Bibliometrics to Altmetrics and its impact on Social Media

Satish Kumar*, Anil Kumar Mishra**,

*Information Scientist, ARIES , Nainital, UK, Email – sklisc@gmail.com
**Senior Technical Assistant 'B', CVRDE, DRDO, Avadi, Chennai, Tamil Nadu, Email – anilmlis@gmail.com

ABSTRACT

In the age of Web 2.0, individuals are strongly connected in the virtual worlds and form diverse online communities. Recently, the transmission of information in these online communities has gained much attention, partly because of the enormous popularity of online social networking sites and their potential marketing impact.

The social media tools available on the internet assist the encouragement of science dramatically, and the change is happening in real time models. Expansion of World Wide Web and newer technologies has improved the ways in which science is communicated and its assessment techniques. The use of conventional metric tools judges the impact of scholarly publications using citation and download counts. They are broadly used to evaluate articles, authors, and disciplines on publishers' platforms. These tools analyze the citation data in normal course of time and make them available.

A barrier to spreading of research is that it depends on the end-user searching for or 'pulling' relevant knowledge from the literature base. Social media instead 'pushes' relevant knowledge directly to the end-user, via blogs and sites such as Facebook and Twitter. The social media is very effective at improving and spreading seems well accepted, but, remarkably, there is no evidence to support this claim.

Key words: *Bibliometrics, Scientometrics, Webometrics, Atlmetrics.*

INTRODUCTION

Over the last two decades, the WWW has revolutionized scientific research, in particular by speeding up the rate of the spread of information. Nowadays, once a paper is electronically published on a journal website, the information can disseminate rapidly in the community, partially due to various scientific blogs and folksonomy websites like CiteULike and Connotea. The propagation of a paper will then be reflected at the level of web usage statistics, in particular, the number of HTML views, i.e. the WWW traffic of the webpage corresponding to the paper.

The impact of research is fundamentally dependent on how well it is disseminated to the end-user. Conventional routes of dissemination involve journal publications, conference presentations and, ultimately although often years later, textbooks. This model of dissemination requires the end-user to search for, or 'pull', the relevant knowledge from the literature base[1].

The rise of open access publication reduces one barrier to effective dissemination by making literature freely available for all who wish to consult it, but it still relies on the end-user pulling out the relevant knowledge[2–3]. The rapid rise in popularity of

web logs (blogs) and social media sites such as Facebook and Twitter, has positioned them as important tools with which to support dissemination.

By having different blog and social media sites, journals allow the end-user to self-select the sort of knowledge they wish to receive. **RSS (Really Simple Syndication)** is another example of how users can self-select information. Although not a pure social media tool, RSS feeds enable the pushing of individualized information and blog contents. RSS permits some user interaction and information sharing.

The fundamental importance of a digital strategy is emphatically stressed by social media promoters[4]. Markers such as the number of 'likes', or the number of Facebook or Twitter followers are cited as measures of research impact, collectively captured by concepts such as '*altmetrics*'[5].

2.0 What is Altmetrics

Altmetrics is the study and use of scholarly impact dealings based on activity in online tools and environments. The term has also been used to describe the metrics themselves–one could propose in plural a ''set of new altmetrics.'' Altmetrics is in most cases a subset of both *scientometrics* and

webometrics; it is a subset of the latter in that it focuses more closely on scholarly influence as measured in online tools and environments, rather than on the Web more generally.

Moreover, the definitions of various terms are not clear and they mean different things to different people. *For instance*, we took 'reach' to be the number of people who have been alerted to the presence of a web page and have the opportunity to view it[6]. Here 'reach' reflects the number of people who could potentially see the blog, either directly because they subscribe to the blog through RSS feed or email alerts, or through following the blog on various social media sites. *For example* Facebook, Twitter, LinkedIn, Google+, Research Blogging etc.. One step closer to impact is engagement, defined here as the number of people who view the web page and then do something in response to viewing it – *for example* they 'like' it, re-tweet it, or they share it with their friends. The concept of '*virality*' attempts to capture a stronger level of engagement and a reflection of the propensity of the message to 'go viral'. Here we use the percentage of engagers who then write a story on the post on Facebook or begin a new tweet. This distinction between terms is important because as few as 16% of Facebook followers actually read a new post and about 1% of people who see and 'like' a Facebook page actually comment on it or start a new story on it[7–8].
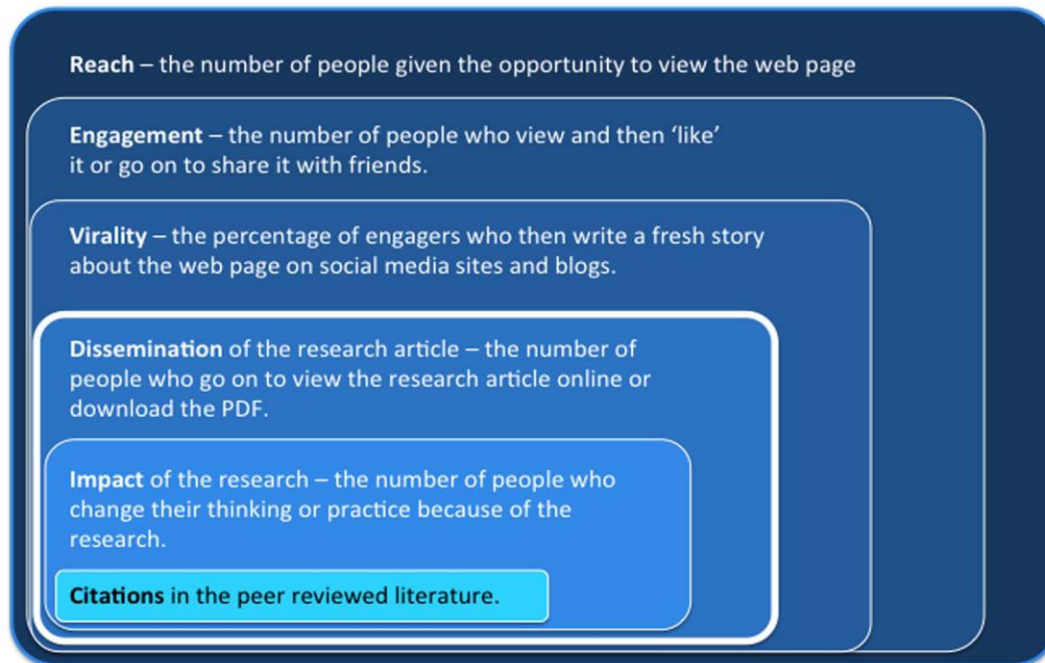
Figure 1

Some of the metrics used for altmetrics are following -

- **Blogs:** The blog (feed) citations are from a manually-curated list of about 2,200 science blogs, derived from the indexes at Nature.com Blogs, Research Blogging and ScienceSeeker.

- **FbWalls:** A licensed Facebook account is used for Wall posts to check for citations.

- **Tweets:** Tweets from a licensed Twitter handle are checked for citations.

- **Google+:** The Google+ Applications Programming Interface (API) is used to identify Google+ posts to check for citations.

- **RH:** Research highlights are identified from Nature Publishing Group journals.

- **MSM:** The ***mainstream media*** citation count is based on a manually curated list of

about 60 newspapers and magazines using links in their science coverage.

- **LinkedIn:** LinkedIn.com posts from the LinkedIn API are checked for citations.

- **Reddits:** Reddit.com posts from the Reddit API are checked for citations.

- **Forums:** Forums are scraped for citations.

- **Pinners:** Pinterest.com is scraped for citations.

- **Q&A:** The Stack Exchange API and scraping of older Q&A using the open source version of Stack Exchange's code are used to get online questions and answers to check for citations.

3.0 Bibliometrics to Altmetrics

The idea of libraries as a collection based centers of metric support goes back to at

least the 1970s, when **Thomson Reuters** made its **Impact Factor** metric available to scholars through **Web of Knowledge.** By serving mediator institutional access to such proprietary tools and metrics, librarians at many universities have provided tenure—track faculty with access to electronic resources while at the same time implicitly or explicitly promoting citation—based impact paradigms.

Altmetric measurements derived from the social web are gradually more advocated and used as early indicators of article impact and usefulness. However, there is a lack of systematic scientific evidence that altmetrics are valid proxies of either impact or utility whereas a few case studies have reported medium correlation between specific altmetrics and citation rates for individual journals or fields.

4.0 General resources

4.1 Google Scholar Citations

This free Google service allows authors to create profiles that manage, calculate, and track citation data such as h-index and i10-index (i.e. number of articles with at least ten citations). Using a statistical model based on author and article metadata to identify related citations, the service offers the option of automatically adding new articles to users' public or private profiles. Google also recently launched a related service, **Google Scholar Metrics**, that gauges the "visibility and influence" of articles and publications from 2007 to 2011, based on Google Scholar citation data.
Access:
http://scholar.google.com/intl/en/scholar/citations.html

4.2   Web of Knowledge

This Thomson Reuters subscription database helped usher in modern bibliometrics with its introduction of the h-index in 1982. Web of Knowledge includes Web of Science, for article and author queries, and Journal Citation Reports, for journal queries. Its metrics include times cited, h-index, impact factor, Eigenfactor, and field-based journal rankings. While many of these metrics have been criticized for not fully representing scholarly value in certain disciplines, they are still considered the gold standard in traditional bibliometrics.
Access: http://www.webofknowledge.com

4.3 Scopus

Scopus is a subscription database known primarily as an option to **Web of Knowledge**, as it offers similar article, author, and journal-level metrics, but uses slightly different algorithms to calculate them. Metrics include standard options such as times cited and h-index, as well as original aid like SJR and SNIP from SCImago. Scopus recently launched "**Altmetric for Scopus**," a third party application that runs within the sidebar of Scopus pages to track mentions of papers across social media sites, science blogs, media outlets, and reference managers.
Access: http://www.scopus.com

4.4 SCImago Journal and Country Rank

SCImago is a free Web site that runs on **Scopus** data to calculate two metrics: **SCImago Journal Rank** (SJR) and **Source Normalized Impact per Paper** (SNIP), which compare directly to Web of Knowledge's Impact Factor. SJR is based on times cited, but also uses an algorithm similar to Google's Page Rank to calculate article influence, which it uses to create rankings.

Using SCImago's online interface, users can compare rankings of up to ten journals at a time, display top journals, and even display countries with influential journals in a discipline.   *Access:* http://www.SCImagojr.com

4.5 Publish or Perish

Anne-Wil Harzing created Publish or Perish (PoP) to assist faculty looking for more diverse bibliometrics. PoP is a free, downloadable program that harvests data from Google Scholar based on author name. Users can manually remove records to refine the data, similar to what is now offered by Google Scholar Citations. PoP can also calculate numerous metrics, including alternatives to the h-index. However, because few people are familiar with non h-index calculations, it is up to users to explain such metrics to larger audiences.   *Access:* http://www.harzing.com/pop.htm

4.6 Bibexcel

Bibexcel is a bibliometric tool developed at the University of Umea (Sweden). Bibexcel is a tool for helping with bibliometric analysis and citation studies in particular. This tool was specifically developed to manage the bibliometric data and build maps, which can be read by software such as Excel, SPSS, UCINET etc. Bibexcel is freely accessible for academic nonprofit use. Bibexcel can read data retrieved from different bibliographic sources, such as ISI Web of Science (WoS), Scopus, and the Procite export format. *Access:* http://www.umu.se/inforsk

4.7 HistCite

HistCite is a software developed to allow the users to aid researchers in visualizing the results of literature searches in the Web of Science. HistCite help the users to analyze and organize the results of a search to obtain various views of the topic's structure, history, and relationships. It is easy, fast, and provides perspectives and information not otherwise available. HistCite is a system designed to help selectively identify the significant (most cited) papers retrieved in topical searches of the Web of Science (SCI, SSCI and/or AHCI). Once a marked list of papers has been created, the resulting Export file is processed by HistCite to create tables ordered by author, year, or citation frequency as well as historiographs which include a small percentage of the most-cited papers and their citation links. *Access:* http://histcite.software.informer.com/12.3/

4.8 Pajek

Pajek is a program, for Windows, for analysis and visualization of large networks having some thousands or even millions of vertices. In Slovenian language the word pajek means spider. It provide tools for analysis and visualization of networks with different views. The latest version of is freely available, for noncommercial use at its home page.   Access: http://vlado.fmf.uni-lj.si/pub/networks/pajek/

5.0 Altmetrics resources

5.1 Altmetrics.org

This free Web site is a central hub for information about the growing altmetrics movement, which it defines as "the creation and study of new metrics based on

the Social Web for analyzing and informing scholarship." Cofounded by prominent figures in the world of bibliometrics, such as **Jason Priem** and **Heather Piwowar**, altmetrics.org maintains links to new online tools for calculating impact. Other prominent features include an altmetrics "manifesto" that argues how altmetrics can improve existing scholarly filters. *Access:* http://altmetrics.org

5.2 Impact Story

Formerly known as **Total Impact**, Impact Story is a free open source tool designed to support URL-based publishing through the aggregation of online altmetrics. Users create collections of materials through online identifiers, such as Google Scholar Profiles, DOIs, and PubMed IDs. Impact Story uses more than a dozen APIs to search for metrics on these collected items, with sources ranging from popular social media to scholarly tools like Mendeley and PLoS. Items are subsequently assigned impact categories, such as generally/highly "saved," "cited," "recommended," or "discussed." This resource is most useful for researchers publishing in nontraditional venues or with scholarship too new to have accumulated traditional citations. Not a comprehensive source for tracing Web impact. *Access:* http://impactstory.it/

5.3 ReaderMeter

ReaderMeter is a free tool that "crowdsources" impact by processing readership data from Mendeley. Created by Dario Taraborelli of the Wikimedia Foundation, it contrasts with traditional bibliometric tools in its focus on readership, not citation. The site functions by compiling reports based on authors' names, which are subsequently processed through the Mendeley API. Each report highlights

information such as an author's "HR-Index," "GR-Index," "Total Bookmarks," and "Top Publications by Readership." ReaderMeter has been by criticized some in the altmetrics community for drawing data exclusively from Mendeley. However, plans exist to integrate data from multiple reference management sites, such as CiteULike.
Access: http://readermeter.org/

6.0 Scholarly peer networks

6.1 Academia.edu

Academia.edu is a free online ***paper-sharing platform*** that encourages academics to increase their visibility and monitor research within and across its scholarly network. With nearly 2 million profiles and 1.5 million uploaded papers, academia.edu has become a popular player in the world of online repositories. Impact metrics for the site are similar to those offered by many blogs, and include profile views, document views, and country-based page traffic. In another increasing trend for scholarly networks, the site also offers features geared toward social interaction, such as user statuses and an "ask a question" tool. *Access:* http://www.academia.edu/

6.2 Mendeley

Mendeley is a relatively recent startup from the same company that created **Last.fm**. It combines a citation manager with a scholarly social network to create a comprehensive research portal. Researchers with profiles can chart views and downloads of their research through the portal, join groups, and view popular articles within their fields. Mendeley has

gained particular traction in the sciences, from which most of its users hail. However, with the integration of Mendeley data into more altmetrics tools, it will likely become popular with other disciplines, too. Mendeley is free with for-cost storage upgrades, and available both online and as a download.
Access: http://www.mendeley.com

6.3 Social Science Research Network (SSRN)

SSRN is an online article repository, recently listed number one in the Web of World Repositories' rankings for 2012. It encompasses three key features: a database of more than 400,000 abstracts, a large electronic paper collection, and 20 specialized subject networks through which registered users can promote their work and connect to free abstracts and articles. Though praised for its ability to facilitate discovery of scholarship, SSRN has also been criticized for the strictness of its policies, which some see as stifling in comparison to emerging scholarly networks. Still, its site-specific metrics for "top papers," "top authors"and"top institutions" remain key to social science faculty.
Access: http://www.ssrn.com

6.4 VIVO

VIVO is a free, downloadable semantic Web application designed to facilitate research collaboration both within and between institutions. Originally developed at Cornell, it invites institutions to upload data related to faculty profiles, which it crawls in order to draw meaningful connections between researchers. VIVO doesn't directly support user-centered metrics, but has the potential to be a powerful tool in collecting

university-level research metrics. To date, only a few large institutions have implemented VIVO, as it requires significant programming knowledge and commitment.
Access: http://vivoweb.org

7.0 Blogs and media

7.1 Citation Culture

This two-year-old blog is the creation of Paul Wouters, director of the Centre for Science and Technology Studies at Leiden University (LU). Authored by Wouter and a fellow LU professor, the blog is dedicated to discussion of academic impact, from citation analysis to the broader evaluation of research across universities. Recent multipart posts have touched on topics, such as humanities bibliometrics and scholarly altmetrics. While information on the site is excellent and detailed, posts are published sparingly, at a rate of one to two per month.
Access:
http://citationculture.wordpress.com/

7.2 Jason Priem's Web site

Jason Priem is a Ph.D. candidate at University of North Carolina-Chapel Hill's School of Information and Library Science and the cofounder of Impact Story. Priem has emerged as one of the strongest advocates for altmetrics, and a champion for library involvement. His interests touch on a variety of altmetrics topics, including the future of scientific communication, the open data movement, and author's rights. As the emerging altmetrics landscape continues to move forward, expect Priem to be at the front. *Access:* http://jasonpriem.org/

7.3 Scholarly Kitchen

Established by the Society for Scholarly Publishing, Scholarly Kitchen is a moderated blog that presents ideas on current topics of scholarly publishing and communication. While not strictly focused on bibliometrics, many of the site's "chefs" boast expertise in the intersection between impact and publishing. The site also offers useful category filters such as "Metrics & Analytics," which includes more than 280 posts and counting. *Access:* http://scholarlykitchen.sspnet.org/

8.0 Bibliometrics research support

8.1 Elsevier Bibliometrics Research Program (EBRP)

EBRP was designed by Elsevier as a way for bibliometrics researchers to gain access to large amounts of data for free. Available data includes publication metadata from Scopus, usage data, and full text data from Science Direct. Researchers apply for the data, and successful applicants receive a dataset specifically designed for their project by Elsevier. Examples of successful projects on the site are especially useful to those who are interested in current altmetrics a topic, such as the relationship between articles downloads and citations. *Access:* http://ebrp.elsevier.com/index.asp

8.2 OII Toolkit for the Impact of Digitised Scholarly Resources

This JISCfunded toolkit was developed by the Oxford Internet Institute to help authors, publishers, and librarians, learn more about measuring the impact of digital scholarship. The Web site is divided into three sections: case studies, quantitative methods, and qualitative methods. The two latter sections define and discuss methodological subcategories, such as bibliometrics/scientometrics and content analysis. Contributions to the toolkit are encouraged in the form of articles and comments, which can be submitted after creating a free user account. *Access:* http://microsites.oii.ox.ac.uk/tidsr/welcome

9.0 Organizations, conferences, and electronic lists

9.1 ACM Web Science Conference

The Web Science Conference is dedicated to the study of socio-technical relationships that shape and engage with the Web. An official ACM conference since 2011, Web Science brings together computer scientists with researchers from the social sciences, humanities, and law. Each conference has included a major workshop on the impact of the Web on scholarly communication—including this year's "Altmetrics12" workshop, run by affiliates of altmetrics.org. *Access:* http://www.websci12.org/

9.2 ASIST SIGMETRICS

This electronic list covers bibliometrics and altmetrics from a LIS perspective. Posts are equal parts information/announcement and discussion of factors related to bibliometrics, such as open access or "gaming" metrics systems. This electronic list is a great option for those interested in bibliometrics culture or in networking with bibliometrics specialists Includes a searchable archive. *Access:* http://web.utk.edu/~gwhitney/sigmetrics.html .

9.3 International Society for Scientometrics and Informetrics (ISSI)

ISSI is a major society dedicated to the study of bibliometrics, particularly in the sciences. Highlighted features include a biannual conference, abstracts of bibliometric journals, and a electronic list. Librarians interested in detailed analyses of bibliometrics should look to this site for a wealth of information. *Access:* http://www.issi-society.info

10.0 Conclusion

Now, There are several social media options that researchers incorporate into their overall 'impact strategy', *for example* listing their research on open access sites such as Mendeley, and joining discussions about research on social media sites such as Twitter or on blogs. Certainly, current measures of dissemination, most notably citations of articles or the impact factor of the journals in which they are published, do not take into account the social media impact of the article. Altmetrics, a sub-domain of webometrics, would challenge traditional indexing databases metrics such as JIF (Journal Impact Factor) and also help to devise new measures to analyze research output in social web space and make them more credible and standard measure. New measurements, such as altmetrics and article-level metrics such as those provided by PLoS, aim to take into account the views, citations, social network conversations, blog posts and media coverage are an attempt to analyse the influence of research across a global community.

Citations and impact factors measure the impact within the scientific community whereas views by social media will also include interested professionals and laypeople and, as such, measure uptake by different audiences. Social influence can produce an effect whereby something that is popular becomes more popular and something that is unpopular becomes even less popular.

References:

1. Lavis JN; Robertson D; Woodside JM; McLeod CB; Abelson J. et al. (2003). How can research organizations more effectively transfer research knowledge to decision makers? *Milbank Quarterly.* 81(2), pp.221–248.

2. Eysenbach G. (2008). Medicine 2.0: Social Networking, Collaboration, Participation, Apomediation, and Openness. *Journal of Medicine Internet and Research* 10(3): e22. Available at http://www.jmir.org/2008/3/e22/doi:10.2196/jmir.1030 (Accessed on 15/01/2015).

3. Eysenbach G. (2007). From intermediation to disintermediation and apomediation: new models for consumers to access and assess the credibility of health information in the age of Web2.0. *Study of Health Technology Information* 129(1), pp. 162–166.

4. Vaughan P. (2011). 23 Reasons Inbound Marketing Trumps Outbound Marketing Infographic Hubspot website. Available: http://blog.hubspot.com/ blog/tabid/6307/bid/28330/23-Reasons-Inbound-Marketing-Trumps-Outbound-Marketing-Infographic.aspx. (Accessed on 15/01/2015).

5. Priem J, Taraborelli D, Groth P, Neylon C. (2011). Altmetrics: a manifest. Altmetrics website. Available:

http://altmetrics.org/manifesto/. (Accessed on 15/01/2015).

6. Allen, Heidi G. [et. al.] (2013). Social media release increases dissemination of original articles in the clinical pain sciences. *PLoS ONE* 8 (7): e68914. (accessed on 18/01/2015).

7. Thelwall, Mike [et. al.] (2013). Do Altmetrics work? Twitter and ten other social web services. *PLoS ONE* 8 (5): e64841. (Accessed on 18/01/2015).

8. Munnolli, Satish S., Pujar, Shamprasad M. (2013). Eugene to Altmetrics: A chase for virtual footprints! *Annals of Library and Information Studies*. Vol. 60, pp.134-139.

9. Roemer, Robin Chin; Borchardt, Rachel. (2013). Institutional Altmetrics and Academic Libraries. *Information Standard Quarterly*. Vol. 25(2), pp.15-19.

10. Harinarayan, N.S and Vasantha Raju, N. (2013). Current Trends in Webometrics Research. *SRELS Journal of Information Management*. Vol. 50, No. 5, pp. 657-665.