



Examining the relationship between academic book citations and Goodreads reader opinion and rating

Kadiresan N^a, Mangkhollen Singson^b and Thiyagarajan S^c

^aResearch Scholar, Department of Library & Information Science, Pondicherry University, Puducherry, E-mail: nkadiresan1992@gmail.com

^bAssistant Professor, Department of Library & Information Science, Pondicherry University, Puducherry, E-mail: manglien@gmail.com

^cCorresponding author: Assistant Professor, Department of International Business, Pondicherry University, Puducherry, E-mail: sthiyags@yahoo.com

Received: 11 March 2020; revised: 16 November 2020; accepted: 24 November 2020

Although the traditional bibliometric citation database is an established academic impact assessment source, in this paper, we examine the role of social media impact on academic books. We identified the highly cited books in Scopus and compared the citations with ratings and reviews on the Goodreads website. R stat was used to extract the data from Goodreads website. We found that there is an uneven distribution of Goodreads rating and reviews. Social science books received the highest number of user's ratings, reviews and citations. The study finds that there is no relationship between citation counts and Goodreads ratings and reviews count in social science books. Although social science books generated the highest number of studies and engagement by the readers, there seems to be no evidence to suggest that this engagement results in an academic citation. Whereas, a correlation was observed between health science books citations and Goodreads overall rating, as with physical science book reviews and Google Scholar citation counts.

Keywords: Sentiment analysis; Web scraping; Citation analysis; Goodreads; Scopus; Google Scholar

Introduction

Traditionally, a scholarly book review consists of an introduction about the content of the book, a brief discussion about the concept of content, and suggestions for further improvement. It is usually an academic critique written by the subject experts in their respective fields. According to Hartley¹, "book reviews should measure or assess two general elements; the scholarly credibility of a newly published book and the author's writing style". Hence, these scholarly book reviews provide a reference for readers, experts, and authors to evaluate the scholarship of the book. However, the process of peer review is time-consuming and an expensive affair.

With the development of Web 2.0, book reviews are contributed not only by the academic fraternity but also by readers across the internet space such as Goodreads (www.goodreads.com) and Amazon (www.amazon.com). The introduction of these online bookstores and social networks has enabled readers to express their views, opinions, and comments without much restriction in format or content otherwise followed in the scholarly book review. At the same time, readers' comments are not confined to popular books of fiction and nonfiction alone, but academic books as well.

In this regard, the present study is an attempt to examine the relationship between academic book reviews, sentiments or opinions posted by readers on Goodreads (which is a popular social book-reading platform and social book cataloguing website) and their citations². Goodreads website enables book lovers to share their views and opinions about the books, review books, rate books, and connect with other readers. However, these book reviews may be highly subjective and may contain a high degree of internet-specific vocabulary³.

Traditionally, citation counts are used for measuring the impact of articles, journals, and evaluation of researchers. Differences in citation count among the various databases and their book reviews could have implications for citation analysis studies. For this study, we have taken books with highest citation counts in Scopus and compared their citation profiles in Google Scholar and with Goodread reviews and ratings.

Review of literature

Quantitative and qualitative indicators have been used for evaluation scholarly output including books. Smith⁴ in his study stated that citation counts are the

raw data for assessment of scientific performance, as they are “unobtrusive measures that do not require the cooperation of a respondent and do not themselves contaminate the response”. Similarly, Garfield⁵ stated that “peer review is probably the best way to assess the significance of books; however, alternate metrics also have been introduced as the replacements or supplementary sources of information”. Although publisher book quality is the established measurement, information experts have argued on considering citation data for their impact assessment as well³⁴.

Many studies have proposed that journal centric citation databases can sometimes be lacking for the impact assessment of book-based disciplines^{6,34}. As a result, online citations and informal scholarly indicators have been adapted for social science research evaluation⁶. Studies suggest that there are more citations to books and monographs than journal articles in some social sciences and many arts and humanities subject areas^{18,32}. One such investigation adopted Google⁷, Google Scholar⁸ and Google Books⁹ for the impact assessment of scientific research and found that there was no comprehensive study adopting Google Books, and Google Scholar for the citation impact of books across different disciplines. Instead, many studies have been conducted to assess the relationship between Google Scholar citation with WoS or Scopus.

Kousha⁹, *et al.* conducted a study on Google Book Search citations with Web of Science citations to 10 selected articles in science, social science, and humanities journals. Findings of the study suggest that Google book citations as per ISI citations accounted for 31% to more than double (212%) of the Web of Science citations in social sciences and humanities and only 3% to 5% in the sciences. They also found that book citations measure demonstrated a similar kind of impact to that of ISI citations. Hence, the Google book search is a useful new source of citation data for the social sciences and humanities. Although books content tends to be much longer and requires more time and effort for subject specialists to evaluate, citations to books can also be a useful indicator of their scholarly impact³². Nevertheless, “academic book reviews may repeat the wider impacts of books, such as educational or cultural influence or in addition to their research values”⁹. Moreover, many experts rely on scholarly book reviews for teaching and research usually¹⁰.

Gorraiz¹¹, *et al.* conducted a relationship study between the number of book reviews and citations to books and found that the correlations could differ between fields and that they were predominantly higher in “literature (0.637), history (0.608) and psychology (0.502) than in biology (0.214), chemistry (0.127) and mathematics (0.123)”. Similarly, Dimitrov¹² *et al.* conducted a comparative behaviour study of the reviewers in Goodreads and Amazon.com by adopting 21,394 books with 2.5 million reviews. The finding of the study shows that the average numbers of reviews per book and book reviews per user in Goodreads are higher than in Amazon.com; however, Amazon book reviews were found to be much longer. On the same line, Kovacs¹³ *et al.* conducted a study on Goodreads.com by analyzing the sentiment of 64 award-winning or shortlisted English language books between 2007-2011 and found that award-winning book attracted more readership. At the same time, the reader's ratings were lower for books that were shortlisted but did not win the award.

Kadiresan¹⁴ *et al.* conducted a study on the 2016 Man Booker prize winner book, “*The Sellout*” on Goodreads website. They found that there was more of a negative sentiment score than a positive sentiment score before and after the announcement of the prize. Zuccala¹⁵ *et al.* investigated Goodreads reader ratings of 8,538 books for measuring the broader impact of scholarly books published in the field of History and found that there was a low correlation of 0.212 between citations and reader ratings. Similarly, González-Fernández-Villavicencio³¹ conducted a study to identify the relation between the most reviews and ratings of books in *Goodreads* and *LibraryThing* and their web impact on Amazon only to find slight evidence between different variables. Hence, the more a book is reviewed, the higher one can assume its importance and impact on the scholarly community.

Objectives of the study

- To examine the relationship between citations metric and Goodreads book reviews and ratings;
- To understand the sentiment polarity of reviews across disciplines;
- To investigate the role of Goodreads in academic books; and
- To identify the gender gap and authorship distribution in book publishing.

Methodology

For the present study, metadata of about 2000 books indexed in Scopus database from 1975 to 2018 were retrieved using a random sampling method. Adopting Scopus subject classification scheme, the retrieved books were categorized into social sciences, health sciences, physical sciences, and life sciences academic books. The retrieved metadata consisted of the author name, the title of the book, publication year, citation count, and ISBN. Citation data to the following books (social sciences, health sciences, physical sciences, and life sciences) were retrieved and downloaded from the Scopus database and manually collected from Google citation. The primary reason for the choice of the two databases i.e. Scopus and Google citation was because of their coverage. Although Google book citation is not a citation index, the citation coverage to digitized books is 1.4 and 3.2 times more than Scopus citations, and their medians were more than twice and three times as high as the median citations from Scopus³⁴. The search strategy string used for retrieving data is *doctyp:bk*. A total of 125 samples from each area were considered for the study.

Thereafter, the book was searched in Goodreads, and their corresponding reviews extracted by web scraping Goodreads site using the R programming language. R packages used for scraping web pages include *data.table*, *dplyr*, *magrittr*, *rvest*, *devtools* and *rselenium*. Further, publisher details and the number of ratings given by the readers were also extracted for the analysis.

A fundamental task in sentiment analysis is polarity detection. The polarity of the review sentiment was analyzed using a bag-of-words approach. In this approach, a set of positive and negative words are predefined, and those words are matched with the reviews to extract the sentiment score. Further, text conversion process such as converting each review into characters, removing special characters, numerals, smileys, and empty spaces was executed. Packages used for sentiment calculation are *stringr* and *tm*. The sentiment score is calculated by the difference in the number of positive words and negative words.

Sentiment score = Number of Positive words - Number of Negative words.

The data was extracted from 7 March 2019 to 31 May 2019. R was used for web scraping, text analytics, sentiment detection, and correlation.

Results

Distribution of reviews across various subjects

Table 1 shows the total number of books with and without reviews in Goodreads. Result of web scraping of review readers comment shows that the majority (54.1%) of the book in social science was with reviews and accounted to 72.5% of the total reviews, followed by physical sciences book (25 reviews), life sciences books (22 reviews), and health sciences books (14 reviews). Kousha³² et al., also found that “Goodreads engagements were numerous enough in the arts (85% of books had at least one), humanities (80%), and social sciences (67%) for use as a source of impact evidence.”

Citations vs Goodreads rating, and review

Table 2 shows that physical science discipline books received the highest citation in Scopus (89522), closely followed by social science (76846), health science (22742), and life sciences (22184, 138683). However, the result suggests that social science received the highest citation in Google scholar and received the highest Goodreads overall rating (493.16) as well as reviews (703). These findings align with Tang²¹ findings that books in social science account to 46% of the overall citation in the U.K, while only 12% of the citations are found in the natural science book. Similarly, Nederhof and van Raan⁷ findings also suggest that citations per

Table 1 — Distribution of reviews across various subjects

Sl. no.		Health Sciences	Life Sciences	Physical Sciences	Social Sciences	Total (n=500)
1	No. of books with reviews	14 (11%)	22 (16%)	25 (19%)	72 (54%)	133 (100%)
2	Total no. of reviews	37 (4%)	114 (12%)	115 (12%)	703 (72%)	969 (100%)
3	No. of books not available in Goodreads	13 (30%)	23 (53%)	5 (12%)	2 (5%)	43 (100%)

Table 2 — Citations, reviews and ratings

Sl. no.	Subjects	Scopus Citations	Google Scholar Citations	Goodreads Overall ratings	Goodreads Reviews
1	Health Sciences	22742	75138	292.82	37
2	Life Sciences	22184	138683	252.6	114
3	Physical Sciences	89522	245320	330.77	115
4	Social Sciences	76846	316796	493.16	703

publication were higher for books. As a result, Hicks²⁶ argued that establishes indexing databases tend to miss 40% of the citations received by books.

Sentiment detection

Table 3 shows the polarity of the reviews from the selected academic books. Social Science (72 reviews) books tend to generate the highest reviews from the users. However, physical science, life science, and health science were considerably low. Wang²⁷ et al., also found that academic books in social science and arts and humanities disciplines account for 80% of the total reviews that tend to receive more attention but also receive low-score evaluations on online social platforms. They also indicated that Librarians tend to review the books critically.

Authorship pattern across various subjects

Table 4 shows the number of authorship patterns among various subjects. The overall results show the predominance of single-authored books (117) over multi-authored books. More than two authorships were not seen in physical sciences and social science book samples selected in this study. Unlike research papers, collaboration in academic books seldom occurs. The present findings of our study also resonate Pillai's²⁸ findings that the "average number of authors per journal articles was 3, and for books, it was 1.69."

Contribution of authors by gender

Table 5 shows the contribution of authors by gender. Some of the earlier studies have also explored gender differences in academic and research fields^{16,24}. Determining the author's gender by looking at the name is difficult because of lack of familiarity with names, gender neutral names also because many authors publish their research work using their initials instead of their full first names. In this study, the gender of authors was guessed from their first name using lists of common male and female first names. There were 154 authors across the disciplines, and it was found that 73% of the authors are male, and only 26% of the authors are female. Danell and Hjerem³³ have previously identified the gap that female are minority in academic fraternity. Moreover, a minimal number in the academic also devotes their time to service-related activities and teaching rather than research^{29,30}.

Correlation between citation and readership

Table 6 shows the result of the correlation between Goodreads ratings, reviews, and citations across the discipline. Findings of the study in the field of health science suggest that there is a significant positive correlation between Goodreads overall users rating and the number of reviews at 5%. With the increase in the user's overall rating of Goodreads, the number of reviews by the user also increases. At the same time, there is a positive correlation between Google Scholar

Table 3 — Classification of reviews based on sentiment polarities

Sl. no.	Subjects	No. of books with reviews	No. of books with positive sentiments	No. of books with neutral sentiments	No. of books with negative sentiments
1	Health Sciences	14	9 (64%)	2 (14%)	3 (22%)
2	Life Sciences	22	17(77%)	2 (9%)	3 (14%)
3	Physical Sciences	25	19 (76%)	3 (12%)	3 (12%)
4	Social Sciences	72	45 (63%)	6 (8%)	21 (29%)

Table 4 — Distribution of authorship pattern across various subject

Sl. No.	Authorship Pattern	Health Sciences	Life Sciences	Physical Sciences	Social Sciences	Total
1	Single author	8 (57%)	20 (91%)	20 (80%)	69 (96%)	117
2	Two authors	4 (29%)	1 (5%)	5 (20%)	3(4%)	13
3	More than two authors	2 (14%)	1 (5%)	0 (0%)	0(0%)	3
Total		14	22	25	72	133

Table 5 — Distribution of author gender across various subject fields

Sl.no.	Gender	Health Sciences		Life Sciences		Physical Sciences		Social Sciences	
		Male	Female	Male	Female	Male	Female	Male	Female
1.	Gender	Male	Female	Male	Female	Male	Female	Male	Female
2.	Frequency	16	7	18	8	23	7	56	19

Table 6 — Correlation between Goodreads reviews, rating and citations

Discipline	Variable	Scopus Citations	No. of Reviews in Goodreads	Overall Ratings in Goodreads	Google Scholar Citations
Health Science	Scopus citations	1			
	No. of reviews in Goodreads	-0.019	1		
	Goodreads overall ratings	0.108	0.178*	1	
	Google scholar citations	0.911**	0.063	0.200*	1
Life Sciences	Scopus citations	1			
	No. of reviews in Goodreads	-0.013	1		
	Goodreads overall ratings	.0126	0.199*	1	
	Google scholar citations	0.451**	0.001	0.144	1
Physical Sciences	Scopus citations	1			
	No. of reviews in Goodreads	0.168	1		
	Goodreads overall ratings	0.106	0.170	1	
	Google scholar citations	0.905**	0.295**	0.090	1
Social Sciences	Scopus citations	1			
	No. of reviews in Goodreads	0.019	1		
	Goodreads overall ratings	0.063	0.119	1	
	Google scholar citations	0.650**	0.120	0.093	1

**& * significant at 1% & 5%.

citation and the overall rating of the book by the readers at 5%. No significant correlation was observed between Scopus citation and Goodreads opinion and rating of the books by the user. In the case of life science books, there is a positive correlation between users' overall rating with the number of reviews at 5%.

Discussion

The present study examined the relationship between Goodreads reader opinion and citation, and at the same time attempts to understand the authorship patterns, gender differences, disciplinary and publisher contribution to books in Goodreads. Book reviews are generally viewed as a scholarly communication process whereby new concepts and new ideas are critically analyzed and discussed¹⁸. It is also an alternative source of reference for readers and authors to aid book impact assessment^{19,20,15}. Reader ratings and reviews on Goodreads serve as an indicator of impact beyond academia. Our findings suggest that social science books generated the greatest number of reviews (57.6%) and citation in Google scholar suggesting reader's inclination to citing, rating and reviewing social science books. However, the social science book citation in Scopus was considerably low. Perhaps, one reason for this weak association could be the lack of social science indexed books by Scopus to their citation reports. The present study is in line with the study of Tang²¹ and

Small and Crane²², who reported that social science books receive higher citations to the book than physical science. Therefore, traditional bibliometric databases should understand the importance of book and monograph citation and find the means to be inclusive¹⁰.

Collaboration in scientific literature is considered an important aspect of the development of science. Kong²³ et al. stated that scientific collaboration could help researchers increase their influence. However, in the case of books, and unlike scientific journal articles, findings of our study suggest that fewer authors collaborate in publishing a book. This is evident in social science, whereby 95.8% of the books are single-authored publications. Perhaps, publishing a book involves the author's long-term commitment to the work, ideas, and royalty.

The gender gap in publishing and academia has been reported by various scholars^{16, 24, 25}. In line with Peñas¹⁶, Brink²⁴ et al., findings of our study suggest a gender gap in publishing as well. Male authors significantly published more books than female authors. According to feminist researchers, the gap could be attributed to female researchers prioritizing most of their time to service-related activities or teaching over research activities^{29, 30}.

Moreover, a study conducted by Zuccala¹⁵ found a weak correlation (0.212) between History book citation count and reader-rating counts. In line with Zuccala¹⁵, our present study suggests no correlation

between social science book citation count and user rating and reviews count. Although social science books generated the highest number of reviews and engagement by the user, there seems to be no evidence to suggest that these reviews translate into academic citations. However, the correlation was observed between Goodreads overall rating and Google Scholar citation count in health science. Similarly, there is a positive relationship between Goodreads reviews and Google scholar citation in physical science books as well.

There are a few limitations to the present study. The sample books considered for the study were limited to the Goodreads website of the English language for the analysis. Popular genre and Goodreads genre were not considered for the study. The sample books were selected randomly, and it is suggested that for future study, the entire academic books might be considered for the study for more insight into the user's behaviour.

Conclusion

Goodreads requires further investigation concerning other popular genres and all the academic books and not confined to just a few selected books alone. However, the few academic books we have studied could help understanding how academic books are perceived and evaluated. For academics, Goodreads possessed the potential to be an alternative metric for scholarly impact beyond traditional citation metrics. More importantly, the Goodreads user's rating and sentiment of reviews could potentially make a strong contribution to a complementary approach to the existing bibliometric metric evaluations. Even though there is a potential for manipulation of 'bot' reviews by publishers and author, Goodreads user reviews and book rating undeniably belongs to the realm of altmetrics. Finally, research in social media platforms as an alternative metric is ever-growing inventory, and with time, an improvement upon the tools and methodology would eventually improve.

References

- Hartley J, Reading and writing book reviews across the disciplines, *Journal of the American Society for Information Science and Technology*, 57(9) (2000) 1194-1207.
- Maity S K, Kumar A, Mullick A, Choudhary V, and Mukherjee A, Understanding Book Popularity on Goodreads, *In Proceedings of the 2018 ACM Conference on Supporting Groupwork*, ACM, 2018, pp. 117-121.
- Zhu Z, Multi-document summarization of online book reviews based on information classification, Nanjing: Nanjing University Press, 2013.
- Smith L C, Citation analysis, *Library Trends*, 30(1) (1981) 83-106.
- Garfield E, Citation indexes for retrieval and research evaluation. *Paper presented at the G7 Consensus Conference on the Theory and Practice of Research Assessment*, Capri, Italy, October 1996.
- Kousha K and Thelwall M, The Web impact of open-access social science research, *Library & Information Science Research*, 29(4) (2007) 495-507.
- Nederhof A J and Van Raan A F, A bibliometric analysis of six economics research groups: A comparison with peer review, *Research Policy*, 22(4) (1993) 353-368.
- Vaughan L and Shaw D, Web citation data for impact assessment: A comparison of four science disciplines, *Journal of the American Society for Information Science and Technology*, 56(10) (2005) 1075-1087.
- Kousha K and Thelwall M, Google book search: Citation analysis for social science and the humanities, *Journal of the American Society for Information Science and Technology*, 60(8) (2009) 1537-1549.
- Kousha K, Thelwall M and Rezaie S, Using the web for research evaluation: The Integrated Online Impact indicator, *Journal of Informetrics*, 4(1) (2010) 124-135.
- Gorraiz J, Gumpenberger C and Purnell P J, The power of book reviews: a simple and transparent enhancement approach for book citation index, *Scientometrics*, 98(2) (2014) 841-852.
- Dimitrov S, Zamal F, Piper A and Ruths D, Goodreads versus Amazon: the effect of decoupling book reviewing and bookselling, *In Ninth International AAAI Conference on Web and Social Media*, April 2015. (accessed on 16 December 2019).
- Kovács B and Sharkey A J, The paradox of publicity: How awards can negatively affect the evaluation of quality, *Administrative science quarterly*, 59 (1) (2014) 1-33.
- Kadiresan N, Gogoi T and Singson M, The Sellout: Readers Sentiment Analysis of 2016 Man Booker Prize Winner. *In 11th International Caliber 2017*. Anna University, Chennai: Information and Library Network Centre, Gandhinagar. 2017, 219-229.
- Zuccala A A, Verleysen F T, Cornacchia R and Engels T C, Altmetrics for the humanities: Comparing Goodreads reader ratings with citations to history books, *Aslib Journal of Information Management*, 67(3) (2015) 320-336.
- Peñas C S and Willett P, Brief communication: Gender differences in publication and citation counts in librarianship and information science research, *Journal of Information Science*, 32(5) (2006) 480-485.
- Huang M H and Chang Y W, Characteristics of research output in social sciences and humanities: From a research evaluation perspective, *Journal of the American Society for Information Science and Technology*, 59(11) (2008) 1819-1828.
- Foster C L, Some thoughts concerning book reviewing, *Educational Research Bulletin*, (1951) 64-66.
- Lindholm-Romantschuk Y, Scholarly book reviewing in the social sciences and humanities: The flow of ideas within and among disciplines (No. 91). Greenwood Publishing Group. 1991.
- Zhou Q, Zhang C, Zhao S X and Chen B, Measuring book impact based on the multi-granularity online review mining, *Scientometrics*, 107(3) (2016) 1435-1455.

- 21 Tang R, Citation characteristics and intellectual acceptance of scholarly monographs, *College & Research Libraries*, 69(4) (2008) 356–369.
- 22 Small H and Crane D, Specialties and disciplines in science and social science: An examination of their structure using citation indexes, *Scientometrics*, 1(5–6) (1979) 445–461.
- 23 Kong X, Mao M, Jiang H, Yu S and Wan L, How does collaboration affect researchers' positions in co-authorship networks? *Journal of Informetrics*, 13(3) (2019) 887-900.
- 24 Brink M V D, Brouns M and Waslander S, Does excellence have a gender? A national research study on recruitment and selection procedures for professorial appointments in The Netherlands, *Employee Relations*, 28(6) (2006) 523-539.
- 25 Manuelraj P and Amudhavalli A, Collaboration pattern amongst health care professionals in India. In *Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting*, Berlin., 2008.
- 26 Hicks D, The difficulty of achieving full coverage of international social science literature and the bibliometric consequences, *Scientometrics*, 44 (2) (1999) 193–215.
- 27 Wang K, Xiaojuan L and Yutong H, Exploring Goodreads reviews for book impact assessment, *Journal of Informetrics*, 13 (3) (2019) 874-886.
- 28 Pillai K G, Authorship patterns in physics literature: An informetric study on citations in doctoral theses of the Indian Institute of Science, *Annals of Library and Information Studies*, 54(2) (2007) 90-94.
- 29 Gander J P, Faculty gender effects on academic research and teaching, *Research in Higher Education*, 40 (2) (1999) 171-184.
- 30 Hancock K J, Baum M A and Breuning M, Women and pre-tenure scholarly productivity in international studies: an investigation into the leaky career pipeline, *International Studies Perspectives*, 14 (4) (2013) 507-527.
- 31 González-Fernández-Villavicencio N, Reading indicators on the social networks Goodreads and LibraryThing and their impact on Amazon, *Zeitschrift für Katalanistik*, 32 (2019) 143–167.
- 32 Kousha K, Thelwall M and Abdoli M, Goodreads reviews to assess the wider impacts of books, *Journal of the Association for Information Science and Technology*, 68 (8) (2017) 2004-2016.
- 33 Danell R and Hjern M, Career prospects for female university researchers have not improved, *Scientometrics*, 94 (3) (2013) 999-1006.
- 34 Kousha K, Thelwall M and Rezaie S, Assessing the citation impact of books: The role of Google Books, Google Scholar, and Scopus, *Journal of the American Society for information science and technology*, 62 (11) (2011) 2147-2164.

CORRIGENDUM

The name of the second author of the article Stemming the rising tide of predatory journals and conferences: A selective review of the literature published in *Annals of Library and Information Studies*, Vol. 67, September 2020, pp. 173-182 stands corrected as Anup Kumar Das instead of Anup Das given in the article.