On the Relationship between Open Access and Altmetrics

Hamed Alhoori, Texas A&M University
Sagnik Ray Choudhury, The Pennsylvania State University
Tarek Kanan, Virginia Tech
Edward A. Fox, Virginia Tech
Richard Furuta, Texas A&M University
C. Lee Giles, The Pennsylvania State University

Abstract
A new and diverse set of online metrics are emerging to capture the effects of the sharing and discussion of research articles on online platforms. In this paper, we investigate whether altmetrics differ between Open Access (OA) and Non-Open Access (NOA) articles. We define a new metric, the *Open Access Altmetric Advantage*, and investigate 14 online data sources (Twitter, Facebook, CiteULike, Mendeley, F1000, blogs, mainstream news outlets, Google Plus, Pinterest, Reddit, Sina Weibo, the peer review sites PubPeer and Publons, policy documents, and sites running Stack Exchange (Q&A)). In eight of the data sources investigated, we found that OA articles receive higher altmetrics than NOA articles; however, we found less significant differences when taking into consideration some influential factors such as journal, publication year, and citation count.
Keywords: altmetrics; social media; open access; research evaluation; Google Scholar
Citation: Alhoori, H., Ray Choudhury, S., Kanan, T., Fox, E., Furuta, R., Giles, L.C. (2015). On the Relationship between Open Access and Altmetrics. In *iConference 2015 Proceedings*.

Acknowledgements: This publication was made possible by NPRP grant # 4-029-1-007 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.
Research Data: In case you want to publish research data please contact the editor.
Contact: {alhoori, furuta}@tamu.edu, {sagnik@psu.edu, giles@ist.psu.edu}, {tarekk, fox}@vt.edu

1 Introduction

Typical research dissemination methods include self-archiving preprint or postprint publications, presenting papers at conferences, publishing in NOA or OA journals, and sharing results with research groups. With recent and continuing research budget cuts, many research institutions have canceled costly subscriptions to journals.[1] Moreover, researchers may not be able to attend many related conferences or follow the vast range of publications available. Freed from subscription barriers, OA articles increase readily-accessible knowledge among researchers and the general public, where the intellectual outcomes of the scholarly community become more visible.

Research evaluation is moving beyond traditional scholarly metrics to include social, cultural, environmental, and economic impacts[2] (Borgman & Furner, 2005; Bornmann, 2013; Moed, 2005). Further, social media platforms are playing an important role in research workflow (Rowlands et al., 2011). For example, scholarly content is increasingly being shared on social media sites, and it is estimated that the number of research articles shared on social media is increasing at the rate of 5–10% per month (Adie & Roe, 2013).

Altmetrics (Priem et al., 2012) are proposed as a new, broader approach to measuring impact, intended to complement traditional citation-based metrics, and are currently under investigation. In several studies, researchers have found low to moderate correlations between citation metrics and altmetrics (Bar-Ilan et al., 2012; Haustein, Peters, Sugimoto, Thelwall, & Larivière, 2014; Thelwall, Haustein, Larivière, & Sugimoto, 2013; Waltman & Costas, 2014), which suggests a complex relationship between altmetrics and scholarly impact.

In this study, we explore the relationship between access approach of scholarly articles (NOA and OA) and altmetrics. We aim to answer the following research questions:

a) Do OA articles receive or generate higher altmetrics than NOA articles?
b) Do NOA and OA articles published in the same journal and year receive different altmetrics counts?

[1] http://www.timeshighereducation.co.uk/news/university-of-montreal-cancels-wiley-blackwell-deal-subscription/2010888.article
[2] http://www.ref.ac.uk/pubs/2011-01/

c) Is there a relationship between scholarly impact (citation count) and social impact (readership count) for NOA and OA articles?

2 Related work

Several studies have investigated whether OA articles receive more citations than NOA articles (known as the OA citation advantage). Lawrence (2001) found a citation advantage for conference articles in the field of computer science that are freely accessible online. Similar results have been reported in other fields, such as philosophy, political science, electrical engineering, electronic engineering, and mathematics (Antelman, 2004), physics (Harnad & Brody, 2004), agriculture (Kousha & Abdoli, 2010), and civil engineering (Koler-Povh et al., 2013). Hajjem et al. (2006) used articles published over a 12-year period from 10 disciplines: administration, economics, education, business, psychology, health, political science, sociology, biology, and law. They found that OA articles had more citations, and that the OA citation advantage ranged from 36% to 172%, according to discipline and year.

Norris et al. (2008b) found disciplinary differences in the citation advantage of OA articles in ecology, applied mathematics, sociology, and economics. Xia et al. (2010) found that multiple OA availability correlates with citation count. McCabe and Snyder (2013) found an OA citation advantage of 8% on average, with differences depending on content quality. Other reasons reported for high citation rates of OA articles include preprint availability, quality bias, and selection bias (Kurtz et al., 2005). Eysenbach (2006) controlled for various confounding variables and found that OA articles were likely to be cited twice as often as NOA articles in the first 4–10 months after publication. Gargouri et al. (2010) reported that OA articles were not subject to a quality bias, finding a high OA citation advantage for both self-selected self-archiving and mandatory self-archiving.

A number of studies have explored the effects of social media on the dissemination of research. Shuai et al. (2012) found that the number of tweets citing preprints on arXiv.org correlated with the number of downloads and early citations. Allen et al. (2013) posted sixteen PLOS ONE articles on Facebook, Twitter, LinkedIn, and ResearchBlogging.org on either a random release date or a control date. They found that the dissemination of research through social media increased the number of views and downloads. Haustein et al. (2014a) found that the coverage and readership of articles published by sampled bibliometricians were higher on Mendeley than on CiteULike. In other recent studies, we found that the altmetrics were related to traditional journal rankings (Alhoori & Furuta, 2014) and countries' scholarly outcomes (Alhoori et al., 2014). Shema et al. (2014) found that articles cited on blogs received more citations. The focus of research to date has been limited to the citation rather than the altmetrics advantage of OA, and studies in this area have not drawn on a wide range of online metrics. The present study explores both of these directions.

3 Data and Methods

We randomly selected 23 NOA and hybrid OA journals from the top 100 journals from all fields as ranked by the 2014 Google Scholar h-index.[3] We used Scopus to download bibliographic information for 42,582 articles published in the selected journals between 2010 and 2014. From the downloaded articles, we selected only those that had DOIs. We then used Google Scholar to determine which of our articles were OA or NOA, because Google Scholar was found to retrieve a higher percentage of OA articles than OAIster and OpenDOAR (Norris et al., 2008a).

We modified a parser for Google Scholar[4] to read our collection of articles, conduct an article title search, and if available retrieve a direct link to the full text of each article (i.e., the search result link adjacent to the article title on the Google Scholar results page).[5] We ran the parser on a computer that did not have a subscription to any journals. In general, for each article, the parser returned one of two results: a web link to the article (e.g., *.html* or *.pdf*) or no link at all. For seven of the journals, we found that Google Scholar returned many links to NOA articles, so we excluded those journals. We removed duplicate articles as well as those retrieved by Google Scholar for years outside the 2010–2014 range, thus reducing the number of journals to 16 and the number of articles to 27,011.

We defined OA articles as those for which the search returned a link, whereas articles for which a link was not returned were flagged as NOA. Using a random sample of 400 articles, we tested whether the title of the returned article link by Google Scholar matched our query title and found an accuracy rate of 99.2%. Using other random samples of 400 NOA articles and 400 OA articles, we checked the

[3] http://scholar.google.com/citations?view_op=top_venues
[4] http://www.icir.org/christian/scholar.html
[5] http://scholar.google.com/intl/en-US/scholar/help.html

accuracy of our classifications of articles as NOA or OA and found accuracy rates of 97.5% and 96%, respectively.

We downloaded each journal's altmetrics from altmetric.com, which comprise mentions of articles on Twitter, Facebook, CiteULike, Mendeley, F1000, blogs, mainstream news outlets, Google Plus, Pinterest, Reddit, Sina Weibo, the peer review sites PubPeer and Publons, policy documents, and sites running Stack Exchange (Q&A). We then matched the articles using DOIs. We removed three sources of altmetrics—Pinterest, Q&A sites, and the policy documents— due to insufficient data.

We defined the *OA Altmetric Advantage* (OAAA) for all types of altmetrics as shown in equation (1). $\overline{OA}$ represents either the average number of articles that received an altmetric (article-based) or the average altmetric across articles (altmetric-based) for OA articles, and $\overline{NOA}$ represents the same for NOA articles.[6]

$$OA\ Altmetric\ Advantage(OAAA) = \frac{\overline{OA} - \overline{NOA}}{\overline{NOA}} \quad (1)$$

We compared NOA articles with OA articles based on altmetric type. We then compared articles with similar altmetric types that were published in the same year. In order to reduce the effects of platform, time, journal ranking (e.g., Impact Factor), and discipline, we extended the comparison by checking articles based on the altmetric type per journal per published year. We then compared articles based on citation count. We used the Mann-Whitney U test to check for significant differences between NOA and OA articles, with regard to the altmetrics advantage. We used Spearman's rank correlation coefficient, ρ (rho), to compare citation count with Mendeley readership. We used Mendeley since we found that it has a high usage and coverage of scholarly activities (Alhoori & Furuta, 2014).

4 Results

Of the 27,011 articles, 6,934 were NOA and 20,077 were OA. Figure 1 provides descriptive statistics of the articles that received various types of altmetrics, with count, percentage, and access type. The vertical axis shows the percentage of NOA (gray columns) and of OA (light-blue columns) articles.

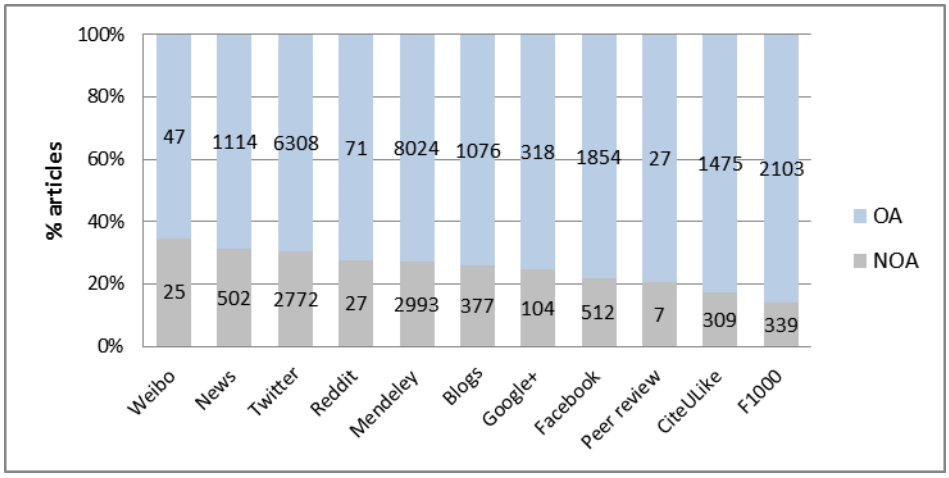

Figure 1: Distribution of NOA and OA articles across online platforms.

We compared the NOA and OA articles that received altmetrics with those that did not, using an article-based approach (Figure 2) and an altmetric-based approach (Figure 3). Figure 2 shows the percentages of $\overline{OA}$ and $\overline{NOA}$ based on article count, and the right side shows an article-based OAAA, which is represented by the red curve. Six platforms did not show any article-based OAAA. However, a clear article-based OAAA is shown for both F1000 and CiteULike.

[6] For example, for a total of 1,000 articles, 400 OA and 600 NOA, and among them 40 OA and 30 NOA with a specific type of altmetrics (e.g., tweet), totaling 800 tweets for OA and 900 tweets for NOA. An article-based approach yields $\overline{OA}$ = 40/400 = 0.1, $\overline{NOA}$ = 30/600 = 0.05, and *OAAA* = (0.1 - 0.05)/0.05*100 = 100%. An altmetric-based approach yields $\overline{OA}$ = 800/400=2, $\overline{NOA}$ = 900/600=1.5, and *OAAA* = (2-1.5)/1.5*100 = 33.3%.

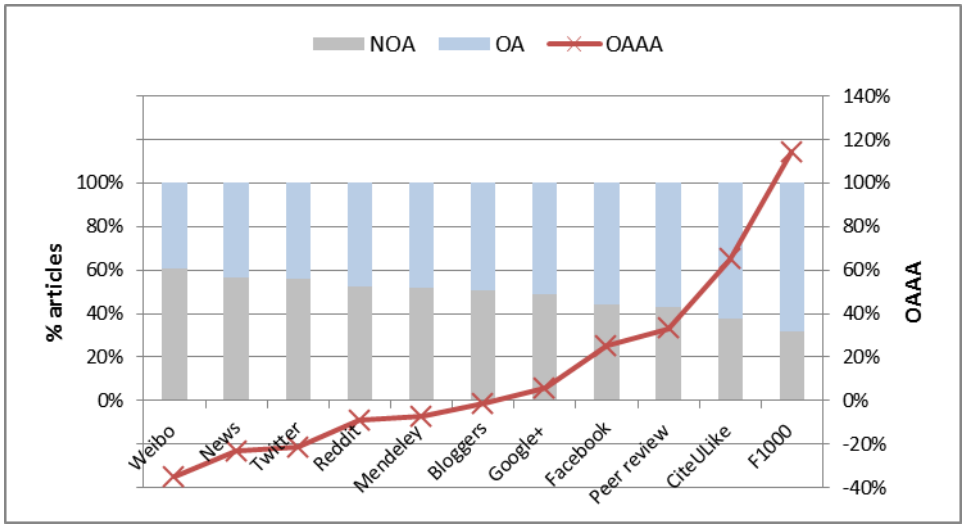

Figure 2: Article-based OAAA.

Figure 3 illustrates the distribution of altmetrics for NOA and OA articles. An altmetric-based OAAA is shown on eight platforms with four of them above 50%. Figure 2 shows that a higher percentage of OA articles received more altmetrics than NOA articles on F1000, CiteULike, Facebook, and peer review sites. Mendeley covers a slightly higher percentage of NOA articles (Figure 2), but the OA articles have 60% more readers (Figure 3). Academic social networks (e.g., F1000, CiteULike, and Mendeley) received high altmetric-based OAAA, whereas there was a clear difference between the general social media sites in terms of altmetrics received by NOA and OA articles. For example, Facebook covered a high percentage of OA articles and showed a high OAAA (105%). On the other hand, Twitter covered a high percentage of NOA articles, but OA articles received more tweets (7%), which might be the effect of publishers sharing NOA articles on Twitter more often than on Facebook. Google Plus, mainstream news outlets, and Weibo did not receive altmetric-based OAAA, which could be due to the effect of high impact articles published in high-ranked NOA journals (Alhoori & Furuta, 2014).

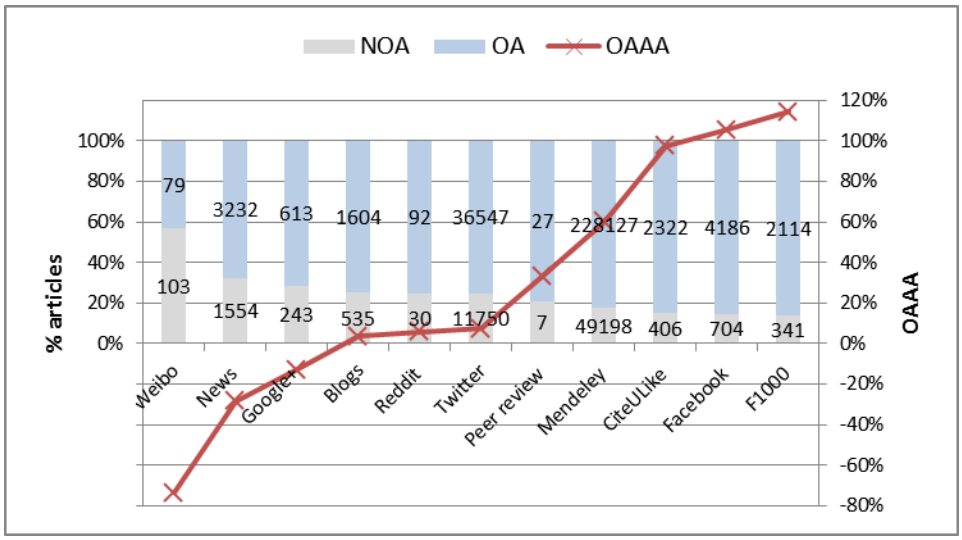

Figure 3: Altmetric-based OAAA.

Table 1 reports significant differences (p-value <0.05) between NOA and OA articles in terms of type of altmetrics and year. CiteULike and F1000 each showed a significant difference between NOA and OA articles for the years 2010–2013. However, no significant difference was found between NOA and OA articles for CiteULike or for F1000 in 2014, which could be due either to insufficient data and/or to a declining OA advantage (Davis, 2009). Twitter and Mendeley showed significant differences between NOA and OA articles in all the years studied, with the exceptions of 2011 for Twitter and 2012 for

Mendeley. The absence of a significant difference in 2011 could be due to missing tweets, as altmetrics.com started accumulating altmetrics in that year. NA values were mainly from insufficient data.

Altmetric type	2010	2011	2012	2013	2014
Blogs	**0.04**	0.43	**0.44**	0.14	0.60
CiteULike	**0.00**	**0.00**	**0.00**	**0.00**	0.96
F1000 reviews	**0.00**	**0.00**	**0.00**	**0.00**	0.60
Facebook	0.11	**0.04**	**0.00**	**0.00**	0.85
Google+	0.51	0.83	0.05	0.05	0.94
Mendeley	**0.00**	**0.00**	0.51	**0.00**	**0.00**
News outlets	0.46	**0.04**	**0.00**	**0.00**	**0.00**
Peer review sites	0.62	0.24	0.15	0.63	0.64
Reddit	0.10	0.75	0.73	0.51	0.53
Twitter	**0.00**	0.05	**0.00**	**0.00**	**0.00**
Weibo	NA	0.53	0.49	0.64	**0.01**

Table 1: Statistical Significance between NOA and OA Articles Across Altmetrics and Years

We checked for significant differences between NOA and OA articles for journals and publication years on platforms that showed OAAA. Table 2 presents an example from Mendeley, which shows a significant difference for eight journals in 2014 but for only two in 2010. This could be because OA articles are available as preprints earlier than NOA articles, whereas in 2011 and 2012 only three and two journals, respectively, showed significant differences. In other platforms, we found similar results for journals showing a significant difference in 2014. However, we found less significant differences within years and journals overall.

Journal name	2010	2011	2012	2013	2014
Accounts of Chemical Research	0.10	0.29	0.25	0.52	0.85
Advanced Materials	0.05	**0.03**	0.30	0.10	**0.00**
American Economic Review	NA	NA	NA	NA	**0.02**
American Journal of Respiratory and Critical Care Medicine	0.45	**0.03**	**0.00**	0.43	**0.01**
Astronomy and Astrophysics	0.97	0.99	0.43	0.85	0.59
Circulation	0.35	0.35	0.44	**0.00**	0.05
Clinical Infectious Diseases	NA	NA	NA	**0.00**	**0.00**
European Heart Journal	0.35	**0.01**	**0.01**	**0.00**	0.14
Gastroenterology	**0.00**	0.68	0.23	**0.00**	0.70
Genes and Development	NA	NA	0.59	0.21	**0.00**
Hepatology	NA	NA	NA	**0.00**	**0.00**
Journal of Clinical Oncology	0.99	0.68	0.72	NA	0.09
Journal of Immunology	0.55	NA	NA	NA	0.27
Journal of the American College of Cardiology	**0.00**	0.64	0.77	0.23	**0.02**
Neuron	NA	NA	NA	**0.01**	**0.00**
Review of Financial Studies	0.23	0.85	0.40	0.07	0.81

Table 2: Statistical Significance between NOA and OA Articles for Readership across Journals and Years

Finally, we compared NOA and OA articles to determine whether there was a correlation between citation count and Mendeley readership, as shown in Figure 4. We selected articles published in 2012 so that they had enough time to accumulate citations and readership. We found a weak significant correlation between citation count and average readership for NOA articles ($\rho = 0.26$). However, we found a

moderate significant correlation between citation count and average readership for OA articles (ρ = 0.56). No correlation was found between readership for NOA articles and readership for OA articles. Further, articles that received more than 80 citations were mostly OA with significant difference, which shows a preference for sharing OA articles over NOA articles in academic social networks.

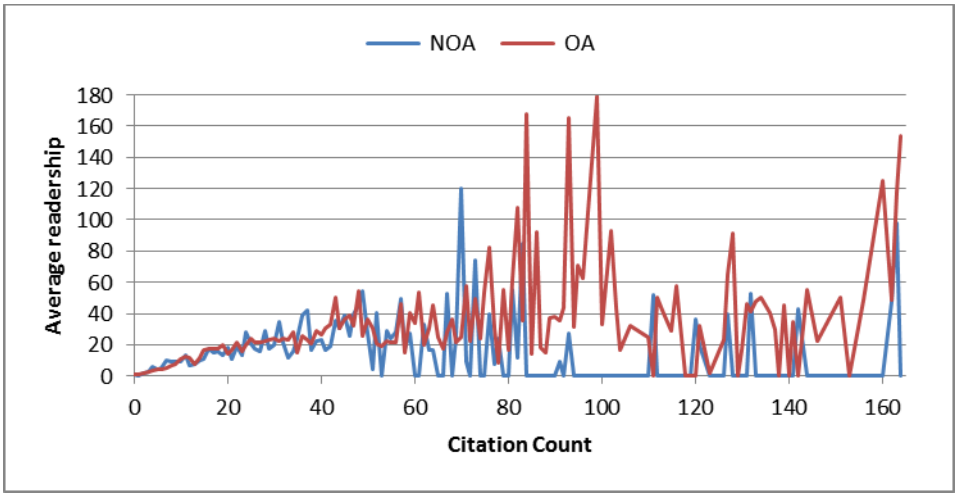

Figure 4: Average Mendeley readership per citation count for NOA and OA articles.

5 Conclusion

In this research, we explored the relationship between altmetrics and NOA and OA articles. On eight online platforms (F1000, Facebook, CiteULike, Mendeley, peer review sites, Twitter, Reddit, and blogs), we found that OA articles received more altmetrics than NOA articles. However, when we investigated the effects of journal, publication year, and citation count, we did not find a clear relationship between OA and altmetrics. We found that academic social networks had a high OAAA. However, the general social media sites differed in terms of the quantity of altmetrics received between NOA and OA articles. For example, Facebook had a high OAAA, whereas Weibo had no OAAA. This study also reported a significant correlation between citations and altmetrics for NOA and OA articles, which was not the case for some previous studies that compared articles in general (Alhoori et al., 2014). We plan to expand this study to include more journals and articles and to explore disciplinary differences. We also plan to investigate whether and to what extent there are differences in altmetrics between green and gold OA articles (Harnad et al., 2008).

6 References

Adie, E., & Roe, W. (2013). Altmetric: enriching scholarly content with article-level discussion and metrics. *Learned Publishing*, *26*(1), 11–17. doi:10.1087/20130103

Alhoori, H., & Furuta, R. (2014). Do Altmetrics Follow the Crowd or Does the Crowd Follow Altmetrics? In *IEEE/ACM Joint Conference on Digital Libraries* (pp. 375–378). doi:10.1109/JCDL.2014.6970193

Alhoori, H., Furuta, R., Tabet, M., Samaka, M., & Fox, E. A. (2014). Altmetrics for Country-Level Research Assessment. In *Proceedings of the 16th International Conference on Asia-Pacific Digital Libraries* (Vol. 8839, pp. 59–64). doi:10.1007/978-3-319-12823-8_7

Allen, H. G., Stanton, T. R., Di Pietro, F., & Moseley, G. L. (2013). Social Media Release Increases Dissemination of Original Articles in the Clinical Pain Sciences. *PLoS ONE*, *8*(7), e68914. doi:10.1371/journal.pone.0068914

Antelman, K. (2004). Do Open-Access Articles Have a Greater Research Impact? *College & Research Libraries*, *65*(5), 372–382. doi:10.5860/crl.65.5.372

Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H., & Terliesner, J. (2012). Beyond citations : Scholars ' visibility on the social Web. In *Proceedings of the 17th International Conference on Science and Technology Indicators* (Vol. 52900, pp. 98–109). Retrieved from http://arxiv.org/abs/1205.5611

Borgman, C. L., & Furner, J. (2005). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, *36*(1), 2–72. doi:10.1002/aris.1440360102

Bornmann, L. (2013). What is societal impact of research and how can it be assessed? A literature survey. *Journal of the American Society for Information Science and Technology*, *64*(2), 217–233. doi:10.1002/asi.22803

Davis, P. M. (2009). Author-choice open-access publishing in the biological and medical literature: A citation analysis. *Journal of the American Society for Information Science and Technology*, *60*(1), 3–8. doi:10.1002/asi.20965

Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS Biol*, *4*(5), e157. doi:10.1371/journal.pbio.0040157

Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., & Harnad, S. (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

Hajjem, C., Harnad, S., & Gingras, Y. (2005). Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. *IEEE Data Engineering Bulletin*, *28*(4), 39–47. Retrieved from http://eprints.ecs.soton.ac.uk/11688/

Harnad, S., & Brody, T. (2004). Comparing the Impact of Open Access vs. Non OA Articles in the Same Journals. *D-Lib Magazine*, *10*(6). Retrieved from http://www.dlib.org/dlib/june04/harnad/06harnad.html

Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., … Hilf, E. R. (2008). The Access/Impact Problem and the Green and Gold Roads to Open Access: An Update. *Serials Review*, *34*(1), 36–40. doi:10.1080/00987913.2008.10765150

Haustein, S., Peters, I., Bar-Ilan, J., Priem, J., Shema, H., & Terliesner, J. (2014). Coverage and adoption of altmetrics sources in the bibliometric community. *Scientometrics*, *101*(2), 1145–1163. doi:10.1007/s11192-013-1221-3

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

Koler-Povh, T., Južnič, P., & Turk, G. (2014). Impact of open access on citation of scholarly publications in the field of civil engineering. *Scientometrics*, *98*(2), 1033–1045. doi:10.1007/s11192-013-1101-x

Kousha, K., & Abdoli, M. (2010). The citation impact of Open Access agricultural research: A comparison between OA and non-OA publications. *Online Information Review*, *34*(5), 772–785. doi:10.1108/14684521011084618

Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E., & Murray, S. S. (2005). The effect of use and access on citations. *Information Processing and Management*, *41*, 1395–1402. doi:10.1016/j.ipm.2005.03.010

Lawrence, S. (2001). Online Or Invisible? *Nature*, *411*(6837), 521. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.28.7519

McCabe, M. J., & Snyder, C. M. (2013). Identifying the Effect of Open Access on Citations Using a Panel of Science Journals. *SSRN Electronic Journal*. doi:10.2139/ssrn.2269040

Moed, H. F. (2005). *Citation analysis in research evaluation*. Retrieved from http://www.springer.com/new+%26+forthcoming+titles+(default)/book/978-1-4020-3713-9

Norris, M., Oppenheim, C., & Rowland, F. (2008a). Finding open access articles using Google, Google Scholar, OAIster and OpenDOAR. *Online Information Review*, *32*(6), 709–715. doi:10.1108/14684520810923881

Norris, M., Oppenheim, C., & Rowland, F. (2008b). The citation advantage of open-access articles. *Journal of the American Society for Information Science and Technology*, *59*(12), 1963–1972. doi:10.1002/asi.20898

Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). Altmetrics in the wild: Using social media to explore scholarly impact. *arXiv:1203.4745*. Retrieved from http://arxiv.org/abs/1203.4745

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

Shema, H., Bar-Ilan, J., & 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. doi:10.1002/asi.23037

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*(11), e47523. 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 Services. *PLoS ONE*, *8*(5), e64841. doi:10.1371/journal.pone.0064841

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. doi:10.1002/asi.23040

Xia, J., Myers, R. L., & Wilhoite, S. K. (2010). Multiple open access availability and citation impact. *Journal of Information Science*, *37*(1), 19–28. doi:10.1177/0165551510389358

7 Table of Figures

8 Table of Tables