Trends of Publications' Citations and Altmetrics Based on Open Access Types

Chifumi Nishioka
Kyoto University
Kyoto, Japan
nishioka.chifumi.2c@kyoto-u.ac.jp

Michael Färber
Karlsruhe Institute of Technology
Karlsruhe, Germany
michael.faerber@kit.edu

ABSTRACT

This paper analyzes trends of citation and altmetrics with respect to different OA types (e.g., gold, hybrid, green). The analysis based on Unpaywall, Altmetric, and COCI shows that articles with a green license obtain more citations than other OA types. Regarding patents, hybrid, green, and bronze articles get more mentions compared to closed and gold articles. In terms of social media (e.g., Twitter and Facebook), bronze articles receive the most mentions.

CCS CONCEPTS

• **Information systems** → **Digital libraries and archives**.

KEYWORDS

bibliometrics, altmetrics, citation, scholarly communication

1 INTRODUCTION

Providing *open access* (OA) to publications has become an important effort in the digital libraries. In general, authors benefit from OA in different ways, such as a higher chance of citation [2] (i.e., Open Access Citation Advantages (OACA)). Although several OA types (e.g., *gold* indicating that the article was published in an OA journal and *green* indicating the self-archiving of the article) have been defined, most previous works examining OACA simply compared OA articles with non-OA articles, ignoring different OA types. In addition, in recent years, *altmetrics*, including the number of the article's mentions in the media, have received much attention, as they reveal the impact of research that cannot be uncovered by traditional citation analysis. However, previous works (e.g., [5]) have not explored to which degree the altmetrics of an article depends on the OA type.

This paper explores temporal trends of citations and altmetrics for each OA type, using Unpaywall, Altmetric, and COCI as datasets.

2 DATASETS AND METHODOLOGY

Datasets. We use the following datasets to identify trends of citation and altmetrics with respect to each OA type.

Unpaywall (https://unpaywall.org) provides a dataset including metadata and OA types of 109,905,121 publications with a Crossref DOI. We use the most recent unpaywall database, released on 11/22/2019. Each publication is assigned to an OA type being either *gold* (i.e., OA in an OA journal), *hybrid* (i.e., OA with the payment of APC), *bronze* (i.e., OA in a publisher site but neither gold nor hybrid), *green* (i.e., OA not in a publisher site, but in a repository), or *closed*.

Altmetric (https://www.altmetric.com/) is a platform which provides altmetrics scores for articles. We use the version from 10/08/2019. Altmetric provides several altmetrics scores, including mentions of articles in media and the number of readers of articles in reference managers. We focus on the mentions of the articles on Twitter and Facebook, in patents, news, and blogs, as well as their combined mentions.

COCI [1] is an open dataset of citation data. We use the dataset dump from 11/22/2018. As the dataset is generated based on citation data given to Crossref by publishers, it is of high quality. However, as some publishers refuse to make citation data available, the dataset contains references for only 27.41% of the publications with a Crossref DOI [3].

We consider all articles from these datasets which meet the following criteria:

(1) The article's type is `journal-article`.
(2) The article has been published between 2000 and 2018.
(3) The article is mentioned at least once in Altmetric.

There are 27,486,634 articles in Unpaywall that meet criteria (1) and (2). For 29.08% of them (7,992,953 articles), we have altmetric scores available. We use this dataset for our analysis. Note that the percentages of gold, hybrid, bronze, green, and closed articles included in our dataset to articles included in Unpaywall is not equally distributed (30.59%, 30.44%, 26.20%, 49.99%, and 24.90%, respectively). Thus, green articles in particular are well covered in our dataset. The percentages of gold, hybrid, bronze, green, and closed articles in our dataset are 16.72%, 6.44%, 10.07%, 19.12%, 47.64%, respectively.

Methodology. Using the 7,992,953 articles that meet all the criteria mentioned above, we compute the mean of each citation and altmetric score with respect to OA types. As shown before, we observe a bias that green articles are more likely to be included in the analysis. Thus, we also calculate the mean of citation count using the 27,486,634 articles and compare the trend with one based on the 7,992,953 articles.
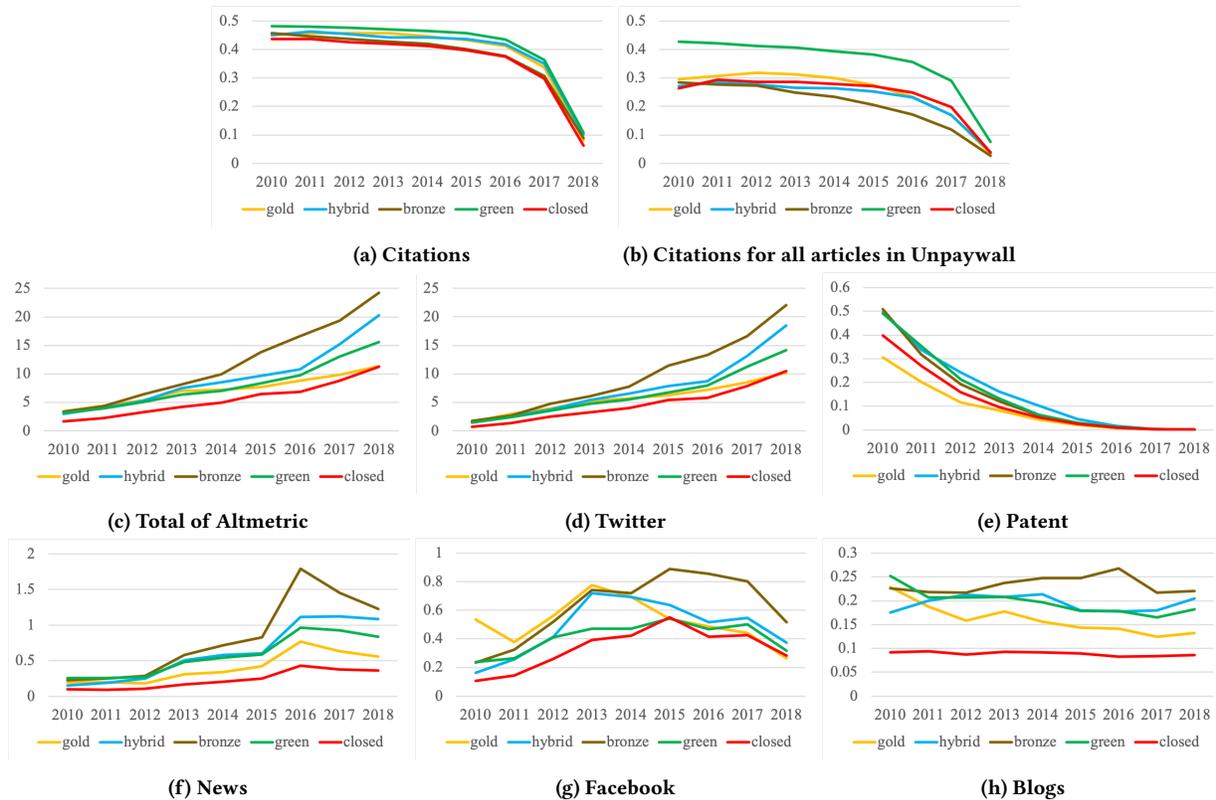
(a) Citations   (b) Citations for all articles in Unpaywall

(c) Total of Altmetric   (d) Twitter   (e) Patent

(f) News   (g) Facebook   (h) Blogs

Figure 1: Trends of citations and altmetrics for different OA types.

3 RESULTS

Figure 1 shows the trends of citations and altmetrics for different OA types. In each chart, the x-axis shows the articles' publication years and the y-axis refers to the mean of citation count or altmetric score respectively (e.g., the number of mentions on Twitter). We can observe several trends.

Citation. As we observe that green articles are more likely to be covered by Altmetric, we compare the trends of articles' citations included in Altmetric (Figure 1a) and those in Unpaywall (Figure 1b). We observe that the number of citations gets higher in Figure 1a compared to those in Figure 1b for articles except green. Thus, gold, hybrid, bronze, and closed articles with higher impact are more likely to be covered by Altmetric. We should consider the bias when conducting an analysis based on Altmetric. Among all OA types, green articles receive the most citations. We can thus confirm the observation made by Piwowar et al. [4].

Altmetrics. Regarding patents, we observe hybrid, green, and bronze articles receiving the most mentions. In addition, the trend of patent shows a decreasing tendency. Thus, patent has the most similar trend to citation compared to other altmetrics. We observe that bronze articles receive most mentions across all channels except for patents, followed by hybrid, green, gold, and closed. As Twitter accounts for 83.06% of total of altmetrics, Figure 1c shows a similar trend to Figure 1d.

4 CONCLUSION

This paper analyzed the trends of citation and altmetrics for each OA type. Across all media types – including news articles and Twitter –, we observed bronze articles receiving the most mentions, followed by hybrid, green, gold, and closed. In the future, we will analyze the time lag between time points of publication and citation.

ACKNOWLEDGMENTS

REFERENCES

[1] Ivan Heibi, Silvio Peroni, and David Shotton. 2019. COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations. *arXiv preprint arXiv:1904.06052* (2019).
[2] Susanne Mikki. 2017. Scholarly publications beyond pay-walls: increased citation advantage for open publishing. *Scientometrics* 113, 3 (2017), 1529–1538.
[3] Chifumi Nishioka and Michael Färber. 2019. Evaluating the Availability of Open Citation Data. In *Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL)*. CEUR Workshop Proceedings, 123–129.
[4] Heather Piwowar, Jason Priem, Vincent Larivière, Juan Pablo Alperin, Lisa Matthias, Bree Norlander, Ashley Farley, Jevin West, and Stefanie Haustein. 2018. The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ* 6 (2018), e4375.
[5] Xianwen Wang, Chen Liu, Wenli Mao, and Zhichao Fang. 2015. The open access advantage considering citation, article usage and social media attention. *Scientometrics* 103, 2 (2015), 555–564.