Uncovering impacts: a case study in using altmetrics tools

Jason Priem[1], Cristhian Parra[2], Heather Piwowar[3], Paul Groth[4], and Andra Waagmeester[5]

[1] University of North Carolina at Chapel Hill
priem@email.unc.edu
[2] University of Trento
parra@disi.unitn.it
[3] National Evolutionary Synthesis Center
hpiwowar@nescent.org
[4] VU University Amsterdam
p.t.groth@vu.nl
[5] Maastricht University
andra.waagmeester@maastrichtuniversity.nl

Abstract. Growing scholarly use of Web tools present an opportunity to track alternative impacts along heretofore invisible paths like reading, bookmarking, and discussing. We present two tools, CitedIn and total-impact, that gather and report these and other "altmetrics" After discussing the tools features, we use a set of 214 articles from a national research center as a demonstration case study. We find that both tools present a meaningful number and variety of altmetrics in a form that could be used for immediate evaluation, and call for more research into the properties and validity of altmetrics.

Keywords: altmetrics, scholarly communication, impact, tools

1 Introduction

The future of scholarly communication is one in which a large part of scholarly communication is conducted online [3]. A key part of the scholarly communication lifecycle is trying to understand the impact of work. The process of understanding impact helps scientists, science administrators and others both find, evaluate, and access scholarly products. Traditionally, this impact assessment has been done primarily through the tracking of formal citations. This is possible because citations counts, for all their occasional ambiguity [2], do reflect use of scholarly products. However, this reflection is of a restricted spectrum; scholarly products are often used by scholars, and others, in ways that do not perturb the citation record [5]. Furthermore, traditional citation does not reflect the rapid nature of communications afforded by the Web. Thus, we need new approaches for measuring impact in this changed world.

Indeed, because of the Web scholarly communication, formerly "underground" uses like reading, bookmarking, sharing, discussing, and rating are beginning to leave online traces. The are becoming visible on Web pages [8, 13],

on blogs [6], in downloads [1, 4], on social media like Twitter [9], and in social reference managers like CiteULike, Mendeley, and Zotero [7]. These alternatives to traditional citation analysis have been labeled altmetrics [11]. Altmetrics offer potential for gathering information on more diverse types of impact, from more diverse scholarly products, including blog posts, slides, datasets, or even tweets. They also have the important benefit of speed; altmetrics typically accumulate in days or weeks rather than the years citations require. This is particular useful in as the research process increases pace where users of scientific content need to understand the impact of it rapidly. To begin to make practical use of altmetrics for measuring impact requires both a greater understanding of the properties and validity of these new metrics, and practical tools for obtaining them [10]. Others have begun the former [12]; here we will pursue the latter, presenting two new tools for gathering and presenting altmetrics.

2 Tools for Altmetrics: CitedIn and total-impact

CitedIn (http://citedin.org) and total-impact (http://total-impact.org) are open-source tools that receive as input a list of identifiers for scholarly products, and output a set of altmetrics for each product. CitedIn accepts only articles with PubMed IDs (PMIDs); total-impact accepts articles identified by PMID or DOI, but also datasets and slides using a variety of identifiers including URL, handle, and accession numbers. Both tools allow users to input identifiers manually; CitedIn also offers a REST API, and total-impact lets users automatically populate the products list using items stored in Mendeley or Slideshare libraries. Once users have uploaded products, CitedIn and total-impact both use calls to open Web APIs to gather data about them; CitedIn also caches available databases. As of September 25, 2011, the data sources used by each are listed in Table 1.

In addition to gathering altmetrics from these sources, both tools also include some additional features. CitedIn lets users input and output data over a REST API, and also reports a "CI-number" that summarizes all almetrics activity in a single value. Total-impact offers persistent URLs for impact report pages; the impact metrics can be refreshed over time. Both tools let users download results as structured text files for further analysis. Output pages for the tools are shown in Figures 1 and 2.

3 Case study: altmetrics for a national research center

We used a set of 214 articles from the National Evolutionary Synthesis Center (NESCent) as a realistic test for the two tools. NESCent was interested in tracking the impact of work they funded in a faster and more comprehensive way than citation analysis allowed – a typical use case for altmetrics. We entered the articles into CitedIn on August 14 2011, and into total-impact September 23 2011, then collected and analyzed the results.

All 214 articles had DOIs, and so were able to be processed by total-impact. Only 174 articles had the PMIDs required by CitedIn, so the CitedIn sample

	CitedIn	total-impact
Data repositories, including locating datasets associated with a given publication	ABS, Ares, Alzgene, Biogrid, BredeWiki, Ctdatabase, cancerCell, ChdWiki, Cosmic, Ctd, Cutdb, Dejavu, HIFTFBS, HNF4, HaemB, Jaspar, Kegg, Mgi, Mint, Mpidb, Nfi Regulome Resource, Oreganno, MIDNCI, BIDReactome, BDB, PleiadesGenes, Gregransbase, Balmer Retinoic, Uniprot, Wikipathways, Wormbase, YTPdb, Zfin	Dryad (downloads of most popular file, package views, total downloads and file views)
Social bookmarking and reference management tools	CiteULike, Connotea, Mendeley	CiteULike, Delicious, Mendeley (groups, readers)
Blogs and social media	Google Blogs, Nature Blogs	Facebook (clicks, comments, likes, shares)
Traditional citation	Google Books mentions	Citation in PubMed Central
Other	PubMed subsets, Citations from Wikipedia (pmid)	Citations from Wikipedia
PLoS ALM	N/A	Connotea, citations (CrossRef, PubMed Central, Scopus), blog mentions(Nature Blogs, ResearchBlogging, Bloglines, Postgenomic), downloads, PubMed activity

Table 1. Data sources for CitedIn and total-impact as of September 2011

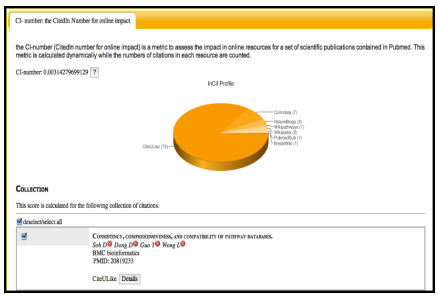

Fig. 1. CitedIn results page (Sept 2011)

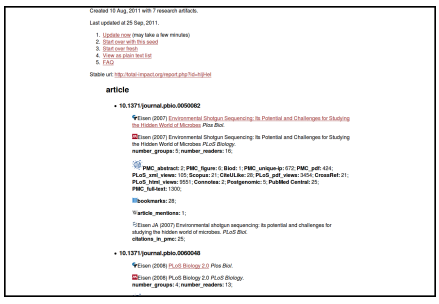

Fig. 2. total-impact results (Sept 2011)

is smaller. Both tools showed that altmetric activity as measured by number of "altmetric events" (bookmarks, downloads, etc.) is relatively widespread across articles: CitedIn found at least one event on 95% of its articles, and total-impact on 85%. There were a mean of 28 and median of 16 events per CitedIn article, with a maximum of 678. Total-impact had a per-article mean of 92 events and a median of 19; the higher mean is due to Dryad dataset downloads, which accumulate more easily than other metrics, reaching a maximum of 2769 on one article. We visualized the activity across articles using heatmaps, shown in Figures 3 and 4 to create a sort of "impact genome." Only altmetrics with nonzero counts are shown, and counts of each altmetric are normalized by that metric's maximum. Articles are arranged so that those with higher mean event counts across all metrics are further left.

Fig. 3. Active CitedIn event types and normalized event counts per article.

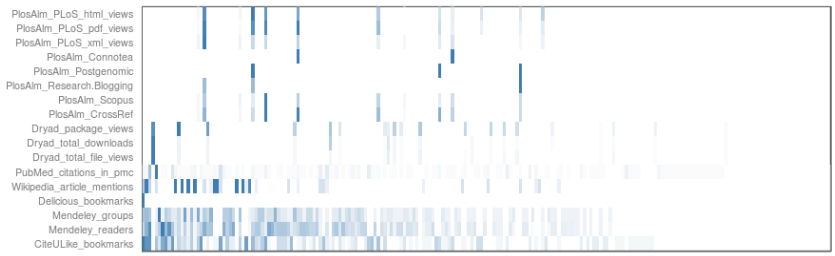

Fig. 4. Active total-impact event types and normalized event counts per article.

4 Conclusion

Altmetrics have potential to improve the speed and breadth of scientific evaluation. CitedIn and total-impact are two tools in early development that aim to gather altmetrics. A test of these tools using a real-life dataset shows that they work, and that there is a meaningful amount of altmetrics data available for use. These tools continue to improve: check out the current versions for up to date capabilities.

The properties and validity of these data, however, are still unclear, and call for additional research. What is the scholarly value of, for instance, a Mendeley

bookmark or a Wikipedia citation? Future work should also investigate how altmetrics for different sets of articles can be compared; this is a particularly tricky problem given the high dimensionality of altmetrics data, and may benefit from better visualization techniques, or statistical approaches like principle component analysis and factor analysis.

- Source code for CitedIn: http://code.google.com/p/citedin
- Source code for total-impact: https://github.com/mhahnel/total-impact
- Source code and data for analysis in this paper:
  https://github.com/jasonpriem/altmetrics-tools-iConference-poster
- The authors of the paper are key developers on CitedIn and Total-Impact

References

1. Bollen, J., Van de Sompel, H., Hagberg, A., Chute, R.: A principal component analysis of 39 scientific impact measures. PLoS ONE 4(6), e6022 (06 2009)
2. Bornmann, L., Daniel, H.D.: What do citation counts measure? A review of studies on citing behavior. Journal of Documentation 64(1), 45–80 (2008)
3. Bourne, P., Clark, T., Dale, R., de Waard, A., Herman, I., Hovy, E., Shotton, D., on behalf of the Force11 community: Force11 White Paper: Improving the Future of Research Communication and e-Scholarship (2011), `http://force11.org/white_paper`
4. Brody, T., Harnad, S., Carr, L.: Earlier web usage statistics as predictors of later citation impact: Research articles. J. Am. Soc. Inf. Sci. Technol. 57(8), 1060–1072 (Jun 2006), `http://dx.doi.org/10.1002/asi.v57:8`
5. Cronin, B.: The Hand of Science: Academic Writing and Its Rewards. Scarecrow Press (2005)
6. Groth, P., Gurney, T.: Studying Scientific Discourse on the Web using Bibliometrics: A Chemistry Blogging Case Study. In: WebSci10 Extending the Frontiers of Society OnLine (2010)
7. Hull, D., Pettifer, S.R., Kell, D.B.: Defrosting the digital library: bibliographic tools for the next generation web. PLoS computational biology 4(10), e1000204 (2008), `http://www.ncbi.nlm.nih.gov/pubmed/18974831`
8. L., V., K., H.: Relationship between links to journal web sites and impact factors. Aslib Proceedings: new information perspectives 54(6), 356–361 (2002)
9. Priem, J., Costello, K.L.: How and why scholars cite on Twitter. Proceedings of the 73rd ASIS&T Annual Meeting 73, `http://jasonpriem.org/self-archived/Priem\_Costello\_Twitter.pdf`
10. Priem, J., Hemminger, B.H.: Scientometrics 2.0: New metrics of scholarly impact on the social Web. First Monday 15(7) (Jul 2010), `http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2874`
11. Priem, J., Taraborelli, D., Groth, P., Neylon, C.: Alt-metrics: a manifesto (2010), `http://altmetrics.org/manifesto/`
12. Shema, H., Bar-Ilan, J.: Characteristics of Researchblogging.org science Blogs and Bloggers. altmetrics 11, `http://altmetrics.org/workshop2011/shema-v0/`
13. Thelwall, M., Vaughan, L., Björneborn, L.: Webometrics. Annual Review of Information Science and Technology 39(1), 81–135 (Oct 2006), `http://dx.doi.org/10.1002/aris.1440390110`