



A probe into 66 factors which are possibly associated with the number of citations an article received

Juan Xie¹ · Kaile Gong¹ · Jiang Li¹ · Qing Ke¹ · Hyonchol Kang¹ · Ying Cheng¹

Received: 7 June 2018 / Published online: 6 April 2019
© Akadémiai Kiadó, Budapest, Hungary 2019

Abstract

The number of citations has been widely used for scientific evaluation of publications and even institutions and individual scientists. However, there is as yet no satisfactory consensus as to when and how citation metrics should be applied. Therefore, it is of great importance to comprehend the factors which influence citations. The purpose of this study is to identify more such factors in order to better understand the dynamics of citations. We first collected articles in Library & Information Science journals indexed by the Chinese Social Science Citation Index. Then, we established reliable schemes to identify and record a total of 66 candidate factors—related to articles, authors, references and citations—which had not been comprehensively studied before. Bivariate analysis was applied to explore the relationship between those factors and citations, and forward stepwise regression was used to select predictive factors. We found that 46 factors were significantly associated with citations, from which six most significant factors were selected by a regression model. Possible strategies were hence proposed for identifying high-quality and high-impact articles.

Keywords Citations · Influencing factors · Author-related factors · Reference-related factors · Stepwise regression

Introduction

As carriers of knowledge, scientific publications play a vital role in scholarly communication. Quick access to high-quality publications is of great importance for scholars. Although peer review has been recognized as the principal mechanism for quality control in most scientific fields (Bornmann 2011), there are many criticisms of the process. Therefore, academia sometimes uses citation count as an alternative measure of quality because of its availability, applicability and objectivity (Garfield 1979). Some studies have shown that the outcomes of bibliometric indicators are generally in line with those of peer reviews (Li et al. 2010; Oppenheim 1995; Raan 2006). However, the use of citations as an alternative to expert judgment of quality has itself been subjected to continuous scrutiny.

✉ Ying Cheng
chengy@nju.edu.cn

¹ School of Information Management, Nanjing University, Nanjing 210093, Jiangsu Province, China

Therefore, it is important to understand the factors which influence the number of citations a scientific publication receives.

Tahamtan and Bornmann (2018) provided a conceptual model of citation process, based on the context of cited documents, processes from selection to citation, and the context of citing documents. Contextual features of the cited documents include document features (e.g., perceived quality, accessibility, title, abstract, etc.), author features (e.g., number of authors, the rank of authors' institutions, etc.), journal features (e.g., the scope and reputation of a journal) and document values (e.g., the perceived utility). Tahamtan et al. (2016)'s comprehensive review has also identified 28 factors contributing to citation count and grouped them into three categories: paper-, journal- and author-related factors. Among previous studies, some have focused on one category of factors, for example, Amara et al. (2015) explored what kind of faculty members achieve high scholarly performance; some researchers have investigated two categories (Wesel et al. 2014); others have taken all three categories into account, e.g., Antoniou et al. (2015) and Leimu and Koricheva (2005). It is now commonly accepted that a number of bibliometric indicators also affect citation count, such as the number of authors, Journal Impact Factor (JIF), the number of pages, and the presence of early citations (Yu et al. 2014). In addition, the influential factors vary across disciplines (Antoniou et al. 2015; Stremersch et al. 2015; Yu et al. 2014).

A variety of regression analyses, including multiple linear regression and negative binomial regression, have utilized different numbers of factors to construct models with different degrees of fit (Yu et al. 2014; Bornmann and Leydesdorff 2015; Stremersch et al. 2015; Haslam et al. 2008; Lokker et al. 2008). Stepwise regression is preferred as a method of selecting significant factors among large candidate sets (Vanclay 2013).

Those studies focusing on a single category of factors have offered detailed explanations of citations within a specific context, but failed to make comparisons across categories. Moreover, although previous studies have considered journal-, author-, and article-related factors in tandem, they only emphasized the quantitative factors directly from scientific databases. Less is known about qualitative factors, such as authors' educational background, etc. In addition, some previous studies reviewed factors related to citation count, but neglected the difference of their importance. Based on these observations, we aim to address the following questions in the present study:

- (1) Are there any heretofore neglected factors which have significant association with the citations of scientific publications?
- (2) Which factors, whether established or newly identified, have the greatest influence on citation count?

The structure of this article is as follows. In the next section we review the literature which involves citation-related factors and frequently used regression models, the third section presents data and methodology, the fourth section reports the results, and in the last section we draw conclusions and discuss potential implications and limitations.

Related work

In the conceptual model of the citation process in Tahamtan and Bornmann (2018), they found that many factors are associated with authors' decision to cite a document, such as the location of the citation context, the features of citing documents, as well as their

authors' and journals' features. In addition, they explained many reasons for citing a document, such as provision of a background for new research, use of the cited authors' methodology, and criticism of a previously published work. In the context of cited documents, document features, author features, and journal features, together with citing authors' positive or negative attitudes toward the documents' value, are all revealed to influence citation process.

The current study intends to uncover more factors which may have significant correlation with citations, and to identify the most correlated ones through comparisons. We organize factors which potentially affect citations into four categories: article-, author-, reference-, and citation-related factors, and review corresponding literature respectively.

Factors associated with citation counts

Article-related factors

It is verified in previous studies that the number of citations an article received is significantly positively correlated to the length of the article (generally operationalized as page count), (Leimu and Koricheva 2005), the number of keywords, the length of abstract (Rostami et al. 2014; Wesel et al. 2014), structured abstract, the number of tables, figures, footnotes, appendixes and formulas (Lokker et al. 2008; Stremersch et al. 2015), open access and inclusion in numerous databases (Lokker et al. 2008), download times (Schlögl et al. 2014), as well as document types, article age, peer reviewing, reviewing times, funds and acknowledgements (Bornmann and Leydesdorff 2015; Vanclay 2013; Rigby 2013). The length of title sometimes has no or negative effects on citations (Rostami et al. 2014). In addition, the number of citations is also associated with the order of an article's appearance in a journal (Yu et al. 2014; Bornmann and Leydesdorff 2015; Dalen and Henkens 2001; Stremersch et al. 2015), i.e., earlier articles usually receive more citations.

Some article-related factors deserve further investigation. For example, the effect of download times used to be analyzed separately by bivariate methods, without comparisons with other factors. For another example, since previous studies only applied absolute order and the number of articles in an issue varies significantly across journals, the relative order of an article is worth exploring.

Author-related factors

Amara et al. (2015) investigated factors which explained why scholars perform differently in terms of research productivity and impact. According to their results, authors' time allocated to research and teaching, financial resources, academic title, and institutional affiliation are all significant in this respect. An author's previous citations and publications increase the citations of a given article (Yu et al. 2014; Stremersch et al. 2015). Co-authorship characteristics, including the number of authors, the first author's *h*-index, and the group's highest and minimum *h*-index, are also related to the number of citations (e.g., Hurley et al. 2013). Other correlated factors include institutional, regional, and international collaborations, as well as interdisciplinary cooperation (Amara et al. 2015).

Numerical factors have been frequently considered in previous studies, such as the number of authors and *h*-index. However, many categorical factors' influence has been underestimated. For example, whether authors' nationality, educational background or academic degree exerts significant influence on citations remains unanswered.

Reference-related factors

The number of citations an article received is associated with the number and impact of references (Stremersch et al. 2015; Bornmann and Leydesdorff 2015), the variety of references (Chakraborty et al. 2014), as well as reference currency, author self-citation and journal self-citations (Roth et al. 2012; Vanclay 2013).

It is revealed that language barrier generates negative impact and prevent people from seeking necessary information (Henderson 2005). Whether an author is proficient in a foreign language or languages (and thus gets access to rich information in those languages) could be indicated by the percentage of foreign-language references. Moreover, different document types carry distinct information or knowledge (Crossick 2016).

Citation-related factors

Early citations, i.e., citations an article receives immediately after publication, are the early feedback of the scientific community about the article's impact, and significantly correlate with total citations (Dalen and Henkens 2005). The time at which a paper receives its first citation represents its speed of dissemination in the scientific community, which is often characterized by the reciprocal of an article's first-cited age (Yu et al. 2014). Citations received in the first 2 years can also be a predictor of citation counts according to Yu et al. (2014). They detected the positive effects of the number of countries, document types, journals and disciplines citing an article in its first 2 years after publication in detail. However, whether these effects can be expanded to a larger citation window requires further study.

Regression models involving citation counts

More and more studies have developed regression models to compare the different effect size of influencing factors and to predict future citations at an early stage. Bivariate analysis was used to investigate the relationship between bibliometric indicators and citation impact (Leimu and Koricheva 2005; Rostami et al. 2014; Schlögl et al. 2014), and to explore whether the above factors have significant effects on citations. Negative binomial regression is a standard model used to account for over-dispersion that would be under-fitted by a Poisson model. For instance, Didegah and Thelwall (2013) employed a zero-inflated negative binomial regression model with eight independent factors and identified JIF and impact of references as the most effective predictors of citation counts. A comparison between negative binomial regression model and ordinary least squares regression was made by Bornmann and Leydesdorff (2015) to prove the stability of the former model's results.

Ordinary least squares regression is a widely used linear regression model in this context, often applied after the logarithmic transformation of bibliometric indicators. For example, through multivariate regression, Antoniou et al. (2015) found three factors' independent influence on the number of citations: study subject, study design and article length. Royle et al. (2013) performed multiple regression to determine the amount of variation in citations attributable to JIF and other factors.

Stepwise regression is preferred as a method of selecting significant indicators among a large candidate set. It is a method that guarantees the validity and importance of the chosen indicators and reduces the additional error introduced by redundant indicators (Yu

et al. 2014). Accordingly, stepwise linear regression has become popular in bibliometric research, where large sets of candidate factors are commonplace. Vanclay (2013), for instance, constructed a stepwise multilinear model beginning with 12 potential factors, five of which were identified to play a significant role in predicting citation impact. Yu et al. (2014) also compared forward selection, backward elimination and bidirectional elimination theoretically. It is a common practice to use forward stepwise regression to select a good subset from a moderate or large number of predictors, and predictors chosen by forward stepwise regression can decrease the residual sum of squares effectively (Taylor and Tibshirani 2015).

Methodology

To address the research questions raised in Introduction, we constructed, labeled, and analyzed a large dataset of full-text journal articles. Detailed information is listed in Tables 1 and 4, while data sources and the holonomic schema are presented in the Appendix. Bivariate analysis was used to detect the relationship between different candidate factors and citations, and stepwise regression analysis was employed to compare the sizes of the effects.

Data source

The Chinese Social Science Citation Index (CSSCI), developed by Nanjing University, is an important tool for inquiry and assessment of the major publications and journals in the humanities and social sciences (Yang et al. 2010). The list of journals in CSSCI is revised every 2 years. The 2014–2015 edition includes 18 journals in Library and Information Science (LIS): *Journal of Academic Libraries*; *Journal of the National Library of China*; *Information Science*; *Information Studies: Theory & Application*; *Journal of The China Society for Scientific and Technical Information*; *Journal of Intelligence, Information and Documentation Services*; *Library*; *Library Work and Study*; *Library Development*; *Library Tribune*; *Research on Library Science*; *Library Journal*; *Library and Information Service*; *Documentation Information & Knowledge*; *Library & Information*; *Data Analysis and Knowledge Discovery* (previously *New Technology of Library and Information Science*); and *Journal of Library Science in China*.

Chinese National Knowledge Infrastructure (CNKI), launched by Tsinghua Tongfang Knowledge Network Technology Company, is a full-text scientific database (Wan et al. 2010). To ensure at least 1 year for citation accumulation, we downloaded full-text articles published between 2005 and 2015 in the 18 journals, obtaining 55,720 articles in total. Forty-nine bibliometric indicators were encoded manually from the full-text articles or accessed automatically from CNKI. The full-text articles published in *Journal of The China Society for Scientific and Technical Information* were not indexed by CNKI prior to 2017; therefore, instead we accessed them from Wanfang Data (another full-text scientific database) and manually searched these articles in CNKI to identify their citation counts.

Baidu Scholar (xueshu.baidu.com), similar to Google Scholar, indexes Chinese scientific publications, as well as overseas literature, across a variety of publishers and disciplines. We used this platform to count the number of databases in which an article was indexed (X_{21}) and to determine whether an article was open-access (X_{22}). Other online resources, such as institutional websites and author's homepages, were used to code author-related factors, e.g., educational background, age and alma mater.

Table 1 Results of data collection and bivariate analysis for article-related factors

Factor	Variable	Frequency	Statistic ^a	<i>p</i>	Literature sources
X_1	Article order: anterior	177	5.83	0.054	
	Article order: medium	188			
	Article order: posterior	201			
X_2	Document type: empirical	185	45.98***	0.000	Mingers and Xu (2010) and Rostami et al. (2014)
	Document type: theoretical	179			
	Document type: case study	56			
	Document type: review	37			
	Document type: other	109			
X_3	Article age	566	-0.07	0.124	Bornmann and Leydesdorff (2015)
X_4	Article length	566	0.27***	0.000	E.g., Antoniou et al. (2015), Bornmann and Leydesdorff (2015) and Haslam et al. (2008)
X_5	Title length	566	0.04	0.401	Stremersch et al. (2015) and Wesel et al. (2014)
X_6	No. of keywords	563	0.15***	0.000	Stremersch et al. (2015)
X_7	Abstract length	563	0.14**	0.001	Wesel et al. (2014)
X_8	Structural abstract: yes	4	1364.00	0.447	Lokker et al. (2008)
	Structural abstract: no	559			
X_9	Category number: G	435	4.18	0.383	
	Category number: TP	54			
	Category number: F	44			
	Category number: D	23			
	Category number: other	10			
X_{10}	Journal impact factor	566	0.13**	0.001	E.g., Bornmann and Leydesdorff (2015) and Haslam and Koval (2010)
X_{11}	Publications in journal year	566	-0.13**	0.003	Yu et al. (2014)
X_{12}	No. of tables and figures	566	0.12**	0.004	Hegarty and Walton (2012) and Stremersch et al. (2015)
X_{13}	No. of footnotes	566	0.02	0.674	Stremersch et al. (2015)
X_{14}	No. of appendixes	566	-		Stremersch et al. (2015)
X_{15}	No. of formulas	566	0.07	0.092	Stremersch et al. (2015)

Table 1 (continued)

Factor	Variable	Frequency	Statistic ^a	<i>p</i>	Literature sources
X_{16}	Peer-review: yes	509	16,582.00	0.076	Bhat (2009)
	Peer-review: no	57			
X_{17}	Time of review	421	-0.07	0.131	Hlimer and Lusk (2009)
X_{18}	Funding: national	97	23.47***	0.000	Amara et al. (2015)
	Funding: Ministry of Education	40			
	Funding: provincial	64			
	Funding: prefectural	37			
	Funding: other	8			
	Funding: none	320			
X_{19}	Acknowledgments: yes	3	1566.50*	0.010	Rigby (2013)
	Acknowledgments: no	563			
X_{20}	Downloads	552	0.77***	0.000	E.g., Schlägl et al. (2014)
X_{21}	Number of databases indexed	566	0.14**	0.001	Lokker et al. (2008)
X_{22}	Open access: yes	122	28,605.50	0.341	Højjem et al. (2005)
	Open access: no	444			

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

^aSpearman correlation or the statistics of MW or KW tests

Sampling

The dataset was then randomly sampled. According to the method introduced by Berenson and Levine (1993), the sample size was calculated via Formula 1 below. P represents the diversity of samples and e represents error. On condition that $P=0.5$ (the maximum) and $e=0.05$, with $Z=1.96$ and $N=55,720$, the required sample size is 382. To enhance the reliability of the study, we sampled 600 articles randomly from the sample set.

$$N = P(1 - P)/(e^2/z^2 + P(1 - P)/N). \quad (1)$$

When extracting data from the sampled articles, we excluded 34 non-scholarly articles, remaining 566 articles in our final sample. We also calculated the percentage of selected articles in each journal and found that the likelihood of being sampled was nearly the same for articles published in different journals.

Data collection

Two authors of this study were trained before data collection and intercoder reliability was calculated by percent agreement and kappa coefficients (Lombard et al. 2002). Two authors conducted a pilot test of the schema informally by extracting data from ten articles together, then discussed problems and revised the schema. The percent agreement of them over all factors was 0.952 (>0.90), and is satisfactory (Lombard et al. 2002). All reference- and citation- related variables, and most of the article- and author-related variables were objective, which can be extracted from databases directly. Thus, the study calculated the kappa coefficients of the other 16 article- and author-related variables which were subjectively coded by the two authors. All 15 coefficients were bigger than 0.700 and the other one was almost 0.700 (kappa for $X_1=0.697$), indicating a substantial agreement between coders (Lombard et al. 2002). Disagreements were settled by negotiation.

Descriptions of selected factors are given below, while the schema for all 66 potential factors and average citations per year is listed in the Appendix. As shown in Tables 1–4, all potential factors ($X1$ to $X66$) studied here are substantial features, rather than subjective judgments like perceived utility and quality. Average citations per year was used as the citation metric to correct year differences.

Article-related factors

We divided article order into anterior, medium and posterior tertiles with respect to the different numbers of articles in different issues. Category number was in accordance with the Book Classification of China (National Library of China 2010). Funding was divided into 6 categories: national funding, funding from Ministry of Education of the PRC, provincial funding, prefectural funding, other, and no funding.

Author-related factors

Since the first author makes the most contribution to a Chinese article in LIS, all author-related features were coded based on the first author and the other authors were neglected. The educational background of the first author was described with reference to his/her highest degree, according to Ministry of Education of the PRC (2005). Academic degree

likewise referred to the highest degree of the first author, categorized as junior college, bachelor's, master's, or doctorate. The classification of universities which the first author graduated from included "Project 985", i.e., thirty-nine leading Chinese universities that meet certain scientific, technical, and human resources standards and receive financial support from the government; "Project 211", i.e., 116 universities in China that meet less stringent requirements and are eligible for government support; "general university"; "foreign university", and "other." Authorship was categorized into teachers, librarians, independent researchers (who are not affiliated to universities), students, officials or engineers. Categories of affiliations included "Project 985" universities, "Project 211" universities, general universities, junior colleges, national libraries, provincial libraries, city libraries, district libraries, military libraries and foreign affiliations.

Reference-related factors

The mean JIF of references was the average impact factors of journals in the references. Reference age was denoted by mean age of references.

Citation-related factors

The first-cited age was denoted by the reciprocal of the number of years from publication to the first citation. A value of zero was used for never-cited articles.

Statistical analysis

We applied bivariate analysis including Spearman Correlations and other Nonparametric Tests to our data, because of the skewed distribution. For ordinal variables and nominal variables, the Mann–Whitney U test is used to determine whether two independent samples were selected from the populations with the same distribution, while Kruskal–Wallis is applied for comparing three or more independent samples. Kruskal–Wallis test indicates that at least one sample differed from the others. An effect can be of no importance, but highly significant (Cumming and Calin-Jageman 2017). Thus, the present study also compared the effect size of correlations, according to Cohen (1988)'s rules, i.e., $r=0.00$ – 0.09 for no correlation, $r=0.10$ – 0.29 for weak, $r=0.30$ – 0.49 for medium and $r=0.50$ – 1.00 for high.

To find the most significant factors, average citations per year was regressed on all variables showing significant effects in previous bivariate analysis. Except for variables of citation-related factors, 7 scale variables with missing data were estimated by linear interpolation in regression analysis, because replacing missing data in early citation-related factors could introduce biases. Records with missing data in 10 categorical variables were deleted automatically. We constructed dummy variables, and scale variables were standardized after logarithmic transformation. Then, we chose the factors that most decreased the residual sum of squares, and entered one factor each time in this procedure. The forward stepwise multiple linear regression model was run with logarithmic transformation of the dependent variable after adding 1.0, and independent variables were entered into the model if $P < 0.10$ using simple regression. SPSS 23.0 was used in both processes.

Results

Bivariate analysis

Article-related factors

As shown in Table 1, we found no significant correlation between the number of citations and article order, category number, timespan of peer review, or the number of footnotes or formulas. In addition, whether the article is OA or has structured abstracts made no difference in terms of citations, whereas document type and inclusion in more databases did. There was a strong relationship between citations and download times ($r=0.74$). The correlation coefficients between citations and the number of keywords ($r=0.15$), words in abstract ($r=0.14$), and number of tables and figures ($r=0.12$), though small, were all significant and positive. However, there were weak relationships between citations and JIF ($r=0.13$) or article length ($r=0.27$). The total number of publications in a journal year, meanwhile, was slightly negatively associated with citations ($r=-0.13$, $p=0.003$). X_{14} , the number of appendixes in an article, equaled 0 in all articles sampled, so the effect of this factor could not be detected.

Author-related factors

As shown in Table 2, it was found that educational background, academic degree, years of research, years of working, average h -index of all authors and authorship category were all significantly associated with citations. However, in this study, whether the first author is a nationally recognized top talent (i.e., a Yangtze Scholar) was not correlated with citations, neither was alma mater nor cross-discipline/province collaboration.

It is identified that the number of citations was associated with gender, productivity, previous citations, previous citations per article, authors' affiliations and academic title. Particularly, the authors' previous total citations and previous citations per article had medium effects ($r=0.42$ and $r=0.49$). Moreover, the h -index of the first author ($r=0.38$) and the highest h -index among authors ($r=0.35$) also correlated moderately with citations, with coefficients higher than that of the minimum h -index among authors ($r=0.20$). Besides, the number of authors had a small but significant effect ($r=0.16$). Cross-institution collaboration connects to citations, but international collaboration does not. There is no evidence that authors' age or nationality correlated to citation count.

Reference-related factors

Table 3 shows that in an article's references, the percentages of conference papers ($r=0.10$), online resources ($r=0.15$) and dissertations ($r=0.12$) as well as monographs ($r=-0.18$) are weakly associated with the number of citations it received. The percentage of journal articles in the references, however, has no effect. Interestingly, the percentage of foreign-language references also slightly correlates with citations ($r=0.17$, $p=0.000$). In addition, the impact of references, i.e., the total number of citations an articles' references received, is associated with citations. There is a slight tendency that articles with more recent references received more citations ($r=-0.15$, $p=0.000$).

Table 2 Results of data collection and bivariate analyses for author-related factors

Factors	Variable	Frequency	Statistic ^a	p	Literature sources
X ₂₃	No. of authors	566	0.16***	0.000	E.g., Bornmann and Leydesdorff (2015) and Haslam and Koval (2010)
X ₂₄	Gender: male	268	34,339.50*	0.021	Ayres and Vars (2000)
	Gender: female	289			
X ₂₅	Author age	548	0.04	0.388	Ayres and Vars (2000)
X ₂₆	Nationality: China	565	106.00	0.279	E.g., Dalen and Henkens (2001), Leimu and Koricheva (2005) and Royle et al. (2013)
	Nationality: USA	1			
X ₂₇	Educational background: library and information science	202	16.57*	0.011	
	Educational background: management science and engineering	24			
	Educational background: business administration	21			
	Educational background: economy	10			
	Educational background: computer science	15			
	Educational background: Chinese literature	20			
	Educational background: other	52			
X ₂₈	Academic degree: doctorate	165	15.20**	0.002	
	Academic degree: master's	190			
	Academic degree: bachelor's	105			
X ₂₉	Academic degree: junior college	1			
	Category of alma mater: 985	226	4.00	0.406	
	Category of alma mater: 211	47			
	Category of alma mater: general	65			
	Category of alma mater: foreign	3			
	Category of alma mater: other	2			

Table 2 (continued)

Factors	Variable	Frequency	Statistic ^a	p	Literature sources
X ₃₀	Authorship: teachers	150	18.89**	0.002	
	Authorship: librarians	295			
	Authorship: independent researchers	19			
	Authorship: students	86			
	Authorship: officials	2			
	Authorship: engineers	13			
X ₃₁	Academic title: professor	75	10.62**	0.014	E.g., Amara et al. (2015), Leimu and Koricheva (2005) and Ayres and Vars (2000)
	Academic title: associate professor	167			
	Academic title: assistant professor	47			
	Academic title: research fellow with PhD	2			
X ₃₂	Top talent: yes	6	2400.00	0.070	
	Top talent: no	560			
X ₃₃	Years of research	564	0.17***	0.000	
X ₃₄	Years of working	218	0.18**	0.008	
X ₃₅	Category of affiliation: 985	193	29.93***	0.000	Amara et al. (2015) and Mingers and Xu (2010)
	Category of affiliation: 211	60			
	Category of affiliation: general university	192			
	Category of affiliation: junior college	22			
	Category of affiliation: national library	9			
	Category of affiliation: provincial library	23			
	Category of affiliation: city library	21			
Category of affiliation: district library	11				
X ₃₆	Category of affiliation: military library	9	0.23***	0.000	Haslam et al. (2008) and Yu et al. (2014)
	Category of affiliation: foreign affiliation	1			
	Productivity	566			

Table 2 (continued)

Factors	Variable	Frequency	Statistic ^a	<i>p</i>	Literature sources
X ₃₇	Previous citations	566	0.42***	0.000	Yu et al. (2014)
X ₃₈	Previous citations per article	566	0.49***	0.000	Yu et al. (2014)
X ₃₉	<i>h</i> index	566	0.38***	0.000	Yu et al. (2014)
X ₄₀	<i>h</i> max	566	0.35***	0.000	E.g., Qian et al. (2017) and Vancley (2013) and Yu et al. (2014)
X ₄₁	<i>h</i> min	566	0.20***	0.000	Hurley et al. (2013)
X ₄₂	<i>h</i> average	566	0.35***	0.000	
X ₄₃	Cross-institution collaboration: yes	78	21,928.50*	0.031	E.g., Didegah and Thelwall (2013) and Sooryamoorthy (2017)
	Cross-institution collaboration: no	488			
X ₄₄	Cross-province collaboration: yes	79	21,602.00	0.079	
	Cross-province collaboration: no	487			
X ₄₅	International collaboration: yes	1	106.00	0.279	Royle et al. (2013) and Didegah and Thelwall (2013)
	International collaboration: no	565			
X ₄₆	Interdisciplinary collaboration: yes	32	9556.00	0.259	
	Interdisciplinary collaboration: no	534			

****p* < 0.001; ***p* < 0.01; **p* < 0.05

^aSpearman correlation or the statistics of MW or KW tests

Table 3 Results of data collection and bivariate analysis for reference-related factors

Factors	Variable	Frequency	Spearman r^a	p	Literature sources
X_{47}	No. of references	566	0.25***	0.000	E.g., Antoniou et al. (2015), Haslam et al. (2008) and Haslam and Koval (2010)
X_{48}	Total citations of references	566	0.18***	0.000	
X_{49}	Citations per reference	566	0.12**	0.004	Didegh and Thelwall (2013)
X_{50}	Mean JIF of references	565	0.12**	0.005	
X_{51}	Reference age	547	-0.15***	0.000	Haslam et al. (2008)
X_{52}	Percentage of foreign-language references	566	0.17***	0.000	
X_{53}	Self-citation of author	564	0.06	0.146	Stremersch et al. (2015)
X_{54}	Self-citation of journal	564	0.14**	0.001	Vanclay (2013)
X_{55}	Percentage of journal articles	566	0.06	0.167	
X_{56}	Percentage of conference papers	566	0.10*	0.023	
X_{57}	Percentage of monographs	566	-0.18***	0.000	
X_{58}	Percentage of online resources	566	0.15***	0.000	
X_{59}	Percentage of dissertations	566	0.12**	0.005	

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 4 Results of data collection and bivariate analysis for citation-related factors

Factors	Variable	Frequency	Spearman r^a	p	Literature sources
X_{60}	Reciprocal of first-cited age	497	0.31***	0.000	Yu et al. (2014)
X_{61}	Citation count in the first year	566	0.76***	0.000	Hlimer and Lusk (2009)
X_{62}	Citation count in the first 2 years	521	0.87***	0.000	Yu et al. (2014)
X_{63}	Citation count in the first 5 years	379	0.97***	0.000	
X_{64}	No. of citing journals in the first year	566	0.66***	0.000	
X_{65}	No. of citing journals in the first 2 years	521	0.78***	0.000	Yu et al. (2014)
X_{66}	No. of citing journals in the first 5 years	379	0.90***	0.000	

*** $p < 0.001$

Among these reference-related factors, the number of references has the highest correlation with citations, but the effect size is still small.

Citation-related factors

There is negative correlation between citations and an article’s first citation age (measured in years), with correlation coefficient 0.31 after reciprocal transformation. The number of citations an article received in the first year since publication correlates strongly with the total citations ($r=0.76$), and the coefficient increases along with the enlargement of citation windows ($r=0.87$ for the first 2 years and $r=0.97$ for the first 5 years). A similar pattern holds for the number of citing journals, where the correlations are 0.66, 0.78 and 0.90 in the first year, first 2 years and first 5 years respectively.

Regression analysis

The results of stepwise multiple linear regression are shown in Table 5. Among the 46 factors which have significant effects in the bivariate analysis, only 6 passed the further test of regression analysis with $R^2=0.948$. The value of R^2 is more satisfactory than

Table 5 Results of stepwise regression

			$R=0.974$ $R^2=0.948$ Adjusted $R^2=0.945$			
			$df_1=1$ $df_2=99$			
Indicators	Coefficient	Std. Err.	β	t	Sig.	VIF
Downloads	0.258	0.037	0.26	7.055	0.000	2.670
Citation count in the first 5 years	0.807	0.040	0.74	20.270	0.000	2.555
Authorship: independent researcher	0.787	0.189	0.10	4.156	0.000	1.039
Educational background: computer science	-0.300	0.110	-0.06	-2.719	0.008	1.018
Percentage of monographs	0.075	0.029	0.06	2.570	0.012	1.072
Affiliation: 211	-0.139	0.078	-0.04	-1.778	0.079	1.032
Constant	-0.124	0.029		-4.321	0.000	

the results obtained by Yu et al. (2014) ($R^2=0.676$), Vanclay (2013) ($R^2=0.450$), or Haslam et al. (2008) ($R^2=0.360$).

The citation count in the first 5 years was found to have the most significant effect on citations ($\beta=0.74$). Download times also exerted a positive influence on average citations ($\beta=0.26$). Interestingly, authors' characteristics, including authorship category, educational background, and category of affiliation, were also found to have positive effects. Authors who are independent researchers, received more citations ($\beta=0.10$). However, authors who work for "Project 211" universities ($\beta=-0.04$) or who have computer science background ($\beta=-0.06$) were indicated to have a negative influence on citations. Unexpectedly, the percentage of monographs in the references was also slightly related to the citations an article received ($\beta=0.06$).

Discussion

The citation preference of social science research is different from that of natural sciences. For example, natural sciences articles have less references than social science ones (Skilton 2006). The percentage of self-citations in social sciences and humanities is 49%, and in natural sciences the percentage is 87% (Larivière et al. 2006).

The factors affecting citations

Author-related factors

Since senior scientists who produce higher-quality publications and hold higher academic titles received more citations (Slyder et al. 2011), we inferred that scientists who conduct research (X_{33}) and work (X_{34}) for longer time were more experienced and hence received more citations. In keeping with the prior observation that faculty members who dedicate more time to research activities at the expense of teaching, administration and consulting activities achieved better publication and citation performance (Amara et al. 2015), the academic title of authors (X_{30}) had a positive correlation with citations. Furthermore, an author's educational background and academic degree reflect academic literacy and improve the performance of publications as well. This is in line with the findings of Ried and McKenzie (2004), who argued that students' academic preparation, as measured by academic degree, was a predictor of citation performance.

There is a strong relationship between citations and the author's h -index as well as the number of authors (Qian et al. 2017). The significant effect of the first author's gender (X_{24}) and affiliation (X_{35}) are also supported by many previous studies (Amara et al. 2015). The association between citations and cross-discipline/country/province collaboration was not supported, but cross-institution collaboration made a difference.

In our study, recognized top talent in the author list did not significantly contribute to an article's citations, though his/her productivity and citation impact did (Dalen and Henkens 2005). This finding may be aroused by the small number of reputational scholars, since there were only 15 Yangtze Scholars in LIS at the time when the dataset was collected.

Reference-related factors

A notable finding in this part was the small but significant influence exerted by the variety of references (X_{56} to X_{59}). Referencing more dissertations, conference papers and online resources were associated with higher citation counts, which has been explained by previous studies (Chung et al. 2012; Crossick 2016; Yi 2004).

We found that the proportion of foreign-language references in an article also associated with citation counts. A study of the usage of English-language resources by Spanish and Latin American literature scholars shed light on the percentage growth of foreign resources in humanities (Nolen 2014).

The number of references, as well as their respective impacts and recency, exerted weak influence on citations, as documented previously (Bornmann and Leydesdorff 2015). This might also imply a quality element: for example, with more references, the authors obtained more subject-matter knowledge from other researchers, which boosted the quality of the original research (Haslam and Koval 2010; Bornmann and Leydesdorff 2015).

Article-related factors

In this study, we found that the article order in a journal issue was not positively associated with the number of citations, which is possibly because of the easy access to publications in digital era. The category number given in Chinese articles had no effect either, which perhaps indicates the increasingly important role that interdisciplinarity has played in LIS (Xu et al. 2016). Although funding (X_{18}) and acknowledgements (X_{19}) are independent in Chinese articles, they highlight the importance of financial and academic supports, both of which facilitate high-quality research (Amara et al. 2015). However, open-access (OA) seems unrelated to citations, possibly because most institutions in China subscribe to scientific full-text databases (e.g., CNKI and Wanfang Data mentioned above) which provide convenient access to rich information sources. Bhat (2009) suggested that peer-reviewed articles and those with longer reviewing time correlate positively with citations, but this study did not support this suggestion.

As documented before (Haslam et al. 2008; Hegarty and Walton 2012), we found that containing more tables and figures, an extensive abstract, more keywords and more pages were slightly associated with citations. Why title length was not associated with citations could be explained by the argument that a valuable title is simple, understandable and informative (Rostami et al. 2014). In addition, it has been observed that articles in journals with high JIFs received more citations.

Although downloads do not inevitably translate to citations, we found that within our sample, article which were downloaded more received more citations. Because once a paper was downloaded, it can potentially have a citation (Jamali and Nikzad 2011). Many scientific databases have the capability to rank search results by download count, so articles being downloaded frequently attract more attention and disseminate more broadly. The positive effect of the inclusion number of databases and the negative effect of the total number of publications in an issue can also be explained as visibility effects (Lokker et al. 2008).

Citation-related factors

We found that the more quickly an article is recognized by scientific community, the more citations it is likely to receive. This effect is sometimes known as the “first mover advantage”: being a pioneer leads to a strong advantage in scientific performance (Sabatier and Chollet 2017). The dynamic Waring model of Glänzel and Schubert (1995) also proved the influence of the number of early citations on those obtained later, which explains the cumulative advantage. The high correlation between current and previous citations of a paper (e.g., citation counts in the first year, 2 years, or 5 years) is of prime importance in increasing visibility, and this process of dissemination can also be accelerated and broadened by the citing journals (Yu et al. 2014). The reasons may be twofold: on one hand, search results in databases are often ranked by previous citation counts; on the other hand, highly-cited articles are frequently seen in others’ references (Hlimer and Lusk 2009).

Regression model

In this model, the six variables exerting significant influence on citations belong to different categories, including examples of author-, reference-, citation- and article-related factors. By examining the factors which correlate to citations, authors can boost the impact of their work, while readers can more effectively select articles with high quality and impact. Since variables were standardized before regression, we compared the coefficient of different variables with average citation count, i.e., dependent variable.

Citation count in the first 5 years exerted the strongest effect in the model, which underscored the cumulative advantage discussed above. With the development of electronic publishing, the metric of download counts has become a popular ranking feature, and its association with citations has been documented in previous bibliometric and altmetric research (Schlögl et al. 2014). In the present study, we first compared the effect of downloads with that of other indicators in a regression model. Apart from the citation counts in the first 5 years after publication, downloads had the strongest correlation with citations. Combined with the results of the present study, this suggests the value of an attractive title and extensive abstract in encouraging more potential readers to download an article (Jamali and Nikzad 2011).

The monograph has a central place in the culture and ecology of research publication in the fields of arts and humanities and is likewise important in most of the social sciences (Crossick 2016). The percentage of monographs in a reference list thus relates to citations. As argued by Crossick (2016), monographs offer the length and space for full examination of a topic, rich ideas, which is not possible in a journal article.

The effects of three author-related factors have also been highlighted. Independent researchers (X_{30}) received more citations. University faculty members’ research time has been occupied by teaching, administration and professional consulting activities (Amara et al. 2015), whereas independent researchers, however, can focus on their research. This result suggests that scholars must devote more time to research activities to increase the probability of success in their academic career. Furthermore, the 39 top-tier universities in China (“Project 985”) received a great deal of funding to promote the development and reputation of Chinese higher education system, and overshadowed “Project 211” universities in all aspects. Therefore, articles published by better universities may have advantages in receiving more citations (Mingers and Xu 2010). We found

that authors with an educational background in computer science might find it difficult to get recognized in the field of LIS. The reasons may be that scholarly practices differ between these two disciplines (Qian et al. 2017), and few articles to date have been published in LIS journals by computer scientists.

Although the dataset of the present study included only eighteen Chinese LIS journals, the findings are in line with previous studies of international journals. First of all, this study proved the significant effects of the rank of affiliations (Stremersch et al. 2015), downloads (Schlögl et al. 2014), and early citations (Yu et al. 2014) on citation counts through regression analysis. Second, it has been found that there are significant relationship between citation counts and the factors discussed above, i.e., document type (Mingers and Xu 2010), article length (Bornmann and Leydesdorff 2015), the number of tables and figures (Stremersch et al. 2015), Journal Impact Factors (Haslam and Koval 2010), publication year (Sin 2011), collaboration (Sin 2011), authors' previous productivity (Yu et al. 2014), academic title (Amara et al. 2015), the number of references (Haslam et al. 2008), etc. Nevertheless, Chinese journals are characterized by local features which may also appear in other non-English journals. For example, no relationship was found between international collaboration and citations (Sin 2011), as there is rare international collaboration in Chinese journal articles. It also characterized Spanish journals (Jesús and José 2004). In conclusion, the present findings from the Chinese journals are not limited to Chinese academia, but also shed lights on international academia.

Conclusions

In this paper, we probed the influence of 66 factors on citations, by drawing sample of articles from eighteen leading Chinese library and information science journals. Bivariate and regression analyses shed light on the relationship between a host of factors and citation counts, and created a more comprehensive picture of citation patterns.

There are at least two limitations in this study. One is our reliance on manual extraction of data, which limited the sample size compared to previous predictive analysis. Though the dependent variable in this randomly sampled set was highly skewed, samples with high citations might be under-represented. In further study, the population can be split into pieces, for example, through investigating the factors related to citations of highly cited articles and factors related to never-cited ones. The other is the limited scope of the sample: our conclusions are based only on Chinese articles in the field of library & information science. Further studies should investigate publications in other fields and other languages, in order to establish a wider applicability of these findings.

Acknowledgements This study is financially supported by a research grant from the National Social Science Foundation of China (Grant Number: 17BTQ014). We would also like to express our appreciation for the reviewers' hard work and helpful suggestions.

Appendix

See Table 6.

Table 6 Data collection schema

No.	Factors	Notes	Type	Source
X_1	Article order	Order in which an article was listed in the issue, divided into anterior, medium and posterior tertiles	Ordinal	CNKI
X_2	Document type	Document type of the article: empirical, theoretical, case study, review or other (including system design, policy analysis, biography, translation and so on)	Nominal	CNKI
X_3	Article age	Number of years since publication	Interval	CNKI
X_4	Article length	Number of pages in the article	Interval	CNKI
X_5	Title length	Number of words in the article title	Interval	CNKI
X_6	No. of keywords	Number of keywords in an article	Interval	CNKI
X_7	Abstract length	Number of words in the article abstract	Interval	CNKI
X_8	Structural abstract	Whether an abstract is structural or not	Nominal	CNKI
X_9	Category number	Category number presented in an article according to the Book Classification of China. G denotes Culture, Science, Education & Sports; TP represents Automation & Computer Engineering, F is Economics; D is Politics & Law; and "other" includes B, C, H, I, J, K and R	Nominal	CNKI
X_{10}	Journal Impact Factor	Journal impact factor in 2015, obtained from Chinese S&T Journal Citation Reports (Expanded Edition 2016)	Interval	CS&T JCR
X_{11}	Publications in journal year	Total number of publications in a journal year	Interval	CNKI
X_{12}	No. of tables and figures	Number of tables and figures in the article	Interval	CNKI
X_{13}	No. of footnotes	Number of footnotes in the article	Interval	CNKI
X_{14}	No. of appendixes	Number of appendixes in the article	Interval	CNKI
X_{15}	No. of formulas	Number of formulas in the article	Interval	CNKI
X_{16}	Peer-review	Whether the article is peer-reviewed	Nominal	Website
X_{17}	Time of reviewing	How many months after publication the article is reviewed	Interval	CNKI
X_{18}	Funding	Funding an article received, classified into 6 categories: national funding, funding from the Ministry of Education, provincial funding (other ministries except the Ministry of Education), prefectural funding, other (e.g., companies and journals), and no funding	Nominal	CNKI
X_{19}	Acknowledgements	Whether the article includes acknowledgements	Nominal	CNKI
X_{20}	Downloads	Number of times the article was downloaded since publication	Interval	CNKI

Table 6 (continued)

No.	Factors	Notes	Type	Source
X_{21}	Number of databases indexed	Number of databases in which the article was indexed, denoted by the number of effective links in Baidu Scholar	Interval	Baidu Scholar
X_{22}	Open access	Whether the article was open-access (OA). If an effective link to the official website of its publishing journal was found in Baidu Scholar, the article could be denoted as OA	Nominal	Baidu Scholar
X_{23}	No. of authors	Number of authors in the article	Interval	CNKI
X_{24}	Gender	Gender of the first author, male or female	Nominal	Homepage
X_{25}	Author age	Age of the first author when the article was published	Interval	Homepage
X_{26}	Nationality	Nationality of the first author	Nominal	Homepage
X_{27}	Education background	Major of the first author, according to the classification criteria from the Ministry of Education of PRC. If a category has less than 5 samples, it will be classified as "other"	Nominal	Homepage
X_{28}	Academic degree	Highest degree attained by the first author: junior college, bachelor's, master's or doctorate	Ordinal	Homepage
X_{29}	Category of alma mater	Category of university the first author graduated from: "project 985", "project 211", general, foreign universities, or other	Ordinal	Homepage
X_{30}	Authorship	Career of the first author: teachers, librarians, independent researchers, students, officials, and engineers. In our sample set, an independent researcher is who pursues scientific study independently without direct affiliation to a university. A teacher is a person who teaches students knowledge, competences or values and at the same time pursues research work in colleges or universities. An engineer also pursues scientific research sometimes and is affiliated to a company	Nominal	Homepage
X_{31}	Academic title	Academic title of the first author: professor, associate professor, assistant professor, or research fellow (with PhD)	Ordinal	Homepage
X_{32}	Top Talent	Whether the first author has been honored as a Yangtze Scholar	Ordinal	Homepage
X_{33}	Years of research	Number of years the first author dedicated to research since first scientific publication. We subtract the publication year of the author's first scientific work from the publication year of the work in our sample set	Interval	Homepage
X_{34}	Years of working	'Years of working' is calculated by the author's first working year (the year the author began graduate study) subtracted by the year the author's first paper has been published in our sample set	Interval	Homepage

Table 6 (continued)

No.	Factors	Notes	Type	Source
X ₃₅	Category of affiliation	Category of the affiliation of the first author: 'project 985' university, 'project 211' university, general university, junior college, national library, provincial library, city library, district library, military library or foreign affiliation	Ordinal	Website
X ₃₆	Productivity	Total number of papers published by the first author before the present article	Interval	CNKI
X ₃₇	Previous citations	Total number of citations received by the first author before the present article	Interval	CNKI
X ₃₈	Previous citations per article	Average citations of the articles published by the first author before the present article	Interval	CNKI
X ₃₉	<i>h</i> index	<i>h</i> -index of the first author	Interval	CNKI
X ₄₀	<i>h</i> max	Maximum <i>h</i> -index of the authors	Interval	CNKI
X ₄₁	<i>h</i> min	Minimum <i>h</i> -index of the authors	Interval	CNKI
X ₄₂	<i>h</i> average	Average <i>h</i> -index of the authors	Interval	CNKI
X ₄₃	Cross-institution collaboration	Whether the authors are from different institutions, according to the primary institution	Nominal	CNKI
X ₄₄	Cross-province collaboration	Whether the authors are from different provinces or states	Nominal	CNKI
X ₄₅	International collaboration	Whether the authors are from different countries	Nominal	CNKI
X ₄₆	Interdisciplinary collaboration	Whether the authors have different disciplinary backgrounds	Nominal	Homepage
X ₄₇	No. of references	Number of references included in the article	Interval	CNKI
X ₄₈	Total citations of references	Total number of citations of the references	Interval	CNKI
X ₄₉	Citation per reference	Average number of citations per reference	Interval	CNKI
X ₅₀	Mean JIF of references	Mean 2015 impact factor of the journals in which the references were published	Interval	CS&T JCR
X ₅₁	Reference age	Mean article age of the references	Interval	CNKI
X ₅₂	Percentage of foreign-language reference	Percentage of foreign articles in the reference list. Foreign references are non-Chinese articles	Ratio	CNKI
X ₅₃	Self-citation of author	Percentage of articles written by the authors in the reference list	Ratio	CNKI
X ₅₄	Self-citation of journal	Percentage of the reference articles published in the journal in which the present article was published	Ratio	CNKI
X ₅₅	Percentage of journal articles	Percentage of journal articles in the reference list	Ratio	CNKI
X ₅₆	Percentage of conference papers	Percentage of conference papers in the reference list	Ratio	CNKI
X ₅₇	Percentage of monographs	Percentage of monographs in the reference list	Ratio	CNKI

Table 6 (continued)

No.	Factors	Notes	Type	Source
X_{58}	Percentage of online resources	Percentage of online resources in the reference list	Ratio	CNKI
X_{59}	Percentage of dissertations	Percentage of dissertations in the reference list	Ratio	CNKI
X_{60}	Reciprocal of first-cited age	Reciprocal of the number of years between the first citation and publication. A value of 0 is given for uncited articles	Interval	CNKI
X_{61}	Citation count in first year	Number of citations within a year of publication	Interval	CNKI
X_{62}	Citation count in first 2 years	Number of citations in the first 2 years after publication. Articles published in 2005–2014 contained information for this variable	Interval	CNKI
X_{63}	Citation count in first 5 years	Number of citations in the first 5 years after publication. Articles published in 2005–2011 contained information for this variable	Interval	CNKI
X_{64}	No. of citing journals in first year	Number of citing journals within a year of publication	Interval	CNKI
X_{65}	No. of citing journals in first 2 years	Number of citing journals in the first 2 years after publication. Articles published in 2005–2014 contained information for this variable	Interval	CNKI
X_{66}	No. of citing journals in first 5 years	Number of citing journals in the first 5 years after publication. Articles published in 2005–2011 contained information for this variable	Interval	CNKI
Y	Average citations per year	Annual citation rate, calculated for each paper by dividing the total citation count by the number of years since publication up to the end of 2016	Interval	CNKI

References

- Amara, N., Landry, R., & Halilem, N. (2015). What can university administrators do to increase the publication and citation scores of their faculty members? *Scientometrics*, *103*(2), 489–530.
- Antoniou, G. A., Antoniou, S. A., Georgakarakos, E. I., Sfyroeras, G. S., & Georgiadis, G. S. (2015). Bibliometric analysis of factors predicting increased citations in the vascular and endovascular Literature. *Annals of Vascular Surgery*, *29*, 286–292.
- Ayres, I., & Vars, F. E. (2000). Determinants of citations to articles in elite law reviews. *Journal of Legal Studies*, *29*, 427–450.
- Berenson, M. L., & Levine, D. M. (1993). *Statistics for business and economics*. Englewood Cliffs, NJ: Prentice Hall.
- Bhat, M. H. (2009). Effect of peer review on citations in the open access environment. *Library Philosophy & Practice*, *6*, 1–6.
- Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology*, *45*(1), 197–245.
- Bornmann, L., & Leydesdorff, L. (2015). Does quality and content matter for citedness? A comparison with para-textual factors and over time. *Journal of Informetrics*, *9*(3), 419–429.
- Chakraborty, T., Kumar, S., Goyal, P., Ganguly, N., & Mukherjee, A. (2014). Towards a stratified learning approach to predict future citation counts. In *ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, London, UK, pp. 351–360.
- Chung, C. J., Nam, Y., & Stefanone, M. A. (2012). Exploring online news credibility: The relative influence of traditional and technological factors. *Journal of Computer-Mediated Communication*, *17*, 171–186.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Crossick, G. (2016). Monographs and open access. *Insights*, *29*(1), 14–19.
- Cumming, G., & Calin-Jageman, R. (2017). *Introduction to the new statistics*. New York: Routledge.
- Dalen, H. P., & Henkens, K. (2001). What makes a scientific article influential? The case of demographers. *Scientometrics*, *50*(3), 455–482.
- Dalen, H. P., & Henkens, K. (2005). Signals in science—On the importance of signaling in gaining attention in science. *Scientometrics*, *64*(2), 209–233.
- Didegah, F., & Thelwall, M. (2013). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of Informetrics*, *7*, 861–873.
- Garfield, E. (1979). *Citation indexing—Its theory and application in science, technology, and humanities*. New York: Wiley.
- Glänzel, W., & Schubert, A. (1995). Predictive aspects of a stochastic model for citation processes. *Information Processing and Management*, *31*(1), 69–80.
- Hajjem, C., Harnad, S., & Gingras, Y. (2005). Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, *28*, 39–47.
- Haslam, N., Ban, L., Kaufmann, L., Loughnan, S., Peters, K., Whelan, J., et al. (2008). What makes an article influential? Predicting impact in social and personality psychology. *Scientometrics*, *76*(1), 169–185.
- Haslam, N., & Koval, P. (2010). Predicting long-term citation impact of articles in social and personality psychology. *Psychological Reports*, *106*(3), 891–900.
- Hegarty, P., & Walton, Z. (2012). The consequences of predicting scientific impact in psychology using journal impact factors. *Perspectives on Psychological Science*, *7*(1), 72.
- Henderson, J. K. (2005). Language diversity in international management teams. *International Studies of Management & Organization*, *35*(1), 66–82.
- Hlimer, C. E., & Lusk, J. L. (2009). Determinants of citations to the agricultural and applied economics association journals. *Review of Agricultural Economics*, *31*(4), 677–694.
- Hurley, L. A., Ogier, A. L., & Torvik, V. I. (2013). Deconstructing the collaborative impact: article and author characteristics that influence citation count. *ASIS&T Annual Meeting, Beyond the Cloud: Rethinking Information Boundaries*. Montreal, Canada, p. 61.
- Jamali, H. R., & Nikzad, M. (2011). Article title type and its relation with the number of downloads and citations. *Scientometrics*, *88*(2), 653–661.
- Jesús, R. R., & José, M. S. M. (2004). Patterns of the foreign contributions in some domestic vs. international journals on Earth Sciences. *Scientometrics*, *59*(1), 95–115.
- Larivière, V., Archambault, É., Gingras, Y., & Vignola-Gagné, É. (2006). The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology*, *57*(8), 997–1004.

- Leimu, R., & Koricheva, J. (2005). What determines the citation frequency of ecological papers? *Trends in Ecology & Evolution*, *20*(1), 28.
- Li, J., Sanderson, M., Willett, P., Norris, M., & Oppenheim, C. (2010). Ranking of library and information science researchers: Comparison of data sources for correlating citation data, and expert judgments. *Journal of Informetrics*, *4*(4), 554–563.
- Lokker, C., McKibbin, K. A., McKinlay, R. J., Wilczynski, N. L., & Haynes, R. B. (2008). Prediction of citation counts for clinical articles at two years using data available within three weeks of publication: Retrospective cohort study. *British Medical Journal*, *336*, 655–657.
- Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research*, *28*(4), 587–604.
- Mingers, J., & Xu, F. (2010). The drivers of citations in management science journals. *European Journal of Operational Research*, *205*(2), 422–430.
- Ministry of Education of the PRC. (2005). *A directory of disciplines and specialties for postgraduates, master and Ph.D.* <http://www.cdgdc.edu.cn/xwyyjsjyxx/sy/gldm/264462.shtml>. Accessed 11 April 2018.
- National Library of China. (2010). *Book classification of China* (5th ed.). Beijing: The Press of National Library of China.
- Nolen, D. S. (2014). Publication and language trends of references in Spanish and Latin American literature. *College & Research Libraries*, *75*(1), 34–35.
- Oppenheim, C. (1995). The correlation between citation counts and the 1992 Research Assessment Exercise Ratings for British library and information science university departments. *Journal of Documentation*, *51*(1), 18–27.
- Qian, Y., Rong, W., Jiang, N., Tang, J., & Xiong, Z. (2017). Citation regression analysis of computer science publications in different ranking categories and subfields. *Scientometrics*, *110*(3), 1–24.
- Raan, A. F. J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, *67*(3), 491–502.
- Ried, L. D., & McKenzie, M. (2004). A preliminary report on the academic performance of pharmacy students in a distance education program. *American Journal of Pharmaceutical Education*, *68*(3), 65.
- Rigby, J. (2013). Looking for the impact of peer review: does count of funding acknowledgements really predict research impact? *Scientometrics*, *94*, 57–73.
- Rostami, F., Mohammadpoorasl, A., & Hajizadeh, M. (2014). The effect of characteristics of title on citation rates of articles. *Scientometrics*, *98*(3), 2007–2010.
- Roth, C., Wu, J., & Lozano, S. (2012). Assessing impact and quality from local dynamics of citation networks. *Journal of Informetrics*, *6*(1), 111–120.
- Royle, P., Kandala, N. B., Barnard, K., & Waugh, N. (2013). Bibliometrics of systematic reviews: Analysis of citation rates and journal impact factors. *Systematic Reviews*, *2*, 74.
- Sabatier, M., & Chollet, B. (2017). Is there a first mover advantage in science? Pioneering behavior and scientific production in nanotechnology. *Research Policy*, *46*, 522–533.
- 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.
- Sin, S. C. J. (2011). International coauthorship and citation impact: A bibliometric study of six LIS journals, 1980–2008. *Journal of the American Society for Information Science and Technology*, *62*(9), 1770–1783.
- Skilton, P. F. (2006). A comparative study of communal practice: Assessing the effects of taken-for-grantedness on citation practice in scientific communities. *Scientometrics*, *68*(1), 73–96.
- Slyder, J. B., Stein, B. R., Sams, B. S., Walker, D. M., Beale, B. J., Feldhaus, J. J., et al. (2011). Citation pattern and lifespan: A comparison of discipline, institution, and individual. *Scientometrics*, *89*, 955–966.
- Sooryamoorthy, R. (2017). Do types of collaboration change citation? A scientometric analysis of social science publication in South Africa. *Scientometrics*, *111*, 379–400.
- Stremersch, S., Camacho, N., Vanneste, S., & Verniers, I. (2015). Unraveling scientific impact: Citation types in marketing journals. *International Journal of Research in Marketing*, *32*(1), 64–77.
- Tahamtan, I., Afshar, A. S., & Ahamdzadeh, K. (2016). Factors affecting number of citations: A comprehensive review of the literature. *Scientometrics*, *107*(3), 1195–1225.
- Tahamtan, I., & Bornmann, L. (2018). Core elements in the process of citing publications: Conceptual overview of the literature. *Journal of Informetrics*, *12*(1), 203–216.
- Taylor, J., & Tibshirani, R. J. (2015). Statistical learning and selective inference. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(25), 7629.
- Vanclay, J. K. (2013). Factors affecting citation rates in environmental science. *Journal of Informetrics*, *7*(2), 265–271.
- Wan, J. K., Hua, P. H., Rousseau, R., & Sun, X. K. (2010). The journal download immediacy index (DII): Experiences using a Chinese full-text database. *Scientometrics*, *82*(3), 555–566.

- Wesel, M. V., Wyatt, S., & Haaf, J. T. (2014). What a difference a colon makes: How superficial factors influence subsequent citation. *Scientometrics*, *98*(3), 1601–1615.
- Xu, H., Guo, T., Yue, Z., Ru, L., & Fang, S. (2016). Interdisciplinary topics of information science: A study based on the terms interdisciplinarity index series. *Scientometrics*, *106*, 583–601.
- Yang, S., Qiu, J., & Xiong, Z. (2010). An empirical study on the utilization of web academic resources in humanities and social sciences based on web citations. *Scientometrics*, *84*(1), 1–19.
- Yi, J. (2004). The development of the China Networked Digital Library of theses and dissertations. *Online Information Review*, *28*(5), 367–370.
- Yu, T., Yu, G., Li, P. Y., & Wang, L. (2014). Citation impact prediction for scientific papers using stepwise regression analysis. *Scientometrics*, *101*(2), 1233–1252.