Informetrics, bibliometrics, altmetrics: What is it all about?

Isabella Peters
ZBW Leibniz Information Center for Economics
& Christian Albrechts University Kiel, Germany
I.Peters@zbw.eu

Judit Bar-Ilan
Department of Information Science
Bar-Ilan University, Israel
Judit.Bar-Ilan@biu.ac.il

ABSTRACT
The aim of this SIG/MET-sponsored panel is to discuss major informetric topics including the impact factor, the h-index, sources of citation data, the Eigenfactor, the making and use of base maps of science, application of informetrics (e.g., for retrieval purposes), altmetrics, and future perspectives on bibliometrics. The panel especially addresses attendees who want to expand their knowledge in this area or got in touch with it only recently.

Keywords
Informetrics, data sources, h-index, impact factor, altmetrics, visualizations.

INTRODUCTION
Informetrics is one of the major subfields of information science, as can be seen in visualizations (Zhao & Strotmann, 2008). However this subfield is not well represented in the ASIS&T Annual Meetings (ASIS&T AM). Although SIG/MET, the special interest group for the measurement of information production and use, has held highly successful pre-conference workshops at the ASIS&T conferences in the last few years, the workshop is more oriented towards researchers already working in this area which is also reflected by the majority of attendees coming from that field. Thus, the lion's share of participants of the ASIS&T AM is hardly exposed to informetric topics.

In spite of this, ASIS&T AM participants, researchers, librarians and information professionals all need basic informetric knowledge. Librarians and information professionals are often faced with decisions regarding subscriptions or cancelling subscriptions. Informetric data can provide support for such decision. Although born out of a problem of collection formation and now belonging to the fundamentals of informetrics, Bradford's law about the distribution of topics about journals is nowadays mostly used in Information Retrieval, so that the link to informetrics is somewhat forgotten. More visible, however, is the connection between informetrics and research evaluation. Researchers are often evaluated based on the number of citations their works receive and on the venues in which their work was published. Increasingly, librarians in their role as providers for digital literacy and data stewardship are also confronted with consultancy in this area. But, not always are they familiar with the actual state-of-the art in critical reflection in informetrics. Last but not least, newly emerging web-based collaborative platforms - such as Mendeley, ResearchGate, FactShare - provide altmetrics (like reader statistics), create social media awareness among researchers and build expert social networks which are also analyzed from the bibliometrics community in order to understand their usefulness for research evaluation (Moed & Halevi, 2014). Although those social media-based numbers are easily retrievable, all too often the user is unaware of the meaning behind them.

This panel addresses all of those aspects by providing a basic understanding of informetric concepts, giving an overview on the different facets of informetrics and its relevance for ASIS&T members. We will especially discuss some of the major topics of general interest to the audience since these are the concepts ASIS&T AM attendees are most often confronted with:

- Sources of informetric analyses
    - Coverage and quality of Web of Science, Scopus, and Google Scholar
    - Social media data
- Application of informetrics
    - Misuse and problems of accessing data
    - Need for local databases to construct meaningful metrics
    - Use of informetric principles in information retrieval and library settings (e.g., collection enhancement)
- Data aggregation and cleaning

ASIST 2014,November 1-4, 2014, Seattle, WA, USA.

    - Main aggregation levels for bibliometric analysis (i.e. author, institution, country)
- Basic, normalized, weighted, and altmetric indicators and their shortcomings (e.g., impact factor, h-index, Eigenfactor)
- Visualization of informetrics analyses (e.g., maps of science)

A major issue when computing different indexes is the data source or sources being used. Citations reported in a specific database come from the references of items indexed by that database. Thus coverage becomes a major issue. Different databases report different values of h-index, as was shown by Bar-Ilan (2008), but the same is true when computing other indicators as well. This is becoming a major issue as Google Scholar becomes more and more widely used for reporting citations. In some fields, Google Scholar citation counts are considerably higher than Web of Science (WoS) or Scopus counts (which are usually more or less comparable).

Providing access to resources and assisting during the evaluation of their relevance is also a crucial role of information providers (Wolfram, 2003). Informetric principles, like Bradford's Law of Scattering (Bradford, 1934), can cater valuable strategies for deciding which resources to subscribe to or how to rank a results list in a retrieval system. For example, Bradford found that journals fall in three different sets (i.e. core, middle, border) in regard to topical relevance which means that researchers have to search in a significantly greater amount of border journals to find the same amount of relevant articles provided by core journals. The applying law is described as $1:n:n^2$, where n is the number of journals in each set. That finding often has practical implications for the acquisition of library resources, but may also result in a "filter bubble" over-representing majority views (when only subscribing to core journals).

To conduct informetric studies researchers also need profound knowledge about how to access diverse data sources as well as how to download and handle data appropriately (e.g. in terms of author disambiguation) so that meaningful results could be derived. Since research assessment is mainly carried out on the number of publications and the number of citations to publications the aggregation of data on different levels (e.g., author, journal, country) is needed to get a broader view on the research landscape and enable comparisons between items of the same level of aggregation.

Further forms of data aggregation can be found in bibliometric indicators (from which we just mention the most prominent):

The idea of the journal impact factor (JIF or IF) was first conceived by Eugene Garfield in 1955 and initially developed as a size-independent measure of citation impact by scientific journals (Garfield, 1955; 1972; 2006). The impact factor of a journal in a given year is the number of citations items published in the journal in the two preceding years received during the given year, divided by the number of articles published in the journal during the two preceding years. The Journal Citation Report (JCR) was first launched in 1975 and it reports the impact factors of journals indexed by the Thomson Reuters Web of Science (originally the ISI Citation Indexes) (Garfield, 2007). No impact factors are reported for journals in the Arts and Humanities. The impact factor reflects the average number of citations an article receives during a one year period, about one and half years after its publication. Citation rates of articles published in journals are highly skewed, thus the average is not the best possible measure of a journal importance. This is just one of many criticisms of the JIF, for example the JIF is often (mis)used to measure the impact of an article and not of the journal. For an extended discussion of some of the problems with the JIF, see Archambault & Larivière (2009). New measures, like the Eigenfactor, the Article Influence (Bergstrom, 2007), the SJR (González-Pereira, Guerrero-Bote & Moya-Anegón, 2010) and the SNIP (Moed, 2010) have been suggested, but the JIF is still by far the most widely used metric. Some of the concerns regarding the impact factor have been recently raised in the San Francisco Declaration on Research Assessment (DORA, 2013) and as of the beginning of April 2014 has been signed by 10,419 individuals and 447 organizations.

A much newer, but still controversial measure is the h-index, devised for the evaluation of individual researchers. The h-index was introduced in 2005 by the physicist, Jorge Hirsch. The definition says that an author had h-index h, if he or she has h publications where each publication has been cited h times or more, but he or she does not have h+1 publications, such that each publication has been cited at least h+1 times. This new index caused great interest in the informetric community, its plusses and minuses were extensively discussed (Bornmann & Daniel, 2007) and many variations of it have been proposed (Alonso, Cabrerizo, Herrera-Viedma & Herrera, 2009). Its main shortcomings, the oversimplification of the two orthogonal dimensions of publication output and citation impact and its inconsistency (Waltman & van Eck, 2012), does not keep the indicator from being used.

Altmetrics are the newest set of metrics in the "informetrics soup", and their aim is to capture "impact" that is not citation based, e.g. counting the number of followers, the number of tweets or the number of readers of an article on a reference manager (Priem, Taraborelli, Groth & Neylon, 2011). This raises many interesting questions about the validity and applicability of these newly proposed measures.

Taking the same line, research on the relationships between citations and downloads has expanded in various studies attempting to understand the relationship between the two as usage phenomenon and as a way to measure research impact. The base assumption is that a sound statistical

analysis of relationship between downloads and citations, and a thorough reflection upon its outcomes, contributes to a better understanding of what both download counts and citation counts measure, or more generally, to more insight into information retrieval, reading, and referencing practices in scientific-scholarly research.

When it comes to making sense of publication and citation data and informetric studies proper displays and visualizations can support that process (Börner, 2010). Especially the visualization of links between different informetric data can give valuable insights into principles of scholarly communication (e.g., citation networks), knowledge flows (e.g., which journals are frequently cited in various disciplines) and development of research topics (e.g., topic analyses). Although graphics provide easy access to results the appropriate interpretation of those visualizations can be challenging.

ISSUES TO BE DISCUSSED AND PANEL FORMAT

The aim of this panel is to discuss major informetric topics including the impact factor, the h-index, sources of citation data, the Eigenfactor, the making and use of base maps of science, application of bibliometrics, altmetrics, and future perspectives on bibliometrics. Reasonable applications, newly emerging alliances with other information services and pitfalls of bibliometric analyses will also be presented. The panel is particularly aimed at the general audience without extensive informetric knowledge. Some of the topics to be discussed are highly controversial, and will allow the panelists to argue both the pros and the cons. We expect the audience to take an active role in these discussions. The panelists will introduce their main arguments on two bibliometric use cases on the website http://www.asis.org/SIG/ SIGMET/activities/panel2014. We invite ASIS&T AM attendees to get in touch with panelists, discuss aspects or post questions of general interest on that website. SIGMetrics will use several channels to promote this activity, e.g., the highly active mailing lists of the ASIS&T community. We hope that we can foster more audience-related discussions in Seattle which will even include perspectives of people who are not able to attend. As such the panel's accompanying website will provide a platform for exchange and will transport bibliometrics topics to the conference and back to the broader audience.

PANELISTS

Panel organizer

Isabella Peters is professor at the ZBW Leibniz Information Center for Economics and Christian Albrechts University Kiel where she is conducting research on user-generated content in digital libraries and on altmetrics. She will moderate panel and website and will infuse the panel discussion with the perspectives discussed on the panel's accompanying website. Isabella will also take charge of reporting on the panel and publishing a summary on the website to keep the audience informed.

Panelists (in alphabetical order)

Judit Bar-Ilan is professor at the Department of Information Science at Bar-Ilan University. She has a solid bibliometric background and will offer her insights on the use of different citation sources for bibliometric analysis. In the past almost all citation based analyses were conducted using the ISI databases (now Thomson Reuters' Web of Science), but now there are additional possibilities, including Elsevier's Scopus and Google Scholar. The differences between the citation sources and the plusses and minuses of each of them will be discussed. Additional citation sources will also be mentioned.

Gali Halevi is a senior analyst and program director at Elsevier. She is responsible for the development of informetric tools for Elsevier's leading products, is the director of their Bibliometric Indicators Research Program and a contributor and member of the editorial board of Research Trends. In her presentation she will provide a collection of research examples highlighting the potentialities and effects of big data on bibliometrics specifically and research assessment in general. She will illuminate newly available datasets and their potentialities for bibliometric research which in turn can be applied to scientific evaluation metrics. Research using large collections and combinations of data open new avenues to analyze scientific output as well as measure and track scientific trends. Several examples are given demonstrating such potentialities taken from different approaches all utilizing potential big datasets.

Stefanie Haustein is a post-doctoral researcher at the Canada Research Chair on the Transformations of Scholarly Communication at the University of Montreal. Her doctoral work focused on the multidimensional evaluation of scholarly journals. She currently works on social media in scholarly communication and making sense of so-called "altmetrics", supported by a grant from the Alfred P. Sloan Foundation. In the panel Stefanie will discuss the inability of a single metric to reflect scholarly impact of researchers, institutions, journals or countries in general. In particular, she will focus on two of the most used and misused bibliometric indicators, the journal impact factor and the h-index. Examples of the shortcomings will be provided, such as the inconsistency of the h-index and the asymmetry between numerator and denominator of the impact factor, which goes back to convenience and cost-efficiency decisions made in the 1960s. It will be shown that single measures are not able to capture the complexity and multiple dimensions of research activity but that multiple metrics are needed to reflect publication output and impact. An outlook on altmetrics as a potential source of broader impact and the challenges involved in providing meaningful metrics will be given.

Andrea Scharnhorst is Head of Research & Innovation at the Data Archiving and Networked Services (DANS) institution in the Netherlands - a large digital archive for research data primarily from the social sciences and

humanities. She is the chair of a European collaboration networks KnoweScape in which information professionals, physicists, sociologists, designer and computer scientists seek to understand the complex nature of knowledge spaces, from library collections over bibliographic databases to Wikipedia. Andrea will discuss the links between information retrieval and bibliometrics/scientometrics and create a common ground for the incorporation of bibliometric-enhanced services into retrieval at the digital library interface, e.g. statistical modelling of scholarship, such as Bradfordizing or network analysis of coauthorship networks.

Jevin West is an Assistant Professor in the iSchool at the University of Washington. He builds models, algorithms and interactive visualizations for improving scholarly communication and for understanding the flow of information in large knowledge networks. Jevin co-founded Eigenfactor.org, a free website and research platform that librarians, administrators, publishers and researchers use to map science and identify influential journals, papers and scholars. As De Solla Price noted, the scholarly literature forms a vast network, where the nodes are the millions of papers published in scholarly journals and the links are the hundreds of millions of citations connecting these papers. New approaches to clustering and visualizing large networks make it possible to map citation networks at unprecedented spatial and temporal scales. In this presentation, Jevin will talk about the opportunities and challenges of these new metrics and maps. In particular, he will talk about how these metrics and maps can be used to better navigate the ever expanding literature.

ACKNOWLEDGMENTS

ASIS&T SIG for the measurement of information production and use (SIG/MET) sponsors this panel.

REFERENCES

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

Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). h-index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, 3(4), 273-289.

Bar-Ilan, J. (2008). Which h-index? A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257-271.

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

Börner, K. (2010). *Atlas of science: Visualizing what we know*. Cambridge, Mass: MIT Press.

Bornmann, L., & Daniel, H.-D. (2007). What do we know about the h index? *Journal of the American Society for Information Science and Technology*, 58(9), 1381-1385.

Bradford, D. C. (1934). Sources of information on specific subjects. *Engineering*, 137, 85-86.

DORA (2013). *San Francisco Declaration of Research Assessment. Putting science into the assessment of research*. Retrieved April 30, 2014 from http://ar.ascb.org/sfdora.html

Garfield, E. (1955). Citation indexes to science: A new dimension in documentation through association of ideas. *Science*, 122, 108-111.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. Journals can be ranked by frequency and impact of citations for science policy studies. *Science*, 178(4060), 471–479.

Garfield, E. (2006). The history and the meaning of the journal impact factor. *JAMA*, 295(1), 90-93.

Garfield, E. (2007). The evolution of the Science Citation Index. *International Microbiology*, 10, 65-69.

González-Pereira, B., Guerrero-Bote, V. P., & Moya-Anegón, F. (2010). A new approach to the metric of journals' scientific prestige: The SJR indicator. *Journal of Informetics*. 4(3), 379-391.

Hirsch, J. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 102(46), 16569-16572.

Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, 4(3), 265-277.

Moed, H. F., & Halevi, G. (2014). Research assessment: Review of methodologies and approaches. *Research Trends*, 36, 3-5.

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2011). altmetrics: A manifesto. Retrieved April 30, 2014 from http://altmetrics.org/manifesto/

Waltman, L., & van Eck N.J. (2012). The inconsistency of the h-index. *Journal of the American Society for Information Science & Technology*, 63(2), 406-415.

Wolfram, D. (2003). *Applied informetrics for information retrieval research*. Westport, Conn: Libraries Unlimited.

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.