DOI: 10.55415/deep-2022-0025.v1

Which factors drive the increased usage of newly published papers?

SONG Chao[1]*

[1] *scdmxy@163.com*
Institute of Science of Science and S&T Management and WISELab, Dalian University of Technology, Dalian 116024, Liaoning, China

Abstract: The usage metrics of academic papers is always the focus of researchers in the field of Scientometrcs. The traditional research topics paid too much attention to the relationship between Citation counts and Usage counts. Few researchers focus on usage metrics itself. Which factors drive the increased usage of papers? Especially for newly published papers. In this paper, we have selected 15 factors to measure the motivation of using newly 4 weeks published papers from WoS Core Collection database. We attempted to explain, statistically, how various factors affect the number of usage counts by Negative Binomial Regression Model. The main results of this study show that Document type and Funding text are two of most importantly positive factors to Usage counts. The Usage counts of open access papers will reduce significantly in WoS database. And distinct or novel title and abstract will increases the likelihood of Usage counts, but more is not always better. Moreover, there is weak effect of Citation counts on Usage counts in a short time.

Keywords: Scientometrics; Altmetrics; Usage; Newly published papers; Negative Binomial Regression Model

Introduction

As for as we known, there are three conventional indicators to measure the quality and quantity of academic ability are funding opportunities, number of articles published, and number of citations. Following the transition from print journals to electronic journals in the past 20 years. This is a fact that printed material will become obsolete much faster without electronic representation. The emergence of online journals and the improvement of retrieval means successfully affect obsolescence. So, in this increasingly digital age, usage metrics as the fourth key indicator has arisen. Nowadays, the number of times a paper used is becoming frequently-used indicator. As a result, usage metrics is becoming a part of the evaluation process of researchers and universities.

Admittedly, citation was always regarded as the most important indicator for a long time. The use of citation frequency to assess the impact of a research article has been well established. We can see it often that metrics of scientific impact are frequently defined as a function of the number of citations received for a long time. However, due to the existence of citation time window, it takes several years to correctly evaluate the value of a published article. It means that citation data are subject to significant publication delays. Obviously, usage and citation have different time dependencies characteristics (Schloegl et al. 2014). By contrast, usage data are not subject to publication delays. Usage data are sensitive to the recent publication record, but citation data are sensitive to the long history. Therefore, although usage metrics is different from citation metrics, we still believed that

DOI: 10.55415/deep-2022-0025.v1

usage metrics has become a necessary and useful complement to citation metrics. In addition, we need to reaffirm that usage metrics is not a subset of Altmetrics, and it has been made clear by Glänzel et al. (2016). For a long time, usage data was not easily accessible by ordinary researchers. We are happy to see that more and more data sets are starting to report the usage data to public or within limits nowadays. Usage metrics reflect quickly readers' attention or interest put into a paper, so it has become increasingly popular in the domain of Scientometrics.

In the early days, researches mainly focused on usage of library journals resources. More than twenty years ago, Line (1993) envisaged already that all volumes of the journals had the same opportunity to be retrieved and used in an electronic environment.In practice, the use of journal data is widely used in library management decisions (Mcdonald 2007; Wan et al.2010; Silton et al.2011; Wical et al.2015). As early as 2005, the academic community began focus their research interests on the relationship between citation and usage. Moed (2005), Coats (2005) and Nicholas et al. (2005) early paid attention to the relationship between the number of journal literature downloads and the frequency of citation. Almost at the same time, Bollen et al. (2005) emphasized that there are still some doubts regarding the validity of the ISI Impact Factor as the sole assessment of journal impact, and suggested the possibility of devising impact metrics based on usage information in general. A view by Brody et al. (2005) held that earlier web usage statistics could predict the later citation impact. But the results by Guerrero-Bote et al. (2014) showed that downloads have limited utility as predictors of citation since it is in the early years when any correlations have the least significance. Anyhow, this is predictable common sensewe that some of effective usage of a paper may lead to being cited. In terms of the research conclusions of most scholars, there is a strong and significant correlation between the citation frequencies and the number of downloads, no matter for journal sample (Schloegl et al.2010; He et al.2017) or at the article level (Markusova et al.2018).

We must have a discussion on this premise that usage data is feasible and valuable, although there is some controversy. Bollen et al. (2006) believed that usage was a part of the future Scientometric and Informetric. Usage data may confer several significant advantages over citation data. For example, usage data can be recorded immediately and widely, even at a very large scale than citation data. However, Thelwall (2012) still pointed that papers downloaded may be intended for teaching purpose rather than research purpose, even never be read. Gorraiz et al. (2014) felt absolutely convinced that taking downloads into consideration as a complementary aspect will broaden our bibliometric citation-restricted horizon and help to better understand the complex processes in scientific communication. Of course, we must admit some facts that there are some deficiencies of usage data. So far, usage data can practically be recorded only at the level of different information data platforms. Hood (2005) noticed early fairly that it may be that high overlap in many different databases gives a paper much more exposure, and therefore a greater opportunity to be used. Moreover, Moed et al. (2016) held that it is questionable whether bulk downloads should be included in the counts. And Doughty (2019) thought that it was supplied by publishers and not sufficiently reliable. All those issues may lead to the conclusions we got by usage metrics are not enough global and universally valid. However, despite that, usage data has emerged as a promising complement to existing indicators of assessment.

DOI: 10.55415/deep-2022-0025.v1

It is without a doubt, usage metrics is a fascinating area for research. On the one hand, just as Elsevier ScienceDirect reported in November 2006, one billion downloads were recorded after 1999 what greatly exceeds the total number of citations published since the 1900s. So, there is a huge amount of information to be discovered in usage metrics. Recent research by Henneken et al. (2017) which published in arXiv showed that download statistics even can be used to describe research activity at different levels and show a strong correlation with socio-economic indicators, like the GDP. On the other hand, we also observe that usage metrics exist short aging characteristic than citation metrics. Schloegl et al. (2011) computed a mean usage half-life of 1.7 years. Wang et al. (2014) found that research papers were downloaded most frequently within a short time period right after published. Nonetheless, Wang et al. (2016) also pointed that citations play an important role in determining the usage count, as to those old papers. As we rescan those views, we can't help but ask, what important information captivates the researchers? In other words, when researchers face with a huge amount of papers, there may be some interesting information in one paper which make they determine whether to choose to use. If we minimize the Matthew Effect of citation counts, we want to make a meaningful attempt to find out which factors drive the increased usage of newly published papers.

In fact, we have found some factors that affect usage metrics from past a few scattered researches. Jamali et al. (2011) found that articles with longer titles were downloaded slightly less than the articles with shorter titles. Zhao et al. (2018) concluded that funded papers attract more usage, but varying in different disciplines. Chi et al. (2017) held a view that higher numbers of co-authors were not associated with higher usage counts or citations. Kurtz et al. (2005) clearly found that citations were a good predictor of downloads, although not the other way around. But Schloegl et al. (2011) came to an uncertain conclusion that citations whether can direct influence downloads. Moreover, Wang et al. (2015) found that OA papers not only have the great advantage of total downloads, but also have the feature of keeping sustained and steady downloads for a long time. By contrast, arXiv-deposited articles received 23% fewer downloads from the publisher's website (Davis et al. 2007).

We take inspiration from all the above researches, but we still think they are still inadequate. Little attention has been paid to which factors drive the increased usage, especially for newly published papers. Therefore, it is essential to do a specific analysis of this problem. This article describes our first systematic explorations in this research area. In our study, we attempt to set up a system of factors which may influence usage metrics and analyse the papers published newly 4 weeks from Web of Science Core Collection.

Methodology

Data Sources and Data Processing

The results of similar questions based on different data sources are sometimes conflicting. For this reason, all the data in this study were collected from Web of Science Core Collection (WoS). The Usage Count is a record of all activity performed by all Web of Science users. Web of Science defines usage as "clicking" or "saving". Therefore, the behaviours of clicking or saving can reflect the interests or motives of users. Because

DOI: 10.55415/deep-2022-0025.v1

most papers published in newly 4 weeks have not been used (unused or zero-used), we selected the papers which usage counts greater than or equal to 1 in order to reduce the interference effect from unused papers. The data set was retrieved and downloaded on May 31, 2019. We obtained the index data of 36,473 papers in total. The papers were retained which document type was "Article" or "Review". Some of them were deleted because of lacking necessary information. 35,341 were the final sample size.

Table 1 shows that we select 16 indicators including variable name, variable definition and variable symbol. Among them, usage is dependent variable, and other 15 indicators are independent variables which may influence usage metrics. With the advent of advance publication, we helplessly assume that those articles are published within 4 weeks without any difference. An analysis at a per-week (even per-day) resolution could provide more insight, but the necessary data were not available in this study.

Table 1 Selection and definition of variables

No.	*Variable name*	*Variable definition*	*Variable symbol*
1	Usage counts	The number of a paper used in the last 180 days	*usage*
2	Title length	The number of title notional words by NLP	*ti_num*
3	Title distinction	The number of title distinctive words by TFIDF algorithm	*ti_tfidf*
4	Abstract length	The number of abstract notional words by NLP	*ab_num*
5	Abstract distinction	The number of abstract distinctive words by TFIDF algorithm	*ab_tfidf*
6	Author collaboration scale	The number of authors	*co_au*
7	Country collaboration scale	The number of countries	*co_country*
8	Organization collaboration scale	The number of organizations	*co_organ*
9	Document type	1= "Article"; 0= "Review".	*dt*
10	Funding text	1= "Funded"; 0= "un-Funded".	*fu*
11	Open access	1= "OA"; 0= "un-OA".	*oa*
12	Reference counts	The number of references	*nr*
13	Page counts	Total pages	*pg*
14	Number of WoS categories	The number of WC in WoS core collection	*wc*
15	Journal impact factor	Journal impact factor of paper	*so_factor*
16	Citation counts	Times cited count of WoS core collection	*tc*

Modelling Methods

Because scientific data rarely conform to normal distribution and present discrete distribution characteristics, the traditional regression models are not applicable. Given the nature of Number of Usage counts, the econometric tools modelling discrete counts can be invoked which were systematically introduced by Sun et al. (2016). These tools include the Poisson regression, the negative binomial regression, the zero-inflated Poisson regression and the zero-inflated negative binomial regression. As a comparison of models, the Poisson regression relies on the strict assumption of equal conditional mean and conditional variance of the dependent variable. The negative binomial regression accommodates the overdispersion by estimating an additional parameter called the overdispersion parameter. The zero-inflated Poisson regression or the zero-inflated

DOI: 10.55415/deep-2022-0025.v1

negative binomial regression is modified on top of the Poisson regression or the negative binomial regression to allow for a situation called zero inflation. This is useful when there is an excessive presence of zeros in the dependent variable. And, the zero-inflated Poisson itself still does not allow for the presence of the overdispersion. We selected the papers which usage counts greater than or equal to 1 in our sample. This indicates that a zero inflation is likely not the case so that the zero-inflated regression is not necessary. With these observations, therefore, we finally choose the negative binomial regression model (NBR) which is applicable to our sample.

The NBR model formula is:

$$P(Y=y)=\frac{\Gamma(\alpha+y)}{\Gamma(\alpha)\Gamma(y+1)}(\frac{1}{1+\theta})^{\alpha}(\frac{\theta}{1+\theta})^{y} \tag{1}$$

The expectation and variance formulas of NBR model are:

$$E(Y)=\alpha\theta \tag{2}$$

$$Var(Y)=\alpha\theta(1+\theta)=E(Y)(1+\theta) \tag{3}$$

Results

An Overview

Before we further analyse the dataset, we present some key descriptive statistics in Table 2 to provide an overview. From an overall perspective, it is obvious that our dataset presents a discrete distribution by range analyse and standard deviations analyse. Discovering the patterns from discrete data makes our follow-up analyse worth looking forward to. In particular, the Max usage count is 199. Thus, some of newly published papers also have the potential to be widely used for a brief period.

Table 2 Descriptive statistics of variables

Variable	N	Mean	SD	Min	Max
usage	35341	3.48	5.85	1	199
ti_num	35341	10.76	3.48	1	34
ti_tfidf	35341	3.08	1.97	0	14
ab_num	35341	127.93	43.36	8	616
ab_tfidf	35341	59.36	17.76	1	164
co_au	35341	6.83	40.54	1	2926
co_country	35341	1.45	1.41	1	112
co_organ	35341	2.74	6.28	1	535
nr	35341	51.44	37.62	0	1021
pg	35341	11.38	6.44	2	359
wc	35341	1.75	0.94	1	6
so_factor	35341	4.00	5.12	0.07	244.59
tc	35341	0.14	0.68	0	51

Fictitious variables are not included; SD stands for the standard deviation.

Correlation Analysis

Due to the highly discrete distributions of most of variables, Spearman correlation was applied in our study instead of Pearson correlation. The Spearman correlation coefficients among variables are summarised in Table 3. According to the results, most of variables passed the significance test. It means that most of them have certain influence on each

DOI: 10.55415/deep-2022-0025.v1

other. Most of correlation coefficients are appropriate, thus, we do not think that the follow-up models are affected by multiple collinearity.

Table 3 Spearman correlations between variables

	usage	*ti_num*	*ti_tfidf*	*ab_num*	*ab_tfidf*	*co_au*	*co_country*	*co_organ*	*nr*	*pg*	*wc*	*so_factor*	*tc*
usage	1												
ti_num	0.061 ***	1											
ti_tfidf	0.094 ***	0.546 ***	1										
ab_num	-0.030 ***	0.255 ***	0.163 ***	1									
ab_tfidf	0.042 ***	0.206 ***	0.223 ***	0.844 ***	1								
co_au	0.082 ***	0.215 ***	0.174 ***	0.207 ***	0.207 ***	1							
co_country	0.018 ***	0.002	-0.033 ***	0.051 ***	0.054 ***	0.238 ***	1						
co_organ	-0.003	0.065 ***	0.024 ***	0.127 ***	0.121 ***	0.473 ***	0.548 ***	1					
nr	0.137 ***	-0.050 ***	-0.022 ***	0.093 ***	0.157 ***	-0.025 ***	0.129 ***	0.077 ***	1				
pg	-0.008	-0.005	-0.017 **	0.140 ***	0.154 ***	-0.082 ***	0.111 ***	0.048 ***	0.501 ***	1			
wc	0.077 ***	0.012 *	0.037 ***	-0.015 **	0.025 ***	-0.018 ***	0.007	-0.005	0.011 *	0.049 ***	1		
so_factor	0.262 ***	0.039 ***	0.078 ***	0.102 ***	0.185 ***	0.282 ***	0.159 ***	0.162 ***	0.297 ***	0.051 ***	0.031 ***	1	
tc	0.010	-0.038 ***	-0.025 ***	0.057 ***	0.052 ***	0.011 *	0.060 ***	0.056 ***	0.070 ***	0.055 ***	-0.020 ***	0.096 ***	1

Fictitious variables are not included; *** p value <0.001, ** p value <0.01, * p value <0.05.

In the previous part, we have looked at correlation coefficients among the selected variables. Now we want to visualize the influence of a single factor on the usage metrics with scatterplots. The scatter plots drawn between usage metrics and other factors were too crowded due to the big dataset. So, it could not be interpreted well. Chetty et al. (2014) used a new way to deal this problem. There are four steps, first, groups the x-axis variable into equal-sized bins; second, computes the mean of the x-axis and y-axis variables within each bin; third, creates a scatterplot of these data points; fourth, draws the population regression line. We draw the binned scatter plots among them as shown in Figure 1. Not surprisingly, there are not linear rules between usage metrics and most of factors. Therefore, further regression analysis is necessary for our sample.

DOI: 10.55415/deep-2022-0025.v1

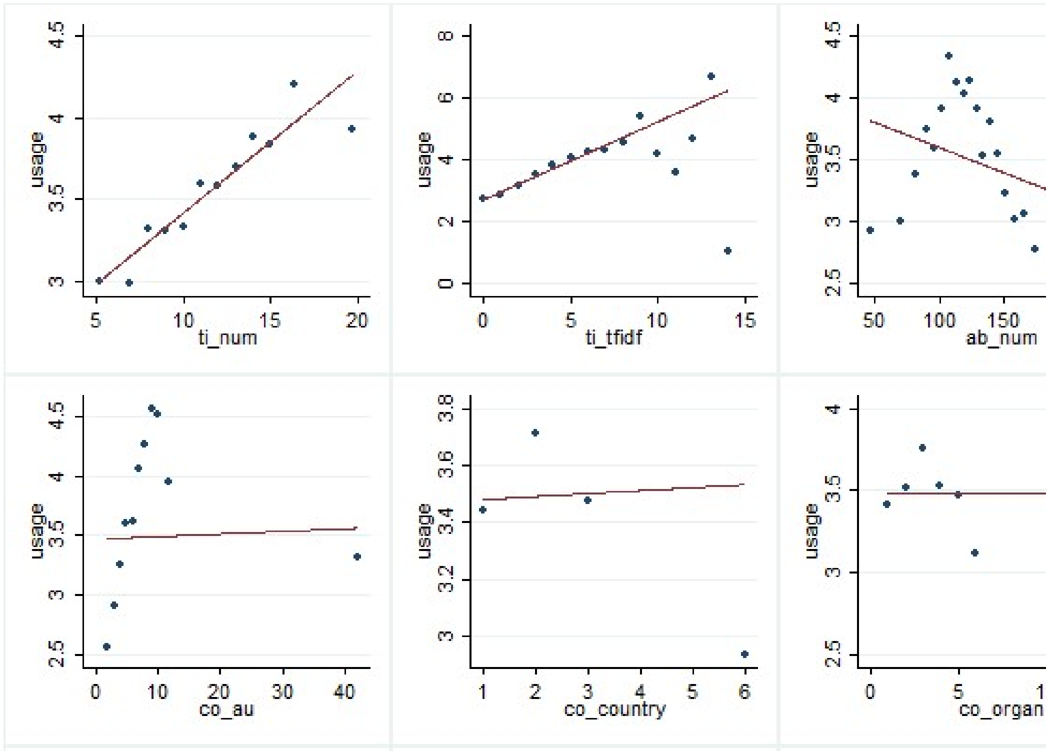

Figure 1 Binned scatterplots between usage metrics and other factors

Regression Analysis

Table 4 reports the regression analysis results. From an overall perspective, all models have satisfactory fitting effect. It shows that regression analysis results are reliable.

Firstly, we emphatically analyse of the title and abstract indicators which reflect the most valuable information of papers in the index database. The Model (1) includes only the title and abstract variables. Title length, Title distinction, Abstract length and Abstract distinction have a statistically significant effect on Usage counts ($p<0.001$). Except Abstract length, the effect on Usage counts are positive. Comparing the four variables, we can find some interesting phenomena. Long title is better than long abstract. And if papers using distinct vocabularies in their titles and abstracts are benefit for researchers to use them. We construct two square variables to test whether the more distinct vocabularies, the better. But unexpectedly, the Model (2) shows that whether Title distinction or Abstract distinction and Usage counts present an inverted U-shape curve.

Secondly, we analyse the impact of scientific cooperation as shown in the Model (3). Author collaboration scale and Organization collaboration scale have a statistically significant effect on Usage counts ($p<0.01$). Country collaboration scale doesn't pass the significance test at the 5 % level. In view of Author collaboration scale and Organization collaboration scale have opposite significant effect on Usage counts, we try to make a bold explanation. Despite large-scale author collaboration can attract new ideas, but a study published by a small number of organizations can present their research power. In other words, large-scale collaboration within organizations may be more attractive to researchers. One possible explanation for this finding is that a few influential academic

DOI: 10.55415/deep-2022-0025.v1

organizations can receive wide attention, so researchers may tend to use papers by many authors collaborating within those organizations.

Thirdly, the relationships between other variables and Usage counts are shown in the Model (4). Most of them have a statistically significant effect on Usage counts ($p<0.001$). However, Citation counts doesn't pass the significance test at the 5 % level. Document type, Funding text, Reference counts, Number of WoS categories and Journal impact factor have a positive significant effect on Usage counts. On the contrary, Open access and Page counts have a negative significant effect on Usage counts. There are several common impressions in Scientometrics. "Article" papers are considered more original, and "Review" papers are more like summaries of the previous research works. Those papers which be funded can access to more academic resources. More references may provide users with other relevant research papers. If a journal was labelled more WoS categories, it may mean that the journal fuses interdisciplinary approaches and attracts the attention of a larger research community. Of course, those journals with high impact factors have been attracting attention. From above views, it is not surprising that those papers can attract more users. In addition, we get some interesting results. "Open Access" do not advantageously to increase Usage counts in our sample. Obviously, most of researchers would like to see full text information of newly published papers available and then can use them freely. As Chen (2017) found, comparing with the pay-for-access WoS, users prefer to visit publisher websites by the free. It is hard to understand that long papers have a negative impact on Usage counts. The Kendall correlation coefficient between Document type and Page counts is significantly negative. Therefore, one possible explanation is that "Review" papers more likely have long pages. As mentioned earlier, due to cumulative effect of Citation counts, it is not surprising that there isn't strongly significant between Citation counts and Usage counts in our sample.

Finally, we incorporate all independent variables into the Model (5) to further support our previous analysis results in this section. Regression coefficients show the important of every influence factor. And we have selected two important factors to summarise this section. "Open Access" is the most negative factor. In other words, the usage counts of open access papers will reduce significantly. By contrast, "Funding text" is the most positive factor. This is an important criterion that can affect researchers use a paper.

Table 4 Models estimation results

	Model (1) *usage*	Model (2) *usage*	Model (3) *usage*	Model (4) *usage*	Model (5) *usage*
ti_num	0.0167***	0.0161***			0.0196***
	(5.85)	(5.65)			(7.79)
ti_tfidf	0.0413***	0.0739***			0.0339***
	(8.21)	(5.98)			(8.08)
ti_tfidf2		-0.00438**			
		(-2.94)			
ab_num	-0.00975***	-0.00969***			-0.00633***
	(-26.41)	(-26.34)			(-19.89)
ab_tfidf	0.0224***	0.0299***			0.0135***
	(24.08)	(13.20)			(17.37)
ab_tfidf2		-0.0000610***			
		(-3.64)			
co_au			0.00191**		0.00268***
			(3.07)		(4.35)

DOI: 10.55415/deep-2022-0025.v1

co_country			0.0224		-0.00524
			(1.88)		(-0.51)
co_organ			-0.0143**		-0.0191**
			(-2.83)		(-2.78)
dt				0.119***	0.131***
				(4.06)	(4.30)
fu				0.289***	0.246***
				(17.01)	(14.79)
oa				-0.509***	-0.481***
				(-20.78)	(-19.05)
nr				0.00427***	0.00411***
				(15.05)	(14.98)
pg				-0.0216***	-0.0193***
				(-14.59)	(-13.22)
wc				0.130***	0.112***
				(16.42)	(14.53)
so_factor				0.0601***	0.0582***
				(20.30)	(20.36)
tc				0.0122	0.0224*
				(1.25)	(2.28)
cons	0.829***	0.570***	1.241***	0.431***	0.190***
	(25.03)	(8.80)	(79.37)	(11.88)	(4.39)
lnalpha cons	-0.352***	-0.354***	-0.282***	-0.500***	-0.556***
	(-22.21)	(-22.48)	(-17.17)	(-31.84)	(-34.75)
N	35341	35341	35341	35341	35341

Z statistic in parentheses; *** p value <0.001, ** p value <0.01, * p value <0.05.

Conclusion and Discussion

In this paper, we have selected some factors to measure the motivation of using newly 4 weeks published papers from WoS database. We attempted to explain, statistically, how various factors affect the number of usage counts.

The conclusions of our analyses are detailed as follows. Title and abstract indicators reflect the main information of papers, so it is better to use an appropriate number of distinct and novel vocabularies to write title and abstract. An influential academic organization can be widely concerned, therefore, those papers which published by a few famous research organizations will have broader exposure. “Article” papers and “Funded” papers are two of most important factors which effect on Usage counts. Open access papers have a great possibility to be used directly, so their usage counts reduce significantly in WoS database. References can provide researchers with relevant studies and increase their interest to keep using. Those papers which have long pages more like are “Review” papers, as a result, them have a negative impact on Usage counts. And those journals which labelled more WoS categories fuse interdisciplinary approaches and attract more usage. Obviously, high IF journals also have been attracting wide attention. The guide function of citation counts in a short time is not fully reflected, so it has only a marginal effect on usage.

We would like to acknowledge the limitations of our study. It is essential to make clear that all findings presented in this article relate to the sample from WoS core collection database. Are the usage characteristics similar in different data platforms? This is one of the questions we want to answer in the future. Moreover, the sample may exhibit

DOI: 10.55415/deep-2022-0025.v1

characteristics that is different from those of older published papers as well. Only analyses of other databases and articles published long enough time may reveal the extent to which the characteristics of usage count have a more general validity. In addition, we can only get the quantity of usage counts from WoS core collection database. But more necessary details about the usage count, such as institutional user address, behaviour of "clicking" or "saving", and specific time are not available. Furthermore, we found that recognizing the actual publication day of per papers was unrealistic in our sample. It was due to this reason have we helplessly assumed that those articles were published within 4 weeks without any difference. Obviously, this assumption was ideal and unconvincing. For some reasons, we don't investigate the difference among different subjects. It is obvious that the characteristics of different subjects, and most likely will, exhibit a completely different pattern to what we have presented in this paper. When it comes to future studies, it is worth to investigate the difference among different subjects. Therefore, the conclusions presented in this study can best be regarded as a case study.

Last but not least, as previously stated, it is a complicated research topic to find out which factors drive the increased usage of newly published papers. Although conclusions are tentative, them can be given a referable status in future studies. We hope that this paper would provide a new perspective to promote the research on usage metrics about which factors drive the increased usage of newly published papers.

References

Bollen, J., & Sompel, H. V. D.. (2006). Mapping the structure of science through usage. Scientometrics, 69(2), 227-258.

Bollen, J., Sompel, H. V. D., Smith, J. A., & Luce, R.. (2005). Toward alternative metrics of journal impact: a comparison of download and citation data. Information Processing & Management, 41(6), 1419-1440.

Brody, T., & Harnad, S.. (2005). Earlier web usage statistics as predictors of later citation impact. Journal of the Association for Information Science & Technology, 57(8), 1060-1072.

Chen, B.. (2018). Usage pattern comparison of the same scholarly articles between web of science (wos) and springer. Scientometrics, 115(2), 519-537.

Chetty, R., Friedman, J. N., Leth-Petersen, S., Nielsen, T. H. & Olsen, T.. (2014). Active vs. Passive Decisions and Crowd-Out in Retirement Savings Accounts: Evidence from Denmark. Quarterly Journal of Economics, 129(3), 1141-1219.

Chi, P. S., & Glänzel, W.. (2017). An empirical investigation of the associations among usage, scientific collaboration and citation impact. Scientometrics, 112(1), 403-412.

Coats, A. J. S.. (2005). Top of the charts: download versus citations in the international journal of cardiology. International Journal of Cardiology, 105(2), 123-125.

Davis, P. M., & Fromerth, M. J.. (2007). Does the arxiv lead to higher citations and reduced publisher downloads for mathematics articles?. Scientometrics, 71(2), 203-215.

Glänzel, W., & Gorraiz, J. (2016).Usage metrics versus altmetrics: confusing terminology? Scientometrics, 102(3), 2161-2164.

Gorraiz, J., Gumpenberger, C., & Schloegl, C. . (2014). Usage versus citation behaviours in four subject areas. Scientometrics, 101(2), 1077-1095.

Guerrero-Bote, V. P., & Moya-Anegón, Félix. (2014). Relationship between downloads and citations at journal and paper levels, and the influence of language. Scientometrics,101(2), 1043-1065.

He, L., & Han, Z.. (2017). Do usage counts of scientific data make sense? an investigation of the dryad repository. Library Hi Tech, 35(2), 332-342.

Henneken, E. A., & Kurtz M J. (2017). Usage Bibliometrics as a Tool to Measure Research Activity . ArXiv. Retrieved from https://arxiv.org/ftp/arxiv/papers/1706/1706.02153.pdf

Hood, W. W., & Concepción Shimizu Wilson. (2005). The relationship of records in multiple databases to their usage or citedness. Journal of the American Society for Information Science and, Technology, 56(9), 1004-1007.

Jamali, H. R., & Nikzad, M.. (2011). Article title type and its relation with the number of downloads and citations. Scientometrics, 88(2), 653-661.

Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., & Murray, S. S.. (2005). Worldwide use and impact of the nasa astrophysics data system digital library. Journal of the American Society for Information Science and, Technology, 56(1), 36-45.

Line, M. B. (1993). Changes in the use of literature with time—obsolescence revisited. Library Trends, 41(4), 665-683.

Markusova, V., Bogorov, V., & Libkind, A.. (2018). Usage metrics vs classical metrics: analysis of russia's research output. Scientometrics, 114(2), 593–603.

Mcdonald, J. D.. (2007). Understanding online journal usage: a statistical analysis of citation and use. Journal of the American Society for Information Science and Technology, 58(1), 39-50.

Moed, H. F., & Halevi, G.. (2016). On full text download and citation distributions in scientific-scholarly journals. Journal of the Association for Information Science and Technology, 67(2), 412-431.

Moed, H. F.. (2005). Statistical relationships between downloads and citations at the level of individual documents within a single journal. Journal of the American Society for Information Science and, Technology, 56(10), 1088-1097.

Nicholas, D., Huntington, P., Dobrowolski, T., Rowlands, I., Hamid, R. J. M., & Polydoratou, P.. (2005). Revisiting 'obsolescence' and journal article 'decay' through usage data: an analysis of digital journal use by year of publication. Information Processing & Management, 41(6), 1441-1461.

Schloegl, C., & Gorraiz, J.. (2010). Comparison of citation and usage indicators: the case of oncology journals. Scientometrics, 82(3), 567-580.

Schloegl, C., & Gorraiz, J.. (2010). Comparison of citation and usage indicators: the case of oncology journals. Scientometrics, 82(3), 567-580.

Schloegl, C., & Gorraiz, J.. (2011). Global usage versus global citation metrics: the case of pharmacology journals. Journal of the American Society for Information Science and Technology,62(1), 161-170.

Schloegl, C., Gorraiz, J., Gumpenberger, C., Jack, K., & Kraker, P. (2014). Comparison of downloads, citations and readership data for two information systems journals. Scientometrics,101(2), 1113-1128.

Silton, & Kate. (2014). Best practices for increasing ejournal usage. Serials Review, 40(3), 158-160.

Sun, Y., & Xia, B. S.. (2016). The scholarly communication of economic knowledge: a citation analysis of Google Scholar. Scientometrics, 109(3), 1965-1978.

Thelwall, M.. (2012). Journal impact evaluation: a webometric perspective. Scientometrics, 92(2), 429-441.

Wan, G., & Liu, Z.. (2010). Knowing your users: the value of article database usage analysis. Learned Publishing, 23(3), 225-236.

Wang, X., Fang, Z., & Sun, X.. (2016). Usage patterns of scholarly articles on web of science: a study on web of science usage count. Scientometrics, 109(2), 917-926.

Wang, X., Liu, C., Mao, W., & Fang, Z.. (2015). The open access advantage considering citation, article usage and social media attention. Scientometrics, 103(2), 555-564.

Wang, X., Mao, W., Xu, S., & Zhang, C.. (2014). Usage history of scientific literature:nature metricsand metrics of nature publications. Scientometrics, 98(3), 1923-1933.

Wical, S. H., & Vandenbark, R. T.. (2015). Combining citation studies and usage statistics to build a stronger collection. Library Resources & Technical Services, 59(1), 33-42.

Wood-Doughty, A., Bergstrom, T., & Steigerwald, D. (2017). Do download reports reliably measure journal usage? Trusting the fox to count your Hens? UC Santa Barbara: Department of Economics. Retrieved from https://escholarship.org/uc/item/1f221007

Zhao, S. X., Lou, W., Tan, A. M., & Yu, S.. (2018). Do funded papers attract more usage? Scientometrics, 115(1), 153-168.