240	9024
event/transaction data	210	2400	Numeric data	216	4317
clinical data	118	3469	Structured questionnaire	155	673
census/enumeration data	109	1019	survey data	127	1315
protein structure	95	190	Seismic:Reflection:MCS	47	185
observational data	30	575	statistical data	41	1352
program source code	10	116	Digital media	40	290
roll call voting data	8	236	EXCEL	25	101

Table 6. Sample 1 by research areas and document types, top 10 areas (n=10,934 items).

with DOI					with URL only				
	# Items		# citations			# Items		# citations	
Research Area	Data set	Data study	Data set	Data study	Research Area	Data set	Data study	Data set	Data study
Criminology & Penology		471		4403	Genetics & Heredity	4658	159	14024	571
Sociology		432		7930	Meteorology & Atmospheric Sciences	91	298	493	2796
Government & Law		352		10399	Biochemistry & Molecular Biology; Genetics & Heredity		353		754
Demography		317		9178	Sociology		286		1994
Health Care Sciences & Services		290		8170	Physics	5	214	10	435
Biochemistry & Molecular Biology	229		458		Business & Economics; Sociology		143		12665
Business & Economics		204		3083	Biochemistry & Molecular Biology; Spectroscopy	129		383	
Environmental Sciences & Ecology; Geology	108		508		Oceanography; Geology	114		353	
Education & Educational Research		69		1881	Demography; Sociology		103		5673
Family Studies		68		2268	Sociology; Demography; Communication		84		393

Results for Sample 2

Sample 2 comprises all items from DCI satisfying the following criteria: two or more citations in DCI, a DOI or a URL and at least one altmetrics score in PlumX (n=301 items). Table 7 shows the general results for this sample. The total number of altmetrics scores is lower than the number of citations for all document types with or without a DOI. Furthermore, the mean altmetrics score is higher for data studies than for data sets.

Tables 8 and 9 show the distributions of data types and subject areas in this sample. Most data with DOI are survey data, aggregate data, event over transaction data, whereas sequence data and images are most often referred to via URL only (see Table 8). Microdata with DOI and spectra with URL only are the data types with the highest altmetrics scores per item. Concerning subject areas the results of Table 9 are very similar to the results of Table 6. Given the small sample size it is, however, notable that in some subject areas, e.g. Archaeology, research data receive more interest in social media (i.e. altmetrics scores), than via citations in traditional publications. This is confirmed by the missing correlation between citations and altmetrics scores for this sample (see Figure 2). Both cases clearly demonstrate that altmetrics can complement traditional impact evaluation. Nevertheless, coverage of research data in social media is still low, e.g. from the nine repositories whose data studies and data sets were cited twice in DCI and had a DOI (see Table 4), only five items had altmetrics scores in PlumX, and only one DOI item of Sample 2 included an ORCID.

Table 7. Citation and altmetrics results of Sample 2 (n=301 items) according to document type. *8 items with URL found in PlumX could not properly be identified (broken URL, wrong item, etc.)

	Document Type	# items	Total Citations	Mean Citations	Maximum Citations	Standard Deviation	Variance
with DOI	Data set	15	173	11.53	52	13.75	189.12
	Data study	179	6716	37.52	1135	107.36	11525.43
	Total	194	6889	35.51	1135	103.40	10691.82
	Document Type	**# items**	**Total Scores**	**Mean Scores**	**Maximum Scores**	**Standard Deviation**	**Variance**
	Data set	15	34	2.27	6	1.75	3.07
	Data study	179	710	3.97	64	7.42	55.09
	Total	194	752	376.00	748	526.09	276768.00
	Document Type	**# items**	**Total Citations**	**Mean Citations**	**Maximum Citations**	**Standard Deviation**	**Variance**
with URL only	Data set	24	172	7.17	46	10.12	102.41
	Data study	31	779	25.13	272	51.67	2669.65
	Repository	44	9677	219.93	3193	662.92	439464.20
	Total*	99	10628	107.35	3193	451.61	203954.50
	Document Type	**# items**	**Total Scores**	**Mean Scores**	**Maximum Scores**	**Standard Deviation**	**Variance**
	Data set	24	428	17.83	378	76.75	5890.23
	Data study	31	664	21.42	213	53.25	2835.65
	Repository	44	3961	90.02	1150	198.53	39415.70
	Total*	99	5319	49.71	1150	139.82	19549.38

Table 8. Citation and altmetrics overview of Sample 2 (n=301 items) according to their data type (Field DY; no aggregated counts, "document type" "repository" (34 items) not included.

Data Type (with DOI)	*# items*	*total citations*	*mean citations*	*total scores*	*mean scores*	*Data Type (with URL only) **	*# items*	*total citations*	*mean citations*	*total scores*	*mean scores*
survey data	110	5276	47.96	353	3.21	miRNA sequence data	15	71	4.73	21	1.40
aggregate data	26	793	30.50	80	3.08	FITS images; spectra; calibrations; redshifts	4	248	62	16	4.00
event/transaction data	19	414	21.79	43	2.26	statistical data	3	333	111	22	7.33
administrative records data	13	125	9.62	58	4.46	Expression profiling by array	3	6	2	4	1.33
clinical data	11	314	28.55	26	2.36	Sensor data; survey data	2	51	25.5	10	5.00
census/enumeration data	8	90	11.25	14	1.75	Quantitative	2	35	17.5	10	5.00
observational data	4	99	24.75	7	1.75	images	1	20	20	3	3.00
Longitudinal data; Panel Data; Micro data	2	79	39.50	46	23.00	images; spectra	1	4	4	102	102.00
roll call voting data	2	178	89.00	3	1.50	table	1	9	9	1	1.00
machine-readable text	1	5	5.00	1	1.00	redshifts; spectra	1	5	5	213	213.00
program source code	1	2	2.00	1	1.00	images; spectra; astrometry	1	2	2	90	90.00

Table 9. Citation and altmetrics overview of Sample 2 according to their subject area.

with DOI				*with URL only*			
Subject Areas	***# items***	***# citations***	***# scores***	***Subject Areas***	***# items***	***# citations***	***# scores***
Sociology	35	1226	213	Genetics & Heredity	26	492	654
Government & Law	28	793	53	Meteorology & Atmospheric Sciences	15	166	28
Criminology & Penology	22	317	42	Astronomy & Astrophysics	9	933	427
Health Care Sciences & Services	14	1498	70	Biochemistry & Molecular Biology; Genetics & Heredity	5	22	557
Environmental Sciences & Ecology; Geology	14	171	33	Cell Biology	4	13	383
Demography	12	433	28	Health Care Sciences & Services; Business & Economics	3	335	68
Family Studies	10	166	26	Genetics & Heredity; Biochemistry & Molecular Biology	2	27	36
Archaeology	10	47	139	Business & Economics	2	35	10
Education & Educational Research	9	661	40	Health Care Sciences & Services	2	423	2
International Relations	9	384	46	Communication; Sociology; Telecommunications	2	51	10

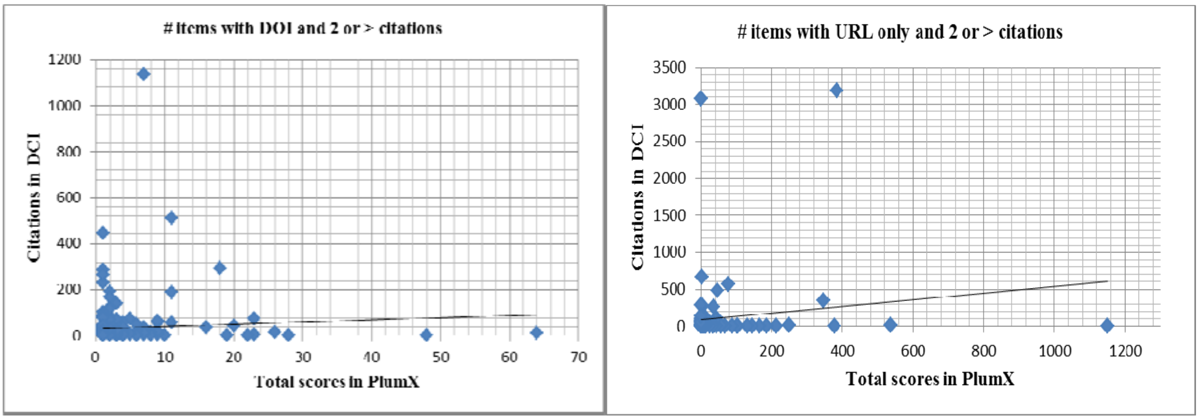


Figure 2. Citations DCI versus scores in PlumX for items with (left) and without (right).

Selected altmetrics scores and comparison of the results of three altmetrics tools

Table 10 shows the general results obtained in PlumX according to PlumX's aggregation groups (i.e., captures, social media, mentions, and usage) for all document types and with or without DOI. While DOIs for data sets seem to be important in order to get captures (mainly in Mendeley), a URL is sufficient for an inclusion in social media tools like Facebook, Twitter, etc.

The top 10 research data-DOIs attracting two or more citations and with at least one entry in PlumX are shown in Table 11. We can observe that cited research data attracts more citations than altmetrics scores, and that there is no correlation between highly cited and highly scored research data.

The comparison of altmetrics aggregation tools also revealed that ImpactStory only found Mendeley reader statistics for the research data: 78 DOIs had 257 readers. Additionally, ImpactStory found one other DOI in Wikipedia. ImpactStory found five items, which have not been found by PlumX, although they all solely relied on Mendeley Data. The Mendeley data scores were exactly the same in PlumX and in ImpactStory. On the other hand, PlumX found 18 items that were not available via ImpactStory. These research data were distributed on social media platforms (mostly shares in Facebook) and one entry has been used via click on a Bitly-URL (Usage:Clicks:Bitly).The tool Altmetric.com found only one from 194 items.

As already reported in Jobmann et al. (2014), PlumX is the tool with the highest coverage of research products found on social media-platforms. Whereas Mendeley is well covered in ImpactStory, no other altmetrics score were found for the data set used in this study.

General Conclusions

Most of the research data still remain uncited (approx. 86%) and total altmetrics scores found via aggregation tools are even lower than the number of citations. However, research data published from 2007 onwards have gradually attracted more citations reflecting a bias towards more recent research data. No correlation between citation and altmetrics scores could be observed in a preliminary analysis: neither the most cited research data nor the most cited sources (repositories) received the highest scores in PlumX.

In the DCI, the availability of cited research data with a DOI is rather low. A reason for this may be the increase of available research data in recent years. Furthermore, the percentage of cited research data with a DOI has not increased as expected, which indicates that citations do not depend on this standard identifier in order to be processed by the DCI.

Table 10. PlumX altmetrics scores for all document types with or without DOI.

		with DOI			*with URL only*			
	Document Type	***Data set***	***Data study***	***Total***	***Data set***	***Data study***	***Repository***	***Total***
	# items	15	179	**194**	24	31	44	**99**
Captures	***Sum***	32	471	**503**	0	0	30	**30**
	Mean	2.13	2.63	**2.59**	0.00	0.00	0.68	**0.28**
	Max	6	48	**48**	0	0	23	**23**
Social Media	***Sum***	1	220	**221**	407	281	3060	**3890**
	Mean	0.07	1.23	**1.14**	16.96	9.06	69.55	**36.36**
	Max	1	58	**58**	366	119	1008	**1008**
Mentions	***Sum***	1	13	**14**	13	62	433	**629**
	Mean	0.07	0.07	**0.07**	0.54	2.00	9.84	**5.88**
	Max	1	4	**4**	12	31	119	**120**
Usage	***Sum***	0	6	**6**	8	321	438	**770**
	Mean	0.00	0.03	**0.03**	0.33	10.35	9.95	**7.20**
	Max	0	6	**6**	4	187	92	**187**
Total entries		**34**	**710**	744	**428**	**664**	**3961**	5319
% Captures		**94.1%**	**66.3%**	**67.6%**	**0.0%**	**0.0%**	**0.8%**	**0.6%**
% Social Media		**2.9%**	**31.0%**	**29.7%**	**95.1%**	**42.3%**	**77.3%**	**73.1%**
% Mentions		**2.9%**	**1.8%**	**1.9%**	**3.0%**	**9.3%**	**10.9%**	**11.8%**
% Usage		**0.0%**	**0.8%**	**0.8%**	**1.9%**	**48.3%**	**11.1%**	**14.5%**

Nevertheless, data studies with a DOI attract more citations than those with a URL. Despite the low number of research data with a DOI in general, surprisingly, the DOI in cited research data has so far been more embraced in the Social Sciences than in the Natural Sciences.

Furthermore, our study shows an extremely low number of research data with two or more citations (only nine out of around 10,000) related to an ORCID. Only three of them had a DOI likewise. This illustrates that we are still a far cry from the establishment of permanent identifiers and their optimal interconnectedness in a data source.

The low percentage of altmetrics scores for research data with two or more citations corroborates a threefold hypothesis: First, research data are either rarely published or not findable on social media-platforms, because DOIs or URLs are not used in references thus resulting in a low coverage of items. Second, research data are not widely shared on social media by the scientific community so far, which would result in higher altmetrics scores[14]. Third, the reliability of altmetrics aggregation tools is questionable as the results on the coverage of research data on social media-platforms differ widely between tools. However, the steadily increasing percentage of cited research data with DOI suggests that the adoption of this permanent identifier increases the online visibility of research data and inclusion in altmetrics tools (since they heavily rely on DOIs or other permanent identifiers for search).

A limitation of our study is that the results rely on the indexing quality of the DCI. Our analysis shows that the categorisation in DCI is problematic at times. This is illustrated by the fact that all items from figshare, which is one of the top providers of records, are categorised

Table 11. Top 10 Research Data with DOI according to the total scores in PlumX.

DOI	*SO*	*PY*	*Captures:Readers:Mendeley*	*Social Media:+1s:Google+*	*Social Media:Shares:Facebook*	*Social Media:Likes:Facebook*	*Social Media:Tweets:Twitter*	*Mentions:Comments:Facebook*	*# total Scores*	*# Citations*
10.5284/1000415	ADS	2012	2		13	45		4	64	13
10.3886/icpsr13580	IUC	2005	48						48	3
10.5284/1000397	ADS	2011			14	12		2	28	2
10.3886/icpsr06389	IUC	2007	25	1					26	14
10.6103/share.w4.111	SHARE	2004			8	15			23	74
10.6103/share.w4.111	SHARE	2010			8	15			23	5
10.3886/icpsr13611	IUC	2006	22						22	3
10.3886/icpsr02766	IUC	2007	20						20	44
10.5284/1000381	ADS	2009		2	3	10	3	1	19	2
10.3886/icpsr09905	IUC	1994	18						18	295
10.3886/icpsr08624	IUC	2010	16						16	36
10.3886/icpsr04697	IUC	2009	11						11	510
10.3886/icpsr06716	IUC	2007	11						11	59
10.3886/icpsr20240	IUC	2008	11						11	190
10.3886/icpsr20440	IUC	2007	3				7		10	3

into “Miscellaneous”. Also, the category “repository” is rather a source than a document type. Such incorrect assignments of data types and disciplines can easily lead to wrong interpretations in citation analyses. Furthermore, it should be taken into account that citation counts are not always traceable.

Finally, citations of research data should be studied in more detail. They certainly differ from citations of papers relying on these data with regard to dimension and purpose. For example, we found that entire repositories are proportionally more often cited than single data sets, which was confirmed by a former study (Belter, 2014). Therefore, it will be important to study single repositories (such as figshare) in more detail. It is crucial to further explore the real meaning and rationale of research data citations and how they depend on the nature and structure of the underlying research data, e.g., in terms of data curation and awarding of DOIs.

[14] figshare lately announced a deal with Altmetric.com which might increase the visibility of altmetrics with respect to data sharing: http://figshare.com/blog/The_figshare_top_10_of_2014_according_to_altmetric/142

Also, little is known about how data citations complement and differ from data sharing and data usage activities as well as altmetrics.

Acknowledgments

This analysis was done within the scope of e-Infrastructures Austria (http://e-infrastructures.at/). The authors thank Dr. Uwe Wendland (Thomson Reuters) and Stephan Buettgen (EBSCO) for granted trial access to Data Citation Index resp. PlumX. 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

Belter, C.W. (2014). Measuring the value of research data: A citation analysis of oceanographic data sets. *PLoS ONE, 9*(3): e92590. doi:10.1371/journal.pone.0092590

Bornman, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics, Retrieved January 1, 2015 from http://arxiv.org/abs/1406.7091

Borgman, C.L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, *63*, 1059-1078.

Costas, R., Meijer, I., Zahedi, Z. & Wouters, P. (2012). The value of research data – Metrics for data sets from a cultural and technical point of view. A Knowledge Exchange Report. Retrieved January 1, 2015 from http://www.knowledge-exchange.info/datametrics.

Jobmann, A., Hoffmann, C.P., Künne, S., Peters, I., Schmitz, J. & Wollnik-Korn, G. (2014). Altmetrics for large, multidisciplinary research groups: Comparison of current tools. *Bibliometrie - Praxis und Forschung*, *3*, Retrieved January 1, 2015 from http://www.bibliometrie-pf.de/article/viewFile/205/258.

Konkiel, S. (2013). Altmetrics . A 21st-century solution to determining research quality. *Information Today*, *37*(4), Retrieved January 1, 2015 from http://www.knowledge-exchange.info/datametrics http://www.infotoday.com/OnlineSearcher/Articles/Features/Altmetrics-A-stCentury-Solution-to-Determining-Research-Quality-90551.shtml.

Piwowar, H.A. & Chapman, W.W. (2010). Public sharing of research datasets: A pilot study of associations. *Journal of Informetrics*, *4*, 148-156.

Torres-Salinas, D., Robinson-Garcia, N. & Cabezas-Clavijo, Á. (2013). Compartir los datos de investigación: Una introducción al 'Data Sharing'. *El profesional de la información*, *21*, 173-184.

Torres-Salinas, D., Martín-Martín, A. & Fuente-Gutiérrez, E. (2013). An introduction to the coverage of the Data Citation Index (Thomson-Reuters): Disciplines, document types and repositories. EC3 Working Papers, 11, June 2013. Retrieved January 1, 2015 from http://arxiv.org/ftp/arxiv/papers/1306/1306.6584.pdf.

Torres-Salinas, D., Jimenez-Contreras, E. & Robinson-Garcia, N. (2014). How many citations are there in the Data Citation Index? *Proceedings of the STI Conference*, Leiden, The Netherlands. Retrieved January 1, 2015 from http://arxiv.org/abs/1409.0753.

Stopped Sum Models for Citation Data

Wan Jing Low, Paul Wilson and Mike Thelwall

W.J.Low@wlv.ac.uk, PaulJWilson@wlv.ac.uk, M.Thelwall@wlv.ac.uk
Statistical Cybermetrics Research Group, School of Mathematics and Computer Science, University of
Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY (UK)

Abstract
It is important to identify the most appropriate statistical model for citation data in order to maximise the power of future analyses as well as to shed light on the processes that drive citations. This article assesses stopped sum models and compares them with two previously used models, the discretised lognormal and negative binomial distributions using the Akaike Information Criterion (AIC). Based upon data from 20 Scopus categories, some of the stopped sum models had lower AIC values than the discretised lognormal models, which were otherwise the best. However, very large standard errors were produced for some of these stopped sum models, indicating the imprecision of the estimates and the impracticality of the approach. Hence, although stopped sum models show some promise for citation analysis, they are only recommended when they fit better than the alternatives and have manageable standard errors. Nevertheless, their good fit to citation data gives evidence that two different, but related, processes drive citations.

Conference Topic
Citation and co-citation analysis

Introduction

Fitting statistical models to citation data is useful both to understand the citation process itself (de Solla Price, 1976) and to identify the factors that affect the citedness of academic papers (Bornmann, Schier, Marx, & Daniel, 2012; Didegah & Thelwall, 2013). For example, negative binomial regression previously has been used to analyse factors underlying patent citations (Maurseth & Verspagen, 2002). The choice of statistical model is not straightforward (Bookstein, 2001), however, because citation data is typically highly skewed (de Solla Price, 1976) with a heavy tail (i.e., with particularly many articles having high citation counts) which makes it difficult to identify and fit the best distribution (Clauset, Shalizi, & Newman, 2009). Nevertheless, it has recently been shown that the distribution of citations to articles from an individual Scopus category and year follows a hooked power law or a discretised lognormal distribution substantially better than a power law (Thelwall & Wilson, 2014a) and that, on this basis, (discretised) ordinary least squares regression on the log of the citation data, after adding 1 to cope with the problem of uncited articles, is applicable and is probably the best available regression method (Thelwall & Wilson, 2014b). It should be noted that although the data is well fitted by the discretised lognormal distribution, it should not be assumed that it was derived from that distribution, as models should not be regarded as literal descriptions of nature (Hesse, 1953). Moreover, it is useful to assess additional statistical models in case a more powerful model can be found as well as to shed light on the processes underlying citation, which are still far from fully understood. This paper investigates stopped sum models for citation data for the first time. These have very different underlying assumptions to the lognormal distribution but can result in similar shaped distributions.

Hence, should citation data fit them well, the results would have both practical and theoretical implications for citation analysis.

Stopped sum distributions

Stopped sum distributions were initially developed by Neyman to model the number of larvae in a field (Neyman, 1939). Neyman viewed the distribution of larvae as resulting from two population waves. The first 'parent' (or primary wave) distribution was followed by a distribution of 'offspring' (or secondary wave), whereby the numbers in the secondary wave would be dependent on the numbers in the primary wave; the overall population being the sum of the populations from the two waves (Johnson, Kemp, & Kotz, 2005, pp. 381–382). The two waves can have completely different statistical distributions. If, for example, the primary wave distribution is Poisson and the secondary wave distribution is negative binomial, the overall distribution is known as a *Poisson stopped sum negative binomial (NB) distribution*. Here stopped sum models are explored due to their potential to model citation data as two waves, the primary wave and secondary wave. Given that the overall number of citations that an article receives might come from a similar two waves process, the primary wave representing citations received shortly after a journal article has been published, and the secondary wave, perhaps overlapping with the first to some extent, representing the citations received as a result of scientists discovering an article because of its previous citations, either directly by following citations or indirectly because more cited articles are ranked more highly in some citation databases.

The stopped sum models for citation counts could also be appropriate if the two waves occurred simultaneously instead of sequentially. For example, for the Poisson stopped sum negative binomial model, one of the wave distributions follows the Poisson distribution and the other wave follows the negative binomial distribution at the same time.

The original model proposed by Neyman (1939) assumed that zero counts in the primary wave will automatically be followed by zero counts in the second wave. Hence, if X follows the Poisson stopped sum NB distribution, P(X=0) is just P(X=0) under the Poisson distribution.

For citation counts of one or more, the stopped sum assumes that this can only be a result of a non-zero citation in the primary wave. For example, a citation count of 3 can only arise as a result of one of the three combinations:

- 3 citations in the primary wave, 0 citation in the secondary wave; or
- 2 citations in the primary wave, 1 citation in the secondary wave; or
- 1 citation in the primary wave, 2 citations in the secondary wave.

The Poisson stopped sum NB distribution will therefore have the following probability mass function (p.m.f.):

$$P(X=y)=\begin{cases} e^{-\lambda} & if\ y=0 \\ \sum_{j=1}^{y}\frac{e^{-\lambda}\lambda^{j}}{j!}*\binom{y-j+\alpha-1}{\alpha-1}p^{\alpha}(1-p)^{y-j} & if\ y\geq 1, \text{and } p=\frac{\alpha}{\mu+\alpha} \end{cases}$$

The other stopped sum distributions that are considered include the NB stopped sum Poisson distribution:

$$P(X=y)=\begin{cases} p^{\alpha} & if\ y=0 \\ \sum_{j=1}^{y}\binom{y+\alpha-1}{\alpha-1}p^{\alpha}(1-p)^{y}*\frac{e^{-\lambda}\lambda^{y-j}}{(y-j)!} & if\ y\geq 1 \end{cases}$$

and the NB stopped sum NB distribution:

$$P(X=y)=\begin{cases} p^{\alpha} & if\ y=0 \\ \sum_{j=1}^{y}\binom{y+\alpha-1}{\alpha-1}p^{\alpha}(1-p)^{y}*\binom{y-j+\theta-1}{\theta-1}q^{\theta}(1-q)^{y-j} & if\ y\geq 1 \end{cases}$$

where $p=\frac{\alpha}{\mu+\alpha}$ in all cases.
The Poisson stopped sum Poisson distribution was considered but because very large AICs were obtained indicating a poor fit for citation data we do not discuss it further here.

Modified stopped sum distributions

In the study made by Neyman in 1939, the restriction of having zero counts in the primary wave resulting in zero counts in the secondary wave was necessary, but in the case of citation analysis, it is feasible that a zero citation count in the first population wave could be followed by a non-zero count in the second. This can occur due to the limitations of the citation database used to analyse the citations. For example, an article may be uncited in Scopus, but cited in Google Scholar, and its Google Scholar citations could attract new second wave citations. Hence a modified stopped sum is also considered, where, for example, 3 citations could arise from 0 citations in the primary wave and 3 citations in the secondary wave. The modified Poisson stopped sum NB distribution for this case has p.m.f.:

$$P(X=y)=\sum_{j=0}^{y}\frac{e^{-\lambda}\lambda^{j}}{j!}*\binom{y-j+\alpha-1}{\alpha-1}p^{\alpha}(1-p)^{y-j}\qquad where\ y\geq 0 \text{ and } p=\frac{\alpha}{\mu+\alpha}$$

Using similar adjustments, the modified NB stopped sum Poisson distribution has p.m.f.:

$$P(X=y)=\sum_{j=0}^{y}\binom{y+\alpha-1}{\alpha-1}p^{\alpha}(1-p)^{y}*\frac{e^{-\lambda}\lambda^{y-j}}{(y-j)!}\qquad where\ y\geq 0 \text{ and } p=\frac{\alpha}{\mu+\alpha}$$

Whilst the modified NB stopped sum NB distribution has p.m.f.:

$$P(X=y)=\sum_{j=1}^{y}\binom{y+\alpha-1}{\alpha-1}p^{\alpha}(1-p)^{y}*\binom{y-j+\theta-1}{\theta-1}q^{\theta}(1-q)^{y-j}$$

$$where\ y\geq 0 \text{ and } p=\frac{\alpha}{\mu+\alpha}$$

Note that the modified Poisson stopped sum Poisson distribution is equivalent to a Poisson distribution, and hence is not considered here.

Research Questions

1. Do stopped sum models fit citation count data better than discretised lognormal and negative binomial models?
2. If so, which stopped sum model produces the most consistent results?

Methods

Data from 20 different subject areas were selected from Scopus in order to assess the models for a wide range of different disciplines. This is important because citation patterns are known to vary considerably between disciplines. This data has previously been analysed in Thelwall and Wilson (2014). Each subject area is a single Scopus category and consists of all documents of type article that were published in 2004, giving ten years for the articles to attract citations.

Fitting statistical models

The models were fitted using the R software (R Core Team, 2014). The MASS package (Venables & Ripley, 2002) was used to fit the negative binomial distribution. As there are no known statistical packages readily available to model the proposed stopped sum distributions, the parameters of the distributions were estimated by maximum likelihood estimations methods. AIC is a commonly used statistic for model selection, the model with the lowest AIC usually being regarded as the model that best fits the data (Bozdogan, 2000).

$$AIC = -2 \times \log(L) + (2 \times p)$$

Hence the AIC may be regarded as a penalised version of the loglikelihood, where L is the likelihood of the model and p is the number of parameters estimated. For example, both the Poisson stopped sum NB and NB stopped sum Poisson will have p=3, as there is one parameter (λ) in the Poisson wave and two parameters (NB mean, μ and size, α) in the NB wave. The NB stopped sum NB model will have p=4 as two parameters (μ and α) are estimated in each of the NB waves. Whilst opinions differ, when selecting the 'best' model, it has been suggested that a difference of 6 between the AICs will be large enough to imply a significant difference between the models (Burnham & Anderson, 2003).

Standard errors

Standard errors were computed to reflect the precision with which the proposed statistical models estimate the relevant parameters (Dodge, 2003, p. 386). For the negative binomial models, standard errors were obtained directly from the model fitting software. For the discretised lognormal, the standard errors were obtained by bootstrapping.

For other models the standard errors were calculated using the Hessian matrix, which is the matrix of the second derivatives of the log-likelihood function. The Hessian matrix can also be obtained whilst estimating the parameters for the corresponding distributions using the optim function in R (R Core Team, 2014). Suppose that L represents the log-likelihood function of a stopped sum distribution with two parameters, say λ and μ, then the Hessian matrix is given by $\begin{pmatrix} \frac{\partial^2 L}{\partial \lambda^2} & \frac{\partial^2 L}{\partial \mu \partial \lambda} \\ \frac{\partial^2 L}{\partial \mu \partial \lambda} & \frac{\partial^2 L}{\partial \mu^2} \end{pmatrix}$, and the standard errors for λ and μ are calculated as the square root of the main diagonal of the inverse of the negative Hessian matrix (Ruppert, 2011, pp. 166–167). At 95% confidence interval can be computed by parameter estimate ± 1.96*standard error.

Results

The modified negative binomial stopped sum negative binomial distribution (NBNB) produced the lowest AIC for 13 out of 20 subjects. The next most successful models are the NB stopped sum NB and the discretised lognormal. The Poisson stopped sum NB and the modified NB stopped sum Poisson each fitted 'best' for only one subject (see Table 3 in Appendix).

Parameter estimates for stopped sum distributions

The estimated parameters for Tourism and Soil will be discussed for the proposed stopped sum distributions. These subjects were selected as they are examples of subjects, which return parameter estimates and errors for all the fitted distributions. From Table 1, when Tourism is fitted with the Poisson stopped sum NB model, one wave follows the Poisson distribution with mean, λ=3.22, whilst the other wave follows a negative binomial distribution with mean, μ=18.77 and size, α=0.57; thus the negative binomial wave has a variance of 640.19, since the negative binomial variance equals $\frac{\mu^2}{\alpha} + \mu$. However, when fitted with the NB stopped sum Poisson model, one wave follows a negative binomial distribution with mean, μ=21.53, size, α=0.98, and variance=495.77, whilst the other wave follows a Poisson distribution with mean, λ=0.01. The estimated means (μ) in both negative binomial waves are relatively larger than the estimated means (λ) in the Poisson waves, suggesting that the majority of citation counts for Tourism derive from the negative binomial wave. This supports the interpretation that the two waves occur simultaneously, instead of sequentially, as mentioned above. It is also interesting to note that the sum of the estimated means from the Poisson waves and negative binomial waves of these stopped sum models are approximately equal to the estimated mean when Tourism is fitted solely with the negative binomial model.

When fitted with the NB stopped sum NB model, the estimated mean for Tourism in the primary NB wave (13.48) is larger than that of the secondary NB wave (8.25), suggesting that the majority of citation counts for Tourism derive from the primary wave. Furthermore, the sum of the estimated means from the NB stopped sum NB model for Tourism is also approximately equal to the estimated mean when Tourism is fitted with the negative binomial model only.

Similar results were obtained for Soil. When citation counts for Soil are fitted with the Poisson stopped sum NB model and NB stopped sum Poisson model, the mean estimates in the NB waves are much larger than those of the Poisson waves, suggesting that the majority of citation counts from Soil derive from the NB wave. Moreover, the sum of the estimated means for the stopped sum models is approximately equal to the estimated mean for the negative binomial model only (which is 16.93).

Table 1. Estimated parameters for the NB, Poisson stopped sum NB, NB stopped sum Poisson and NB stopped sum NB models.

	Negative binomial		*Poisson stopped sum NB*			*NB stopped sum Poisson*			*NB stopped sum NB*			
Sub.	*mu*	*size*	λ_1	*mu2*	*size2*	*mu1*	*size1*	λ_2	*mu1*	*size1*	*mu2*	*size2*
Tour.	21.53	0.98	3.22	18.77	0.57	21.53	0.98	0.01	13.48	1.30	8.25	0.10
Soil	16.93	0.74	2.27	16.09	0.56	16.87	0.74	0.06	13.78	0.82	3.46	0.04

Table 2 compares estimated parameters for the NB distribution against those of the modified stopped sum distributions. For the modified versions, the estimates of the Poisson stopped sum NB are similar to those of the NB stopped sum Poisson distributions. Similarly to the stopped sum distributions, Tourism and Soil depends largely on the wave that derives from the NB distribution, as the λ estimates are relatively lower than the mu estimates. Furthermore, the sum of the two mu estimates for the modified NB stopped sum NB distributions (21.533 and 16.931) are also similar to the estimates from the NB distribution.

Table 2. Estimated parameters for the NB, modified Poisson stopped sum NB, modified NB stopped sum Poisson and modified NB stopped sum NB models.

	Negative binomial		*Modified Poisson stopped sum NB*			*Modified NB stopped sum Poisson*			*Modified NB stopped sum NB*			
Subj.	*mu*	*size*	λ_1	*mu2*	*size2*	*mu1*	*size1*	λ_2	*mu1*	*size1*	*mu2*	*size2*
Tour.	21.53	0.98	1.41	20.12	0.75	20.12	0.75	1.41	14.75	0.35	6.79	1.17
Soil	16.93	0.74	0.11	16.82	0.72	16.81	0.72	0.11	4.92	0.08	12.01	0.75

Standard errors for stopped sum distributions

Figures 1 and 2 show the mean and size estimates for the primary and secondary waves of the modified NB stopped sum NB distributions. Visual, Literature and Rehab were excluded as standard errors could not be obtained as a result of a singular hessian matrix.

Although the modified NB stopped sum NB distribution gave the lowest AIC, the model produced very large standard errors, resulting in large confidence intervals, as shown in Figures 1 and 2, indicating that this modified NB stopped sum NB model is impractical. This result could possibly be due to the nature of citations, which differs from that of the larvae studied by Neyman. With larvae and their offspring it is clear which wave of population a larvae originates from, this is not the case with citations – usually it will be far from clear cut which wave a given citation might belong to, which in turn leads to difficulty estimating the mean number of citations for that wave, and hence the large associated standard errors.

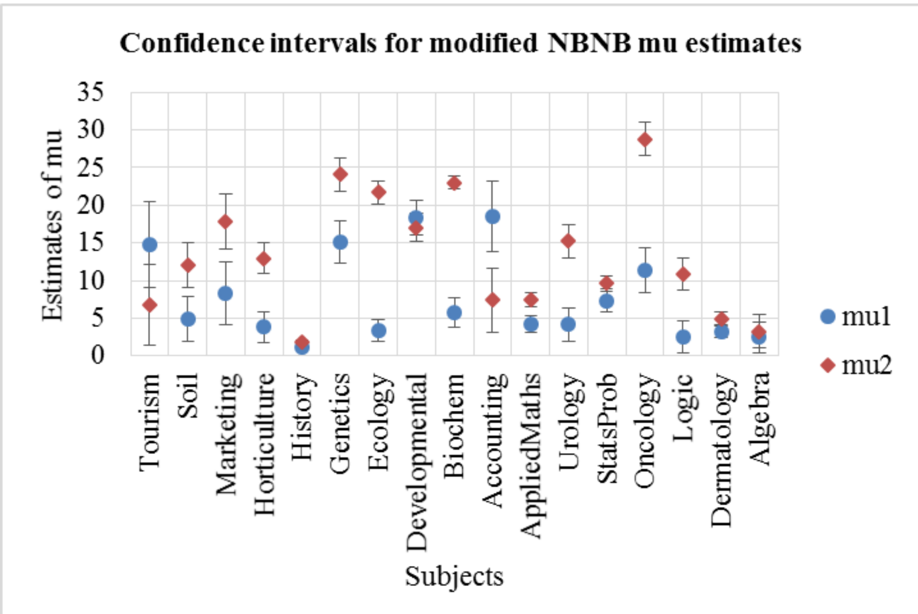


Figure 1. Mean (mu) estimates for the modified NB stopped sum NB distribution for both primary and secondary waves with 95% confidence intervals.

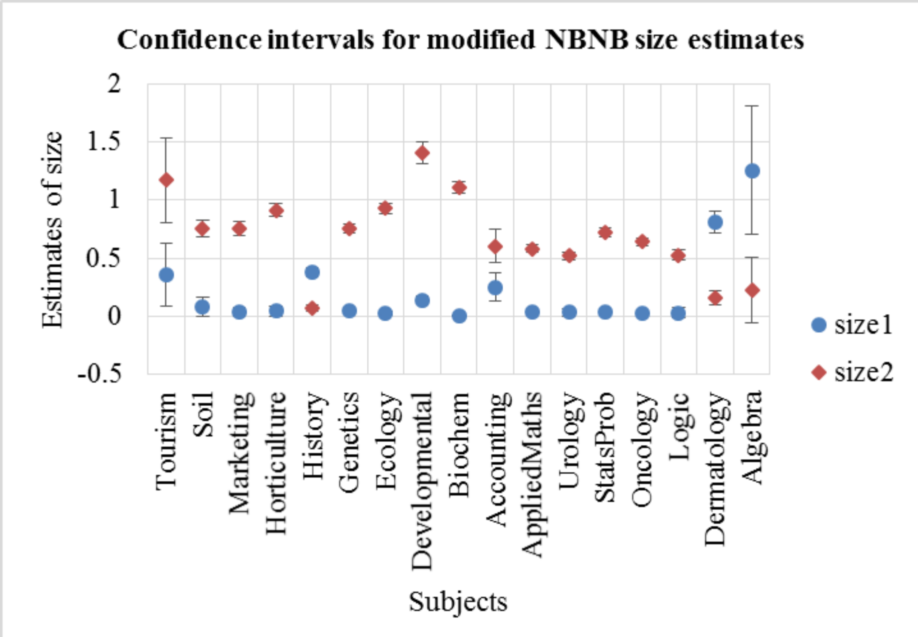


Figure 2. Size estimates for the modified NB stopped sum NB distribution for both primary and secondary waves with 95% confidence intervals.

A further examination of the modified NBNB stopped sum model was carried out with simulations using some known fixed parameters, and similar results were obtained. Moreover, simulations were carried out on all the other stopped sum models and similar results were also obtained for the NBNB stopped sum distribution. Hence it can be concluded that both the stopped sum and modified NBNB stopped sum models are impractical when modelling data with no covariates. Further studies should be conducted to see if adding covariates would change the reliability of the model.

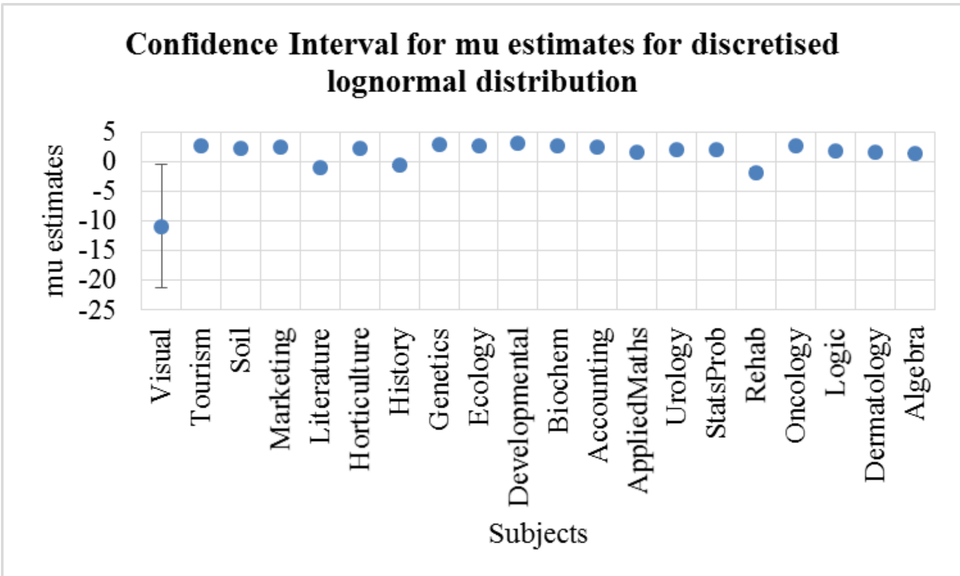


Figure 3. Mu estimates for the discretised lognormal distribution with 95% confidence intervals.

On the other hand, the 95% confidence interval for all subjects except Visual for the discretised lognormal distribution (Fig. 3) are much narrower compared to that of the modified NB stopped sum NB distribution. This indicates that the discretised lognormal distribution is more suitable in practice.

Conclusions

This paper tested stopped sum distributions for modelling citation data for the first time and also introduces a modification to allow the 'waves' to occur simultaneously rather than sequentially. However, given that the standard errors for the stopped sum distribution tend to be very large it is doubtful whether these distributions are useful for citation data even though they produce the lowest AIC. For example, out of all the tested distributions, the modified NB stopped sum NB distribution produced the lowest AIC, but the large standard errors suggests that it is an unsuitable model as its parameter estimates are too unreliable for predictions or conclusions based upon the model to be meaningful.

Overall, the results suggest that for covariate free data, the discretised lognormal distribution is much more suitable for regressing citation data from a single subject and year. Nevertheless, on a theoretical level, the good fits found for some of the stopped sum models give evidence that there are (at least) two important and separate processes that govern the citing practices of authors. For one of these processes, existing citations are irrelevant for new citations, and for the other, they are relevant.

References

Bookstein, A. (2001). Implications of ambiguity for scientometric measurement. *Journal of the American Society for Information Science and Technology*, *52*(1), 74–79. doi:10.1002/1532-2890(2000)52:1<74::AID-ASI1052>3.0.CO;2-C

Bornmann, L., Schier, H., Marx, W., & Daniel, H.-D. (2012). What factors determine citation counts of publications in chemistry besides their quality? *Journal of Informetrics*, *6*(1), 11–18.

Bozdogan, H. (2000). Akaike's Information Criterion and recent developments in information complexity. *Journal of Mathematical Psychology*, *44*(1), 62–91. doi:10.1006/jmps.1999.1277

Burnham, K. P., & Anderson, D. R. (2003). *Model selection and multi-model inference: A practical information-theoretic approach* (2nd ed., p. 520). Springer.

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*(4), 661–703. doi:10.1137/070710111

De Solla Price, D. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, *27*(5), 292–306. doi:10.1002/asi.4630270505

Didegah, F., & Thelwall, M. (2013). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of Informetrics*, *7*(4), 861–873. doi:10.1016/j.joi.2013.08.006

Dodge, Y. (2003). *The Oxford dictionary of statistical terms*. (S. D. Cox, D. Commenges, A. Davison, P. Solomon, & S. Wilson, Eds.) (1st ed., p. 498). Oxford: Oxford University Press.

Hesse, M. B. (1953). Models in Physics. *The British J. for the Philosophy of Science*, *4*(15), 198–214.

Johnson, N. L., Kemp, A. W., & Kotz, S. (2005). *Univariate discrete distribution* (3rd ed., p. 672). Wiley-Interscience.

Maurseth, P. B., & Verspagen, B. (2002). Knowledge spillovers in Europe: A patent citations analysis. *Scandinavian Journal of Economics*, *104*(4), 531–545. doi:10.1111/1467-9442.00300

Neyman, J. (1939). On a new class of "contagious" distributions, applicable in entomology and bacteriology. *The Annals of Mathematical Statistics*, *10*(1), 35–57. doi:10.1214/aoms/1177732245

R Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Ruppert, D. (2011). *Statistics and data analysis for financial engineering* (p. 638). New York: Springer.

Thelwall, M., & Wilson, P. (2014a). Distributions for cited articles from individual subjects and years. *Journal of Informetrics*, *8*(4), 824–839. doi:10.1016/j.joi.2014.08.001

Thelwall, M., & Wilson, P. (2014b). Regression for citation data: An evaluation of different methods. *Journal of Informetrics*, *8*(4), 963–971. doi:10.1016/j.joi.2014.09.011

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (Fourth.). New York: Springer. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4

Appendix

Table 3. AIC for all subjects for each distribution

Subjects	*Discretised lognormal*	*Negative binomial*	*Poisson stopped sum NB*	*NB stopped sum Poisson*	*NB stopped sum NB*	*Modified Poisson stopped sum NB*	*Modified NB stopped sum Poisson*	*Modified NB stopped sum NB*	*Number of articles*
Visual	7902	7928	7916	7930	7865	7920	7920	**7865**	4096
Tourism	4956	4980	4980	4982	4969	4964	4964	**4955**	608
Soil	33470	33344	33458	33345	33287	33344	33344	**33282**	4347
Marketing	**12917**	13073	13025	13073	12941	13015	13015	12932	1550
Literature	11624	11635	**11618**	11637	11622	104485	11624	25449	5000
Horticulture	23058	23093	23165	23095	23001	23067	23067	**22992**	3009
History	19797	19994	19849	19996	19824	19880	19880	**19795**	5000
Genetics	45622	46014	45997	46002	45474	45982	45982	**45471**	5000
Ecology	42787	42343	42441	42335	42253	42366	42793	**42240**	5000
Developmental	40985	41604	41340	41558	40979	41385	41385	**40956**	4541
Biochem	42901	43690	43540	43638	**42675**	43659	43659	42680	5000
Accounting	9927	9933	9924	9931	9914	9929	9929	**9896**	1178
AppliedMaths	33504	33739	33704	33741	33460	33685	33685	**33441**	5000
Urology	38932	38621	38793	38623	**38560**	38623	38623	38563	5000
StatsProb	**36696**	37416	37177	37418	36742	37186	37186	36706	5000
Rehab	28086	27531	27622	27533	27628	27483	**27483**	28322	5000
Oncology	42577	42620	42679	42607	**42196**	42660	42684	42225	4646
Logic	32258	32044	32164	32046	32012	32045	32045	**32010**	4547
Dermatology	19608	19774	19671	19776	19675	19692	19692	**19606**	3184
Algebra	**2968**	2991	2973	2993	2977	2978	2978	2972	528

Table 4. Estimated parameters of negative binomial distribution with the stopped sum distributions

	Negative binomial		*Poisson stopped sum NB*			*NB stopped sum Poisson*			*NB stopped sum NB*			
Subjects	*mu*	*size*	*lambda1*	*mu2*	*size2*	*mu1*	*size1*	*lambda2*	*mu1*	*size1*	*mu2*	*size2*
Visual	0.66	0.17	0.28	1.61	0.34	0.66	0.17	0.00	0.60	0.19	0.26	0.00
Tourism	21.53	0.98	3.22	18.77	0.57	21.53	0.98	0.01	13.48	1.30	8.25	0.10
Soil	16.93	0.74	2.27	16.09	0.56	16.87	0.74	0.06	13.78	0.82	3.46	0.04
Marketing	26.13	0.63	2.63	24.97	0.43	26.02	0.62	0.12	20.34	0.76	6.16	0.01
Literature	0.79	0.32	0.40	1.18	0.33	0.79	0.32	0.00	0.41	9.22	1.16	0.31
Horticulture	16.72	0.83	2.52	15.15	0.54	16.71	0.83	0.01	14.27	0.94	2.62	0.02
History	2.90	0.30	0.75	4.08	0.27	2.90	0.30	0.00	1.26	0.75	3.12	0.12
Genetics	39.23	0.61	2.71	38.78	0.50	38.96	0.60	0.28	24.30	0.80	15.85	0.04
Ecology	25.02	0.86	2.52	24.17	0.79	24.73	0.84	0.31	22.61	0.76	2.60	0.32
Developmental	35.45	0.93	4.03	31.86	0.60	34.56	0.86	0.90	17.95	1.52	17.73	0.12
Biochem	28.81	0.84	3.21	26.60	0.61	28.08	0.79	0.75	22.86	1.12	6.09	0.01
Accounting	25.89	0.64	2.46	25.36	0.50	25.66	0.63	0.26	12.93	0.87	14.03	0.12
AppliedMaths	11.71	0.50	1.68	12.20	0.39	11.71	0.50	0.00	8.20	0.63	4.28	0.03
Urology	19.39	0.51	1.80	20.69	0.50	19.47	0.51	0.00	15.49	0.56	4.60	0.03
StatsProb	16.93	0.54	2.12	16.62	0.36	16.93	0.54	0.00	10.50	0.77	7.21	0.03
Rehab	9.29	0.23	0.83	14.56	0.37	9.28	0.23	0.00	0.83	89.55	14.56	0.37
Oncology	40.23	0.55	2.34	41.68	0.53	39.94	0.54	0.33	25.50	0.68	16.33	0.05
Logic	13.40	0.53	1.67	14.21	0.49	13.37	0.53	0.00	11.59	0.56	2.19	0.02
Dermatology	8.07	0.65	1.79	7.44	0.37	8.06	0.65	0.01	1.83	41.25	7.39	0.36
Algebra	5.75	0.90	1.90	4.46	0.37	5.74	0.90	0.01	1.94	42.31	4.41	0.36

Table 5. Estimated parameters of negative binomial distribution with the modified stopped sum distributions

	Negative binomial		*Modified Poisson stopped sum NB*			*Modified NB stopped sum Poisson*			*Modified NB stopped sum NB*			
Subjects	*mu*	*size*	*lambda1*	*mu2*	*size2*	*mu1*	*size1*	*lambda2*	*mu1*	*size1*	*mu2*	*size2*
Visual	0.66	0.17	0.04	0.62	0.14	0.62	0.14	0.04	0.60	0.19	0.06	0.00
Tourism	21.53	0.98	1.41	20.12	0.75	20.12	0.75	1.41	14.75	0.35	6.79	1.17
Soil	16.93	0.74	0.11	16.82	0.72	16.81	0.72	0.11	4.92	0.08	12.01	0.75
Marketing	26.13	0.63	1.02	25.11	0.50	25.11	0.50	1.02	8.35	0.03	17.78	0.76
Literature	0.79	0.32	11.82	11.99	0.00	0.72	0.24	0.07	4.65	2.71	3.85	0.00
Horticulture	16.72	0.83	0.50	16.24	0.73	16.18	0.72	0.53	3.82	0.05	12.90	0.91
History	2.90	0.30	0.20	2.70	0.21	2.70	0.21	0.20	1.08	0.38	1.82	0.07
Genetics	39.23	0.61	0.43	38.81	0.57	38.81	0.57	0.43	15.12	0.04	24.12	0.75
Ecology	25.02	0.86	0.00	23.60	0.91	18.21	0.80	0.00	3.36	0.02	21.67	0.93
Developmental	35.45	0.93	2.56	32.89	0.69	32.89	0.69	2.56	18.40	0.14	17.04	1.41
Biochem	28.81	0.84	0.69	28.12	0.76	28.12	0.76	0.69	5.79	0.01	23.02	1.11
Accounting	25.89	0.64	0.34	25.55	0.60	25.55	0.60	0.34	18.48	0.25	7.40	0.60
AppliedMaths	11.71	0.50	0.28	11.44	0.44	11.44	0.44	0.28	4.26	0.04	7.45	0.58
Urology	19.39	0.51	0.02	19.37	0.51	19.37	0.51	0.02	4.17	0.03	15.21	0.52
StatsProb	16.93	0.54	0.78	16.16	0.41	16.15	0.41	0.78	7.19	0.04	9.74	0.72
Rehab	9.29	0.23	0.09	9.19	0.21	9.19	0.21	0.09	5.71	0.00	25.74	0.20
Oncology	40.23	0.55	0.00	45.66	0.54	34.70	0.57	0.00	11.43	0.02	28.81	0.64
Logic	13.40	0.53	0.04	13.37	0.52	13.37	0.52	0.04	2.52	0.03	10.88	0.53
Dermatology	8.07	0.65	0.60	7.48	0.47	7.48	0.47	0.60	3.22	0.81	4.85	0.16
Algebra	5.75	0.90	0.84	4.91	0.55	4.91	0.55	0.84	2.48	1.25	3.27	0.23

Differences in Received Citations over Time and Across Fields in China

Siluo Yang[1], Junping Qiu[1], Jinda Ding[2] and Houqiang Yu[1]
[1] *58605025@qq.com*
School of Information Management, Wuhan University, Wuhan 430072, China
[2] *djdhyn@126.com*
Department of library, information and archives, Shanghai University, Shanghai 200444, China

Abstract
We analyse and compare the difference in discipline level of the received citations over a period of time and across fields in China by implementing the diachronous methods of bibliometrics. The citations of 896,645 papers from the Chinese Citation Database (1994 to 2013) that comprised four disciplines, namely, Philosophy, Library and Information Science (LIS), Physics, and Mechanical Engineering, are collected. Results indicate the following conclusions. First, the received citations strongly differ across various fields and over time. Second, the average of the received citations after a given year has an identical change. The number initially increases rapidly, and then declines slightly in the recent years. Uncitedness rate decreases in the early stage of the study period, whereas the rate stabilises or increases slightly in the recent years. Third, the average of the received citations peak after seven and nine years in mechanical engineering and philosophy, respectively, whereas both physics and LIS peak after three years. The span from the year of publication to the cited peak is relatively stable in LIS for 20 years. However, the span decreases in the early stage of the study period, and then stabilizes in the recent years for the other three disciplines. Recently, all four disciplines indicate relatively consistent citation trends. These results highlight the recent evolution of Chinese research systems towards relatively steady states.

Conference Topics
Citation and Co-citation Analysis; Country-level Studies

Introduction

Citing is a fundamental academic behavior among scholars. Citing shows the use of previous research, presents the processes of scientific inheritance and communication, and manifests respect for other scientific researchers (Yang et al., 2010). In the 20th century, citing other works became common in writing scholarly or scientific papers (Kaplan, 1965). Analysis of citing behavior is an important field and method in information science. At present, citation analysis is widely used to evaluate scientific works, initiate scholarly communication, analyse academic behavior, and process information retrieval (Hirsch, 2005; Hammarfelt, 2011; Ketzler & Zimmermann, 2013; Ding et al., 2014).

Information scientists have extensively investigated the distributions and changes of citing behavior (Finardi, 2014). According to the general theory of human behavior, we design the framework of citation behavior analysis. Figure 1 shows a four-dimensional model of citing behavior analysis. This model integrates analytical dimensions in terms of level (who), method (how), perspective (when), and content/topic (what and why). The combination of different dimensions can display the citing behavior in multiple functions and aspects. According to the analysis perspective, citing behaviors mainly include synchronic and diachronic distributions that fundamentally designate and refer to completely different characteristics of scientific literature (Nakamoto, 1988). Synchronic analysis is generally more common than other analytical approaches to citing behaviors (Heistermann et al., 2014). Line and Sandison (1974) proposed the diasynchronous analysis, a kind of synchronous analysis, which studies the synchronous distribution of cited documents at different time periods. Larivière et al. (2008) studied the evolution of yearly synchronous scores computed from 1900 to 2004. Their study showed the increase in average and median ages of cited literature, whereas the price index decreases over time. However, Egghe (2010) argued that

"Larivière' results do not have a special informetric reason but that they are just a mathematical consequence of a widely accepted simple literature growth model."

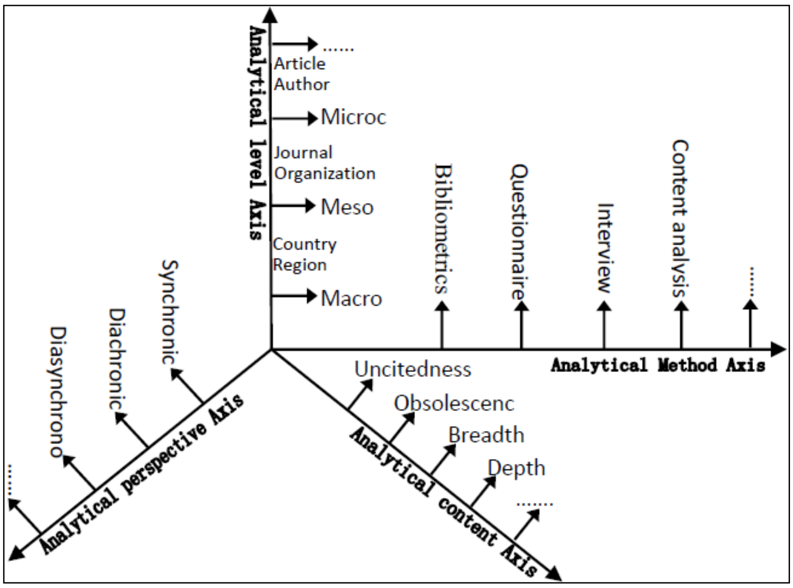


Figure 1. Four-dimensional model of citing behavior analysis.

Diachronic analysis consists of analyzing the distribution of citations gained over time by a publication within a given year by subsequent literature. However, this analysis is generally ignored because of the unavailability of data and the difficulty in implementation. Nevertheless, diachronic analysis has certain advantages, including its appropriateness for citation distribution (Bouabid & Larivière, 2013). Some papers focused on citation distribution and its evolution based on diachronic analysis. First, Finardi (2014) plotted the mean received citations against the time gap (in years) between the publication of the cited article and received citations. Afterwards, he established that citations follow different trends in various fields or disciplines. Some scholars studied the time gap between the publication of a scientific work, as well as the first citation it received (Bornmann & Daniel, 2010). Egghe et al. (2011) proposed a first-citation-speed index, which is utilised for a set of papers, based on the number of publication times and the initial citation. Bouabid and Larivière (2013) recently used a diachronous model to study life expectancy changes and to identify variations in life expectancy between countries and scientific fields based on the citations received by papers.

Second, studies focused on one intriguing aspect of citation analysis, which is the distribution of uncitedness. Schwartz (1997) defined uncitedness as the inability of papers to be cited in citation indexes within five years after their publication. Stern (1990) claimed that although most papers are eventually cited, a number of papers in various scientific disciplines are never cited. Pendlebury (1991) established that the lowest rates of uncitedness occurred among physics and chemistry papers. Garfield (1998) opined that knowing the number of uncited papers and clearly defining these prior to interpretation are important. Egghe et al. (2011) discovered that Nobel laureates and Fields medalists cover a large fraction (10% or more) of uncited publications. A positive correlation was found between the h-index and the number of uncited articles as well.

Lastly, some researchers investigated changes in citing behavior in the context of the overall situation. Larivière et al. (2008) studied the evolution of the aging phenomenon, particularly on how the age of cited literature has changed in over 100 years of scientific activity. They discovered that the average and median ages of cited literature underwent several changes during the period. Evans (2008) showed that as more journal issues are offered online, fewer

journals and articles are cited, and a large part of these citations refer to a small number of journals and articles. Larivière et al. (2009) challenged the conclusion of Evans (2008) and argued that the dispersion of citations is, in fact, increasing. Yang et al. (2010) studied citing behavior by employing three measures of citation concentrations using the Chinese Citation Database (CCD). The concentration of citations was claimed to be declining, and cited papers are broad and diverse. In our view, the diachronic analysis of citation behaviour has two main aspects: the citation change of papers **published** in different years and the citation change of papers **cited** in different years. However, scholars have yet to analyse received citations over a long period of time and across various fields in China.
Since 1978, when the reforms and opening up policies were implemented, China has experienced unprecedented changes. Chinese science exhibited remarkable progress as well. With the popularity of the Internet and development of computer networks in recent years, social environment and scientific research underwent significant changes (Zhou et al., 2009; Yang, 2010). In China, What is the exact general distribution of citation? What are the advancements in citation behaviour in Internet era? Are there differences in citation behaviour across various scientific fields in China?
Our research aims to discover the citation distribution trends over time in different scientific fields in China. Specifically, we focus on the following: (1) the general differences of citation distributions among disciplines, (2) the citation or uncitedness characteristic of papers **published** in different years (For example, papers published in 2000, 2001, 2002... are cited respectively after 5 years, that is, 2004, 2005, 2006...), and (3) the citation characteristic of papers **cited** in different years (For example, a paper published in 2000 is cited in 2000, 2001, 2002...).

Methods and data

Data sample

China has the following citation databases: Chinese Science Citation Database, Chinese Social Sciences Citation Index, Chinese Humanities and Social Science Citation Database, Chinese Science and Technology Paper Citation Database, CQVIP Citation Database, and CCD. In this study, we used CCD as our data resource. CCD collects all references for the China National Knowledge Infrastructure (CNKI) and performs deep data excavation on the citation relationship between studies. Furthermore, CCD provides a citation statistical analysis function based on authors, institutions, publishers, and journals. CCD is one of the products of CNKI (http://www.cnki.net/), and the database covers 6,642 journals while its web version has more than 8200 journals. CCD only contains Chinese journals. Tsinghua University and Tsinghua Tongfang Holding Group first launched CNKI in June 1999. CNKI is the key project of the national informatization construction in China, which established the most comprehensive system of academic knowledge resources (CNKI, 2014). CNKI comprises more than 90% of the knowledge resources in China, which is the broadest in titles and type coverages, as well as the most in-depth in years of coverage in the country. The oldest paper dates back to 1979. This database is updated daily.
We analysed publications and citations from 1994 to 2013, which spans 20 years, to identify publishing and citing patterns at the discipline level. This period was chosen because it is recent and 20 years is sufficiently long in performing the comparisons. All papers from 1994 to 2013 were collected in July 2014. The papers covered four disciplines based on the classification system of CNKI: philosophy, library and information science (LIS), physics, and mechanical engineering. These disciplines, respectively, represent the humanities, social sciences, science, and engineering. The LIS is somewhat peculiar given its evolution towards forms of publication and citation that are closer to the hard sciences. However, we are highly

familiar with this subject because many related research also use LIS as an example. We considered citation types including journals, books, dissertations, meetings, and newspapers. To verify the consistency of the data, we downloaded the data again after a week. We consulted the database provider several times regarding data access issues (i.e., the exact time of database upgrade per day and the range and scope of the citation database). The database is only appropriate for a country, and only reflects the situation in China. Thus, results may differ when international databases are used for comparison.

Methodology

Three aspects of related indicators of received citations across fields and over time are presented. The three aspects involve six equations.

Generally, the papers published in year i were cited in year j. Both i and j are from 1994 to 2013, and $j>=i$. P_j represents the number of papers published in year i. C_j represents the number of citations in year j, which were obtained from the papers published in year i. We analyse the general situation of the papers cited and published every year and analysed them using the following equations.

1) The average number of citations obtained by each paper from the published year to year m (m equals to 2013 in this study), and the average number of citations obtained by each paper in each year.

F1: $\frac{\sum_{j=i}^{m} C_j}{P_i}$ expresses the average number of citations obtained by each paper from the published year to year m.

F2: $\frac{F1}{n}$ expresses the average number of citations obtained by each paper in each year, where n represents the distance between published years i and m, that is, $n = m-i+1$.

2) Percentage of uncited papers within a given time period.

F3: $(1-\frac{P^c}{P_i}) \times 100$, P^c is the number of papers cited at least once within a given time period after publication. The time span of one, two, or all years are set. In the case of three years, all papers published in 2003 are referred to as P_{2003}. We attempted to determine how many of the papers are uncited after three years (between 2003 and 2005). The time period ends in 2005 for the three-year perspective (including the publication year).

3) Time evolution of the average received citations.

We obtain Equation 4 by the methodology described in Finardi (2014).

F4: $MEAN_k = \frac{C_j}{P_i}$ expresses the average number of citations in year j, which were obtained by the papers published in year i. That is, the received citations of each paper in year j after being published for x ($x=j-i+1$) years (including the published year). At a constant value of x, which can be changed or assigned between 0 and 19 in the empirical analysis, we can obtain a series of $\mathrm{MENN_k}$. For example, if we set x equals to 3, then we have $MEAN_1 = \frac{C_{1996}}{P_{1994}}$, $MEAN_2 = \frac{C_{1997}}{P_{1995}}$... $MEAN_{18} = \frac{C_{2013}}{P_{2011}}$, where k is from 1 to N and N is dependent on x that equals to 2013-1993-x+1(x is the time distance between the published and cited years i and j, respectively).

F5: $AMEAN_x = \frac{\sum_{k=1}^{N} MEAN_k}{N}$ expresses the average of means among different occurrences from papers published in several years. By this equation, any possible bias because of the use of citations received in a single year may be avoided. The final result is the plot of $AMEAN_x$ vs. x.

F6: $CAC_x = \frac{\sum_{j=i}^{i+x} C_j}{P_i} \Big/ x$ expresses the cumulating average number of citations that each paper has received during x years, beginning its publication in year i (including the published year). For example, if i equals 2000 and x equals 3, the number of citations received at 2000, 2001, and 2002 from the papers published in 2000 will be summed, and then the cumulating average values of received citations of each paper per year will be calculated.

Result and discussion

Overview

A total of 896,645 papers in philosophy, LIS, physics, and mechanical engineering that were published from 1994 to 2013 were collected. The upper left curve in Figure 2 shows that 41,793,391 papers were published across all fields in CCD for the past 20 years (1994 to 2013). The number of papers steadily increased each year, from 927,684 in 1994 to 3,478,490 in 2013. The curve shows that the growth pattern is an S-shape and has three stages (i.e., slow, rapid, and slow growth). The growth of scientific papers slowed down after 2008. The progress of LIS and philosophy papers remains consistent with those of the other fields. However, a downward trend in physics and a highly irregular trend in mechanical engineering in the recent years are observed. Instead of using typical journals, we selected sample papers in the selected disciplines by an artificial category classification of the database. Numerous papers in China are being published in international journals, especially those in the science and technology field, resulting in changes in the growth rate in Chinese journals.

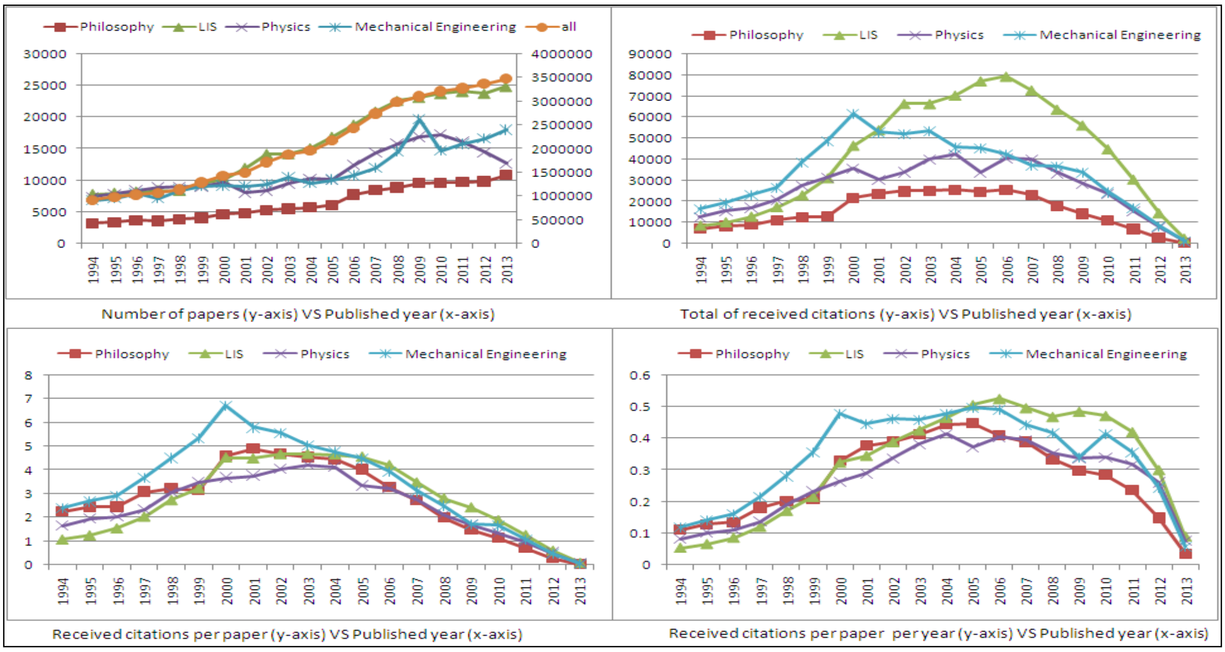


Figure 2. Overall situations of received citations across four disciplines.

Figure 2 shows the overall situations of received citations across four disciplines in CCD. The curves of the received citations exhibit an arch shape (i.e., the middle is high and the end is low). A paper published a long time ago generally has increased chances of receiving citations because of the cumulative phenomenon. However, Figure 2 exhibits the trend of received citations in all four subjects as from increasing in the early periods of the study period to decreasing in the recent years. This phenomenon is caused by two reasons. First, the number of published papers and references for each paper increases each year. The rapid updating of information and the increase in the received citations of each paper can lead to the increase in the number of citations (Price, 1965). Therefore, the cumulative effect of received citations is weakened. Second, people are generally interested in and use the latest research as reference. Researchers strive to make their papers novel. Thus, papers published in previous years have become irrelevant. Figure 2 also exhibits that the received citations of each paper each year (bottom right corner of the figure) eliminate the accumulation phenomenon and display the advantages of papers published in the recent years. The curves of the total received citations and the number of papers published in a specific year are generally consistent. LIS indicated the largest number of papers and received citations in the recent years, whereas philosophy recorded the lowest.

Citation and uncitedness characteristics of papers published in different years

Figures 3 and 4 show the average of the received citations after a paper is published in a given year. In the case of five years window, all papers published in 2000 were taken as the research sample; we determined the average number of times that these papers were cited in 2004. For clarity of presentation, Figure 3 displays only the received citations in four fields after 1, 2, 5, and 10 years. The curves exhibit an identical change (i.e., an initial rapid increase and then a slight decline in the recent years) and indicate that the average of the received citations (published in the recent years) failed to increase. The rapid growth of the average of the received citations in the early stages of the study period changes to a relatively stable development phase because of the slow growth in the number of published papers, the development of the Internet, and the widespread use of open-access and e-print materials. However, whether a special informetric reason or merely a mathematical consequence of a simple literature growth model exists, this phenomenon requires further validation and investigation (Egghe, 2010). The average of the received citations exhibits significant differences among the four disciplines in various time spans. The maximum value was attained by LIS after one, two, and three years compared with the other three disciplines in each publication year. However, this value slowly decreased, and LIS attained the minimum value each year after 10 years. Physics and mechanical engineering show the exact opposite of LIS. That is, after 10 years, the maximum value of the average of the received citations was achieved.

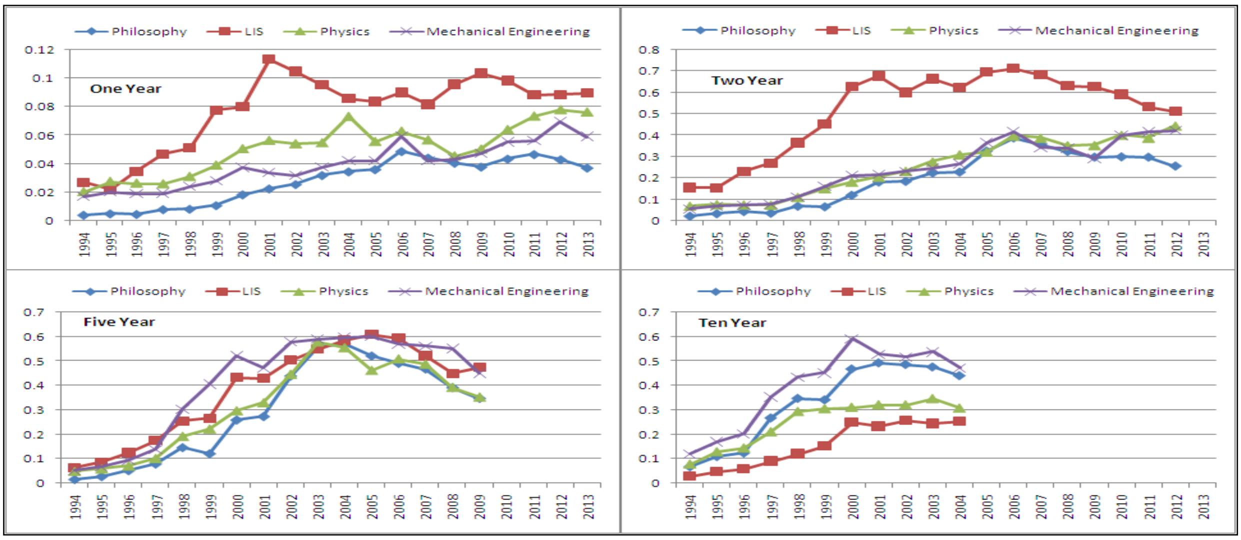


Figure 3. Received citations of each paper each year (y-axis) vs. published year (x-axis) (Part I).

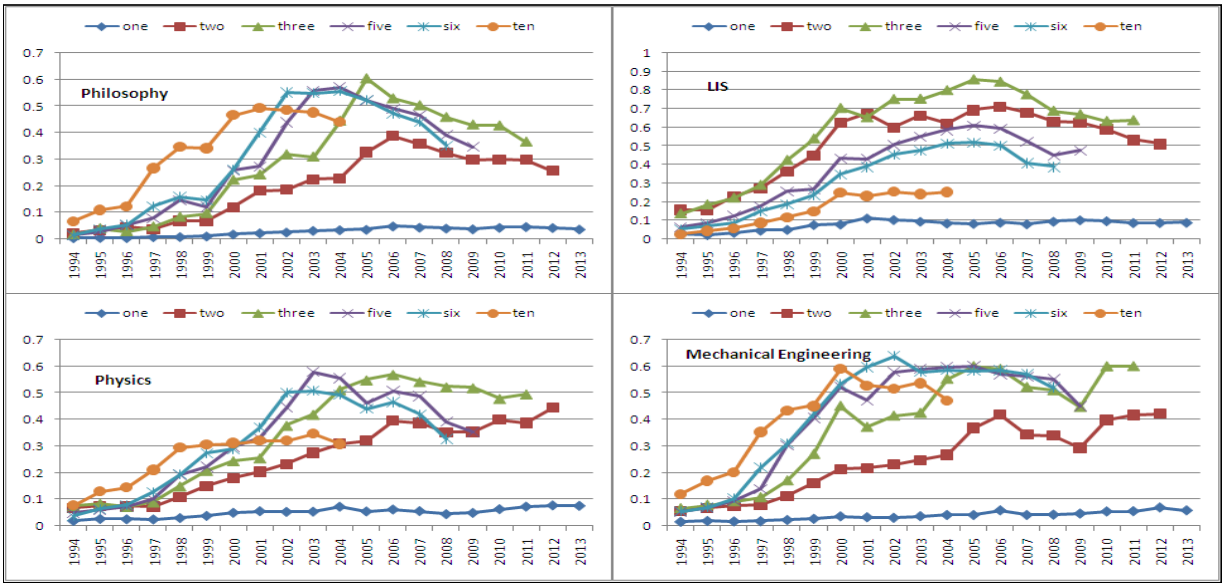


Figure 4. Received citations of each paper each year (y-axis) vs. published year (x-axis) (Part II).

Figure 4 illustrates the received citations by discipline and clarifies the situations of various time spans in each field. Philosophy, physics, and mechanical engineering papers published in the early stages of the study period received more citations in six and ten-year windows than in the recent years. Generally, recently published papers have more citations of papers published from the last three years, which implies that the life expectancy of scientific literature is generally becoming shorter. Papers on LIS (published almost all year) received more citations in two and three-year windows.

The uncitedness results are presented in four citation windows representing one, two, five, and all years after the publication year. Figures 5 and 6 show that the uncitedness rate generally decreases in the early stages of the study period, and then stabilises or increases slightly in the recent years. This phenomenon is due to the following reason. First is the emergence of databases and networks that provided researchers with additional opportunities to find articles for citation and that allowed equal access to all documents. However, the development of databases has entered a period of relative stability in recent years and the uncitedness rate changes slowly as well. Second, the steady increase in the number of published articles and references for each paper decreases the uncitedness rates in the early stages of the study period. However, the rates of both published articles and references relatively stabilised in the recent years. Third, CCD, which is used and promoted in a wide range of areas, was established in 1999. As CCD became increasingly stable, its data updates became timely in recent years. After the reform, the opening up, and the development of science and technology, research conditions and environments significantly improved. The state of scientific research has become steady in recent years in China.

A number of studies showed that the uncitedness rate is lowest in the sciences, high in the social sciences, and highest in arts and humanities (Hamilton, 1991). However, Figures 5 and 6 display contrasting results. The uncitedness rates in LIS are significantly lower than the other three disciplines in the one-, two-, and five-year citation windows in almost all publication years. A possible reason for this phenomenon is the privileges and required expertise in accessing and using documents (especially online information retrieval) in LIS. Papers published in the recent year exhibit high uncitedness rates for Philosophy in the one-, two-, and five-year citation windows. However, the low uncitedness rates in the all-year citation window showed more documents being cited in this discipline.

Figure 6 shows the uncitedness situation by discipline. The curves exhibit the same trend for all four disciplines. The uncitedness rates in the one-year window are relatively stable, while in the two-year window, the uncitedness rates decrease rapidly and decline sharply in the five-year window. However, the all-year window is special because different results were obtained for papers in different publication years. For example, papers published in 1994, 2000, and 2008 are in the all-year citation window, particularly 20, 14, and 6, respectively. Consequently, the two curves of the five- and all-year windows move gradually closer.

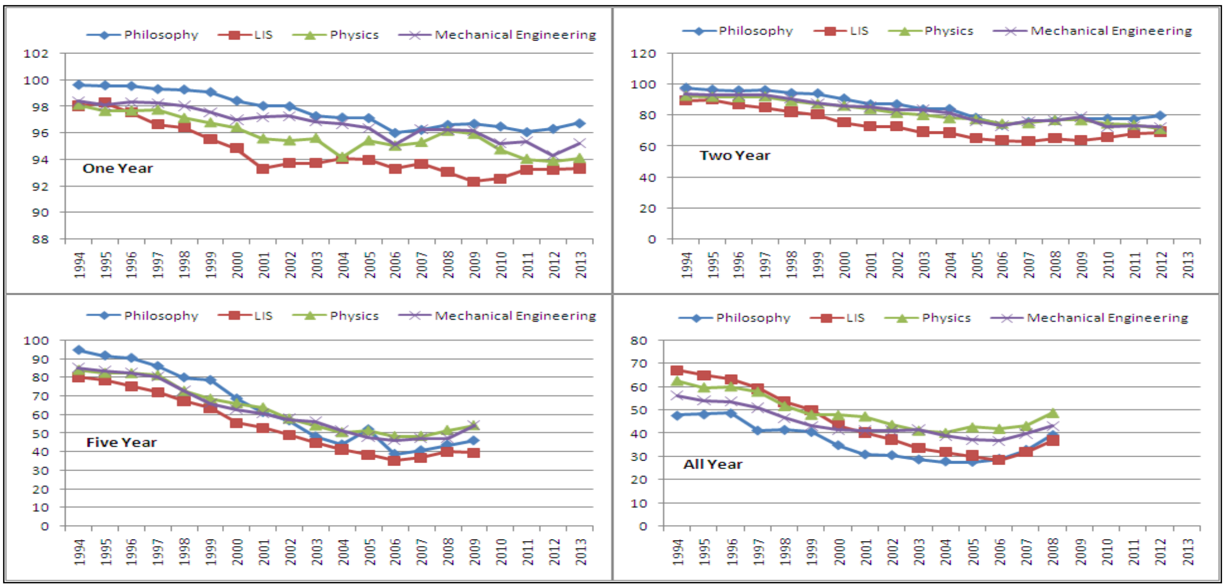


Figure 5. Number of uncited articles (y-axis) vs. published year (x-axis) (Part I).

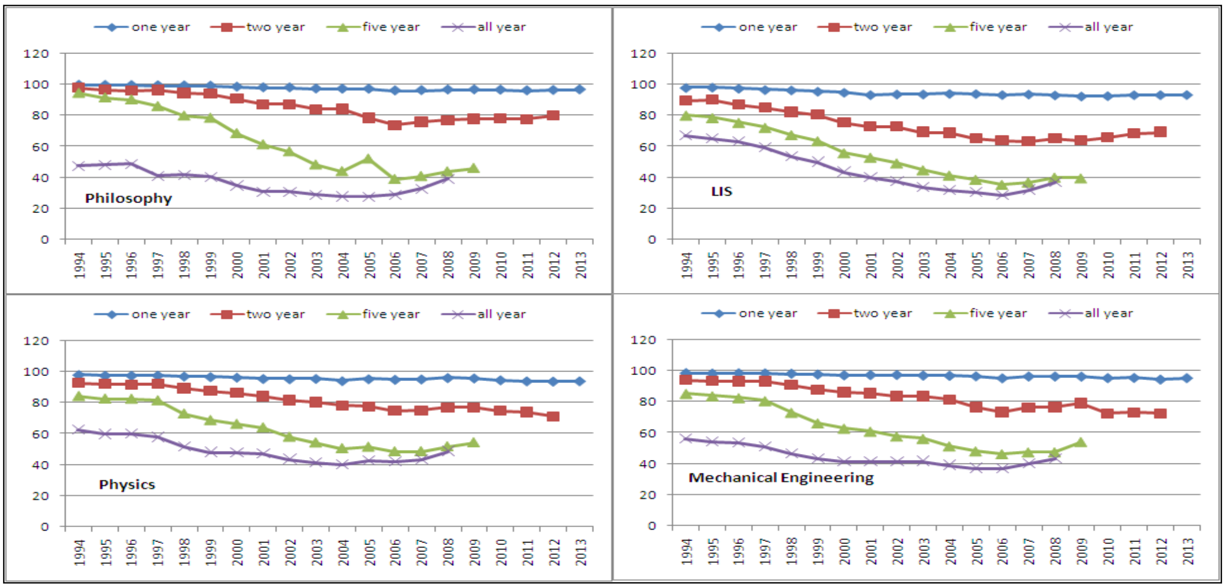


Figure 6. Number of uncited articles (y-axis) vs. published year (x-axis) (Part II).

Citation characteristics of papers cited in different years

Figure 7 shows the mean of the received citations of each paper after a given time period using Equations 5 and 6. The average value avoids possible biases that are caused by using the received citations in a single year. The curve shows the values from 1 to 20 years after publication.

Figure 7 presents the average of the received citations over time. The typical citation curve starts with a rapid increase during the initial years followed by a peak, and then a slow but steady decrease (Larivière et al., 2008). LIS and physics had a similar trend in terms of the average of the received citations. These disciplines peaked at three years after publication, as observed by Finardi (2014) and Bouabid & Larivière (2013). However, physics steadily decreases and LIS rapidly decreases, which created a steep curve. The times of cited peak values are distinct among different disciplines. The trend of mechanical engineering presents a peculiar behaviour because a peak is not exhibited. Instead, the received citations increase in the first three to five years and then stabilise at high values. Citations of mechanical engineering papers continue for a long time after their publication. Figure 7 also suggests that philosophy has a different citation path, with the continuous growth from one to eight years, peak at nine years, and a subsequent slight decrease. This trend is because philosophy information can be accessed and used for a long time, with slow obsolescence.

Figure 7 shows that notable differences exist between the trends of the mean of the received citations in different fields. Consequently, we can conclude that clear differences exist among other specific fields of natural and social sciences. However, further evidence must be obtained by using longer time periods and increasing the number of disciplines compared with that in this study. The maximum values of the average of the received citations peaked after seven years in mechanical engineering and nine years in philosophy. The journal impact factor (IF) only considers citations received in the first two or five years after publication (i.e., 2-years IF or 5-years IF). Thus, high citation values are not captured in the IF computation. The following reasons can explain the particular trends in mechanical engineering and philosophy. Papers published in both disciplines increased from 1994 to 2013, resulting in a parallel growth in the number of citations. Moreover, referring to old literature is preferred in both disciplines, resulting in stable citation curves.

The curves at the right of Figure 7 represent the cumulating value. The curves of the right and left categories in Figure 7 are relatively consistent. However, the curves on the right are smoother than the curves on the left, and the corresponding peaks lag for several years because of the average cumulative effect. For example, in the case of x=3 (x-axis) in Equation 6, we calculated the number of received citations published after one, two, and three years, and then calculated the average values of the received citations of each paper each year.

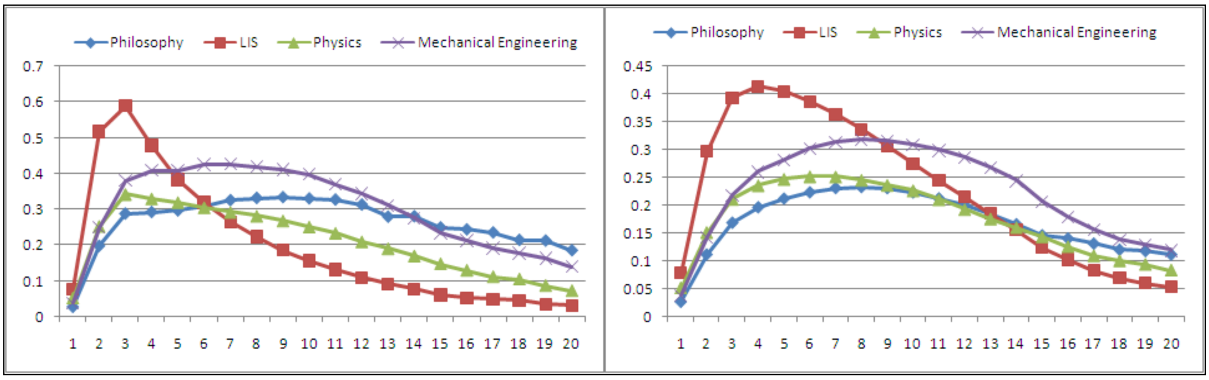


Figure 7. $AMEAN_x$ **(y-axis) vs.** x **(x-axis).**

Figure 8 shows the received citations of each paper each year within the identified time period. We selected the publication years of 1994, 1998, 2002, and 2008 as representatives. The data for the other years of publication showed the same trends. However, these were not included in this paper. The trend in LIS is completely different from those of the other three disciplines. LIS presents a peak at two or three years, which slightly decreases in all cited years. The curves of the other three disciplines are relatively consistent. The received citations of papers published in 1994, 1998, and 2002 increase tremendously and peak in 2006 before slightly decreasing. However, a big difference is observed in the received citations for papers published in different years (i.e., 1994, 1998, and 2002). We can conclude that the early publication years tend to have late citation peaks. For example, the received citations of philosophy papers published in 1994 exhibited their peak 14 years after publication (2007), whereas papers published in 2002 exhibited their peak six years after publication (2007). In general, all four disciplines possess a relatively consistent citation trend in recent years.

Figure 9 shows the situation of the received citations by discipline. Philosophy papers published in the early part of the study period still received many citations. These old papers are not excluded from the science system. Thus, they remain to have a relevant contribution. The citation curves in LIS are consistent in the different cited years. However, the curves of the other three disciplines exhibit a similar trend; papers in these three disciplines became more quickly obsolete in general in recently. Furthermore, many curves peak between 2006 and 2008.

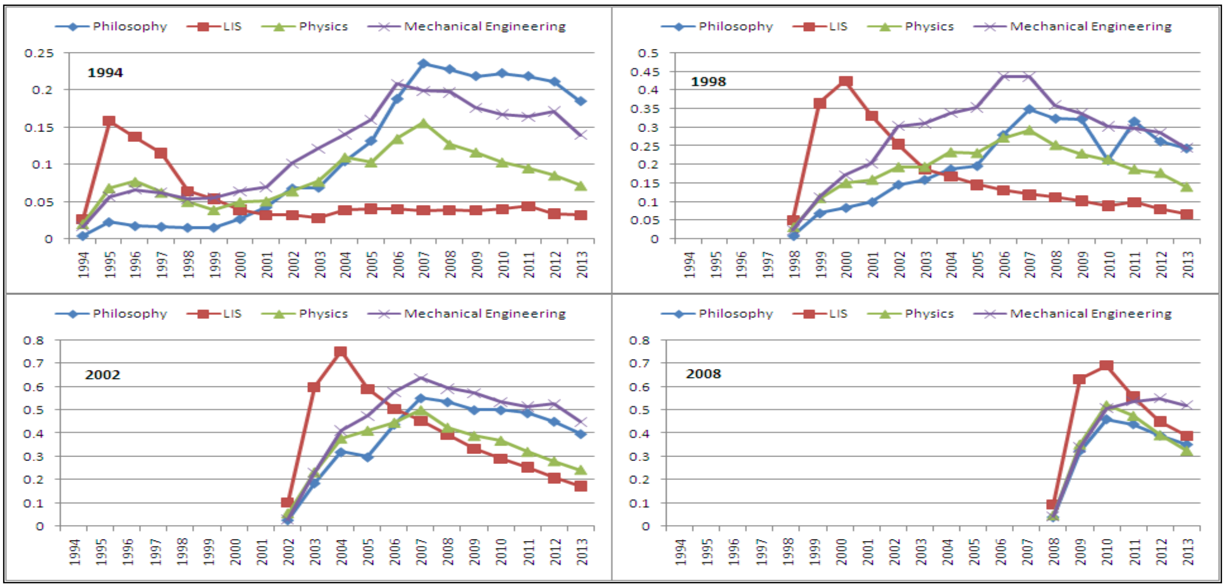


Figure 8. Received citations of each paper each year (y-axis) vs. cited year (x-axis) (Part I).

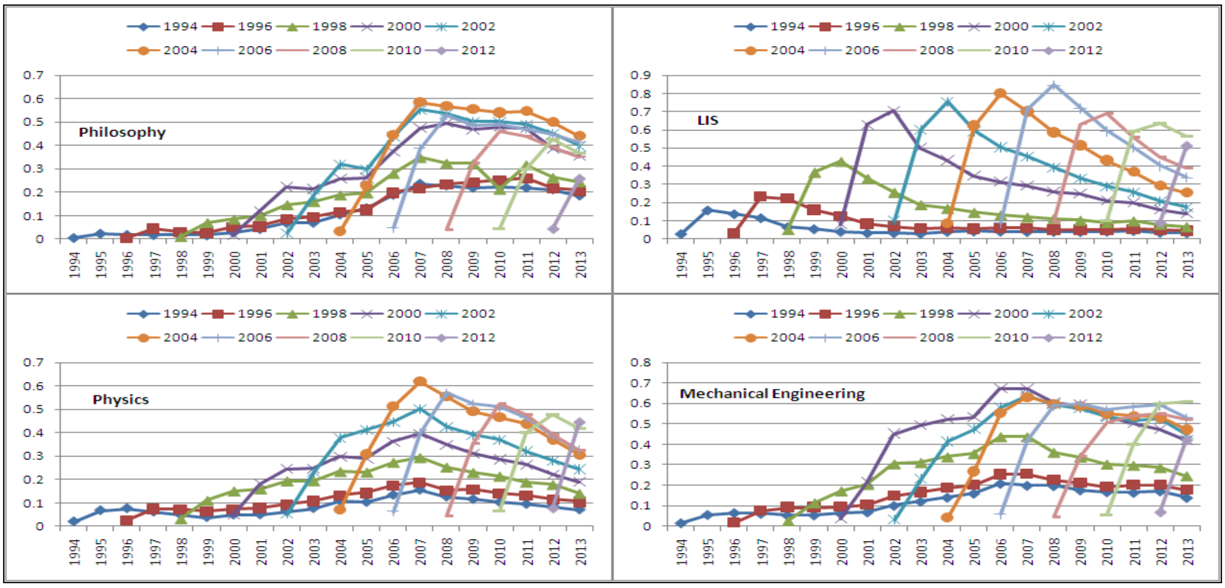


Figure 9. Received citations of each paper each year (y-axis) vs. cited year (x-axis) (Part II).

Conclusion and further research

A total of 896,645 papers on philosophy, LIS, physics, and mechanical engineering, which were published from 1994 to 2013, were collected. This study analysed the differences of these papers in terms of the received citations across fields and over time in China. The following conclusions were derived from the results. First, the growth of published papers is generally S-shaped and undergoes three stages (i.e., slow growth and rapid growth). The curves of the received citations of each paper exhibit an arch shape (i.e., the middle is high and the end is low). The cumulative phenomenon of received citations is not obvious. Second, the average of the received citations in a given year window changes identically, initially increases rapidly, and then slightly decreases in the recent years. The average of the received citations exhibits significant differences among the four disciplines in various time spans. In one-, two-, and three-year windows, a maximum value is observed in LIS in each published year. The value slowly decreases until the LIS obtains a minimum value within the 10-year windows. However, physics and mechanical engineering exhibit an exactly opposite change. Third, the uncitedness rate generally decreases in the early stages of the study period, but stabilises or increases slightly in recent years. The uncitedness rates in the one-year window are relatively stable, but decreases rapidly in the two-year window and drops sharply in the five-year window. Fourth, notable differences exist among the trends of the mean of the received citations of the different fields. The maximum values of the average of the received citations peak after seven years for mechanical engineering, nine years for philosophy, and three years for both physics and LIS. These results are similar to those obtained by Finardi (*2014*) and Bouabid & Larivière (*2013*). Lastly, citation characteristics of papers cited in different years. LIS citations are completely different from those of the other three disciplines. LIS citations peak at two or three years and then slightly decrease in all cited years. The curves of the other three disciplines are similar. Papers published in the early stages of the study period have a later cited peak. In the recent years, all four disciplines possess a relatively constant citation trend. Generally, Chinese research systems evolve into a relatively steady state from a rapid growth and then change in the early period.

This study has analysed comprehensively the received citations across fields and over time in a systematic manner. As a result, consistent conclusions are drawn. For future research, we intend to perform the following. First is we will measure the received citations at the discipline level by implementing diachronous methods. We will consider synchronic methods and combine the two methods. Aside from the discipline level, other levels (e.g., journals, authors, countries, papers, agencies) will also be analysed. We intend to study citations based on literature units and analyse large-scale samples using probability statistics. Second is we will increase the number of disciplines. We will choose additional representative samples from other disciplines for a comprehensive statistical analysis. Furthermore, we will select other document databases such as international document databases, to verify the pattern and characteristic changes in the received citations. Third is we will increase the level of examination and improve the measured indicators of distribution and evolution of the received citations. The measurement methods of the received citations can be enhanced, and an in-depth analysis of the specific distribution of highly cited papers will be conducted. Lastly, a detailed and in-depth study will be implemented to check the factors that affect citation evolution and examine the cause and effect of these changes (e.g., the effect of the growth in number of papers on received citations). Furthermore, we will determine how to handle the trend and changes in the distribution of the received citations.

Acknowledgments

This research is funded by A Foundation for the Author of National Excellent Doctoral Dissertation of PR China.

References

Bornmann, L. & Daniel, H. (2010). The citation speed index: A useful bibliometric indicator to add to the h index. *Journal of Informetrics*, *4*(3), 444-446.

Bouabid, H., & Larivière, V. (2013). The lengthening of papers' life expectancy: a diachronous analysis. *Scientometrics*, *97*(3), 695-717.

CNKI, http://oversea.cnki.net/kns55/support/ en/about_cnki.aspx

Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X. & Zhai, C. (2014). Content-based citation analysis: The next generation of citation analysis. *Journal of the American Society for Information Science & Technology*.

Egghe, L. (2010). A model showing the increase in time of the average and median reference age and the decrease in time of the Price Index. *Scientometrics*, *82*(2), 243-248.

Egghe, L. Bornmann, & L. Guns R. (2011). A proposal for a first-citation-speed-index. *Journal of Informetrics*, *5*(1), 181-186.

Egghe, L. Guns, R. & Rousseau, R. (2011). Thoughts on uncitedness: Nobel laureates and Fields medalists as case studies. *Journal of the American Society for Information Science and Technology*, *62*(8), 1637-1644.

Evans, J.A. (2008). Electronic publication and the narrowing of science and scholarship. *Science*, *321*(5887), 395-399.

Finardi, U. (2014). On the time evolution of received citations, in different scientific fields: An empirical study. *Journal of Informetrics*, *8*(1), 13-24.

Hamilton, D. (1991). *Research Papers: Who's Uncited Now?* Science, 251:25. Available: http://www.garfield.library.upenn.edu/commentaries/tsv12(14)p10y19980706.pdf

Hammarfelt, B. (2011). Citation analysis on the micro level: The example of Walter Benjamin's Illuminations. *Journal of the American Society for Information Science and Technology*, *62*(5), 819-830.

Heistermann, M., Francke, T., Georgi, C., & Bronstert, A. (2014). Increasing life expectancy of water resources literature. *Water Resources Research*. *50*(6), 5019-5028.

Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. *PNAS*, *102*(46), 16569-16572.

Kaplan, N. (1965). The norms of citation behavior: Prolegomena to the footnote. *American Documentation*, *16*(3), 179-184.

Ketzler, R. & Zimmermann, K.F. (2013). A citation-analysis of economic research institutes. *Scientometrics*, *95*(3), 1095-1112.

Larivière, V. Archambault É. & Gingras, Y. (2008). Long-term variations in the aging of scientific literature: from exponential growth to steady-state science (1900-2004). *Journal of the American Society for Information Science and Technology*, *59*, 288-296.

Larivière, V., Gingras, Y. & Archambault, É. (2009). The Decline in the Concentration of Citations, 1900–2007. *Journal of the American Society for Information Science and Technology*. *60*(4), 858-862.

Line, M.B., & Sandison, A. (1974). "Obsolescence" and changes in the use of literature with time. *Journal of Documentation*, *30*, 283-350.

Nakamoto, H. (1988). Synchronous and dyachronous citation distributions. In L. Egghe & R. Rousseau (Eds.), Informetrics 87/88 (pp. 157-163). Amsterdam: Elsevier.

Price, D.J. (1965). Networks of scientific papers. *Science*, *149*, 510-515.

Schwartz, C. A. (1997). The rise and fall of uncitedness. *College & Research Libraries*, *58*(1), 19-29.

Yang, S. Ma, F. Song Y. & Qiu, J. (2010). A longitudinal analysis of citation distribution breadth for Chinese scholars. *Scientometrics*, *85*(3), 755-765.

Zhou, P.,Thijs, B. & Glanzel, W.(2009).Is China also becoming a giant in social sciences? . *Scientometrics*, *73*(3), 593-621.

The Rise in Co-authorship in the Social Sciences (1980-2013)

Dorte Henriksen[1]

[1] *dh@ps.au.dk*
Danish Centre for Studies in Research & Research Policy, Department of Political Science, Business and Social Sciences, Aarhus University, Denmark.

Abstract
This paper examines the rise in co-authorship in the Social Sciences over a 33-year period. We investigate the development in co-authorship in different research areas and discuss how the methodological differences in these research areas and changes in academia affect the tendency to co-author articles. The study is based on bibliographic data about 4.5 million peer review articles published in the period 1980-2013 and indexed in the 56 subject categories of the Web of Science's (WoS) Social Science Citation Index (SSCI). Results show that in the majority of the subject categories we can document a rise in the mean number of authors and that there are disciplinary differences in how much the number of authors has increased. The most substantial rise in the mean and median number of authors has happen in subject categories, where the research often is based on the use of experiments, large data set, statistical methods and/or team-production models.

Conference Topic
Citation and Co-citation Analysis

Introduction
This paper explores the rise in co-authorship in the social sciences. The study is based on all the articles registered from 1980-2013 in the Web of Science's (WoS) Social Science Citation Index (SSCI). Several studies have examined the rise in the number of authors in different research fields. The studies vary in design, but the majority of the bibliometric studies can be categorized as studies either based on bibliographic data from a national database (Lariviere, Gingras, & Archambault, 2006; Ossenblok, Verleysen, & Engels, 2014) or a selection of journals (Cronin, Shaw, & La Barre, 2003; Fisher, Cobane, Vander Ven, & Cullen, 1998; Hudson, 1996; Norris, 1993; White, Dalgleish, & Arnold, 1982). The study by Wuchty, Jones, and Uzzi (2007) is one of the few studies, that examined the increase in research collaboration by using bibliographic data about research articles from multiple fields collected from the subject categories in WoS. However, their study is based on a sample of research articles and not an exhaustively data collection of the research articles indexed in WoS. Furthermore, Wuchty et al. (2007) do not clarify how many articles in their study that are indexed in either Science and Engineering, Social Sciences or Arts and Humanities. This paper is the first study of the rise of co-authorship in the social sciences to use a large sample of time series data based on all of the publications in SSCI, thus the study cover multiple fields of the social science. The study is therefore not bias by national publication tendencies or the selection of journals. The disadvantage of a data set restricted to articles from SSCI is that other publication types and a substantial share of journals are excluded (Hicks, 2005; Ossenblok et al., 2014; Piro, Aksnes, & Rørstad, 2013). However, we believe that the larger data sample compensate for these data limitations. Hence, the objective of this

paper is to document the rise in co-authorship in the social sciences and discuss the factors that could have influenced this evolution.

The increasing focus on authorship can partly be attributed to the growing importance of and attention paid to a researcher's publication record, which is influential in the considerations for employment, promotion, funding and increases in salary (Biagioli, 2012; Costa & Gatz, 1992; Weingart, 2005). Thus, there is a tendency to measure and assess researchers' based on their quantitative research output instead for the content of this output. This creates incentives to "game" the system to improve one's resume by co-producing publications. This is especially the case, when the performance-based research funding systems use whole counts instead of fractionalizing (Butler, 2003; Ossenblok et al., 2014), so the reward for producing a publication does not have to be shared. Hence, the instrumental uses of performance-based funding systems affect the researchers' publishing behavior, including their definitions, perceptions and practices of authorship (e.g. Ossenblok, Engels, & Sivertsen, 2012). However, the rises in co-authorship and research collaboration are also affected by other factors that influence the research community. The rise can be a result of the increasing tendency to perform large scale research projects executed as team-production models. These projects require greater human and financial resources, a larger data collection effort and often more advanced technical and statistical analyses, hence leading to more specialization and division of labor in the research process (Beaver, 2001; Birnholtz, 2006; Cronin et al., 2003; Fisher et al., 1998; Hudson, 1996; Moody, 2004; Rennie, Yank, & Emanuel, 1997; White et al., 1982). These types of projects are often associated with natural and medical sciences, where there is a strong tradition for working in the fore mention team-production model. However, the increasing tendency to work with large scale data set, the rise in using quantitative methods and in some cases experiments have generated a similar team-production model in the social sciences (Cronin et al., 2003; Hudson, 1996; Moody, 2004). Furthermore, studies have found that researchers in the more quantitative research areas of social science is more likely to collaborate (Fisher et al., 1998). Others have pointed at the increasing mobility of researchers that has made it possible and desirable to expand inter-institutional collaborations (Melin, 2000; White et al., 1982) while the development of communication and information technology have enabled geographically disperse researchers to collaborate, by making it easier to communicate, analyze and exchange data (Beaver, 2001; Fisher et al., 1998; Melin, 2000). Furthermore, the growing number of people working in academia has created more collaboration opportunities (Fisher et al., 1998; Lee, 2000; Melin, 2000), especially the increase in PhD students have created more opportunities for research advisors to collaborate and co-author with their students (Fisher et al., 1998; Price, Dake, & Oden, 2000). However, this tendency has given rise to issues regarding honorary or gift authorship in academia and some studies suggest that research advisors may be inappropriately demanding co-authorship with their students (Rennie et al., 1997). This is disputed by Costa and Gatz (1992), who found that students willingly are giving their advisors inappropriate authorship credit even though the advisors do not fulfill the journal guidelines and requirements for co-authorship. However, they do suggest that the willingness to offer co-authorship can be affected by a power imbalance between advisors and advisees, especially because of the increase in PhD students being subsidized by grants held by their advisors. In this paper we will document the evolution of co-authorship and research collaboration by presenting evidence for the increase in the number of authors per publication.

Method

The bibliometric data used in this study were collected from the Centre for Science and Technology Studies (CWTS) enhanced version of Thomson Reuters' WoS database in December 2014. We collected bibliographic information for 4,466,134 articles from 99,752 journal issues published in 1980 to 2013 and registered in WoS' SSCI 56 subject categories. These 56 subject categories have in our analysis been grouped into 6 overall subject categories. The grouping of the categories is based on the topics of each subject category described in the SSCI scope notes (SSCI, 2012). Hence, there are differences in how many categories there has been group together, and the similarity of the research areas. The Social Sciences, Interdisciplinary group consist of a variety of subject categories and do not have the similar thematic relationship as the other groups.

- **Management, Planning & Geography** (Geography, Planning & Development, Urban Studies, Environmental Studies, Management, Transportation)
- **Political Sciences, Business and Law** (Criminology & Penology, Business, Business, Finance, Economics, Public administration, International Relations, Law, Political Science
- **Psychology** (Psychology, Mathematical, Psychology, Psychoanalysis, Psychology, Experimental, Psychology, Social, Psychology, Educational, Psychology, Applied, Psychology, Biological, Psychology, Clinical, Psychology, Developmental, Psychology, Multidisciplinary, Psychiatry
- **Social Health Sciences** (Public Environmental & Occupational Health, Substance Abuse, Gerontology, Health Policy & Services, Rehabilitation, Education, Special, Nursing, Ergonomics)
- **Social Sciences, Interdisciplinary** (Social Sciences, Biomedical, Family Studies, Information Science & Library Science, Social Sciences, Interdisciplinary, Hospitality, Leisure, Sport & Tourism, Industrial Relations & Labor, Social Sciences, Mathematical Methods, Communication, Linguistics, Ethics, History & Philosophy of Science, History of Social Sciences, History)
- **Sociology & Anthropology** (Anthropology, Area Studies, Social Work, Education & Educational Research, Women's Studies, Demography, Social Issues, Sociology, Ethnic Studies, Cultural Studies)

Our study limits the relevant types of publications to journal articles, though we know that the publication pattern in the social sciences is more varied (Lariviere et al., 2006; Ossenblok et al., 2014), thus letters, book chapters and books are an essential part of the scholarly communication in some fields of the social sciences. Unfortunately, the Thomson Reuters Book Citation Index (BCI), part of the WoS core collection, do not have as systematic and exhaustively bibliographic information about books compared to the SSCI's information about journal articles. The BCI do only cover the time period from 2006-present, while SSCI have bibliographic data from 1900 to present, so by choosing to only include journal articles we can set a larger time frame for this study.

Results

In the follow subsections we will present the data showing the increase in number of authors per publication. For each group we will present a figure demonstrating the development in the different subject categories[1]. Our data show that the fields of social sciences have experienced a mean 114 percent increase in the number of authors during the last 33 years, hence there have been added 1,2 authors more to each publication. However, there are large differences in how much the number of authors has risen, with the lowest increase being in the History subject category with a minimal change (0.1 authors) to the highest mean increase in Psychiatry (3 authors).

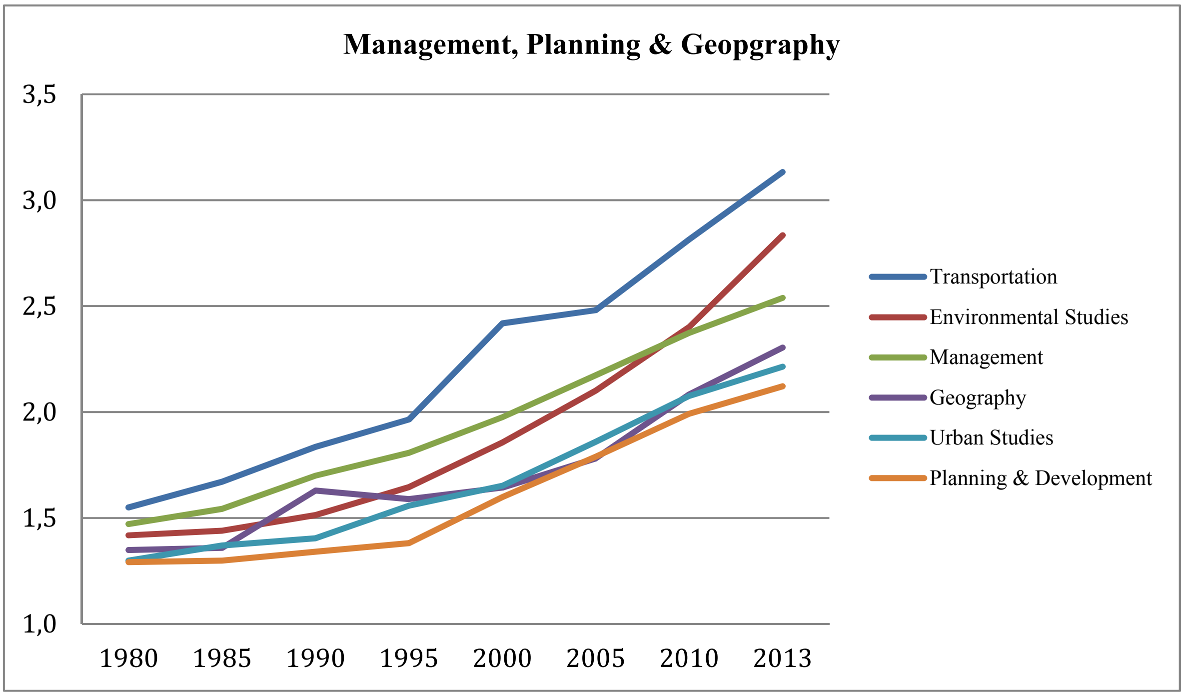


Figure 1. The mean number of authors per publication from 1980-2013 of the group Management, Planning & Geography.

The six categories group as Management, Planning & Geography consist of 373,372 publications. Figure 1 shows the evolution in numbers of authors. The mean numbers of author have increase 71% to 102% or 0.8 – 1.6 authors during the 33 year time period. The mean numbers of authors in 1980 are in the range of 1.3-1.6 authors and have increased in 2013 to 2.1-3.1 authors. The median number of authors is 1 in all categories in 1980. In 2013 the median number of authors has risen to 3 in the category Transportation, while the remaining categories have a median of 2. Even though the category Transportation does not cover civil engineering per se, the close relation with the above mentioned research field can explain some of the increase in co-authorship in this category. The subject categories in this group have all similarities to research fields

[1] We have in this article, because of the space limit, decided to present the development of co-authorship in six figures. The data behind the study will be presented in more details at the conference and are also available if requested.

in science and technology, and are probably influenced by collaboration and publication tendencies dominating these fields.

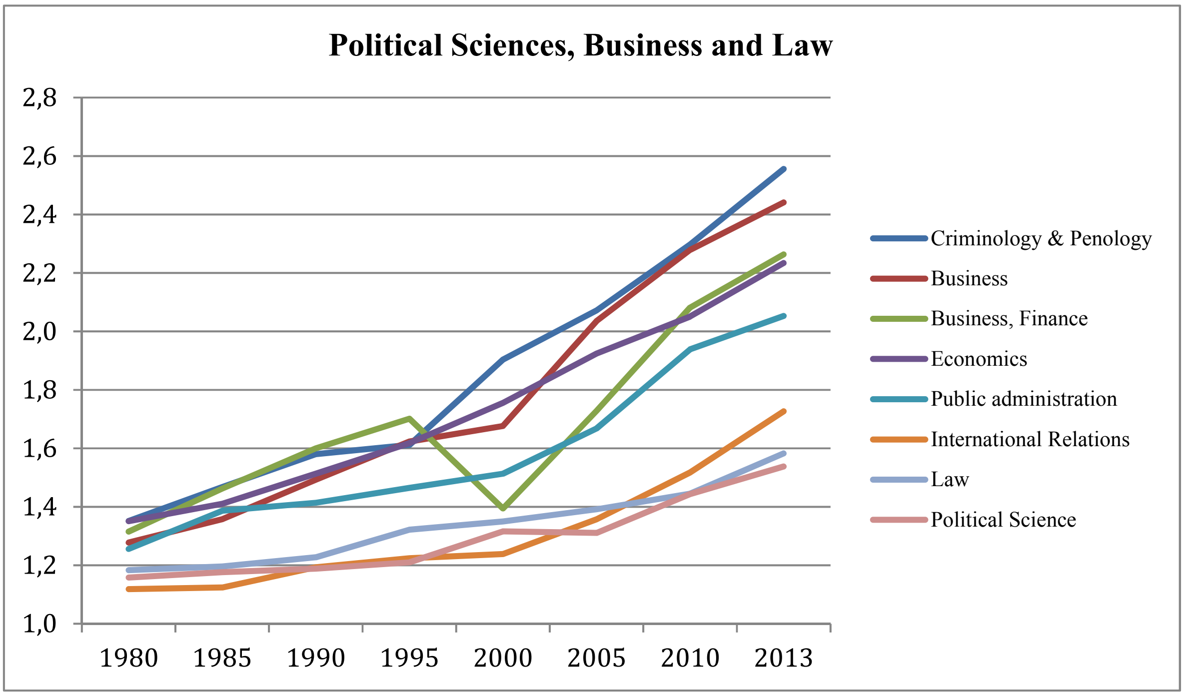


Figure 2. The mean number of authors per publication from 1980-2013 of the group Political Sciences, Business and Law.

The 1,011,725 publications belonging to the Political Sciences, Business and Law show a rise between 38% to 89% in the mean numbers of authors. The mean numbers of authors are between 1.1 - 1.4 in 1980 to 1.5 - 2.6 in 2013 (see figure 2). In Business, Business, Finance, Economics, Criminology & Penology and Public Administration have the median number of authors increase from 1 to 2 authors during the 33 years, while in the remaining categories the median number of authors is 1 during the time period. The greater rise in mean number of authors in the categories Criminology & Penology, Business, Business, Finance, Economics, and Public Administration could be because of the greater use of statistics and register/survey data (Fisher et al., 1998; Hudson, 1996). Political Science is the category in this group with the highest amount of publications (n = 172,625) and covers a broad range of research, thus the lower increase and mean number of authors is probably because areas of Political Science have similarities with research fields in the humanities. The same is the case for the category Law that draws on methods often associated with humanities, such as text analysis.

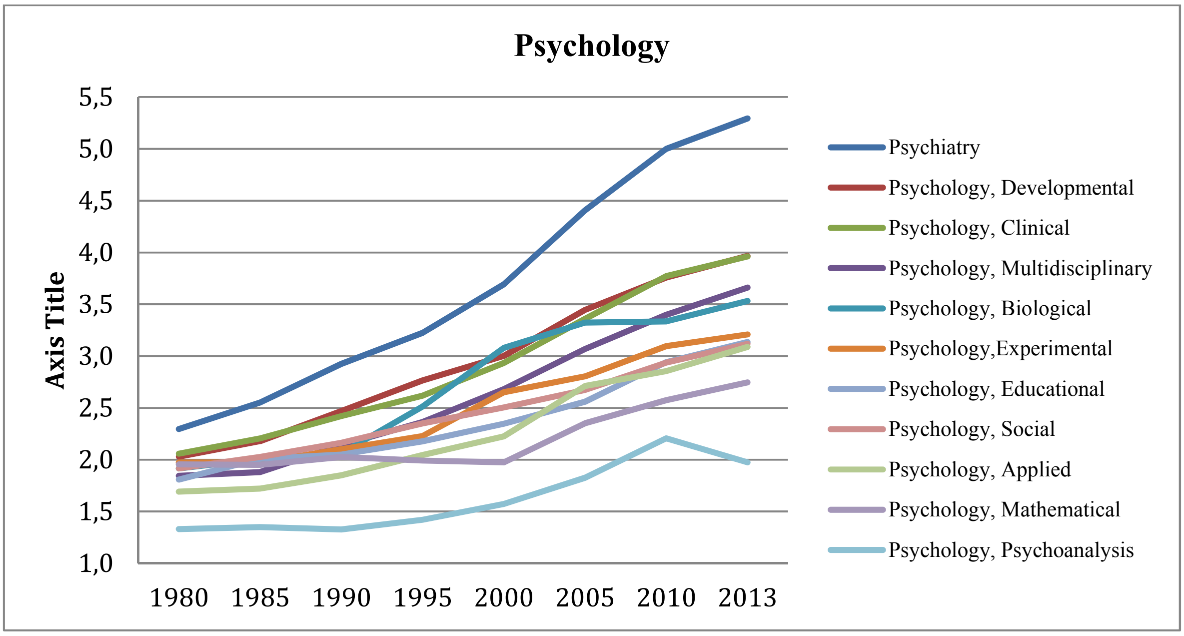


Figure 3. The mean number of authors per publication from 1980-2013 of the group Psychology.

We have collected 1,101,234 publications categorized as Psychology. During the 33 years the mean increase in number of authors is in the range from 0.6 to 3 authors or from 40 % to131%. The mean numbers of authors in 1980 are between 1.4-2.3 authors, this have in 2013 increased to between 2-5.3 authors. The categories of Psychology have all increased the number of authors in the byline during the 33 years, though it is not a constant increase as can been seen in figure 3. The category with the lowest increase is Psychoanalysis, the subject category with a publication and collaboration behavior closest to the humanities, and a mean number of authors in 2013 at 2 authors. Psychoanalysis is the only subject category in the Psychology group where the median have remain constant at 1. In the other end of the scale we have Psychiatry, a subject category with close relations to the medical research fields and therefore a similar collaboration and publication pattern. The mean number of authors in this category is 5.3 authors and the median is 5. Psychology, Mathematical have constantly had a median at 2, while Psychology, Applied have had an increase in the median number of authors from 1 to 3 and Psychology, Clinical have had an increase in median authors from 2 to 4. Psychology, Experimental, Psychology, Social, Psychology, Educational, Psychology, Development, Psychology, Biological and Psychology, Multidisciplinary have had an increase in the median number authors from 2 to 3 authors.

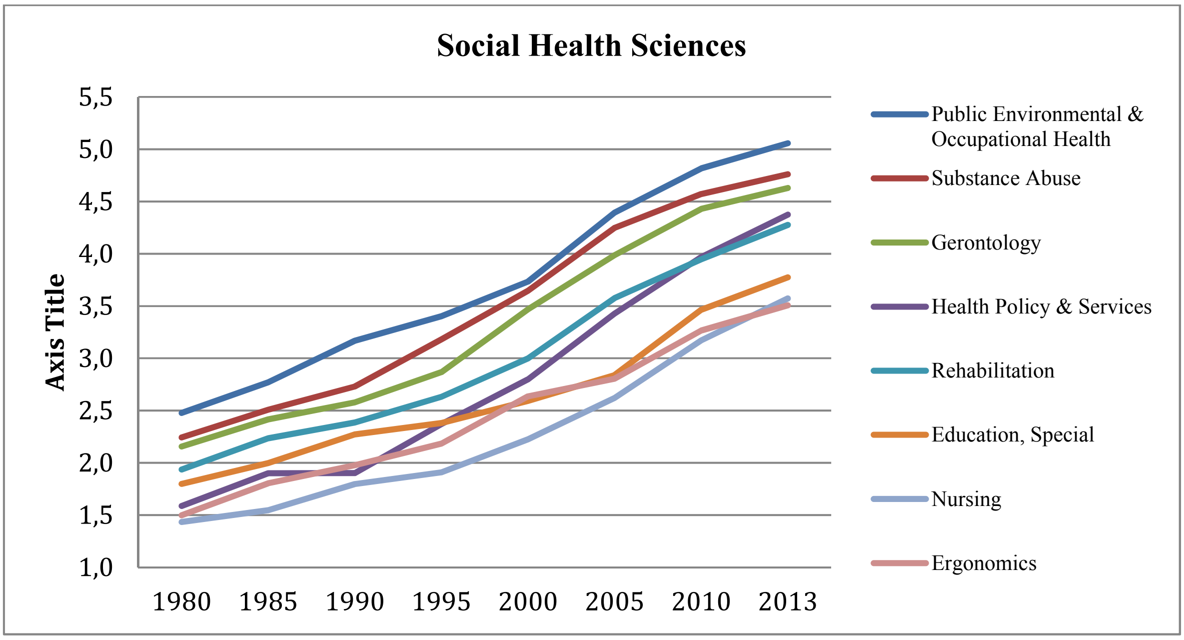


Figure 4. The mean number of authors per publication from 1980-2013 of the group Social Health Sciences.

The data of the categories Social Health Sciences is based on 824,125 publications. The mean number of authors per publication in the Social Health Sciences categories has risen between 104% to 176% or 2-2.6 authors. Figure 4 shows how there have been a substantial increase in all seven subject categories during the 33 years. The median number of authors in 1980 is 1 in the categories Ergonomics, Health Policy & Services and Nursing and 2 in the categories Rehabilitation, Public Environmental & Occupational Health, Substance Abuse, Gerontology and Educational, Special. In 2013 the median numbers of authors have risen to 3 authors in Ergonomics, Nursing and Education, Special and to 4 in the remaining categories. The mean numbers of authors in the Social Health Sciences are between 1.4-2.5 authors in 1980 and have risen to 3.5-5.1 authors in 2013. The average numbers of authors are general quite high in Social Health Sciences compared to other subject categories in the Social Sciences and the subject categories have a publication and collaboration pattern similar to the health and life sciences.

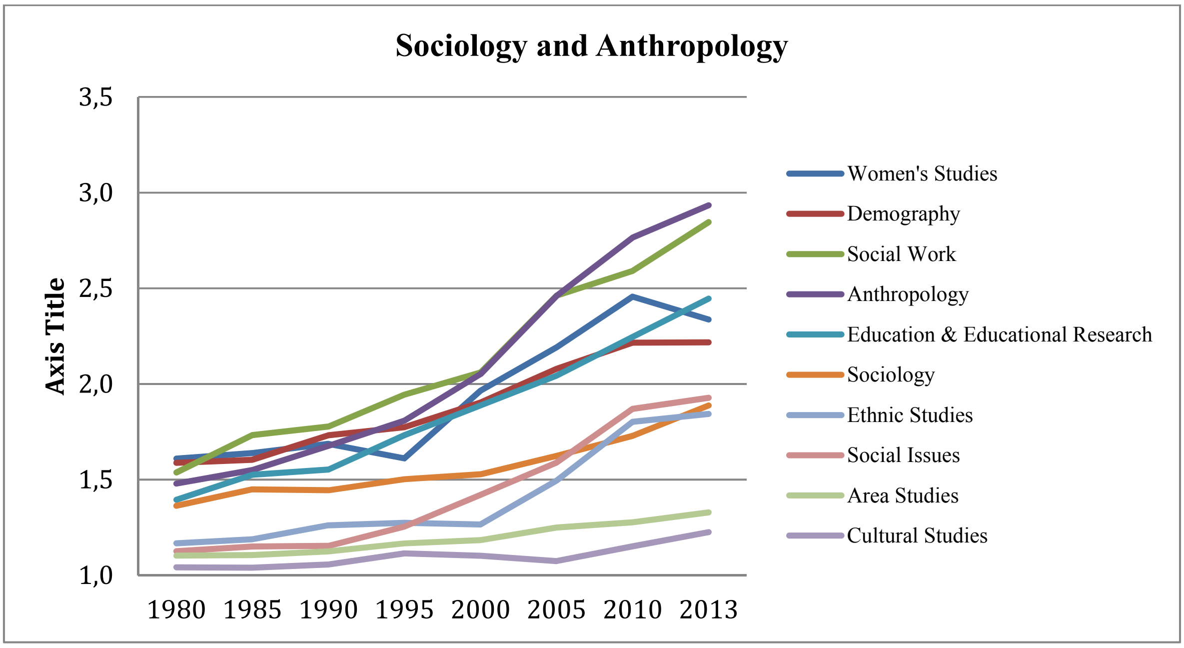


Figure 5. The mean number of authors per publication from 1980-2013 of the group Sociology and Anthropology.

In our data set we have 514,504 publications categorized in the 10 subject categories of Sociology and Anthropology, and the mean percentage increases in numbers of authors are between 17% to 98%. In Figure 5 is the increase in number of authors demonstrated. There have been minimal changes in the mean number of authors in the subject category Cultural Studies and Area Studies, while the categories Social Issues and Ethnic Studies have increased with 0.6-0.8 authors. All of these fore mention categories have a median at 1 in the whole time period. The median has risen to 2 authors for Education & Educational Research, Anthropology, Social Work, Sociology, Women's Studies and Demography. These categories, except Sociology, have a mean number of authors between 1.4-1.6 authors in 1980, which has increased to 2.2-2.9 in 2013. The mean number of authors has only increased with 0.5 for Sociology.

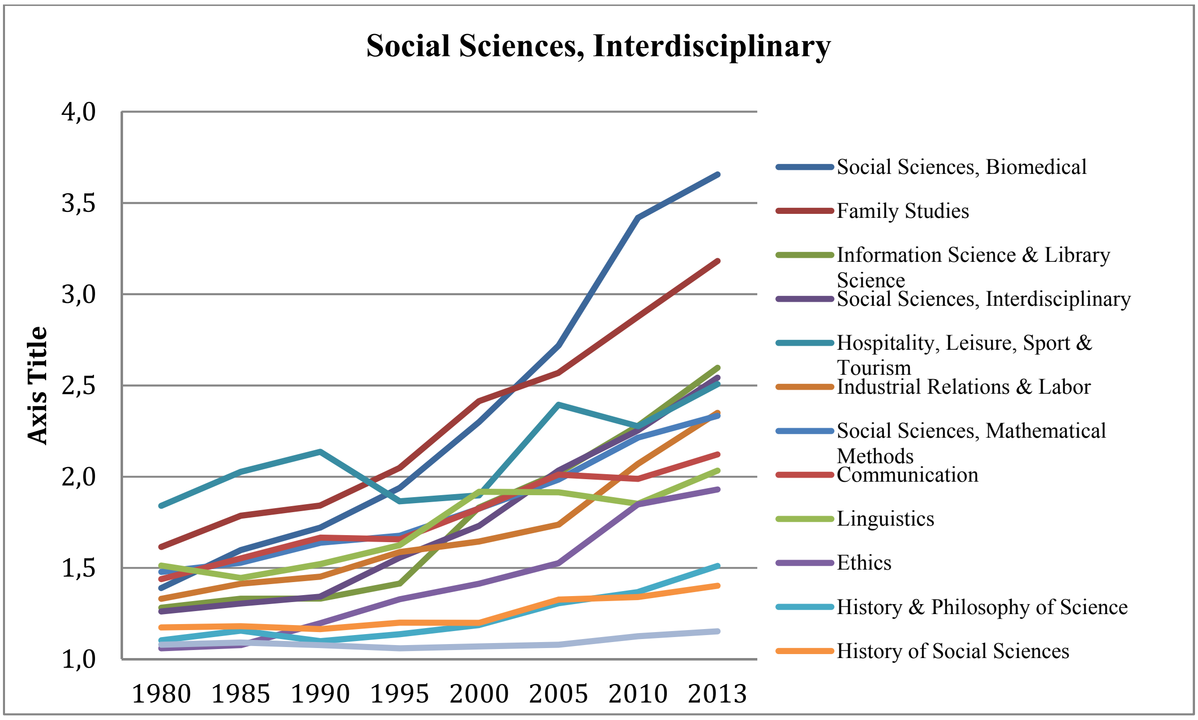


Figure 6. The mean number of authors per publication from 1980-2013 of the group Social Sciences, Interdisciplinary.

694,752 publications are indexed in the categories in the group Social Sciences, Interdisciplinary. The mean increases in numbers of authors are between 6.8%-163% or between 0.1-2.3 authors. Figure 6 demonstrates how much the increase in the numbers of authors varies from 1980 to 2013. The category History hardly had any changes in the mean number of authors and the median remain constantly at 1 during the time period. The median also remains at 1 author in the categories History of Social Science, History & Philosophy of Science and Ethics, while the mean rises from 1.1-1.2 authors in 1980 to 1.4-1.9 authors in 2013. The median increases from 1 to 2 authors in the categories Communication, Information Science & Library Science, Industrial Relations & Labor, Linguistics, Social Sciences, Interdisciplinary and Social Sciences, Mathematical Methods and the mean numbers of authors increases from 1.3-1.5 authors to 2-2.5 authors. The median is constant at 2 authors in Hospitality, Leisure, Sport & Tourism, where mean number of authors rise from 1.8 authors to 2.5 authors. The median increases from 1 author in 1980 to 3 authors in 2013 in the categories Family Studies and Social Sciences, Biomedical and the mean numbers of authors rises from 1.4-1.6 authors to 3.2-3.7 authors. In this very mixed group we can see how the categories with research closest to the humanities such as History, History of Social Science, History & Philosophy of Science and Ethics have a lower rise in the number of mean authors, while the categories Family Studies and Social Sciences, Biomedical, that both are methodological close to the life and medical sciences have had a substantial high rise in number of authors.

Discussion

In this study we document the evolution of co-authorship in the social sciences and find that the majority of research fields have had substantial increases in the numbers of authors per publication. During the 33 years the increase is equal to one author or more in

31 out of 56 subject categories, and in further five subject categories, the increase is nearly 1 author (0.9). We detect a similar increase when we include the median increase in the number authors, where the median number of authors has increased by one or more authors in 42 out the 56 subject categories. The increases in the number of authors have not happened in the same degree in all areas of the social sciences and illustrate how heterogeneous the research fields of social sciences are. The articles indexed in the four subject categories History, Cultural Studies, Area Studies and History of Social Sciences have only had a mean increase in the number of authorship between 0.1-0.2, and could be categorized as status quo during the 33 years. The percentage increases in the mean number of authors in the subject categories varies from 6.8% (History) to 175.6% (Health Policy & Services).

The results of this study confirm that there is an increasing tendency to co-author and collaborate and is in line with the tendency detected in previous studies of co-authorship and collaboration (e.g. Bebeau & Monson, 2011; Fisher et al., 1998; Ossenblok et al., 2014). Namely that the number of authors per publication has increased in the social sciences and that the largest increases have occurred in the fields with use of experiments, large data set, statistical methods and/or team-production models, such as the Social Health Sciences and parts of Psychology. A good example in our study of how the methodological differences affect the collaboration patterns is the subject categories group as Psychology. The subject categories Psychology, Psychoanalysis and Mathematical are both examples of research domains dominated by theory building and abstract concepts and with methodological relationships to research fields often defined as belonging to the humanities. The opposite are Psychiatry and Developmental Psychology, where the research are more experimental and empirical, and often sampled in collaboration with other researchers. Hence, the greatest rises in number of authors have occurred in subject categories containing research fields using quantitative methods and with a close relationship to the medical and life sciences or the natural sciences. An additional explanation for the rise in co-authorship in the majority of the subject categories is the increasing tendency for supervisors to co-author with students (Costa & Gatz, 1992; Fisher et al., 1998; Price et al., 2000).

Conclusion

As mentioned in the introduction, most of the bibliometric studies about co-authorship and research collaboration in the social sciences have been focusing on the trends and patterns in particular research fields or countries and have been based on data collected from a selection of journals in one or few research fields or national databases. In this study we use a larger sample of articles to confirm there is a rise in co-authorship in the majority of the research fields in the social sciences, and that in more than half of the subject categories the mean number of authors has increased by one or more authors.

Few of these studies undertake a deeper investigation of the rise of co-authorship and research collaboration (Costa & Gatz, 1992; Fisher et al., 1998), and the explanations offered for the rise is often speculative and anecdotal or borrowed from the “hard” sciences. We have discussed some of the factors that influence the researchers’ collaboration behavior and the rise in co-authorship. However, our explanations are based on the fore mention studies, and we therefore suggest that the next step is a thoroughly investigation of the effects of these factors in the fields we have documented a rise in co-authorship.

Acknowledgement

I thank my supervisor Jesper Schneider for comments and subtracting data from the CWTS database. Furthermore, I thank Søren Midtgaard for his suggestions and comments.

References

Beaver, D. D. (2001). Reflections on scientific collaboration, (and its study): past, present, and future. *Scientometrics, 52*(3), 365-377. doi: 10.1023/a:1014254214337

Bebeau, M., & Monson, V. (2011). Authorship and Publication Practices in the Social Sciences: Historical Reflections on Current Practices. *Science and Engineering Ethics, 17*(2), 365-388. doi: 10.1007/s11948-011-9280-4

Biagioli, M. (2012). Recycling Texts or Stealing Time?: Plagiarism, Authorship, and Credit in Science. *International Journal of Cultural Property, 19*(3), 453-476. doi: 10.1017/s0940739112000276

Birnholtz, J. P. (2006). What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology, 57*(13), 1758-1770. doi: 10.1002/asi.20380

Butler, L. (2003). Modifying publication practices in response to funding formulas. *Research Evaluation, 12*(1), 39-46. doi: Doi 10.3152/147154403781776780

Costa, M. M., & Gatz, M. (1992). Determination of Authorship Credit in Published Dissertations. *Psychological Science, 3*(6), 354-357. doi: DOI 10.1111/j.1467-9280.1992.tb00046.x

Cronin, B., Shaw, D., & La Barre, K. (2003). A cast of thousands: Coauthorship and subauthorship collaboration in the 20th century as manifested in the scholarly journal literature of psychology and philosophy. *Journal of the American Society for Information Science and Technology, 54*(9), 855-871. doi: 10.1002/asi.10278

Fisher, B. S., Cobane, C. T., Vander Ven, T. M., & Cullen, F. T. (1998). How many authors does it take to publish an article? Trends and patterns in political science. *Ps-Political Science & Politics, 31*(4), 847-856. doi: 10.2307/420730

Hicks, D. (2005). The Four Literatures of Social Science. In H. Moed, W. Glänzel & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 473-496): Springer Netherlands.

Hudson, J. (1996). Trends in multi-authored papers in economics. *Journal of Economic Perspectives, 10*(3), 153-158.

Lariviere, V., Gingras, Y., & Archambault, E. (2006). Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. *Scientometrics, 68*(3), 519-533. doi: DOI 10.1007/s11192-006-0127-8

Lee, W. M. (2000). Publication trends of doctoral students in three fields from 1965-1995. *Journal of the American Society for Information Science, 51*(2), 139-144. doi: 10.1002/(sici)1097-4571(2000)51:2<139::aid-asi5>3.0.co;2-1

Melin, G. (2000). Pragmatism and self-organization - Research collaboration on the individual level. *Research Policy, 29*(1), 31-40. doi: Doi 10.1016/S0048-7333(99)00031-1

Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review, 69*(2), 213-238.

Norris, R. P. (1993). Authorship Patterns in Cjnr - 1970-1991. *Scientometrics, 28*(2), 151-158. doi: Doi 10.1007/Bf02016897

Ossenblok, T. L. B., Engels, T. C. E., & Sivertsen, G. (2012). The representation of the social sciences and humanities in the Web of Science-a comparison of publication patterns and incentive structures in Flanders and Norway (2005-9). *Research Evaluation, 21*(4), 280-290. doi: DOI 10.1093/reseval/rvs019

Ossenblok, T. L. B., Verleysen, F. T., & Engels, T. C. E. (2014). Coauthorship of journal articles and book chapters in the social sciences and humanities (2000–2010). *Journal of the Association for Information Science and Technology, 65*(5), 882-897. doi: 10.1002/asi.23015

Piro, F. N., Aksnes, D. W., & Rørstad, K. (2013). A macro analysis of productivity differences across fields: Challenges in the measurement of scientific publishing. *Journal of the American Society for Information Science and Technology, 64*(2), 307-320. doi: 10.1002/asi.22746

Price, J. H., Dake, J. A., & Oden, L. (2000). Authorship of health education articles: Guests, ghosts, and trends. *American Journal of Health Behavior, 24*(4), 290-299.

Rennie, D., Yank, V., & Emanuel, L. (1997). When authorship fails - A proposal to make contributors accountable. *Jama-Journal of the American Medical Association, 278*(7), 579-585. doi: 10.1001/jama.278.7.579

SSCI. (2012). Scope Notes 2012. Social Science Citation Index. Retrieved 10-01-2015, 2015, from http://ip-science.thomsonreuters.com/mjl/scope/scope_ssci/#BF

Weingart, P. (2005). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics, 62*(1), 117-131. doi: 10.1007/s11192-005-0007-7

White, K. D., Dalgleish, L., & Arnold, G. (1982). Authorship Patterns in Psychology - National and International Trends. *Bulletin of the Psychonomic Society, 20*(4), 190-192.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036-1039. doi: 10.1126/science.1136099

The Recurrence of Citations within a Scientific Article

Zhigang Hu[1], Chaomei Chen[2] and Zeyuan Liu[1]

[1] *huzhigang@dlut.edu.cn, liuzy@dlut.edu.cn*
WISE Lab, Dalian University of Technology, 116024 Dalian (China)

[2] *cc345@drexel.edu*
College of Computing and Informatics, Drexel University, 19104 Philadelphia (USA)

Abstract

Although listed at the tail of a scientific article only once, a reference is usually cited repeatedly inside the full text of the article. In this research, we investigated the universality of recurring citations in Journal of Informetrics. About 1/4 references are repeatedly cited. For these repeatedly cited references, their citation location and citation context for the first and subsequent times are examined separately. Normally, recurring citations of a same reference tend to be located in the same section instead of different ones. It proves that, even if a reference is cited for multiple times in a single citing paper, it is still focus on the same topic in the same section most of the time. We also explored the reason why recurring citations are happening. By comparing the contexts of two kinds of citations, the first-time citations and the succeeding citations, we found that, for a specific reference, its first-time citation is usually not as intentional as the succeeding citations. Just because of the relative importance of the succeeding citations compared to the first-time citation, recurring citations are reasonable and necessary.

Conference Topic

Citation and co-citation analysis

Introduction

Citations are essential components for scientific articles. Traditional citation analysis is more like reference analysis, since only references listed at the tail of the article are researchers' concern. Citations, which indicate the locations and context where references are cited, are almost ignored in previous research. The reference analysis is much easier and effective most of the time, but in the meanwhile, some important information might be neglected. For example, where are these references are cited inside the citing papers? How are the citations distributed among different sections? By investigating the citation location and the citation context, however, we can understand not only the pattern how references are cited, but also the reason why authors cite it like that.

Nowadays, full-text citation analysis, which is about how references are cited in the body of citing papers, is just beginning (Ding, Liu, Guo, & Cronin, 2013; Hu, Chen, & Liu, 2013; Liu, Zhang, & Guo, 2013; Zhang, Ding, & Milojević, 2013). During to the increasingly availability of structured full texts such as XML-formatted articles, researchers began to turned their attention from references to citations in the body of articles. For example, Ding et al. have examined the distribution of references across text and find that most highly cited works appear in the Introduction and Literature Review sections of citing papers (Ding et al., 2013). Hu et al. visualize the location distribution of citation instances, especially those to highly-cited references. The results show that citations are usually distributed very uneven inside the full texts of scientific articles (Hu et al., 2013).

In full-text citation analysis, recurring-citation is an interesting issue. Recurring-citation refers to the phenomenon that a reference is cited more than once in a citing paper. Take this paper for example, we cite the reference of (Hu et al., 2013) in the first sentence of last paragraph for the first time, and then cite it again in the last sentence of the same paragraph for the second time. In this paper, we call this reference a repeatedly cited reference, or a reference with recurring citations. Recurring-citation is a common fact in citation behaviour. In our previous research, we find that, sometimes, a reference might be repeatedly cited as many as nine times in a single paper (Hu et al., 2013).

In this research, we will investigate the phenomenon of recurring citations. Our concern is the universality and the pattern of recurring-citations, including: (1) how common recurring citations are in scientific articles? (2) where the recurring citations of a single reference are usually located inside the paper? (3) what the difference is between its first-time citation and the succeeding ones? In the end, the reason why recurring-citation is necessary will also be discussed.

Data and Methods

To detect recurring citations of a reference inside a citing paper, the full text of the citing paper need to be observed. There are two types of full texts: one is in unstructured format such as PDFs, which is human-friendly; the other is in structured format such as XMLs/HTMLs, which is machine-friendly. Compared with PDFs, structured full texts, e.g., XMLs, are much easier to process by computer. For example, XML-formatted full texts can be parsed directly using an existing function: *xml_parse()* in PHP. Thus, it is very straightforward to identify citations inside a citing paper. Nowadays, structured XML-formatted full texts are available and downloadable in almost every bibliographic database, such as Elsevier, Springer, John & Wiley, and especially, the open access online journals like PLoS ONE. In this study, the data of full texts was sourced from Elsevier ConSyn (http://consyn.elsevier.com), a content syndication system developed by Elsevier. Since 2011, Elsevier ConSyn provided downloadable articles in XML format.

In Elsevier ConSyn, we retrieved and downloaded all the full texts of 350 articles published in Journal of Informetrics (JOI) from 2007 to 2013. Journal of Informetrics is chosen as the case in this study because it is published by Elsevier and belongs to the field of library and information science. By our own developed program, we parsed these XML-format full texts and extract all the citations inside them. Since each citation instances is clearly marked with a XML tag, i.e. <*ce:cross-ref*>, they can be recognized and extracted easily. All the attributions of each citation, including its location and its citees, were recorded and import into database tables.

By looking into citations' citees, we achieved the cited times of each reference inside each citing paper. If the cited times is equal to one, it means the reference is one-time cited inside this citing paper. While if cited more than once, the reference is considered as repeatedly cited or recurrently cited. In this research, we will count the frequency of each type of reference, e.g., once-cited, twice-cited, triple-cited, etc. In this way, the universality and intensity of recurring-citation can be estimated accordingly.

For repeatedly cited references, their citation locations will be studied. The location of citation can be measured by, from macro to micro scales: character, word, sentence, paragraph and section. In this study, we chose the measurement at the largest scale: section. We will calculate the count of citations in each section and see how citations are distributed in different sections. Generally, a scientific article is made up of four sections, namely *Introduction*, *Data and Methods*, *Results,* and *Discussion and Conclusions*. It is called IMRaD structure usually (see e.g. Agarwal & Yu, 2009; Swales, 1990) To some extent, citation location can reveal the citation motivation. If we are aware of the section where a citation is located, the role of the citation can be figured out to some extent. For instance, if a reference is cited in the section of *Data and Methods*, usually section II, it is probably a helpful citation relevant in the aspect of methodology; while if it located in the section III or the section of *Results*, the citation is more likely about comparable results.

Besides the location distribution of recurring citations, we also examined the difference between a reference's first-time citation and the succeeding ones. We extracted the context when a reference is cited for the first time and when it is cited again in the following parts. The first-time and the succeeding citation contexts will be compared in terms of the count of their citees inside. The more citees/references a citation contains, the less important each citee/reference is. The citation with many citees/references, such as the one in the first sentence of the second paragraph, is called perfunctory citation (Cano, 1989; Oppenheim & Renn, 2004; Pham & Hoffmann, 2003; Voos & Dagaev, 1976), which means authors decide not to cite the citees/references seriously in an excluded way. In this research, we are interesting in which one, the first-time citation or the succeeding citation, is more likely to be perfunctory citation for a multiple-cited reference.

Results and Discussion

The universality of recurring citations

Firstly, we examined how common recurring citations are in the Journal of Informetrics. Among all the 11,327 references inside the 350 articles, 8,417 (74.3%) of them were cited once in a single citing article. The other 2,910 references (account for 25.7%) were cited twice or more, including 1,726 (15.2%) twice-cited references, 613 (5.4%) triple-cited references, and 571 (5.0%) references cited for four times or more. Although one-time citation is the main citation pattern undoubtedly, the phenomenon of recurring-citation cannot be ignored in both frequency and intensity.

Figure 1 shows the frequency distribution of references of each kind, companied with a distribution graph in double logarithmic coordinates. As it shown in the best fitting line, the frequency distribution of multiple citations follows a power law ($y= 21557\ x^{-3.479}$, $R^2=0.9679$), which is a very common law in the field of bibliometrics, such as the distribution of scientific productivity (Lotka, 1926) or keywords (Zipf, 1949). Obviously, it is not accidental that the frequency distribution of recurring citations is in this pattern.

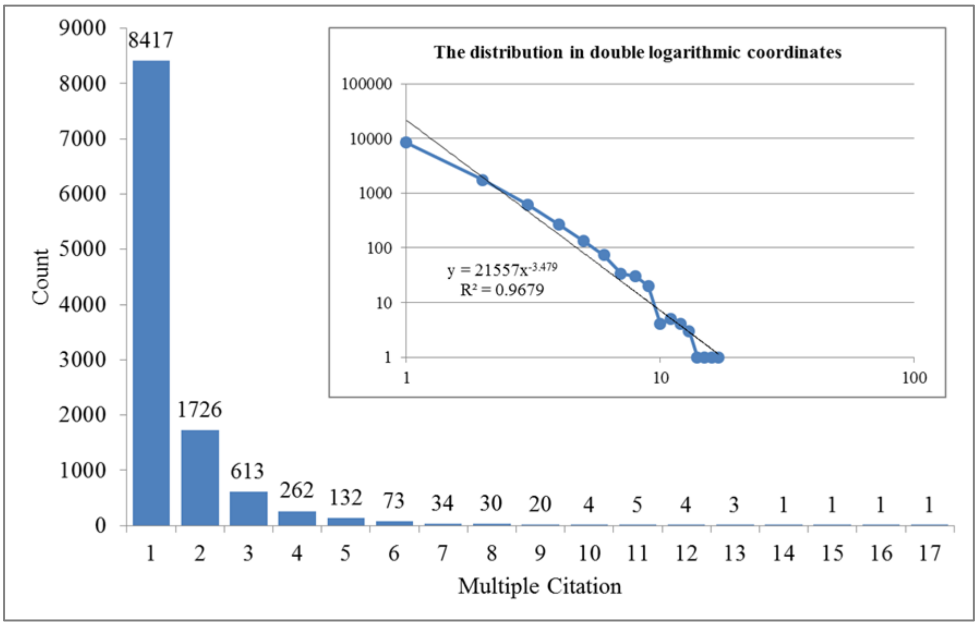


Figure 1. The count of references by their multi-citation times.

The locations of recurring citations

The location pattern of recurring citations is the focus of this research. In this part, we will investigate the location distribution of multi-citation by section. In Journal of Informetrics, 92 articles (26.3% of all) adopt IMRaD structure, which is most used form to organize articles in our research. Thus, we selected all these 92 four-section articles in IMRaD structure as cases, and explored how citations are distributed in the four different sections.

As shown in Figure 2, among all the 3035 citations in these 92 articles, 1238 (40.8%) citations are located in Section I, or the section of *Introduction*; 760 (25.0%) of them are located in Section II (or *Methods*); 769 (25.3%) citations is in the sections of *Results*; and 268 (8.8%) in *Discussion and Conclusions*. This mode of section distribution of citations meets our expectation on citation locations, since it is the widely accepted fact that authors are likely to cite most in the section of *Introduction*.

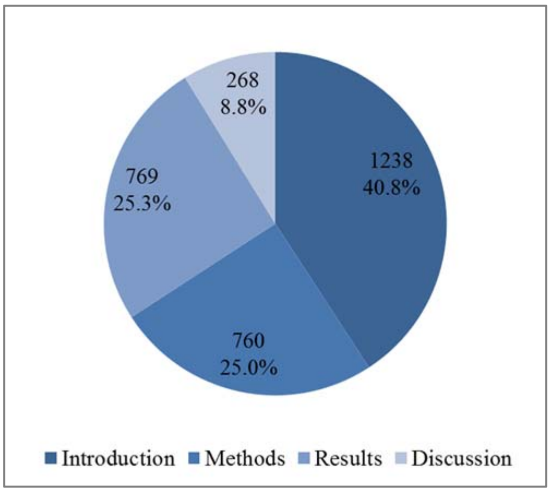


Figure 2. The count of citations in each section.

Based on the section distribution of citations, we are then able to investigate the section combination distribution of a reference's recurring citations. For each repeatedly cited reference, its recurring citations could be located in any sections, either the same section or the different sections. Since twice cited references are the simplest and most common (59.3%) types of repeatedly-cited references, they were chosen for calculating combined-section distribution.

For each twice-cited reference, we recorded the located sections of both citations. The counts of the 10 types of section combinations are shown in Table 1. Among all the 796 twice-cited references, the most common ones are those cited in Section I for both the first and the second time. 224 (28.2% of all) references belong to this type. 124 (15.6%) references are cited in Section I for the first time and Section II for the second time. References that are cited in section IV twice are least common (18 or 0.8%). Totally, 444 (55.8%) references are cited in the same section twice, while 350 (44.2%) ones are cited in the difference sections.

Table 1. The combined section distribution of the twice citations of references

Located Section of the second citation / Located section of the first citation	*Sec I*	*Sec II*	*Sec III*	*Sec IV*
Sec I	**224** 28.2%			
Sec II	**124** 15.6%	**90** 11.3%		
Sec III	**68** 8.6%	**42** 5.3%	**112** 14.1%	
Sec IV	**64** 8.1%	**24** 3.0%	**28** 3.5%	**18** 2.3%

Although more twice cited references are cited in the same section, we cannot say that a reference's multiple citations tend to be located in the same sections except that the expected proportion of the multiple citations located in the same section is calculated and compared. Thus, we assume that a reference's twice citations are located independently and randomly, just like two arbitrary citations in the article. Under this hypothesis, the expected distribution of section combinations of twice citations can be calculated as follow:

$$
\begin{aligned}
&(Sec\ I, Sec\ I) : (Sec\ I, Sec\ II) : (Sec\ I, Sec\ III) : (Sec\ I, Sec\ IV)\\
&: (Sec\ II, Sec\ II) : (Sec\ II, Sec\ III) : (Sec\ II, Sec\ IV)\\
&: (Sec\ III, Sec\ III) : (Sec\ III, Sec\ IV)\\
&: (Sec\ IV, Sec\ IV)\\
&= 40.8\%\times40.8\% : 40.8\%\times25.0\%\times2 : 40.8\%\times25.3\%\times2 : 40.8\%\times8.8\%\times2\\
&\quad : 25.0\%\times25.0\% : 25.0\%\times25.3\%\times2 : 25.0\%\times8.8\%\times2\\
&\quad : 25.3\%\times25.3\% : 25.3\%\times8.8\%\times2\\
&\quad : 8.8\%\times8.8\%
\end{aligned}
$$

=16.6% : 20.4%: 20.7% : 7.2% : 6.3% : 12.7% : 4.4% : 6.4% : 4.5% : 0.8%

Figure 4 shows the expected and observed proportions of the section combinations of each kind. If the expected values match the observed well, it means the twice citation are located independently and randomly indeed; otherwise, it means that there is a certain tendency in how to cite a reference twice. In Figure 4, we have not seen the match between the expected and observed values. For example, based on our initial hypothesis, the proportion of (Sec I, Sec I) should be 16.6%, not even closed to 28.2% as observed; the proportion of (Sec I, Sec III) should be 20.7%, while the observed value is 8.6%, which is much lower. Neither of them presents the match as assumed.

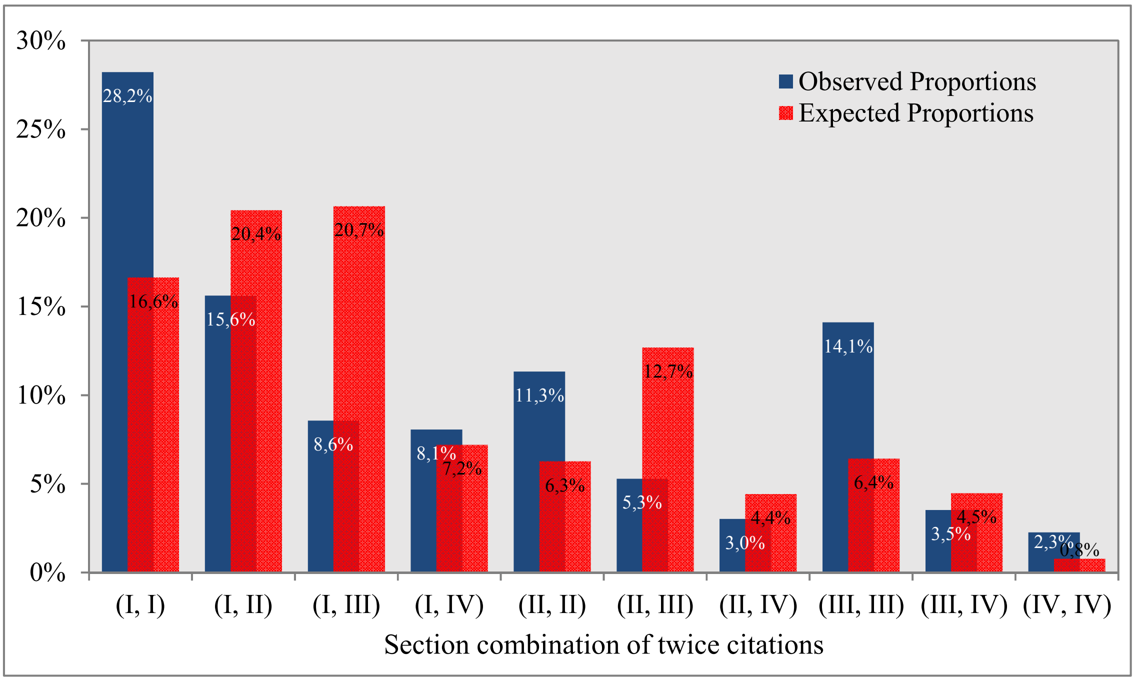


Figure 3. The distribution of the expected and observed proportions of located section combinations of twice citations.

According to the comparison between the expected and observed values, these 10 combinations can be divided into two classes: the above-expectation combinations and the below-expectation ones. The combinations of (Sec I, Sec I), (Sec II, Sec II), (Sec III, Sec III), (Sec IV, Sec IV) and (Sec I, Sec IV) belong to the former. Their observed proportions are higher than expected significantly. The other 5 combinations, i.e., (Sec I, Sec II), (Sec I, Sec III), (Sec II, Sec III), (Sec II, Sec IV) and (Sec III, Sec IV), belongs to the latter. From this division, we can see that, the references with twice citations located in the same section are preferable to those with twice citations located in different sections. The only exceptions are the references cited inside (Sec I, Sec IV), which have an above-expectation proportion (2.3% v.s. 0.8%), though its twice citations located in the different sections.

Why do authors tend to cite a reference multiple times inside the same section? The explanation could be simple. Normally, a reference is only helpful for a single topic, usually existing in a concentrated part of an article, such as a section. Few references are necessary for several different topics, or in different sections. That is why references are preferred to be cited in a single section. This explanation also interprets why the combination of (Sec I, Sec IV) is an exception. The first and the fourth section, although farthest away with each other, are actually discussing about the same topic at the same level, i.e., the hindsight and foresight of research questions.

The context of recurring citations

We have revealed how common recurring citations are and where these recurring citations are usually located, and now we will examine their contexts. Firstly, the citation contexts of repeatedly cited references for the first and the succeeding times were extracted separately. There are totally 11,448 first-time citation contexts and 5,469 succeeding ones extracted. We will explore the difference between these two groups of citation contexts in terms of citation intensity, which can be estimated by how many citees they contained.

The count of citees contained in each citation context is calculated one by one. Averagely, a citation contains 1.94 citees, or put it another way, authors cite 1.94 references once at a time. As it shown in Figure 6, although most citations (64.2% of all) cite only one single citee/reference, there are still more than 1/3 of citations contain two or more citees/reference. 1457 (8.7%) citations cite even five or more references once.

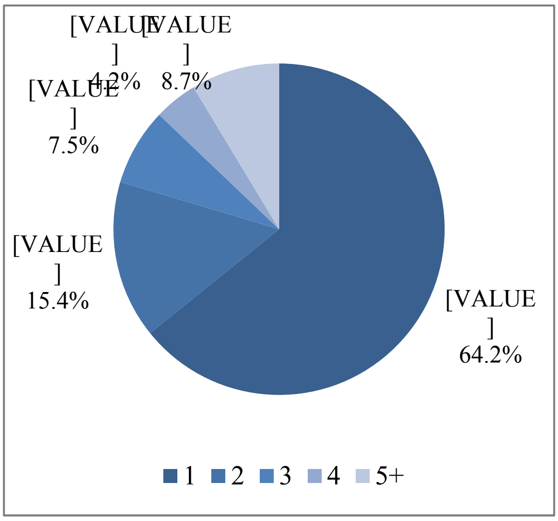


Figure 4. The distribution of citations by the count of contained citees.

Separately, the counts of citees contained by the first-time and succeeding citations are investigated. The first-time citations contain 2.13 citees on the average, while the succeeding citations contain 1.94 citees. Figure 5 shows the specific distribute of both of them by their count of contained citees. For the first-time citations, totally 38.5% of citations cited two citees or more; while for succeeding citations, only 30.1% did. It means the first-time citations are more likely to be perfunctory citations than the succeeding citations. In other words, authors normally cite a reference more casually and perfunctorily for the first time; and then cite it again in the following paragraphs more formally and solemnly. In other words, usually, authors just mention a reference in the beginning, and then seriously use it when citing it later again.

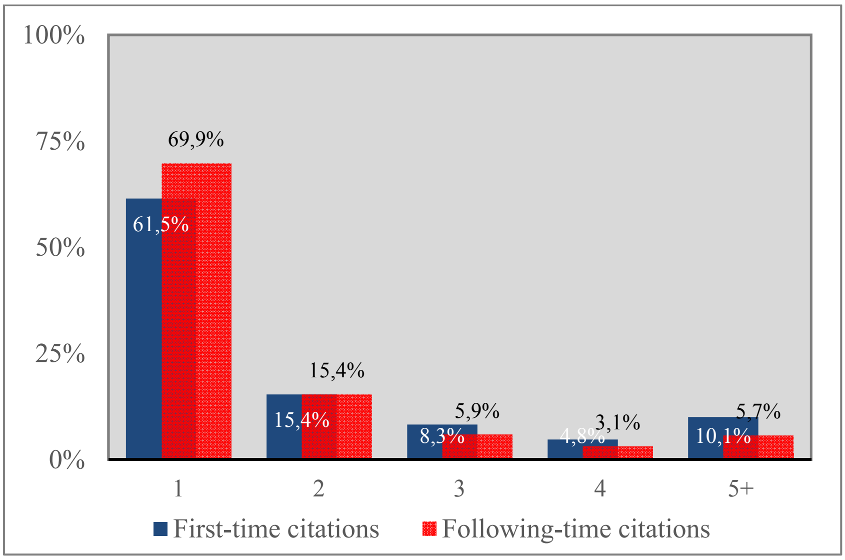


Figure 5. The distribution of the first-time and succeeding citations by their count of contained citees.

Conclusions

Recurring citations are common in scientific publications. In Journal of Informetrics, about 1/4 references are repeatedly cited in citing papers. Although not the mainstream of citation pattern, recurring-citation is undoubtedly a phenomenon that cannot be ignored in full-text citation analysis, an increasing hot research field in recent year.

In this study, we investigate the recurring-citation phenomenon in two perspectives: the citation location and the citation context. In citation location analysis, we find that a reference's recurring citations tend to be located in the same section or closely with each other. It shows that a reference is only cited in a single topic normally. When the topic switches, the reference has little chance to be cited again.

The context of recurring citations contexts are also examined in terms of their citation intensity. As it shown in the result, for a repeatedly cited reference, its first-time citation is usually kind of perfunctory. The reference is always cited accompanied with other references together. When it is cited another time in the following part of the citing paper, the citations are more exclusively and solemnly. Precisely because the succeeding citations are usually more importantly, recurring citations are reasonable and necessary inside scientific articles.

Acknowledgments

The research was supported by the Natural Science Foundation of China (NSFC) under Grant 71103022, 61301227 and 71473028.

References

Agarwal, S., & Yu, H. (2009). Automatically classifying sentences in full-text biomedical articles into Introduction, Methods, Results and Discussion. *Bioinformatics*, *25*(23), 3174–80. doi:10.1093/bioinformatics/btp548

Cano, V. (1989). Citation behavior: Classification, utility, and location. *Journal of the American Society for Information Science*, *40*(4), 284–290.

Ding, Y., Liu, X., Guo, C., & Cronin, B. (2013). The distribution of references across texts: Some implications for citation analysis. *Journal of Informetrics*, *7*(3), 583–592. doi:10.1016/j.joi.2013.03.003

Hu, Z., Chen, C., & Liu, Z. (2013). Where are citations located in the body of scientific articles? A study of the distributions of citation locations. *Journal of Informetrics*, *7*(4), 887–896.

Liu, X., Zhang, J., & Guo, C. (2013). Full‐text citation analysis: A new method to enhance scholarly networks. *Journal of the American Society for Information Science and Technology*, *64*(7), 1852–1863.

Lotka, A. J. (1926). The frequency distribution of scientific production. *Journal of the Washington Academy of Science*, *16*, 317–323.

Oppenheim, C., & Renn, S. P. (2004). Highly cited old papers and the reasons why they continue to be cited. *Journal of the American Society for Information Science*, *51*(5), 225–231.

Pham, S. B., & Hoffmann, A. (2003). A new approach for scientific citation classification using cue phrases. *AI 2003 Advances in Artificial Intelligence*, 759–771.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. *Journal of Advanced Composition* (Vol. 11, p. 272). Cambridge: Cambridge University Press.

Voos, H., & Dagaev, K. S. (1976). Are all citations equal? Or did we op. cit. your idem? *Journal of Academic Librarianship*, (1), 20–21.

Zhang, G., Ding, Y., & Milojević, S. (2013). Citation content analysis (CCA): A framework for syntactic and semantic analysis of citation content. *Journal of the American Society for Information Science*, *64*(7), 1490–1503.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge: Addison-Wesley Press.

Do Authors With Stronger Bibliographic Coupling Ties Cite Each Other More Often?

Ali Gazni[1, 2] and Fereshteh Didegah[3]

[1]*ali.gazni@isc.gov.ir*
[1]Islamic World Science Citation Center, Shiraz (Iran)
[2]Regional Information Center for Science and Technology, Shiraz (Iran)

[3]*fdidgah@gmail.com*
Statistical Cybermetrics Research Group, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, (UK)

Abstract

Author bibliographic coupling is extended from bibliographic coupling concept and holds the view that two authors with more common references are more related and have more similar research interests. This study aims to examine the association between author bibliographic coupling strength and citation exchange in Information Science & Library Science and more specifically, in imetrics. The results show that there is a positive and significant association between these two factors in Information Science & Library Science and also in imetrics; however, the correlation is more significant among imetricians. This confirms the Merton's norm of universalism versus constructivists' particularism. A closer investigation of bibliographic coupling and citation networks among thirty highly cited imetricians shows that Thelwall, M. is in strong bibliographic coupling and citation relationships with the majority of authors in the network. He and Bar-Ilan have the strongest ABC and citation relationships in the network. Rousseau, R., Glänzel, W., Bornmann, L., Bar-Ilan, J., and Leydesdorff, L. are also in strong ABC relations with each other as well as other authors in the network.

Conference Topic

Citation and co-citation analysis

Introduction

Bibliographic coupling (BC), first introduced by Kessler in 1963, refers to the number of common references between two articles. The more the number of common references between two articles, the more intellectually related they are.

In contrast with co-citation analysis (CA) requiring strength signals (number of citations), BC could help in research fronts detection even with weak signals (Glänzel & Czerwon, 1996). Kuusi and Meyer (2007) claimed that BC has never been used for exploring technology foresight and rare studies used it for research evaluation purposes. However, they used BC for anticipating technological breakthroughs. Yan and Ding (2012) compared different types of networks, including citation based and non-citation based networks at institutional level, and found that BC and AC networks have high similarity and also found that AC has a high similarity with citation networks. Boyack, Börner and Klavans (2009) applied BC to mapping the structure and evolution of research publications in Chemistry. Soó (2014) proposed age-sensitive BC, so if two documents share recent references, they are more related than those sharing older references. Hence, not only the number of common references, but also their age, influences the extent of relatedness between two research works. Van Raan (2005) also reported that intellectual relatedness between two documents could be better obtained through using common references that are more recent. BC is an effective way for science mapping, research fronts detection and information retrieval (See Jarneving, 2007, 2005; Morris, Yen, Wu, & Tesfaye, 2003; Qiu, 2007). Peters, Braam and van Raan (1995) investigated chemical engineering publications and found that publications with common citations to highly cited papers are more related. White et al. (2004) claim that intellectual ties based on shared references could serve as a better predictor for citations between authors than social ties.

Author bibliographic coupling (ABC), first proposed by Zhao and Strotmann (2008), is extended from BC concept and holds the view that two authors with more common references have more similar research interests. They mentioned that BC is fixed when two articles are published but ABC is constantly evolving over time as the two authors' oeuvre grows. Ma (2012) stated that ABC has an advantage in providing a more comprehensive and concrete map of intellectual structure of the fields and detecting their research fronts in comparison to author co-citation analysis (ACA). The very few studies on ABC did only an author coupling analysis of intellectual structure of few subject fields. For example, using a combination of ACA and author bibliographic coupling analysis (ABCA), Zhao and Strotmann (2014) sought to predict future research trends in information science (IS). They studied research fronts and knowledge bases of IS and also the structural evolution of IS between two 5-year periods (2001-2005 and 2006-2010). They found ABCA an appropriate method to investigate authors' specific research interests in IS and suggested using ACA and ABCA together to better investigate intellectual structure of a subject domain. The same combined method was used in Byun and Chung (2012) to study the research trends of authors in social welfare science; they also suggested using both ACA and ABCA together to investigate traditional and future research trends of a specific domain.

The extent to which two authors are coupled through common references is measured by ABC strength which has different methods to calculate it: Simple, minimum and combined methods (Ma, 2012). Rousseau (2010) also proposed a simple method for calculating the relative ABC by dividing the number of common references between two authors by the total number of their references. Frequency of common references was simply used to measure ABC strength in this study.

No research on the association between ABC strength of two authors and number of citations exchanging between them is found, so this study seeks to examine this relationship in Information Science & Library Science (IS&LS) and more specifically, in imetrics. Therefore it aims to examine the correlation between ABC strength measured by the number of common references between two authors and the number of citations exchanged between them.

Research questions

According to the normative theory of citation, citations are indicators of the cognitive or intellectual influence of a scientific work (Merton, 1973). In a scientific paper, citations can be concept markers (Small, 1978), however, and can transfer knowledge and help with its enlargement (Merton, 1988). As a result, methods like CA have been used for mapping intellectual structure in science (Small, 2004), where BC is used for the same purpose. Hence, common references between pairs of documents, authors, journals or institutions show the extent to which they are related. For instance, two authors who share a larger number of common references are likely to do research on a narrow area and exchange a high number of citations. Counting citations between two authors with different BC strengths, not only could support Robert K. Merton's norm of universalism versus constructivists' particularism, but also shows any possible difference by the number of common references as a measure of relatedness and types of authors (i.e. highly cited vs. less cited authors).

The theories of citation, normative view vs. social constructivist view, will be examined through answering these questions. The normative theory of citation holds that citations reflect the scientific quality and merits of research outputs because citers use them to reward the works of their colleagues (Small, 2004; White, 2004; MacRoberts & MacRoberts, 1987; Merton, 1973) whereas the social constructivist theory holds that authors use the references to support their own claims and points made. This latter theory emphasises factors affecting citations other than the quality and content of the cited article (White, 2004; Baldi, 1998; Gilbert, 1977).

Given that BC shows relatedness, a positive association between the number of common references and number of citations between two authors will confirm that citations are made for the matter of 'relatedness' and are not perfunctory.

To reach the research goals, this study seeks to answer these questions:

1. Do two authors with a higher number of common references cite each other more often?
2. Is the above association stronger for highly cited authors than other authors?

Methodology

Data collection:

Documents published during 1990-2012 in the journals of Information Science & Library Science (IS&LS) were extracted from Thomson Reuters Web of Science (WoS). This time period is current and consists of a reasonable number of years for investigating the relationship between number of common references and citations exchanged between authors. WoS indexes the mainstream of research and the most prestigious journals in different fields of science; however, a large number of journals in WoS come from a small number of international publishers (Didegah & Gazni, 2011).

Author names disambiguation:

The author names were disambiguated by improving Gazni & Thelwall (2014) method, resulting in 98.2% precision and 92.7% recall. The co-authorship network of authors was used for the improvement. For example, A is a disambiguated author and B is his/her co-author. The papers written by both A and B as co-authors were appended to A's articles. Author names' disambiguation will improve the accuracy of research on author level analysis by distinguishing one name that belongs to several different people and conflating the name variants of a single person.

Calculations:

To make the processing manageable, a random sample of 385 authors with any properties out of all authors who have at least one paper in the journals of IS&LS during 1990-2012 was chosen. The number of common references between these 385 authors and all other authors in the field were counted, where the joint papers were eliminated either for counting the number of common references or for counting the number of citations made and received between each pair of authors. Only citations made and received from the journals in the field were processed for either counting the number of citations between authors or counting the number of common references among them. A list of authors who have at least one common reference with the authors in the sample, and also exchanged citations with them, was created for each author in the sample. For a closer investigation of the association between the number of common references and citations between pairs of authors and also of ABC networks, a sample of highly cited authors in imetrics was taken into account. For this purpose, thirty highly cited imetricians introduced in Abrizah and colleagues (2014) were selected for further analysis. The main reason for taking this sample into account is that these are prolific authors in a specific domain, publishing for a long time and have an excellent knowledge of the domain, its publications and researchers. This is while in the sample of authors from IS&LS, there may be less prolific authors, such as students who publish for a short period of time and then disappear from the research area, and their unfamiliarity with the area will affect their reference and citation behaviours. Therefore, a sample of thirty highly cited imetricians is a consistent sample for showing the association between ABC strength and citation exchange between pairs of authors.

Results

The association between number of common references (BC strength) and number of exchanged citations between pairs of authors in IS&LS

Spearman correlation was tested for the association between the number of common references and the number of citations exchanging between pairs of authors. The results show positive significant correlations between the number of times two authors cited each other and the number of common references between them. The correlation was tested for different groups of pairs of authors with one to 300 common references; it is stronger for the groups of authors with 300 common references than those with a single common reference (Table 1). Therefore, as the number of common references between two authors increases, the number of citations between them also increases. Table 1 shows the increase trend; however, the correlation fluctuated as the number of common references increases but tends to increase. To put it in another way, when the bibliographic coupling strength is stronger between two authors, they tend to cite each other more often. Author bibliographic coupling strength shows how strongly two authors are intellectually related. So, more intellectually related authors cite each other more often. This result confirms the normative theory of citation holding the view that authors cite relevant works, and citations reflect scientific merit and quality.

Table 1. Spearman correlation between ABC strength and number of citations in IS&LS.

No of common refs	Spearman correlation
1	0.31
10	0.36
20	0.35
30	0.38
40	0.37
50	0.37
60	0.39
70	0.38
80	0.36
90	0.39
100	0.4
150	0.46
200	0.47
250	0.58
300	0.61

ABC strength and citation relationship among thirty highly cited authors in imetrics

Thirty highly cited authors in imetrics identified in Abrizah and colleagues (2014) were chosen for a closer investigation of research goals. The main research question on the association between ABC strength and number of exchanged citations was also examined for this group of highly cited authors. Spearman correlation test shows a strong positive association between the number of common references and the number of citations between the authors (Spearman's rho= 0.771, p-value< 0.001), once more confirming the significance of content relevance in citation behavior and normative view of citations.

Moreover, all ABC relations are mapped between each pair of highly cited authors (See Fig. 1). Based on the results, all thirty authors are in BC relationships with all or some of other authors in the network except for Griffith, BC. During 1990-2012, he has published 4 papers in imetrics and has no common references with any of the highly cited authors.

Thelwall, M. is in strong BC relationships with all other authors except with Vanleeuwen, T.N. (only one common reference) and VanRaan, A.F.J. (three common references). He and Bar-Ilan, J have shared the highest common references in the network (4,527 common references) and they have exchanged a large number of citations in the network (118 citations). Thelwall, M. has more than 100 common references with 18 authors in the network. He is also in a strong BC relationship with Vaughan, L. (2,725 common references). Thelwall, M. has also exchanged the highest number of citations in the network with Vaughan, L. (195 citations). He has also strong BC ties with seven others, Leydesdorff, L., Ingwersen, P., Rousseau, R., Cronin, B., Glänzel, W., and Egghe, L., respectively.

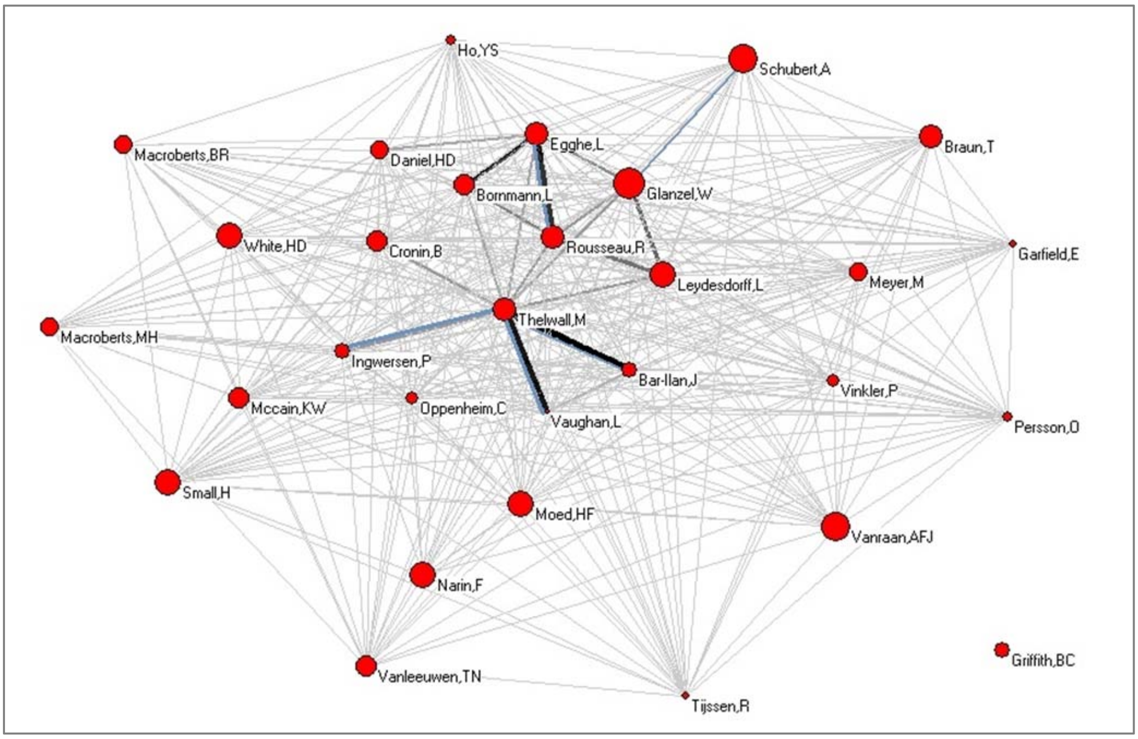


Figure 1. ABC among highly cited authors in imetrics; the black lines show ABC relations and the width of the lines shows ABC strength between pairs of authors; the blue lines show the strongest citation relations in the network and the width of the lines shows the number of citations exchanged between pairs of authors; the size of vertices shows the number of other highly cited authors in the network that each author is in an ABC relation with.

Another strong ABC relationship, and also citation relationship, is seen between Rousseau, R and Egghe, L. (2,270 common references and 175 exchanged citations). Rousseau, R is also in strong BC relationships with other authors in the network. He has strong BC ties with Leydesdorff, L., Bornmann, L., Glänzel, W., and Thelwall, M., respectively.

Glänzel, W., Bornmann, L., Bar-Ilan, J., and Leydesdorff, L. are also in strong BC relationships with other authors in the network. They also have strong citation relationships with each other as well as other highly cited authors.

The correlation between ABC strength and citation exchange in imetrics in comparison with IS&LS

The correlation between the number of common references and the number of citations for top thirty imetricians was examined first amongst themselves and then between them and all other authors in IS&LS with whom they are in BC or citation relationships. As shown in Figure 2, a stronger relationship exists between the authors in the first group than in the second one and regarding the top thirty imetricians, the correlation varies from one author to another one.

For each highly cited imetrician, the proportions of common references with each in-group authors was estimated. Fig. 3 shows that each highly cited author is in a BC relationship with 27 other in-group authors. For example, about 24% of references of each author are common with one other author. The author distribution of the number of common references with other authors demonstrates a core-scatter shape.

Core references in imetrics

We tried to go further than author couples for common references and identified a number of common references between three and more authors. The thirty highly cited authors in imetrics were examined for this purpose.

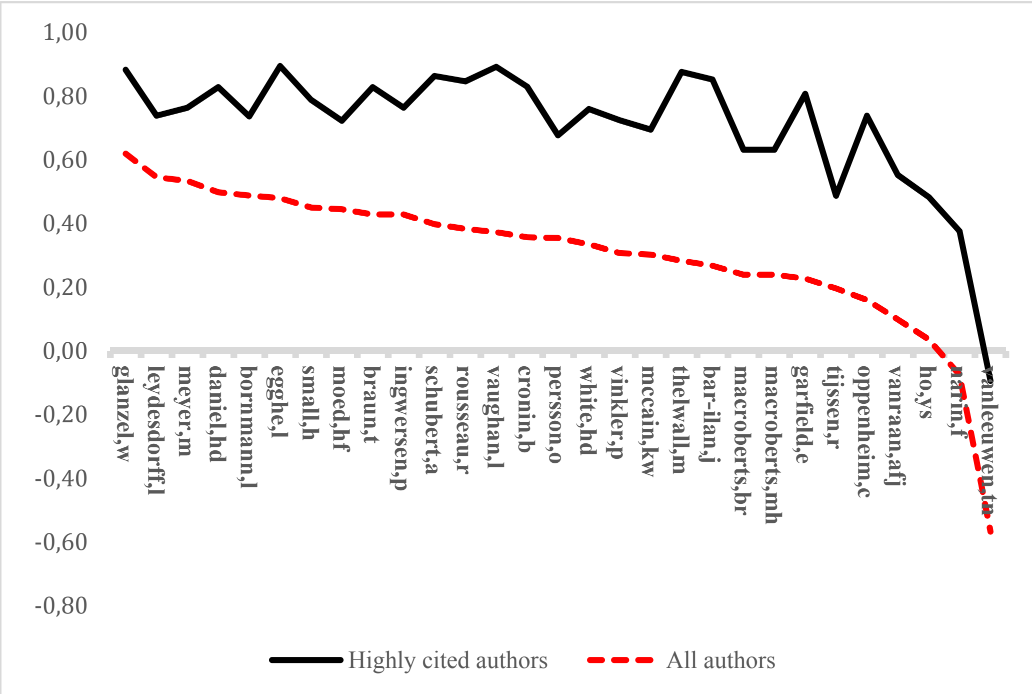


Figure 2. ABC strength and citation correlation between highly cited authors and all authors in IS&LS.

The interesting result is that seventeen highly cited imetricians have one reference in common. The common reference is Hirsch's paper on H-index (Hirsch, J.E. (2005): An index to quantify an individual's scientific research output. *Proceedings of the national academy of sciences of the United States of America*, 102 (46)). Egghe, L., Rousseau, R., and Bornmann, L. have cited this paper more than thirty times in their publications showing that the H-index is one of their common research interests. It is interesting to note that Egghe, L. and Rousseau, R also have the strongest citation relationship with each other in the network (seventeen5 citations have been exchanged between them) and these two imetricians are also in a strong citation relationship with Bornmann, L. with Bornmann, L. being the fourth top author in citation relationships with both Egghe, L. and Rousseau, R. The strong citation relationships between these authors are mainly due to their similar research interests, one of which is H-index. Twelve highly cited authors have simultaneously five references in common which are listed in Table 2. Eleven authors have nine references and ten authors have eleven references in common.

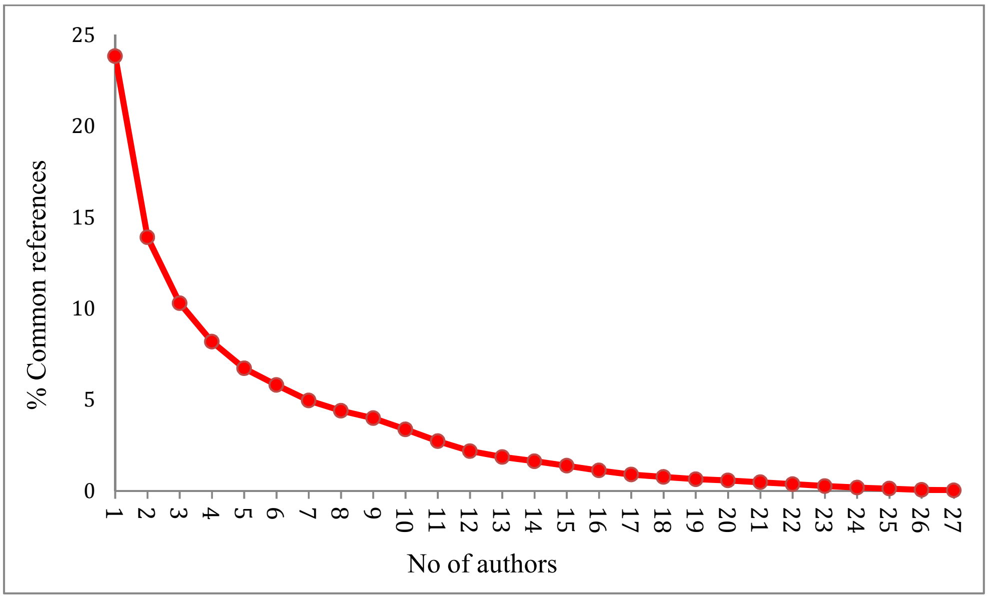


Figure 3. The proportion of common references between each of thirty highly cited imetricians and other in-group authors.

Table 2. Five common references between twelve highly cited imetricians.

VanRaan, A.F.J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, 67(3).
Meho, L. & Cronin, B. (2006). Using the h-index to rank influential information scientists. *JASIS&T*, 57(9).
Glänzel, W., Thijs, B., & Schlemmer, B. (2003). Better late than never? On the chance to become highly cited only beyond the standard bibliometric time horizon. *Scientometrics*, 58(3).
Macroberts, B.R. & Macroberts, M.H. (1996). Problems of citation analysis. *Scientometrics*, 36(3).
Moed, H.F., Vanleeuwen, T.N., & Debruin, R.E. (1995). New bibliometrics tools for the assessment of national research performance- database description, overview of indicators and first applications. *Scientometrics*, 33(3).

Discussion and conclusion

This study examined the association between author bibliographic coupling strength and the number of times authors cited each other. The results of the study on authors in IS&LS showed that there is a positive and significant correlation between ABC and exchanged citations between two linked authors confirming that authors are citing related authors and relevant research works in their field (Table 1). This finding opposes the social constructivist view holding that authors cite others for some other reasons than relevance or rewarding the cited author, but it confirms the normative theory of citations. A group of thirty highly cited authors in imetrics were also examined for this purpose. The result of the association between ABC and the number of citations shows a positive strong correlation between ABC and exchanged citations between imetricians. Therefore, highly cited authors in imetrics are in strong BC relationships with whom they also have strong citation relationships.

The number of common references between pairs of authors was accepted as a measure of relatedness between them. Therefore relatively, the higher number of common references between two authors, especially in a long-term period, could show the extent to which they are working in similar research areas; however, authors may change their research interests over time due to changes in the research fields. The higher number of citations between two authors with higher number of common references, when they are not co-authors, could probably show that they cite each other since they may work on similar research areas and also for the matter of relevancy.

ABC relations between the thirty highly cited imetricians were examined and mapped and strong relationships were determined. Thelwall, M. and Bar-Ilan, have the strongest ABC relationship in the network; they are also in a strong citation relationship. Rousseau, R., Glänzel, W., Bornmann, L., Bar-Ilan, J., and Leydesdorff, L. are also in strong ABC relations with each other as well as other authors in the network. In an investigation of the number of common references in groups of two and more imetricians, smaller groups have a larger number of references in common while larger groups have fewer numbers of common references. For example, seventeen imetricians have only one reference in common while some two-author groups have more than a thousand common references. The latter groups presumably work on narrow research areas. Larger groups with fewer number of common references suggest membership in a wider research area. The results show that a maximum of seventeen authors have one reference on H-index in common. Authors citing this single paper are also in strong citation relationship with each other.

Comparing the correlation between number of common references and number of exchanged citations for highly cited imetricians and all authors in IS&LS related to Fig. 2 shows that number of common references between imetricians increases the probability of higher citations between them more than that of IS&LS. Moreover, ABC relationship or common references with each single author may result in different number of citations with him/her.

Intuitively, considering the core-scatter distribution of citations to papers in the science network, an author probably has common references with a large number of other authors, while he/she probably has more common references with a fewer number of other authors (Fig. 3). The author presumably has more related research interests with the latter group of authors where some of them may belong to the same research community.

The number of common references and citations between pairs of authors could be also influenced by the number of papers published by the authors. For example, two authors may have five common references whilst the first author only published a single paper during his/her entire research life and the second one published more than twenty papers. The first author will have fewer common references with any other authors in the field than the second author and he/she will have less opportunity to cite other authors due to his/her short research life. So authors' research lifetime in the science network (e.g. newcomers, students, faculty members and professional researchers) does matter. Authors with a longer research life have more chances to know other researchers in similar research fields and they also have extra opportunities to focus on more specific and narrow research topics, compared to authors with a shorter research lifetime. Hence, a stronger association between the number of common references and citations exchanged between authors is found for the former group.

Science network and its attributes are continuously changing over time and a research specialty may appear or disappears after a while; authors may also change their research interests during their research lifetime. In the current study, a longer time span is used to show that clustering authors, based on more recent common references, may be replaced by a shorter one, which could result in a stronger relationship between the bibliographic coupling network and the citation network. According to the results of current studies, authors with a longer research lifetime and more citations demonstrate a stronger relationship between their number of common references and citations. However, even weak ties in bibliographic coupling networks could also be used for research front detection purposes. Bibliographic coupling is not enough for mapping intellectual structure of science and measuring relatedness by itself. Thus, as with previous studies, it is better to be combined with other methods, such as co-citations, to realise better results.

References

Abrizah, A., Erfanmanesh, M., Rohani, V. A., Thelwall, M., Levitt, J. M., & Didegah, F. (2014). Sixty-four years of informetrics research: productivity, impact and collaboration. *Scientometrics*, *101*(1), 569-585.

Baldi, S. (1998). Normative versus social constructivist processes in the allocation of citations: a network-analytic model. *American Sociological Review*, *63*, 829-46.

Boyack, K. W., Börner, K., & Klavans, R. (2009). Mapping the structure and evolution of chemistry research. *Scientometrics*, *79*(1), 45-60.

Byun, J-H & Chung, E-K (2011). Domain Analysis on Electrical Engineering in Korea by Author Bibliographic Coupling Analysis. *Journal of Information Management*, *42*(4), 75-94.

Didegah, F., & Gazni, A. (2011). The extent of concentration in journal publishing. *Learned Publishing*, *24*(4), thirty3-310.

Gazni, A. & Thelwall, M. (2014). The long-term influence of collaboration on citation patterns. *Research Evaluation*, doi: 10.1093/reseval/rvu014.

Gilbert, G. N. (1977). Referencing as persuasion. *Social Studies of Science*, *7*, 113-122.

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

Jarneving, B. (2005). Acomparison of two bibliometric methods for mapping of the research front. *Scientometrics*, *65*, 245–263.

Jarneving, B. (2007). Bibliographic coupling and its application to research front and other core documents. *Journal of Informetrics*, *1*, 287–307.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American documentation*, *14*(1), 10-25.

Kuusi, O., & Meyer, M. (2007). Anticipating technological breakthroughs: Using bibliographic coupling to explore the nanotubes paradigm. *Scientometrics*, *70*(3), 759-777.

Ma, R. (2012). Author bibliographic coupling analysis: A test based on a Chinese academic database. *Journal of Informetrics*, *6*(4), 532-542.

MacRoberts, M. H. & MacRoberts, B. R. (1987). Another test of the normative theory of citing. *Journal of the American Society for Information Science*, *38,* 305-306.

Merton, R. K. (1973). The normative structure of science. The sociology of science: Theoretical and empirical investigations, 267.

Merton, R. K. (1988). The Matthew effect in science II: cumulative advantage and the symbolism of intellectual property. *Isis*, 79, 606-23.

Morris, S. A., Yen, G., Wu, Z., & Tesfaye, B. (2003). Time line visualization of research fronts. *Journal of the American Society for Information Science and Technology*, *54*(5), 413–422.

Peters, H. P., Braam, R. R., & van Raan, A. F. (1995). Cognitive resemblance and citation relations in chemical engineering publications. *Journal of the American Society for Information Science*, *46*(1), 9-21.

Qiu, J. P. (2007). *Methods of citation analysis*. In Informetrics. Wuhan: Wuhan University Press.

Rousseau, R. (2010). Bibliographic coupling and co-citation as dual notions. In L. Birger (Ed.), *A Festschrift in honour of Peter Ingwersen, special volume of the e-zine of the ISSI, June 2010*, 173–183.

Small, H. (1978). Cited documents as concept symbols. *Social Studies of Science*, *8*(3), 327-340.

Small, H. (2004). On the shoulders of Robert Merton: towards a normative theory of citation. *Scientometrics*, *60*, 71-79.

Soós, S. (2014). Age-sensitive bibliographic coupling reflecting the history of science: The case of the Species Problem. *Scientometrics*, *98*(1), 23-51.

Van Raan, A. F. (2005). Reference-based publication networks with episodic memories. *Scientometrics*, *63*(3), 549-566.

Yan, E., & Ding, Y. (2012). Scholarly network similarities: How bibliographic coupling networks, citation networks, co-citation networks, topical networks, co-authorship networks, and co-word networks relate to each other. *Journal of the American Society for Information Science and Technology*, *63*(7), 1313-1326.

White, H. D. (2004). Reward, persuasion, and the Sokal Hoax: a study in citation identities. *Scientometrics*, *60*(1), 93-120.

Zhao, D., & Strotmann, A. (2008). Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic-coupling analysis. *Journal of the American Society for Information Science and Technology*, *59*(13), 2070–2086.

Zhao, D., & Strotmann, A. (2014). The knowledge base and research front of information science 2006–2010: An author co-citation and bibliographic coupling analysis. *Journal of the Association for Information Science and Technology*, *65*(5), 995-1006.