Scientometrics (2021) 126:9457–9476
https://doi.org/10.1007/s11192-021-04178-5

Usage, content and citation in open access publication: any interaction effects?

Carlos Vílchez-Román[1] · **Arístides Vara-Horna**[2]

Received: 1 March 2021 / Accepted: 1 October 2021 / Published online: 24 October 2021

Abstract

This study aims to validate an empirical model, at document level, that explains the interaction among content, usage, and citation within open access publications. The PLoS site was the information source for this study. Using an R API (Application Programming Interface) for PLoS ONE, 776,465 records were downloaded on August 24, 2018. Those records (from 2006 to 2018) were organized according to the PLoS' thematic areas. The empirical framework was validated using mediation analysis. For computing the parameters' significance, bootstrapping with 500 replications for the general model and each thematic area was used. When usage was included as the mediating variable within the model, the total effects of cognitive and social variables got better predictive capability, as expressed by the explained variance of citation ($R^2 = 0.282$) and usage ($R^2 = 0.333$). The same trend was observed for the indirect effects after carrying out the mediation analysis by categories. Promotion campaigns of scientific publication should reinforce the widespread adoption of easy-to-use social media because, besides the velocity and variety of diffusion channels, the extended use guarantees that journal's papers will reach increasing audiences. This is one of the first studies that analyze the interaction effects of variables at the article-level within open access publications.

Keywords Open access publications · Mediation analysis · Social media · Altmetrics

Mathematics Subject Classification 62

JEL Classification D80

Introduction

With the expansion of subscription-based services and Application Programming Interfaces (APIs) developed for programs like Python and R, scholars have found an alternative (to citation-based approaches) for analyzing the use of research results. The analysis of

✉ Carlos Vílchez-Román
cvilchez@pucp.edu.pe

1 CENTRUM Católica Graduate Business School (CCGBS), Lima, Peru

2 Universidad San Martín de Porres (USMP), Lima, Peru

the repercussion of research results in digital environments like those provided by social media is known as altmetrics (Priem et al., 2010). Altmetrics is a discipline that studies the impact and (academic and non-academic) use produced by a scientific publication and the pieces of research related to it (Williams, 2017). This classification of the academic and the non-academic side of research use highlights the interest in fields or topics with different levels of specialization: general users prefer popular topics while scholars are more interested in very specialized fields (Roemer & Borchardt, 2015). On the academic side, altmetrics focus primarily on the "reads" or "downloads" of a paper's full text, or the inclusion of paper information into bibliographic management software (e.g., Mendeley or Zotero), or usage indicators available in databases like Scopus, or Web of Science, and to a lesser degree in mentions on social media. Regarding the non-academic side, altmetrics analyzes, mainly but not exclusively, the dissemination and use of research papers, on social media (e.g., mentions in Facebook's public pages or Twitter accounts), as well as the presence in blogs or news sites because non-academic users are more interested in trending topics like climate change or weight loss (Roemer & Borchardt, 2015).

Based on the data availability, research managers can determine the reception of research on the users' side, more than the academic impact measured by citation counting. This understanding includes how the measures mentioned above interact among them to identify the drivers of research impact. For this reason, this study introduces an empirical model for examining the interaction between research use and impact. Studies from an altmetric perspective increase every year, as can be corroborated in two open access multidisciplinary databases: Dimensions and Google Scholar (GS). In the first one, there were 11 documents published in 2012 with the term altmetrics in the title and abstract fields. In 2018, 2019, and 2020, the number went up to 186, 182, and 263. In GS, the number of documents with the term altmetrics in the title field increased from 54 in 2012 to 160 in 2018. However, this growing importance of altmetrics as a research topic has not produced theory-driven studies that test conceptual or empirical models because most of the investigations are descriptive.

Altmetrics can be explained as a descriptive orientation (different from a theory-driven one) and a discipline because it has not passed more than ten years since the publication of the Altmetrics Manifesto (Priem et al., 2010), while disciplines like bibliometrics or scientometrics have more than fifty years of history and research tradition.

Literature review

Given the massive volume of altmetric data generated faster than citations and the variety of information gathered from several digital channels, the altmetric zeitgeist should be understood as the big data trend's bibliometric side. There are now literally terabytes and terabytes of altmetric data, but there is no empirical framework that organizes altmetric data neither clarifies the relationships among the different altmetric activity manifestations. Therefore, in the altmetric landscape, we have reached the same conclusion that the one observed in other areas characterized by high volumes of data produced very fast and from different information sources: big data is not enough, it is always necessary to give a sense of our data by adopting a theoretical or empirical perspective. Nevertheless, there were some attempts to explore the relationship among altmetric indicators and predict the academic impact based on those indicators.

Bivariate correlation-based studies on altmetrics

Besides the descriptive studies on the academic and social impact measured by altmetrics (Gontijo & De Araujo, 2021; Maricato & Vilan-Filho, 2018), other studies have explored the relationship among those variables. The first group of studies explored the associations among altmetric indicators and citations in multidisciplinary databases like Dimensions, Scopus, and Web of Science (WoS). Researchers have reported low to moderate correlation with citation according to the Spearman's rho: $\rho = 0.373$ and 0.201 for articles posted in specialized web pages and blogs (Thelwall et al., 2013), $\rho = 0.167$ for mentions in Facebook pages, $\rho = 0.113$ and 0.148 for mentions in Twitter accounts (Haustein et al., 2015; Wei & Noroozi, 2020), $\rho = 0.540$ and 0.650 for downloads (Guerrero-Bote & Moya-Anegon, 2014; Moed & Halevi, 2016) and $\rho = 0.494$ and 0.701 for readings in the Mendeley bibliographic management software (Thelwall, 2017; Wei & Noroozi, 2020). In other words, previous studies have analyzed the relationship between citation and research use (academic and non-academic use).

Regression-based studies on altmetrics

Besides the correlation-based perspective, other studies predicted Scopus and WoS' citations using altmetric indicators as predicting variables, like those using regression models. There are two perspectives regarding regression-based studies examining altmetric indicators: one from the informetric and scientometric side and the other from the medical sciences approach. In the first group, authors predicted citations in Scopus and WoS using linear regression models and negative binomial regression, the latter because of the log-normal distribution of citations (Bornmann & Haunschild, 2018; Thelwall & Nevill, 2018; Thelwall & Wilson, 2014; Wang et al., 2020a, b). For example, Thelwall and Nevill examined publications with higher Altmetric scores, taking into account citations in Scopus and indicators in 27 broad fields (e.g., accounting, artificial intelligence, economics and econometrics, or history). For the prediction analysis, they used the ordinary least squares (OLS) linear regression model. According to the results, among the associated factors with citations, the most important ones are CiteULike ($\beta = 0.65$), Mendeley ($\beta = 0.55$), blogs ($\beta = 0.49$), tweets ($\beta = 0.36$) and news ($\beta = 0.26$), given that the regression coefficients obtained statistical significance (Thelwall & Nevill, 2018).

Researchers from medical sciences have analyzed the contribution of associated factors that are more relevant to the general public, like mentions in blogs or posts in social media. In this sense, the primary associated factors identified were bloggers, Twitter and Mendeley (Maggio et al., 2018; Smith et al., 2019), as well as public policy documents, Google+ and Wikipedia (Sathianathen et al., 2020). In most of these studies, altmetric indicators were obtained from the data provider Altmetric.com. In the first study (Maggio et al., 2018), researchers analyzed 2486 articles with altmetric indicators, published in seven health professions education journals, and indexed in WoS. Those authors used two regression models: a negative binomial regression with citations as the predicted variable and an OLS linear regression model with access counts as the predicted variable. They found three variables with the highest incident rate ratio: blogs ($\beta = 1.13$, $p < 0.05$), Twitter and Mendeley (both with $\beta = 1.01$, $p < 0.001$). In another research (Sathianathen et al., 2020), 2033 articles published in 10 urology journals between June 2016 and June 2017 were analyzed using a multivariate linear regression model with a forward stepwise regression method. They found four statistically significant variables with the highest prediction power: public policy documents ($\beta = 28.7$), Google+ ($\beta = 18.2$), Wikipedia ($\beta = 12.9$) and blogs ($\beta = 7.8$).

Alternative analytical approaches for altmetrics

The no inclusion of interaction effects in the regression models reported in the literature can be explained by the lack of theoretical or empirical models that take into account the mutual effect among variables or because researchers consider that indirect effects are not necessary for understanding the dynamics of altmetrics. On the other side, even though altmetric indicators are measured directly, they can be grouped into latent variables by adopting structural approaches, considering that mentions in Facebook or Twitter describe a higher level of social interaction rather than just the presence in social media. Therefore, it is necessary to explore alternative analytical approaches for a discipline of growing importance for research managers and people interested in the output of research activities.

In the area of information science, few studies have worked with mediation analysis for examining the effect of online reviews on product sales (Li et al., 2019) or the mediating effect of discussion frequency on networking heterogeneity when using social media (Strauβ et al., 2020). Regarding citation as the predicted variable, we found two recent publications that adopt this methodological approach (Ebrahimy et al., 2016; Wu et al., 2021; Zhang et al., 2019). In one of the studies, researchers adopted the social capital theory to explain the associated factors with the academic impact of supply chain management scholars ($n = 450$). The academic impact was calculated from the citation impact factors of journals where the scholar published their works and social capital was measured as the number of followers in ResearchGate. Mediation analysis suggests that the scholar's social capital has a strong explanatory power ($\beta = 0.66$, $p < 0.001$) on the relationship between research skill and academic impact (Zhang et al., 2019). A more recent study explored the interaction among innovation, chief technology officer (CTO)'s gender and characteristics, firm characteristics, industry and year effects within an innovation corporate environment. Data from 5508 female CTOs were obtained from the BoardEx database and the firm's information was extracted from Compustat and Google Patents. Innovation was operationalized by three indicators: count of a firm's patent applications in a year, number of citations received on all firm's patents in a given year (including self-citations), and number of outside citations. According to this study, transformational leadership mediates the relationship ($\beta = 0.66$, $p < 0.001$) between female CTOs and firm's patent citation (Wu et al., 2021).

No study included interaction effects (i.e., mediating or moderating variables) for predicting the research impact from altmetric indicators, only one recent study used the structural equation modeling for examining the mediating effect of usage between mentions in social media and citation in Scopus (Vílchez-Román et al., 2020). For the empirical model tested in this work, the social media component was the driver for the mediated process that explained the relationship among the analyzed variables. The altmetric records ($n = 3186$) used in this research were restricted to PLoS publications from five South American countries: Bolivia, Chile, Colombia, Peru, and Venezuela. Therefore, the question remains as to whether the model will work at a larger scale than the Andean countries.

Research question

From the literature review, we found one study that examined the relationship between mentions in social media and citation by including a mediating effect (Vílchez-Román et al., 2020). Even though the structural model introduced in that study got an acceptable explained variance and appropriate goodness-of-fit indicators, it worked for open access publications from five Andean countries. It is unknown whether the model will work at a larger scale, for example by including international publications. To have a broader perspective on the interaction among altmetric indicators we included a content-related variable and framed the following research question: what is the relationship among mentions in social media, the content of the abstract, usage and download and received citations in Scopus of papers published in open access?

Hypotheses

Concerning the interaction between social media and usage/download as driver of citation, there is not enough empirical evidence for establishing a causal path among those variables, because citation can also be explained by reading papers downloaded by bibliographic management software like Mendeley or Zotero (Akella et al., 2021; Thelwall & Nevill, 2018) or searching academic references in WoS (Wang et al., 2020a, b). Furthermore, previous studies have established an association between papers reading and usage/download (Haustein & Larivière, 2014) and social media activity (Cho, 2021). In this sense, the use of bibliographic management software (this includes reading the papers downloaded with that software) can also be a plausible explanation for the mediating role of the usage/download with the citation in Scopus or WoS.

Therefore, omitting the mediating effect of the use of bibliographic management software (mostly operationalized as Mendeley's readings) can be a limitation of an empirical model that explain the interaction among social media activity, usage/download and citation in multidisciplinary databases. In order to explore the relationships stated in the research question from a mediating effects perspective, we disaggregated the question into three statements to understand to what degree the download and usage mediated the relationship between social media and content with the citation in Scopus (see Fig. 1).

H1 Social media indicators for open access publications have a medium and positive association with Scopus' citation.

H2 Abstract's content of open access publications has a medium and positive association with citation.

H3 Download and usage of open access publications mediate the association between social media indicators and abstract's content with the citations in Scopus.

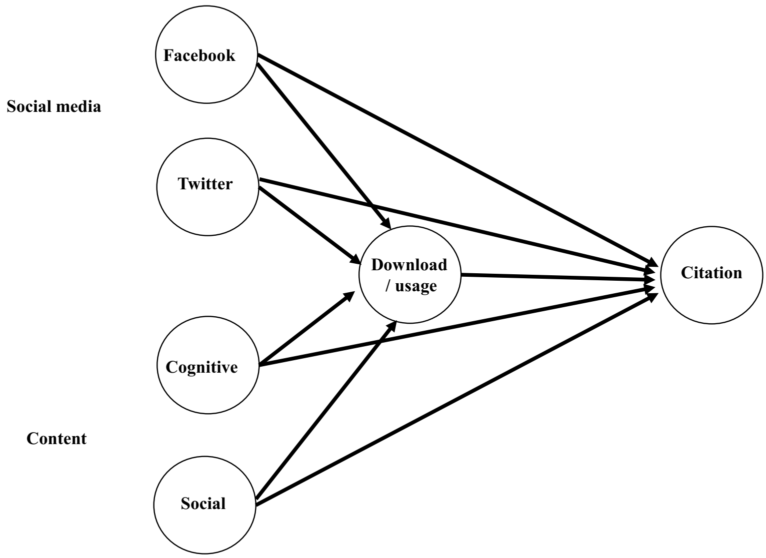

Fig. 1 Empirical framework with mediation effects

Material and methods

In this study, we validated an empirical model at the document level that explains the interaction among content, usage, and citation in Scopus, according to the information available from the Public Library of Science (PLoS). In August 24, 2018, we downloaded 776,465 PLoS records for the period 2006–2018. Downloaded publications were divided in the following categories: Biology and life sciences = 221,845; Computer and information sciences = 23,947; Earth sciences = 19,613; Ecology and environmental sciences = 21,149; Engineering and technology = 18,524; Medicine and health sciences = 174,666; People and places = 37,939; Physical sciences = 79,580; Research and analysis methods = 140,943; Science policy = 3513; Social sciences = 34,746.

We chose PLoS as the information source because it hosts journals with worldwide papers, published in open access. This publishing model means that any researcher could read and use them free. Likewise, given our hypotheses' nature, we considered that the data available from this source were appropriate to test them. However, we did not include metrics on papers "reading" because these data were not available from this information source. Despite its broad coverage, we must mention that PLoS journals are focused on medicine, science, and technology and fully published in English; in that sense, studies from humanities and non-English speaking countries are underrepresented in this information source.

Data collection and preprocessing

We used the R program rplos for downloading the article-level data (e.g., download and citation counts, and the title and abstract for each publication). Rplos is an Application

Programming Interface (API) developed to access and download the information available from PLoS website, which publishes articles from the following open access journals: PLoS ONE, PLoS Biology, PLoS Computational Biology, PLoS Genetics, PLoS Medicine, PLoS Neglected Tropical Diseases and PLoS Pathogens. We recorded that information as comma-separated values (CSV) files, one for each of the 11 categories mentioned at the beginning of this section. Given that our empirical model's two predictor variables come from the publication's content, synthesized in the abstract, in the preprocessing step, we excluded those PLoS records that did not have an abstract, even though they got citations in the Scopus database. After filtering, 758,419 records were considered for the analysis (Biology and life sciences = 216,872; Computer and information sciences = 23,348; Earth sciences = 19,225; Ecology and environmental sciences = 20,693; Engineering and technology = 18,120; Medicine and health sciences = 170,673; People and places = 36,535; Physical sciences = 78,373; Research and analysis methods = 138,535; Science policy = 2753; Social sciences = 33,292). Regarding textual information, we converted the abstract's content into lowercase to maintain consistency during content analysis. Since we were interested in examining the abstract's full content, we did not use any stopword file during textual data processing.

Variables measurement

We operationalized the abstract's content as the percentage of words related to cognitive and social dimensions. Scientific activity creates new knowledge (the cognitive component) and is mainly carried out by research teams or scholars sharing data and resources (social component). For calculating the percentage of cognitive and social related terms, we used the internal dictionary of the Linguistic Inquiry and Word Count (LIWC) software, version 2015 (Pennebaker et al., 2001). The validity and reliability of the LIWC dictionary have been established previously (Graybeal et al., 2002; Mehl & Pennebaker, 2003). LIWC's dictionary's cognitive component of LIWC's dictionary includes six elements (causation, insight, discrepancy, inhibition, tentative, and certainty), while the social component five elements (communication, references to people, friends, family, and humans). Even though there are more complex methods for processing large volumes of text, like the topic modeling technique known as latent Dirichlet allocation (LDA), most of them use the bag of words approach by which words are extracted from their context to be analyzed with the text mining algorithms (Blei et al, 2003; Ekinci & Omurca, 2020; Hu et al., 2014), the same way the LIWC carries out its content analysis. For this reason, we decided to work with the two LIWC's categories for psychological process mentioned above, since previous studies also used the cognitive and social categories for analyzing scientific texts, like the abstract of papers, manuscript reviews, or science students' writings (Bjekić et al., 2014; Bornmann et al., 2012; Smith-Keiling & Hyun, 2019).

Social media activity had two elements: mentions in Facebook's public pages and Twitter accounts. It was measured as the count of mentions in each one. The download and usage were operationalized as the count of downloads and readings for each publication. Finally, the citation was operationalized as the citations received in Scopus.

Before testing the empirical model with mediation analysis, we randomly deleted duplicated records (because several publications were classified in more than one category) using a SPSS macro and computed central tendency and dispersion measures. In this way the records deletion maintained the same proportion of records to be analyzed for all PLoS categories. Then we obtained a correlation matrix to confirm variables of the model that were associated. The final dataset had 235,384 records. In order to assure the replicability of the data, we ran the same path analysis with the filtered dataset ($n = 758{,}419$ records) and found similar results as those reported in this study, when considering the positive or negative direction and the statistical significance of the coefficients for direct, indirect and total effects (see "Appendix A"). The filtered and analyzed datasets are available on the links included in the supplemental material (see "Appendix B").

Statistical procedure: mediating analysis

Mediation analysis is a procedure to determine the mechanism through which independent variables influences dependent variables (Hayes, 2013). In that sense, researchers assume that the independent variable affects the mediator, which in turn, affects the dependent variable. Therefore, we can assume that the relationship between the independent and dependent variable is indirect (Baron & Kenny, 1986).

We worked with the standard method for computing OLS regression coefficients and obtained the z estimates for direct and indirect effects, with a confidence interval of 95%, and the R^2 for predicted and mediator variables. We followed established guidelines to determine whether the mediation effect was complete or partial (Baron & Kenny, 1986; Hayes, 2013; Iacobucci, 2008; Jose, 2013). We used bootstrapping with 500 replications and the type bias-corrected percentile to obtain each path coefficient's statistical significance. For handling missing values, we applied the full information maximum likelihood. We compared the OLS mediation analysis results with a negative binomial regression to identify differences. Finally, we obtained the direct, indirect, and total effects for the general model and specific model for each topic. This analysis was carried out with JASP 0.14.1 (JASP Team, 2020).

Results

Descriptive analysis

Most of the variables included in the empirical model showed a high dispersion, except for the two variables related to the abstract's content: cognitive (9.62 ± 3.09) and social (3.81 ± 2.36) for all records included in the dataset. The descriptive analysis for thematic areas, citation, social media indicators, and download/usage showed higher values for the dispersion measure used in the study, but again social and cognitive areas showed low dispersion (see Table 1).

Bivariate correlation analysis on altmetric indicators

Regarding correlation analysis, contrary to expected, social media mentions showed moderated and low statistical significance ($r = 0.383$). We also obtained moderated values for measures related to social and cognitive content ($r = 1.31$). However, we got higher values for the correlation between citation and usage ($r = 0.477$), a previously reported relationship in the literature (Ortega, 2015; Schlögl & Gorraiz, 2010; Schlögl et al., 2014). It must be noted that from these three cited works, only Ortega (2015) used article-level data. Therefore, despite the high dispersion of citation, social media indicators, and download/

Table 1 Descriptive statistics (mean and standard deviation for thematic areas)

Thematic area	Papers	Citation	Twitter	Facebook	Usage	Social	Cognitive
Biology and life science	13,344	20.75	3.86	4.70	5534.04	3.88	9.48
		40.37	30.04	62.07	10,354.59	2.27	2.94
Computer and information sciences	1247	23.38	4.19	3.92	5726.49	4.39	10.47
		48.54	14.16	36.26	7019.85	2.32	3.04
Earth sciences	802	17.63	10.23	11.34	6315.84	3.60	9.33
		34.47	30.09	41.33	10,541.40	2.38	2.86
Ecology and environmental sciences	4183	16.62	5.42	6.05	4842.95	3.66	10.10
		36.70	18.86	32.06	9014.60	2.03	2.89
Engineering and technology	982	18.62	4.88	4.23	5257.44	3.54	9.99
		53.38	15.32	21.53	6684.27	2.37	3.21
Medicine and health sciences	37,095	17.48	2.77	3.93	4603.88	3.59	9.21
		35.09	24.07	61.87	7883.58	2.06	2.70
People and places	12,793	14.38	4.24	6.69	4696.16	4.26	9.23
		32.57	20.42	61.57	9692.81	2.76	2.78
Physical sciences	25,529	14.95	2.75	4.83	4313.59	3.23	9.42
		30.13	26.00	111.46	7999.81	2.04	2.92
Research and analysis methods	107,317	17.37	1.91	2.47	4396.39	3.56	9.25
		38.40	15.71	31.79	6075.24	1.95	2.77
Science policy	1732	20.28	15.89	7.91	7236.92	3.55	11.01
		87.30	78.74	120.07	18,922.05	2.84	3.55
Social science	30,360	12.95	9.34	10.37	5878.98	5.29	11.64
		41.60	59.66	156.571	21,412.82	3.34	3.98
Total	235,384						

For all variables, the arithmetic mean is displayed above the standard deviation

usage, we decided that the moderated correlations provided empirical support for the mediation analysis. The variation was low and moderate in the two variables that have not been analyzed previously in similar frameworks: cognitive and social content (see Table 2).

Table 2 Correlation matrix of variables included in the model (Pearson correlations)

	Twitter	Facebook	Social	Cognitive	Usage	Citation
Twitter	–					
Facebook	0.383 ***	–				
Social	0.067 ***	0.020 ***	–			
Cognitive	0.048 ***	0.017 ***	0.131 ***	–		
Usage	0.557 ***	0.350 ***	0.071 ***	0.039 ***	–	
Citation	0.088 ***	0.038 ***	0.029 ***	0.012 ***	0.477 ***	–

***$p < .001$

Regression-based mediating analysis on altmetric indicators

Even though the estimates for direct effects were negative but significant for social indicators, when usage was included as mediating variable within the model, total effects (i.e., direct and indirect effects) of cognitive and social variables obtained better predictive capability (see Table 3 and Fig. 2) and acceptable values for the explained variance of citation ($R^2 = 0.282$) and usage ($R^2 = 0.333$). The same trend was observed when we obtained the regression estimates for the indirect effects after carrying out the 11 thematic areas' mediation analysis. The stronger mediation effects were observed in the path social content → download/usage → citation for earth sciences ($\beta = 0.053$, $p < 0.00$), computer and information sciences ($\beta = 0.028$, $p < 0.00$), and research and analysis methods ($\beta = 0.021$, $p < 0.00$). From 44 mediation effects included in the empirical framework, nine did not get statistical significance, despite most of the variables showed high dispersion. According to these results, it was observed a complete mediation for cognitive and social, but partial for social media related indicators (see Table 4).

Discussion

Our results provided empirical support for the model introduced in this study: the social content (rather than the cognitive one) is the driver of a mediated process that predicts the download and usage of open access publications, and their later citation in Scopus. We found that Twitter mentions have a negative and significant direct effect on the citation, but cognitive and social content have no direct effect on citation. However, when usage

Table 3 Parameter estimates of mediation analysis for the general model

	Estimate	SE	*z*-value	*p*	95% confidence interval	
					Lower	Upper
Direct effects						
Twitter → Citation	− 0.298	0.000	− 106.665	0.000	− 0.298	− 0.298
Facebook → Citation	− 0.047	0.000	− 50.876	0.000	− 0.047	− 0.047
Social → Citation	0.015	0.000	0.530	0.596	0.015	0.015
Cognitive → Citation	− 0.010	0.000	− 0.475	0.635	− 0.010	− 0.010
Indirect effects						
Twitter → Usage → Citation	0.406	0.000	200.356	0.000	0.406	0.406
Facebook → Usage → Citation	0.050	0.000	84.488	0.000	0.050	0.050
Social → Usage → Citation	0.349	0.000	19.823	0.000	0.349	0.349
Cognitive → Usage → Citation	0.069	0.000	5.132	0.000	0.069	0.069
Total effects						
Twitter → Citation	0.108	0.000	37.737	0.000	0.108	0.108
Facebook → Citation	0.002	0.000	2.232	0.026	0.002	0.002
Social → Citation	0.364	0.000	10.882	0.000	0.364	0.364
Cognitive → Citation	0.059	0.000	2.291	0.022	0.059	0.059

Delta method standard errors, normal theory confidence intervals

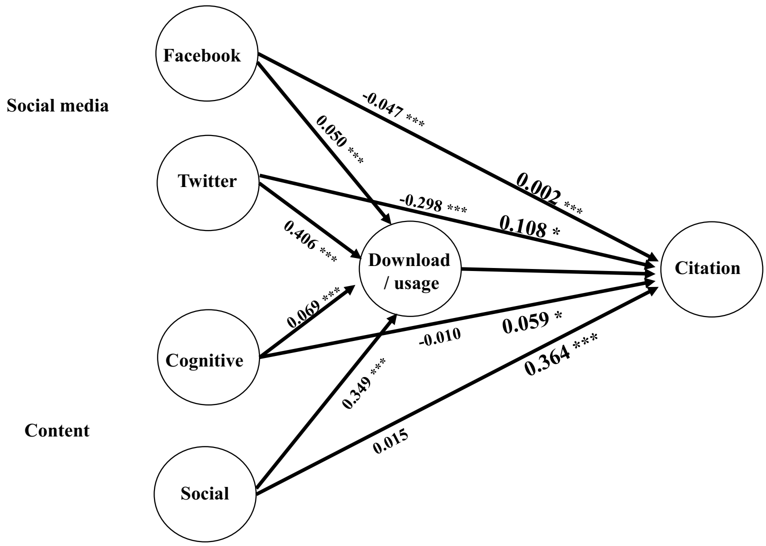

Fig. 2 Path coefficients for the mediation effects model. Note: Total effects in larger font size, direct and indirect effect in smaller font size, $^{***}p<0.001$, $^{*}p<0.05$

mediates the relationship among variables (e.g., social media and content related indicators), the mediation coefficients obtain higher values and get statistical significance. Our empirical framework's total effects demonstrate that mediation effects apply to Twitter mentions and the social content of the abstract, as predictors of citations in Scopus. Moreover, our findings corroborate that it is not enough to promote the article's content increasing mentions in social media because before citing an academic contribution, scholars need to download and read papers.

These findings are consistent with those reported in the academic literature. Regarding to our first hypothesis ("*Social media indicators for open access publications have a medium and positive association with Scopus' citation*"), Dehdaridad (2020) analyzed 47,961 articles in the area of life sciences and biomedicine, published between 2014 and 2016. The researcher retrieved those records from Medline and found that early tweet counts predict later citation counts, both in the count ($\beta=0.017$) and the logit ($\beta=0.203$) models.

In another study, researchers used sentiment analysis for exploring the relationships among altmetric indicators. They studied 6,482,260 tweets linked to 1,083,535 papers indexed in Scopus and confirmed the predictive capability of the positive ($\beta=0.254$), neutral ($\beta=0.338$) and negative ($\beta=0.456$) tweets on citations (Hassan et al., 2020). Despite those studies did not include interaction effects, they constitute partial evidence for supporting the empirical model introduced in this study.

Concerning to the third hypothesis ("*Download and usage of open access publications mediate the association between social media indicators and abstract's content with the citations in Scopus*"), a research team studied the relationships among subject matter, citations, downloads and the Altmetric Attention Score (AAS) within rheumatology. They worked with altmetric data for 1460 papers, published from 2010 to 2015 and found that the regression model for downloads and citations achieved an $R^2=29\%$ (Chen et al., 2020). Another recent study, examined the relationship between usage and citation in 7669

Table 4 Parameter estimates of mediation effects for each category

	Estimate	SE	z-value	p	95% confidence interval	
					Lower	Upper
Biology and life science						
Twitter → Usage → Citation	0.009	0.000	44.947	0.000	0.003	0.019
Facebook → Usage → Citation	0.003	0.000	32.324	0.000	0.002	0.004
Social → Usage → Citation	0.015	0.002	7.373	0.000	0.008	0.024
Cognitive → Usage → Citation	0.005	0.002	2.944	0.003	0.001	0.010
Computer and information sciences						
Twitter → Usage → Citation	0.010	0.002	6.505	0.000	0.005	0.019
Facebook → Usage → Citation	0.005	0.001	9.102	0.000	−0.002	0.010
Social → Usage → Citation	0.028	0.009	3.202	0.000	0.011	0.052
Cognitive → Usage → Citation	0.006	0.007	0.906	0.365	−0.007	0.019
Earth sciences						
Twitter → Usage → Citation	0.014	0.001	13.808	0.047	0.000	0.010
Facebook → Usage → Citation	−0.002	0.001	−3.252	0.023	−0.004	−0.001
Social → Usage → Citation	0.053	0.009	5.881	0.086	−0.002	0.061
Cognitive → Usage → Citation	0.011	0.007	1.590	0.123	−0.002	0.042
Ecology and environmental sciences						
Twitter → Usage → Citation	0.014	0.001	22.135	0.000	0.010	0.018
Facebook → Usage → Citation	−0.001	0.000	−2.254	0.024	−0.002	0.000
Social → Usage → Citation	0.007	0.005	1.257	0.201	−0.002	0.019
Cognitive → Usage → Citation	0.003	0.004	0.800	0.424	−0.004	0.009
Engineering and technology						
Twitter → Usage → Citation	0.015	0.002	9.665	0.000	0.007	0.036
Facebook → Usage → Citation	0.001	0.001	1.206	0.228	−0.001	0.004
Social → Usage → Citation	0.017	0.010	1.740	0.082	0.000	0.036
Cognitive → Usage → Citation	0.001	0.007	0.163	0.871	−0.011	0.018
Medicine and health sciences						
Twitter → Usage → Citation	0.009	0.001	65.505	0.000	0.005	0.017
Facebook → Usage → Citation	0.002	0.000	47.173	0.000	0.001	0.004
Social → Usage → Citation	0.016	0.001	12.465	0.000	0.011	0.022
Cognitive → Usage → Citation	0.004	0.001	4.568	0.000	0.002	0.007
People and places						
Twitter → Usage → Citation	0.017	0.001	49.659	0.000	0.012	0.024
Facebook → Usage → Citation	0.000	0.000	5.924	0.000	0.000	0.001
Social → Usage → Citation	0.010	0.001	6.322	0.000	0.005	0.015
Cognitive → Usage → Citation	0.002	0.001	1.464	0.143	−0.001	0.007
Physical sciences						
Twitter → Usage → Citation	0.011	0.000	60.547	0.000	0.005	0.019
Facebook → Usage → Citation	0.001	0.000	22.602	0.000	0.000	0.003
Social → Usage → Citation	0.014	0.002	9.165	0.000	0.010	0.019
Cognitive → Usage → Citation	0.005	0.001	4.702	0.000	0.002	0.008

Table 4 (continued)

	Estimate	SE	z-value	p	95% confidence interval	
					Lower	Upper
Research and analysis methods						
Twitter → Usage → Citation	0.013	0.000	99.342	0.000	0.009	0.019
Facebook → Usage → Citation	0.002	0.000	39.208	0.000	0.001	0.004
Social → Usage → Citation	0.021	0.001	21.614	0.000	0.017	0.024
Cognitive → Usage → Citation	0.008	0.001	12.216	0.000	0.007	0.010
Science policy						
Twitter → Usage → Citation	0.005	0.000	14.961	0.000	0.003	0.008
Facebook → Usage → Citation	0.001	0.000	2.389	0.017	−0.001	0.005
Social → Usage → Citation	0.016	0.007	2.443	0.015	0.005	0.045
Cognitive → Usage → Citation	0.005	0.001	1.282	0.200	−0.001	0.022
Social science						
Twitter → Usage → Citation	0.007	0.000	85.387	0.000	0.004	0.013
Facebook → Usage → Citation	0.000	0.000	33.554	0.000	0.000	0.003
Social → Usage → Citation	0.002	0.000	2.058	0.040	−0.003	0.006
Cognitive → Usage → Citation	−0.002	0.000	−2.583	0.010	−0.004	−0.000

preprints and 15,721 non open access papers published in 220 journals. According to their results, usage ($\beta = 0.375$) has a moderate and positive effect on citations in WoS (Wang et al., 2020a, b). Again, even though these regression models did not include mediating effects, the magnitude of the observed associations provides partial support to the empirical model developed for this study. We did not make a direct comparison with other mediating effects models because, at the time of this writing, they were not available in the academic literature. Therefore, we focused our discussion on those studies reported in the literature.

We provided compelling evidence for an empirical model that expands our understanding of the mediating effect of download and usage of publications between social media indicators and abstract content of a publication. According to the hypothesis testing results, from three hypotheses, just one was accepted for the general model, while for the thematic areas, the mediating effect was significant in most cases. This result provided empirical support to the model developed for this study. Moreover, we provided robust evidence that supports the idea that the abstract's content (a proxy for relevance) and the social media indicators (a proxy for interaction) have a moderate association with the citation in Scopus when mediated by download and usage. Indeed, PLoS' family of journals does not represent the worldwide scientific publications, but since this is one of the most prominent journals, our findings pose several theoretical and practical challenges.

Implications for theory

Theoretically-driven models of research use give prominence to relevance (measured as content in our empirical framework). In contrast, our findings challenge this long-time held perception and provide support for an explanation that demonstrates the powerful effect of the usage as mediating variable, at least according to the data available from article-level metrics. On the other side, for developing well-grounded models, theoretical frameworks must consider the dynamic nature of the analyzed process, which is not restricted to direct relationships but includes interaction effects, like mediation or moderation. Thus, future theoretical models can improve their predicting capability.

This study's results can provide initial support for building a conceptual model on how usage mediates the relationship between social media and content with citation. Regarding information use theories, classic models of information searching and use originated before the internet era. In that sense, they are more oriented to identifying information needs, searching, evaluating, and using information obtained from printed-based resources. Recent information use theories consider the high diversity of channels and platforms, granting more prominence to the use of databases and electronic resources.

Among the information search and use models, the information-seeking behavior model matches the empirical framework introduced in this research because it postulates the use of information sources, mediated by computer-based systems, as well as the evaluation and adoption of a piece of information obtained after analyzing or reading it (Wilson, 1999). Following the main stages of the model, in the beginning, the user searches for a piece of information, looking for data in different sources; in a similar way, a potential researcher browses web pages, finds mentions of a publication in social media, and then scans the title and abstract of that paper. If the content of the abstract or mentions in Facebook public pages capture the information seeker's interest, s/he downloads the PDF file and reads the full-text of that publication. When the potential researcher perceives that the downloaded publication contributes, s/he includes a reference citing the source. Given that the information required for testing the main assumptions of the proposed theoretical model is based on altmetric activity, this conceptual model would only apply at the document (publication) level, not at the user level.

Implications for practice

Our evidence suggests that scientific publication promotion campaigns should reinforce the widespread adoption of easy-to-use social media because, besides the velocity and variety of diffusion channels, the extended use guarantees that journal's papers reach increasing audiences. Likewise, given that the abstract's content improved its predicting capability when the moderating effect was included, we recommend increasing the adoption of open access strategies because in this way, the usage of a publication will increase. Finally, our study demonstrates the usefulness of mediation analysis to validate exploratory conceptual models based on article-level metrics. Since altmetric or citation data had high dispersion on many occasions, and altmetrics is an emerging field with most of the conceptual models still in the exploratory phase, the use of mediating analysis for validating theoretical frameworks could be the appropriate decision.

Limitations and future research

The study has two main limitations. First, our empirical framework included just six variables because the API's current implementation only extracted the model's information. In this sense, our analytical approach should be replicated with studies that include more variables than those mentioned above. Second, despite the mediating role of usage (measured by downloads and reads of each paper) was validated, few studies reported correlations between the social media indicators and citation. Thus, this interaction effect should be verified in future studies, maybe with data exported from well-known subscription-based databases (e.g., Web of Science). Therefore, the analytical approach should be validated with new data or information extracted from multidisciplinary databases.

Conclusion

Researchers usually include elements from existing models to build conceptual models because this approach contributes to knowledge growing. In this way, theory building is a consequence of recognizing and incorporating previous contributions. However, when there are not previous theoretical frameworks for building or testing explanations on the analyzed variables, researchers should begin with a data-driven exploratory approach, which can be a reasonable choice. For this reason, in this study, we explored the interaction effects among article-level metrics and found that usage mediates the relationship between social media and content with Scopus's citation. In this sense, this is one of the first studies that analyzes the associated factors with citation, using a predictive model that incorporates interaction effects.

In brief, we developed an empirical framework and tested it with an analytical approach appropriate for testing exploratory models. Our results provide compelling evidence on the mediation model robustness; however, to get a sound empirical framework, future studies must work in the model fit by incorporating more indicator variables to the constructs choice and research use. This model could be improved by comparing these initial results with findings of new studies based on data extracted from multidisciplinary databases like Scopus or WoS.

Appendix A

See Tables 5 and 6.

Table 5 Parameter estimates of mediation analysis for the general model (filtered dataset, $n = 758{,}419$)

	Estimate	SE	*z*-value	*p*	95% confidence interval	
					Lower	Upper
Direct effects						
Twitter → Citation	−0.007	0.000	−194.079	0.000	−0.008	−0.006
Facebook → Citation	−0.001	0.000	−84.225	0.000	−0.001	−0.000
Social → Citation	−0.000	0.000	−0.646	0.518	0.002	0.001
Cognitive → Citation	−0.000	0.000	−2.887	0.004	−0.001	−0.000
Indirect effects						
Twitter → Usage → citation	0.010	0.000	370.820	0.000	0.009	0.012
Facebook → Usage → Citation	0.001	0.000	151.277	0.000	0.000	0.002
Social → Usage → Citation	0.008	0.000	29.948	0.000	0.005	0.010
Cognitive → Usage → Citation	0.000	0.000	4.762	0.000	0.000	0.002
Total effects						
Twitter → Citation	0.003	0.000	78.234	0.000	0.002	0.004
Facebook → Citation	0.000	0.000	11.078	0.000	−0.000	0.000
Social → Citation	0.007	0.000	15.379	0.000	0.006	0.008
Cognitive → Citation	0.000	0.000	0.082	0.934	−0.000	0.000

Delta method standard errors, normal theory confidence intervals

Table 6 Parameter estimates of mediation effects for each category (filtered dataset, $n = 758{,}419$)

	Estimate	SE	*z*-value	*p*	95% confidence interval	
					Lower	Upper
Biology and life science						
Twitter → Usage → Citation	0.013	0.000	187.280	0.000	0.009	0.016
Facebook → Usage → Citation	0.001	0.000	67.227	0.000	0.000	0.002
Social → Usage → Citation	0.012	0.000	24.563	0.000	0.010	0.015
Cognitive → Usage → Citation	0.002	0.000	6.473	0.003	0.001	0.004
Computer and information sciences						
Twitter → Usage → Citation	0.005	0.000	53.381	0.000	0.003	0.007
Facebook → Usage → Citation	0.000	0.000	20.184	0.000	0.00	00.001
Social → Usage → Citation	0.013	0.001	11.308	0.000	0.008	0.018
Cognitive → Usage → Citation	0.000	0.000	−0.487	0.626	−0.002	0.001
Earth sciences						
Twitter → Usage → Citation	0.009	0.000	54.973	0.000	0.006	0.015
Facebook → Usage → Citation	0.000	0.000	9.729	0.000	−0.000	0.001
Social → Usage → Citation	0.013	0.002	8.493	0.000	0.009	0.019
Cognitive → Usage → Citation	−0.000	0.001	−0.427	0.670	−0.003	0.002
Ecology and environmental sciences						
Twitter → Usage → Citation	0.010	0.000	63.363	0.000	0.005	0.015
Facebook → Usage → Citation	0.003	0.000	37.041	0.000	−0.000	0.005
Social → Usage → Citation	0.004	0.001	2.645	0.008	−0.001	0.010
Cognitive → Usage → Citation	0.000	0.001	0.767	0.443	−0.002	0.004
Engineering and technology						
Twitter → Usage → Citation	0.006	0.000	53.307	0.000	0.004	0.010
Facebook → usage → Citation	0.000	0.000	13.337	0.000	0.000	0.003
Social → Usage → Citation	0.015	0.001	10.432	0.000	0.010	0.020
Cognitive → Usage → Citation	−0.000	0.001	−0.645	0.519	−0.003	0.001
Medicine and health sciences						
Twitter → Usage → Citation	0.014	0.000	182.484	0.000	0.010	0.020
Facebook → Usage → Citation	0.002	0.000	71.904	0.000	0.000	0.004
Social → Usage → Citation	0.009	0.000	15.315	0.000	0.004	0.013
Cognitive → Usage → Citation	0.001	0.000	2.928	0.003	−0.000	0.003
People and places						
Twitter → Usage → Citation	0.008	0.000	78.855	0.000	0.005	0.013
Facebook → Usage → Citation	0.000	0.000	12.544	0.000	−0.000	0.000
Social → Usage → Citation	0.006	0.000	7.604	0.000	0.003	0.010
Cognitive → Usage → Citation	−0.000	0.000	−0.111	0.911	−0.002	0.002
Physical sciences						
Twitter → Usage → Citation	0.013	0.000	144.560	0.000	0.009	0.019
Facebook → Usage → Citation	0.002	0.000	71.574	0.000	0.000	0.006
Social → Usage → Citation	0.003	0.000	3.421	0.000	−0.002	0.010
Cognitive → Usage → Citation	0.000	0.000	0.879	0.379	−0.002	0.003
Research and analysis methods						
Twitter → Usage → Citation	0.012	0.000	151.189	0.000	0.009	0.017

Table 6 (continued)

	Estimate	SE	z-value	p	95% confidence interval	
					Lower	Upper
Facebook → Usage → Citation	0.002	0.000	67.361	0.000	0.000	0.005
Social → Usage → Citation	0.008	0.000	11.268	0.000	0.002	0.016
Cognitive → Usage → Citation	0.002	0.000	4.743	0.000	− 0.000	0.005
Science policy						
Twitter → Usage → Citation	0.003	0.000	19.392	0.000	0.001	0.004
Facebook → Usage → Citation	0.000	0.000	8.310	0.000	0.000	0.001
Social → Usage → citation	0.009	0.004	2.387	0.017	0.000	0.022
Cognitive → Usage → citation	0.000	0.003	0.240	0.810	− 0.004	0.008
Social science						
Twitter → Usage → Citation	0.007	0.000	88.356	0.000	0.004	0.012
Facebook → Usage → Citation	0.000	0.000	36.716	0.000	0.000	0.003
Social → Usage → Citation	0.002	0.000	2.544	0.011	− 0.002	0.006
Cognitive → Usage → Citation	− 0.002	0.000	− 2.543	0.011	− 0.004	− 0.000

Appendix B

Links to the Comma-Separated-Values (CSV) files used in the study.

Filtered dataset ($n = 758{,}419$ records)
https://doi.org/10.6084/m9.figshare.16652707.

Analyzed dataset ($n = 235{,}384$ records)
https://doi.org/10.6084/m9.figshare.16652758.

References

Akella, A., Alhoori, H., Kondamudi, P. R., Freeman, C., & Zhou, H. (2021). Early indicators of scientific impact: Predicting citations with altmetrics. *Journal of Informetrics, 15*(2), 101128. https://doi.org/10.1016/j.joi.2020.101128

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182. https://doi.org/10.1037//0022-3514.51.6.1173

Bjekić, J., Lazarević, L. B., Živanović, M., & Knežević, G. (2014). Psychometric evaluation of the Serbian dictionary for automatic text analysis—LIWCser. *Psihologija, 47*(1), 5–32. https://doi.org/10.2298/PSI1401005B

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*(4–5), 993–1022.

Bornmann, L., & Haunschild, R. (2018). Do altmetrics correlate with the quality of papers? A large-scale empirical study based on F1000Prime data. *PLoS ONE, 13*(5), e0197133. https://doi.org/10.1371/journal.pone.0197133

Bornmann, L., Wolf, M., & Daniel, H.-D. (2012). Closed versus open reviewing of journal manuscripts: How far do comments differ in language use? *Scientometrics, 91*(3), 843–856. https://doi.org/10.1007/s11192-011-0569-5

Chen, W. M. Y., Bukhari, M., Cockshull, F., & Galloway, J. (2020). The relationship between citations, downloads and alternative metrics in rheumatology publications: A bibliometric study. *Rheumatology, 59*(2), 277–280. https://doi.org/10.1093/rheumatology/kez163

Cho, J. (2021). Altmetrics analysis of highly cited academic papers in the field of Library and Information Science. *Scientometrics, 126*(9), 7623–7635. https://doi.org/10.1007/s11192-021-04084-w

Dehdaridad, T. (2020). Could early tweets counts predict later citation counts? A gender study in Life Sciences and Biomedicine (2014–2016). *PLoS ONE, 15*(11), e0241723. https://doi.org/10.1371/journal.pone.0241723

Ebrahimy, S., Mehrad, J., Setareh, F., & Hosseinchari, M. (2016). Path analysis of the relationship between visibility and citation: The mediating roles of save, discussion, and recommendation metrics. *Scientometrics, 109*(3), 1497–1510. https://doi.org/10.1007/s11192-016-2130-z

Ekinci, E., & Omurca, I. (2020). Concept-LDA: Incorporating Babelfy into LDA for aspect extraction. *Journal of Information Science, 46*(3), 406–418. https://doi.org/10.1177/0165551519845854

Gontijo, M. C. A., & de Araújo, R. F. (2021). Impacto acadêmico e atenção on-line de pesquisas sobre inteligência artificial na área da saúde: Análise de dados bibliométricos e altmétricos. *Encontros Bibli: Revista Eletrônica De Biblioteconomia e Ciência Da Informação, 26*, 01–21. https://doi.org/10.5007/1518-2924.2021.e76249

Graybeal, A., Seagal, J. D., & Pennebaker, J. W. (2002). The role of story-making in disclosure writing: The psychometrics of narrative. *Psychology and Health, 17*(5), 571–581. https://doi.org/10.1080/08870440290025786

Guerrero-Bote, V. P., & Moya-Anegón, F. (2014). Relationship between downloads and citations at journal and paper levels, and the influence of language. *Scientometrics, 101*(2), 1043–1065. https://doi.org/10.1007/s11192-014-1243-5

Hassan, S.-U., Aljohani, N. R., Idrees, N., Sarwar, R., Nawaz, R., Martinez-Camara, R., Ventura, S., & Herrera, F. (2020). Predicting literature's early impact with sentiment analysis in Twitter. *Knowledge-Based Systems, 192*(15), e105383. https://doi.org/10.1016/j.knosys.2019.105383

Haustein, S., & Larivière, V. (2014). A multidimensional analysis of Aslib proceedings—Using everything but the impact factor. *Aslib Journal of Information Management, 66*(4), 358–380. https://doi.org/10.1108/AJIM-11-2013-0127

Haustein, S., Costas, R., & Larivière, V. (2015). Characterizing social media metrics of scholarly papers: The effect of document properties and collaboration patterns. *PLoS ONE, 10*(3), e0120495. https://doi.org/10.1371/journal.pone.0120495

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*. The Guilford Press.

Hu, Z., Fang, S., & Liang, T. (2014). Empirical study of constructing a knowledge organization system of patent documents using topic modeling. *Scientometrics, 100*(3), 787–799. https://doi.org/10.1007/s11192-014-1328-1

Iacobucci, D. (2008). *Mediation analysis*. Sage.

JASP Team. (2020). *JASP, version 0.14.1* (computer software). https://jasp-stats.org/download.

Jose, P. E. (2013). *Doing statistical mediation and moderation*. The Guilford Press.

Li, X., Wu, C., & Mai, F. (2019). The effect of online reviews on product sales: A joint sentiment-topic analysis. *Information Management, 56*(2), 172–184. https://doi.org/10.1016/j.im.2018.04.007

Maggio, L. A., Leroux, T. C., Meyer, H. S., & Artino, A. R., Jr. (2018). #MedEd: Exploring the relationship between altmetrics and traditional measures of dissemination in health professions education. *Perspectives on Medical Education, 7*(4), 239–247. https://doi.org/10.1007/s40037-018-0438-5

Maricato, J. M., & Vilan-Filho, J. L. (2018). The potential for altmetrics to measure other types of impact in scientific production: academic and social impact dynamics in social media and networks. *Information Research*, *23*(1), paper 780. Retrieved from http://InformationR.net/ir/23-1/paper780.html.

Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and conversations. *Journal of Personality and Social Psychology, 84*(4), 857–870. https://doi.org/10.1037/0022-3514.84.4.857

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. https://doi.org/10.1002/asi.23405

Ortega, J. L. (2015). Relationship between altmetric and bibliometric indicators across academic social sites: The case of CSIC's members. *Journal of Informetrics, 9*(1), 39–49. https://doi.org/10.1016/j.joi.2014.11.004

Pennebaker, J. W., Francis M. E., & Booth, R. J. (2001). *Linguistic Inquiry and Word Count LIWC 2001 Manual*. Retrieved from https://www.researchgate.net/publication/246699633_Linguistic_inquiry_and_word_count_LIWC.

Priem, J., Taraborelli, D., Groth, P. & Neylon, C. (2010). *Altmetrics: A manifesto*. Retrieved from http://altmetrics.org/manifesto.

Roemer, R. C., & Borchardt, R. (2015). Issues, controversies, and opportunities for altmetrics. *Library Technology Reports, 51*(5), 20–30.

Sathianathen, N. J., Lane, R., Condon, B., Murphy, D. G., Lawrentschuk, N., Weight, C. J., & Lamb, A. D. (2020). Early online attention can predict citation counts for urological publications: The #UroSoMe_Score. *European Urology Focus, 6*(3), 458–462. https://doi.org/10.1016/j.euf.2019.10.015

Schlögl, C., & Gorraiz, J. (2010). Comparison of citation and usage indicators: The case of oncology journals. *Scientometrics, 82*(3), 567–580. https://doi.org/10.1007/s11192-010-0172-1

Schlögl, 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. https://doi.org/10.1007/s11192-014-1365-9

Smith, Z. L., Chiang, A. L., Bowman, D., & Wallace, M. B. (2019). Longitudinal relationship between social media activity and article citations in the journal Gastrointestinal Endoscopy. *Gastrointestinal Endoscopy, 90*(1), 77–83. https://doi.org/10.1016/j.gie.2019.03.028

Smith-Keiling, B. L., & Hyun, H. I. F. (2019). Applying a computer-assisted tool for semantic analysis of writing: Uses for STEM and ELL. *Journal of Microbiology & Biology Education, 20*(1), 1–6. https://doi.org/10.1128/jmbe.v20i1.1709

Strauβ, N., Alonso-Muñoz, L., & Zúñiga, H. G. (2020). Bursting the filter bubble: The mediating effect of discussion frequency on network heterogeneity. *Online Information Review, 44*(6), 1161–1181. https://doi.org/10.1108/OIR-11-2019-0345

Thelwall, M. (2017). Are Mendeley reader counts useful impact indicators in all fields? *Scientometrics, 113*(1), 1721–1731. https://doi.org/10.1007/s11192-017-2557-x

Thelwall, M., & Nevill, T. (2018). Could scientists use Altmetric.com scores to predict longer term citation counts? *Journal of Informetrics, 12*(1), 237–248. https://doi.org/10.1016/j.joi.2018.01.008

Thelwall, M., & Wilson, P. (2014). Regression for citation data: An evaluation of different methods. *Journal of Informetrics, 8*(4), 963–971. https://doi.org/10.1016/j.joi.2014.09.011

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. https://doi.org/10.1371/journal.pone.0064841

Vílchez-Román, C., Huamán-Delgado, F., & Alhuay-Quispe, J. (2020). Social dimension activates the usage and academic impact of Open Access publications in Andean countries: A structural modeling-based approach. *Information Development*. https://doi.org/10.1177/0266666920901849

Wang, Z., Chen, Y., & Glänzel, W. (2020a). Preprints as accelerator of scholarly communication: An empirical analysis in Mathematics. *Journal of Informetrics, 14*(4), 101097. https://doi.org/10.1016/j.joi.2020.101097

Wang, Z., Glänzel, W., & Chen, Y. (2020b). The impact of preprints in Library and Information Science: An analysis of citations, usage and social attention indicators. *Scientometrics, 125*(2), 1403–1423. https://doi.org/10.1007/s11192-020-03612-4

Wei, M., & Noroozi Chakoli, A. (2020). Evaluating the relationship between the academic and social impact of open access books based on citation behaviors and social media attention. *Scientometrics, 125*(3), 2401–2420. https://doi.org/10.1007/s11192-020-03678-0

Wilson, T. D. (1999). Models of information behaviour research. *Journal of Documentation, 55*(3), 249–270. https://doi.org/10.1108/EUM0000000007145

Williams, A. E. (2017). Altmetrics: An overview and evaluation. *Online Information Review, 41*(3), 311–317. https://doi.org/10.1108/OIR-10-2016-0294

Wu, Q., Dbouk, W., Hasan, I., Kobeissi, N., & Zheng, L. (2021). Does gender affect innovation? Evidence from female chief technology officers. *Research Policy, 50*(9), 104327. https://doi.org/10.1016/j.respol.2021.104327

Zhang, Y., Wu, Y. J., Goh, M., & Liu, X. (2019). Supply chain management scholar's research impact: Moderated mediation analysis. *Library Hi Tech, 39*(1), 118–135. https://doi.org/10.1108/LHT-07-2017-0149