

Analysis of Over 50000 Research Papers' Altmetrics Published in Scopus Over a Period of 24 Weeks

Kaoru Matsumoto ^{*}, Masao Mori [†]

Abstract

In this paper, we report the outcomes of a trial analysis of information collected from the weekly acquisition of altmetrics index values for research papers that were published in Scopus [1]. While numerous applications of altmetrics have been explored, there is a lack of methodical evidence regarding their reliability, validity, and contextual usage, and only few large-scale studies have been conducted [2]. Additionally, there is a lack of understanding regarding the generation and utilization of these metrics by stakeholders, and the evaluation of index values remains uncertain. By accumulating altmetrics data and conducting time-series analysis, it may be possible to identify an optimal approach for utilizing these index values.

Keywords: altmetrics, PlumX, Scopus, research evaluation

1 Introduction

In this study, we use PlumX [3], one of the altmetrics, which consists of five metrics, and PlumX metrics are generated using the history of the Internet activity (Table 1).

Table 1: PlumX metrics

Categories	Description
Social Media	SNS posts. (The tweets, Facebook likes, etc. that reference the research)
Mention	Media coverage. (Measurement of activities such as news articles or blog posts about research)
Capture	App registrations. (Number of registrations to Mendeley and other apps. Indicates that someone wants to come back to the work)
Usage	Downloads. (PDF or HTML files. A way to signal if anyone is reading the articles or otherwise using the research)
Citation	Traditional citation indexes. (Conventional methods)

PlumX is an internet-based system developed by Plum Analytics, which allows users to access the publisher's article publication page for viewing the indicators. In addition, if you possess an Elsevier Scopus subscription, you can automatically retrieve the data programmatically through the Web API.

^{*} University of Toyama, Toyama, Japan

[†] Tokyo Institute of Technology, Tokyo, Japan

2 Methods

The data collection tool for PlumX was developed in Python 3.9.6 utilizing pybliometrics [4], a Scopus API wrapper developed by Rose Michael et al. The tool extracts PlumX metrics from Scopus by reading a paper list generated in SciVal [5]. The paper list was created using the parameter that the papers were published in 2022 and had at least one author affiliated with a Japanese organization updated weekly [6].

Data collection commenced on April 14, 2022, and tracked the changes in PlumX metrics immediately after publication in Scopus. Since the Scopus database is constantly updated, it is challenging to ascertain which papers were registered at what time. Therefore, we compared the week to the previous week and registered the newly added papers as Week_1. Thereafter, we collected and accumulated PlumX data for papers every week, and continuously tracked changes in values for the 50576 papers registered in Scopus by August 25 through Week_24. After reviewing the gathered data, we discovered a few system errors that led to incorrect values being outputted. Consequently, these were rectified by inserting the values obtained in the previous week. Furthermore, 486 papers' data were invalidated and deleted after being registered once, leaving 50090 papers' data to be analyzed as valid data.

3 Results

3.1 Overall Trends

At the time of Scopus registration (Week_1), Social Media had the highest average value of 6.94, followed by Capture at 1.07, Citation at 0.14, Mention at 0.17, and Usage at 0.04 (Figure 1). Although these values should be zero, however, there is a time lag of one to two months between the publication of the paper and their registration with Scopus, and the values reflect the behavior on the Internet during that time.

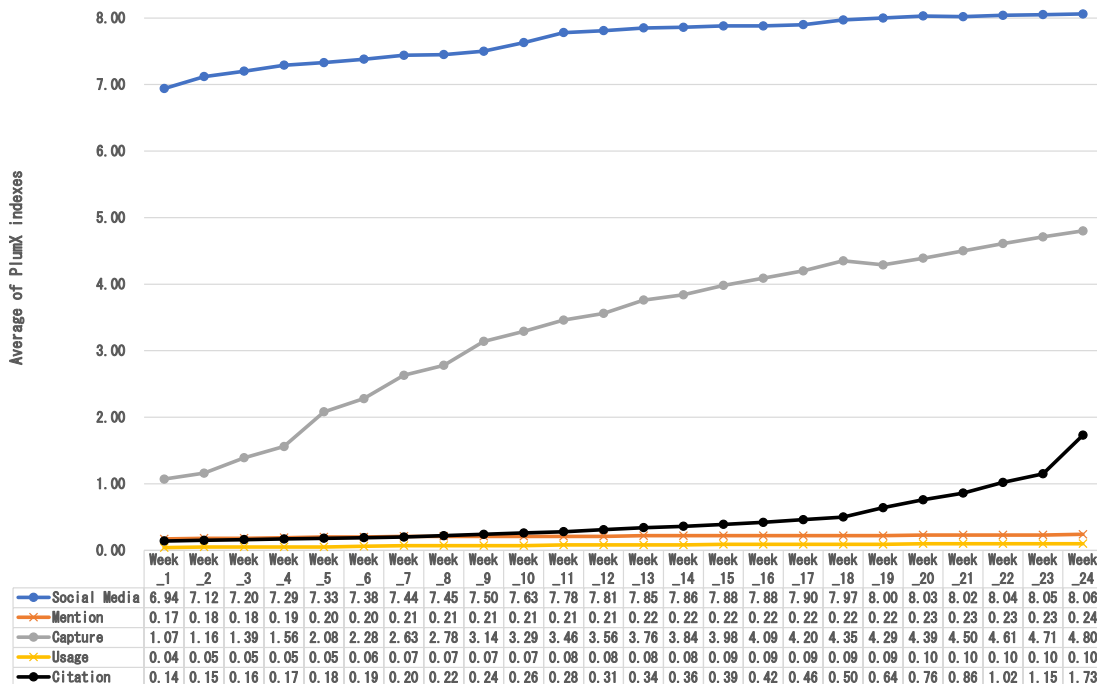


Figure 1: Time series of 24-week change in PlumX average

Table 2 shows the time-series change in the descriptive statistics of PlumX. We identified papers with a high citation count in the first week after their Scopus registration, as these papers may contain erroneous data, and proceeded to investigate the underlying factors contributing to the observed high citation counts. According to Elsevier, Scopus manages the publication year of the journal as the reference year, so there were cases where the publication year and the issue year were different. Thus, we examined the frequency distribution of citations in the first week and the actual issue year of each paper to estimate the extent of the difference between publication and issue years.

Table 2: Descriptive statistic of PlumX metrics (Week_1 & 12 & 24)

		Social Media	Mention	Capture	Usage	Citation
Week_1 (n = 50,090)	Max.	24,144	201	747	394	67
	Ave.	6.94	0.17	1.07	0.04	0.14
	Min.	0	0	0	0	0
	SD	134.12	2.46	5.30	2.71	1.02
Week_12 (n = 50,090)	Max.	25,030	209	7,475	412	101
	Ave.	7.81	0.21	3.56	0.08	0.31
	Min.	0	0	0	0	0
	SD	155.16	2.75	38.90	3.69	1.50
Week_24 (n = 50,090)	Max.	24,623	242	3,446	415	436
	Ave.	8.06	0.24	4.80	0.10	1.73
	Min.	0	0	0	0	0
	SD	163.39	3.14	21.56	4.15	9.03

Table 3: Frequency of Citation and Publication Year (Week_1)

Citation	Frequency	Ratio (%)	Citation	Pub. Year	Frequency	Ratio (%)
0	46,656	93.14	5 - 9	2013	1	0.54
1	2,308	4.61		2015	1	0.54
2	561	1.12		2019	4	2.16
3	219	0.44		2020	43	23.24
4	87	0.17		2021	93	50.27
5	67	0.13		2022	43	23.24
6	57	0.11	10 - 19	2020	16	33.33
7	23	0.05		2021	19	39.58
8	20	0.04		2022	13	27.08
9	19	0.04	20 - 29	2020	5	35.71
10 - 19	48	0.10		2021	7	50.00
20 - 29	14	0.03		2022	2	14.29
30 - 39	7	0.01	30 - 39	2021	6	85.71
40 - 49	1	0.00		2022	1	14.29
50 - 59	2	0.00	40 - 49	2021	1	100.00
60 -	1	0.00	50 - 59	2021	2	100.00
Total	50,090	100.00	60 -	2021	1	100.00

Due to the difficulty in automating this verification process, it was performed manually, with only papers having a citation count of 5 or higher were examined (Table 3). In several cases, papers issued in 2021 had the year of publication in the following year because the procedure crossed the year. In some cases, the issue year was old, but the publication year was new. While the total number of cases with a citation count of 1 or higher in the first week is less than 7%, it is important to consider that there may be a time lag of 1 to 2 years and errors in the analysis.

3.2 Correlations

Figures 2 and 3 show the correlation coefficients among the five PlumX metrics. Thick arrows in the figures indicate strong correlations and thin arrows indicate weak correlations. Figure 2 shows the correlation coefficients for Week_12, and Figure 3 shows the correlation coefficients for Week_24. In the first half, the correlations between Social Media and Mention, Mention and Citation, and Capture and Citation are strong (Figure 2). In the second half, the correlation between Mention and Capture becomes strong, and the correlation between Capture and Citation becomes strong accordingly (Figure 3). On the other hand, Usage consistently showed a weak correlation with the other metrics, a trend that did not change in the second half of the period.

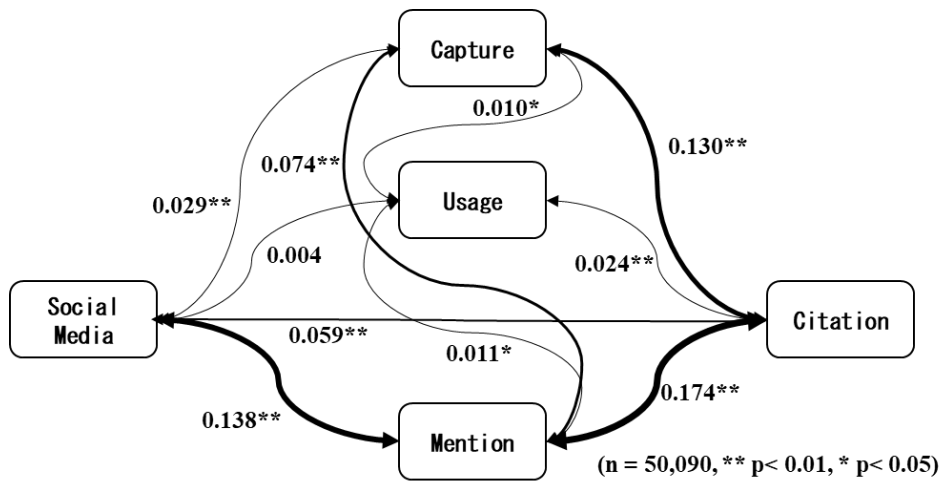


Figure 2: Correlation coefficients between PlumX metrics (Week_12)

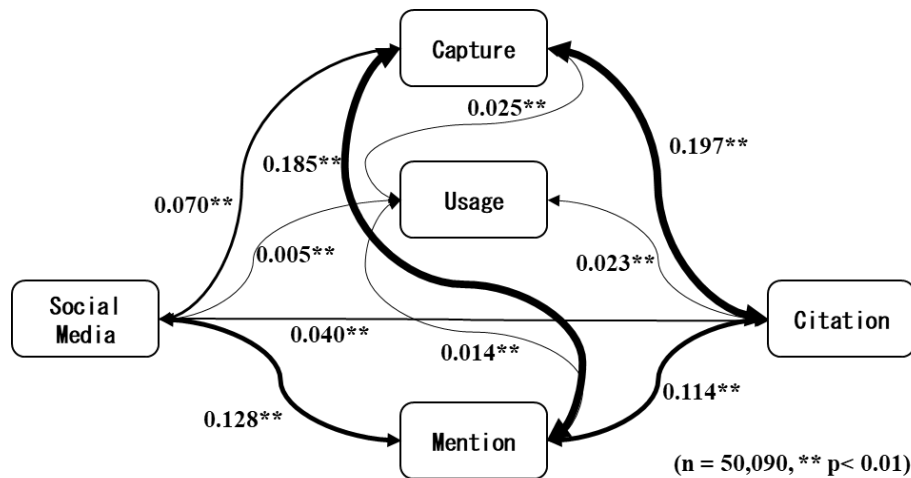


Figure 3: Correlation coefficients between PlumX metrics (Week_24)

To observe changes in the correlations between metrics over time, we analyzed the 24-week changes in average correlation coefficients. However, the consistently weak average correlation coefficients for Usage, ranging from 0.001 to 0.044, this metric was excluded from the analysis, and we focused on the remaining four metrics.

Figure 4 shows the time-series changes in the correlations between Social Media, Mention, Capture, and Citation. It shows that correlation with Capture dropped sharply between Week_5 and 18. Examination of the original data showed that the Capture value of the paper Elsevier ID: 2-s2.0-85129657706 recorded a very large value from Week_5 to 18, from 0 until Week_4, to 7361 in Week_5, followed by 7475 until Week_18, to 41 in Week_19, and to 0 in the latest value, which may be due to the influence of this paper. This anomalous behavior may be an influence of this specific publication, which belongs to the Book Series - Chapter classification, rather than the more prevalent Journal - Article type. As a result, it may be beneficial to analyze these metrics separately according to the publication type or to consider the possibility of a system error causing a sudden change in the data. Furthermore, it is important to acknowledge that the PlumX values for numerous publications have low initial values after registration, so some papers with large initial values may have a significant impact on the correlation coefficient, potentially leading to extreme values.

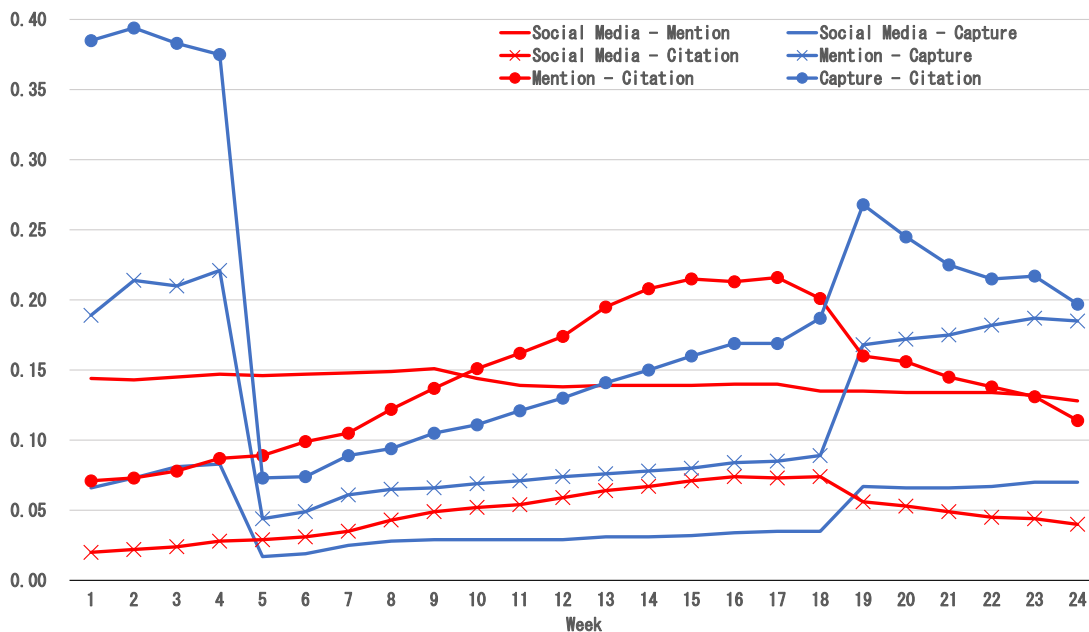


Figure 4: Time series of 24-week average correlations between PlumX metrics

Focusing on indicators not affected by outliers, the correlation between Mention and Citation reaches the top around Week_17, suggesting that Mention may require a boost by Week_17 to gain Citation. Hence, even if Mention increases post Week_17, Citation is not expected to be increased rapidly. It can be inferred from the Citation average in Figure 1, which exhibits a sharp rise beginning from Week_18.

Regarding the correlation between Social Media and Mention, the values were already large from the first week and have not changed significantly since then, continuing to decrease slightly. Currently, research universities in Japan are actively engaged in research publicity to publicize promising research and encourage citation. Therefore, as soon as a new paper is pub-

lished, research is publicized in the mass media and social media research publicity. And then, the values of Social Media and Mention increase simultaneously, and the correlation coefficient also expect to increase.

4 Conclusions

In this paper, we collected data on PlumX metrics from papers published in 2022 over a period of 24 weeks. Subsequently, we conducted a preliminary analysis to investigate the trends in PlumX metrics. The results showed that Usage, which indicates the number of papers downloaded, had a weak correlation with Citation over the 24-week duration. We also found that the PlumX metrics may have been influenced by publicity related to research conducted by Japanese research universities. Nevertheless, we also identified that certain papers published more than a year ago were included in our data, and some papers exhibited extreme and sudden fluctuations in values. While the impact of such outliers on the overall results was minimal when average values were used, their presence could potentially distort correlation coefficients, which are particularly sensitive to outliers.

This analysis represents an initial step, and we intend to collect additional data to monitor trends over a long period and many papers. We would like to continue collecting and analyzing data for a few years. We also believe that analysis needs to be conducted, focusing on differences in academic disciplines and publication type. Ultimately, we would like to ascertain how papers that have received high citation numbers or high attention have come.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number JP22K12732.

References

- [1] Elsevier B.V., “About Scopus - Abstract and citation database | Elsevier,” 2023; www.elsevier.com/solutions/scopus.
- [2] P. Wouters and R. Costas, “Users, narcissism and control: Tracking the impact of scholarly publications in the 21st century,” 2012, pp.847-857; www.researchgate.net/publication/268271559_Users_Narcissism_and_Control-Tracking_the_Impact_of_Scholarly_Publications_in_the_21_st_Century.
- [3] Plum Analytics, “About PlumX Metrics,” 2023; plumanalytics.com/learn/about-metrics/.
- [4] Rose Michael E. and John R.Kitchin, “pybliometrics: Scriptable bibliometrics using a Python interface to Scopus,” *SoftwareX* 10 (2019) 100263, 2019, doi:10.1016/j.softx.2019.100263.
- [5] Elsevier B.V., “SciVal | Navigate the world of research with a ready-to-use solution | Elsevier solutions,” 2023; www.elsevier.com/solutions/scival.
- [6] K. Matsumoto and M. Mori, “Development of altmetrics data collection tools and data analysis trials,” 11th Proceedings of the Meeting on Japanese Institutional Research (MJIR2022), MJIR, 2022, doi:10.50956/mjir.11.0_106.