

Type of Submission: Presentation

Tracking Digital Footprints to Academic Articles: An Investigation of Publication Referral Data

Author1 Name

Author1 Institution, Country. email

Author3 Name

Author3 Institution, Country. email

Author2 Name

Author2 Institution, Country. email

Author4 Name

Author4 Institution, Country. email

ABSTRACT

In this study, we investigate the patterns of visits to academic articles on *PeerJ*, an open access peer-reviewed scientific mega-journal. About 3.4 million web traffic referrals of 1,432 *PeerJ* papers are collected and analyzed. We find that at least 57% of the visits are from external referral sources, among which Google, followed by Facebook and Twitter, contribute the most in directing people to the articles on *PeerJ*. In terms of academic search engines, NCBI, focusing on similar disciplines with the publication scope of *PeerJ*, is the largest referral source to *PeerJ* articles. Web of Science is ranked the second. In addition, our correlation analyses indicate that articles with more social attentions tend to have more visitors, which demonstrate the role of social media in enhancing visitors of academic articles.

KEYWORDS

article usage, altmetrics, referral, information seeking, open access, *PeerJ*, social media.

INTRODUCTION

The advent of Article-Level Metrics (ALMs) and “altmetrics” (Priem, Taraborelli, Groth, & Neylon, 2010) has brought opportunities to evaluate the impact of academic outputs in a broader range. From views, saves, discussions, recommendations to citations, there is a natural accession of increasing interest in and level of engagement with the scholarly works (Lin, & Fenner, 2013). Haustein, Bowman, & Costas (2015) propose a conceptual framework which looks into the “scholarly acts” associated with various types of “altmetrics” traces. In this study, we analyze the referral sources of academic articles on *PeerJ* in order to understand the behaviors associated with the articles visits (Davis & Solla, 2003). In addition to investigating how people find scholarly articles, we pay particular attention to the role of social media and news outlets in attracting readers of academic articles in online scholarly communication.

DATA

PeerJ, an open access peer-reviewed journal covering research in the biological and medical sciences, provides article-level metrics on the webpage of each article (including the number of views, downloads, unique visitors, and referral sources that bring 3 or more unique visitors to *PeerJ*). Data in this study were harvested from *PeerJ* on November 30, 2015. We collected bibliographic metadata and usage data for all published articles on *PeerJ*, including DOI, publish date, citations, visitors, views, downloads, as well as Twitter, Facebook, and Google+ activity counts. In addition, we also harvested web referral data for each article, including social referrals and other referrals. Data were parsed and processed into a SQL database for our analyses.

METHODS AND RESULTS

To analyze the referral sources, we took a four-step approach. First, we analyzed the referral sources in their original forms. In the second step, considering that many referrals belong to the same second-level domain (SLD), i.e., the same company or organization, and that many referrals use more than one top-level-domains (TLD), we merged all the subsidiaries of each company or organization into one referral source. In the third step, we integrate referrals sources into nine categories. There are 12440 unique referral sources to all the *PeerJ* articles that we collected. The nine categories include:

- General Search Engine
- Academic Resource, including both academic search engines (e.g., Google Scholar, NCBI, Web of Science, etc.), and academic publishers’ sites (e.g., Nature, Science, etc.)
- Social Network
- Bookmark or typed URL
- Mail System
- News & Blog
- RSS
- Institute (university, college, research institutes, or research organization website)
- Others (technology websites such as Github and Python.org, scientific data repositories such as Dryad, and other referrals that cannot be classified into the eight categories above)

In the second step, we were able to analyze what companies and organizations the referral sources were; In the third step, we examined the types of referral sources. In addition to the two steps above, we conducted a fourth step and investigated the top referrals in each category. Due to the limited space, we only present the results of our second step in Figure 1 below.

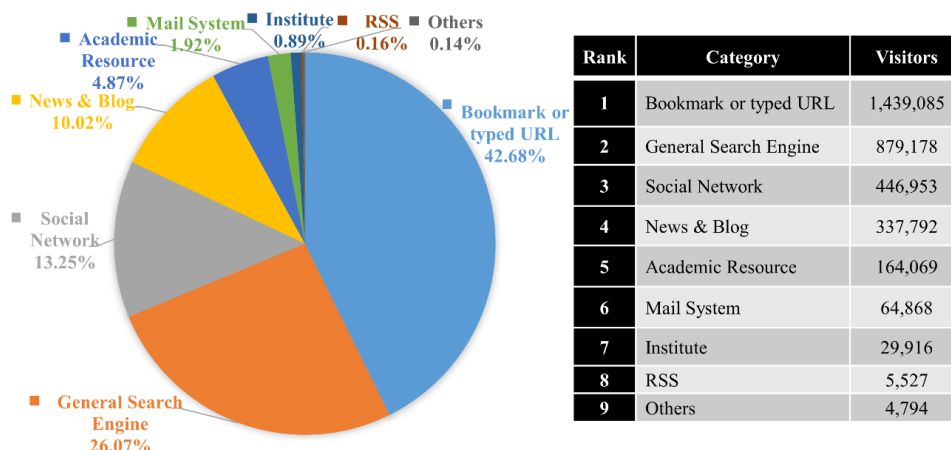


Figure 1 Categories of referrals

To gain a deeper understanding of the role of social media in bringing more readers to academic articles, we conducted Spearman correlation analyses between the number of visitors resulted by social media (Facebook and Twitter) shares and the number of mentions on these two platforms. Table 1 summarizes the descriptive statistics and correlation analysis results. Social media mentions on Facebook and Twitter are both positively and significantly correlated with the resulted numbers of article visitors.

Table 1 Summary statistics and Spearman correlation analysis results

Social media	N	Mean	Standard deviation	Correlation with visitors
Facebook	943	102.87	694.666	.688**
	871	27.62	91.815	
Tweet	1114	62.62	260.930	.833**
	1213	14.98	41.107	

Note: ** $p < .01$

DISCUSSIONS AND CONCLUSION

Our study indicates that general search engines are commonly used to find scholarly articles. Aided by the open movements, general search engines indexing scholarly articles make it easier for readers to find scholarly articles online. Regarding academic search engines, Web of Science is used not as frequently as NCBI when searching for articles published in the disciplines covered by *PeerJ*. Besides search engines, the top referrals are found to be either social networking platforms, or news outlets and blogs.

This study has also demonstrated that social network discussions and news reports do attract more visitors to scholarly articles. Although here the visitors don't necessarily mean readers, visiting scholarly articles is already one step further than reading only the social media and news media content mentioning the scholarly articles. With potentially more academic publishers providing referral data of article visits, opportunities are likely to emerge for researchers to conduct further qualitative studies to explore scholars' behavior and enrich the meaning of web-based metrics.

REFERENCES

- Davis, P. M., & Solla, L. R. (2003). An IP-level analysis of usage statistics for electronic journals in chemistry: Making inferences about user behavior. *Journal of the American Society for Information Science and Technology*, 54(11), 1062-1068.
- Haustein, S., Bowman, T. D., & Costas, R. (2015). Interpreting "altmetrics": viewing acts on social media through the lens of citation and social theories. *arXiv preprint arXiv:1502.05701*.
- Lin, J., & Fenner, M. (2013). Altmetrics in evolution: defining and redefining the ontology of article-level metrics. *Information Standards Quarterly*, 25(2), 20.
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: A manifesto. Retrieved July 1st 2018 from <http://altmetrics.org/manifesto/>