Webology (ISSN: 1735-188X)
Volume 19, Number 2, January, 2022

Comparing the h-Index (Hirsch Index) of Nobel Laureates Scopus and Google Scholar Profiles in Chemistry and Economics Sciences

Abhay Maurya
Department of Library & Information Science, Mizoram University, Aizawl, Mizoram, India, Email: abhaymaurya17@hotmail.com

Amit Kumar
Department of Library & Information Science, Mizoram University, Aizawl, Mizoram, India, Email: amit85kr@gmail.com

Abstract

The present study analyze the data regarding h-Indices and the number of citations in respect of Nobel Laureates in Chemistry and Economic Sciences extracted from Scopus and Google Scholar. The Pearson correlation and Spearman correlation have been used for analysis for all Nobel Laureates. Data from the profiles (Google Scholar and Scopus) are taken for the period 2014 to 2021 (30th Nov., 2021), and noble laureates from 2014-2021, during this period, 22 individuals have conferred the Nobel Prize in Chemistry while 15 economists were given the Nobel Prize in Economic Sciences.

Keyword: Chemistry, Economic science, h index, Citation, Scopus, Google scholar, Pearson Correlation, and Spearman Correlation

1. Introduction

Until the previous decade, most of the global universities relied upon the ISI Citation Indexes and the Journal Citation Reports deciding upon promotional avenues of researchers. Despite the criticism received by the Citation Indexes, the ISI Citation Index was the most highly used to gather information on citations. Presently several versions of citation indexes are available which include the Web of Science developed by Eugene Garfield and maintained by Clarivate Analytics, Scopus developed by Elsevier, and Google Scholar. Each index has its unique policy relating to the collection of papers. This paper is an attempt to understand the difference between Scopus and Google Scholar using the h-index of Nobel Laureates in Chemistry and Economic Sciences from 2014 till 2021 (30th Nov., 2021).

In 2005 J E Hirsch published an article titled '*An index to quantify an individual's scientific research output'* published in the *Proceedings of the National Academy of Sciences of the United States of America* where he had proposed the h-index as a quantitative metric which has its foundations of the number of publications and the number of citations. Hirsch had defined the h-index using the words "*A scientist has index h if h of his or her Np papers have at least h citations each and the other (Np – h) papers have ≤h citations each*."

Hirsch had argued that the *h* index should be preferred to other criteria like the total number of publications, the total number of citations and citations per paper and had included certain caveats:

- ✓ Any single number cannot describe the profile of any researcher requiring the consideration of other factors.
- ✓ The h-values of the authors would be different across different fields of knowledge.
- ✓ In case an author publishes a few papers with high citation, the h-index would not do justice to his/her accomplishments and has advocated considering the entire career of the individual[1].

2. Literature Review

The Science Citation Index was developed by Eugene Garfield (1961) in 1963 and uses the citation data from the year 1961. The Web of Science was the only citation database that was used by scholars until 2004 when Scopus was launched on the 3rd day of November 2004 (J. Bar-Ilan, 2007) and Google Scholar was launched on 18th November 2004 (D. Payne, 2004). With 27 million abstracts, 230 million references from 200 million web pages, the Scopus database is considered the largest abstract and citation database (Elsevier, 2021). While Scopus contains citation data for the items present in its database, Google Scholar, though free to use, does not contain information regarding the volume of its database and the age of its coverage. Google Scholar has in its database only those articles whose abstract is free to use (Google Scholar,2021), while access to the full text depends upon the author. Google Scholar also sources its data from various sources which include preprint servers, personal websites, and others.

Despite being free to use, Google Scholar has had its share of criticism. While Jim Giles has noted that various members of the faculty and the students have started using Google Scholar and has predicted that Google Scholar would replace the more sophisticated search tools (J. Giles, 2005), certain other scholars and researchers have opined on the contrary stating that only 113 ARL libraries have been linked to Google Scholar suggesting the lower levels of enthusiasm among the librarians than their clients (L. Mullen & K. Hartman, 2006). Based upon their studies that analyzed citation counts of JASIST articles published between 1985 and 2000, scholars have observed that the citation counts in Google Scholar were considerably higher than in Web of Science or Scopus (K. Bauer & N. Bakkalbasi, 2005). This observation led the researchers to conclude that researchers should consult other citation databases like Scopus or Web of Science besides Google Scholar, especially in case of recent productions, or new domains. During the searching procedure for the h-index for this study, it was observed that Google Scholar has more items than Scopus.

According to Jacso, P (2006) and certain other researchers have also criticized Google Scholar for its inconsistencies. The inability of Google Scholar to identify the publication year and its inability to attribute the citations to the correct publications have made scholars to conclude that Google Scholar cannot substitute the Web of Science. The researchers have noted the deflation in citation counts in Google Scholar, as it has in its database non-scholarly articles and calculates preprints and journal versions separately. Researchers have noted that Google Scholar does not work with Boolean and range operators, a problem that the database has inherited from Google (J. Bar-Ilan, 2005). The inability of Google Scholar to identify the publication year and its inability to attribute the citations to the correct publications have made scholars to conclude that Google Scholar cannot substitute the Web of Science (P. Jacso, 2006). However, this conclusion contradicts the conclusion made by (D. Pauly & K. Stergiou, 2005). Certain scholars who have found that Google Scholar was more useful in evaluating researchers in the domain of computer sciences (E. Rahm & A. Thor, 2005). Susan Gardner and Susanna Eng (2005) have compared Google Scholar with several databases concerning Social Science. The authors concluded that, despite the shortcomings observed by previous studies, Google Scholar can show significant improvement. Peter Jocso (2005) analyzed the citations available in Garfield's work and observed that while Web of Science lists 83 citations, Scopus had listed 76 citations and Google Scholar had 82 citations listed in its database. However, observing that only 33 items were common in all the databases, the author concluded that there are features other than citation counts. Similar results have been obtained by (K. Kousha & M. Thelwall, 2006) when the authors have observed that the overlap between Google Scholar and Web of Science was very low and only 33% in the domain of Chemistry. One of the major disadvantages of the h-index is its inability to account for any specific field of research making a comparison of h-indices of two scientists of different fields of research very difficult. This difficulty arises due to differences in the publication rates and the citation rates across the different fields of knowledge. Researchers have observed that on average, articles in life sciences receive 6 times more citations than articles in mathematics or computer sciences (R. Adler, J. Ewing & P. Taylor, 2009). This makes comparing citation outputs between scientists of two domains difficult. Scholars have pointed the need to find accurate ways to measure the research performance of researchers other than using the h-index, by using relevant measures like an average number of citations per publication for any particular scientific field of study (A. Bletsas & J. Sahalos, 2009).

In his seminal work, Hirsch (2005) has mentioned that researchers in life sciences have high h-values compared to researchers in other sciences like physics. Hirsch (2005) has quoted the example of a moderately productive scientist in physics as having an h equal to the number of years of service while scientists in biological sciences have higher values of h creating differences in the values of h-index in different fields. Hirsch (2005) has further stated that researchers engaged in domains other than science will not have the same values of h-index as the top stratum of those working in highly topical areas. As such, all these HCRs also have high values of h indices, and those high values of h-indices in the life sciences are much more than in physics. These results lead to the conclusion that h indices in biological sciences tend to be higher than in physics, besides indicating that the difference appears to be much higher at the high end than on average.

Several researchers including Batista, P et. al. (2005) has observed that the number of citations received by any publication is dependent on the number of authors. This observation has been the basis for proposing a complementary to the Hirsch Index. Juan E Iglesias and Carlos Pecharroman (2007) proposed a formula to calculate the h-index accounting for the number of citations

$$h = \sqrt[3]{\frac{N_p}{4}\, X^{2/3}}$$

where, Np represents the total number of papers published; and
χ is the average number of citations per paper for the researcher.

Based on the above formula J. Iglesias and C. Pecharromán (2007) suggested using as a normalizing factor for the h-index and having the following expression:

$$f_i = \left({X_{Physics}}/{X_i}\right)^{2/3}$$

where χi is the average number of citations per paper of scientific field i, and
χ physics stands as the reference category.

Thus, the normalized h-index is given by the formula:

$$h_{normalized} = f_i X\, h = \left({X_{Physics}}/{X_i}\right)^{2/3} X\, h$$

The normalization method is applied to a real dataset that comprise h-index values of HCRs affiliated with Spanish institutions. The results show that the h values become more homogeneous after correction with the normalizing factor (Malesios, C. & Psarakis, S., 2012).

3. Research Gap

The study is unique in itself as after gone through the extensive review of literature it is found that there are numerous studies have been undertaken by scholars on conceptual aspects of h-index but the study of h-index with both Google Scholar and Scopus profiles in two disciplines are still not taken by the scholars. So the paper fills this research gap and opens the path for other scholars to undertake the research on it.

4. Objectives of the Study

The objectives of the study lie in understanding the differences in the value of the h-index across various citation databases. We have considered Scopus and Google Scholar and have analyzed the h-index of Nobel Laureates in Chemistry from 2014 till 2021. Further, this study also compares the h-index of chemistry with that of economic sciences for the said period.

5. Methods Employed

The h-indexes of the Nobel Laureates in Chemistry and Economic Sciences have been extracted from Scopus and Google Scholar. The data has been fed into a spreadsheet in Microsoft Excel and all visualizations have been done using Microsoft Excel.

4.1 Data set

Tang, Zhang, et. al. (2008) have proposed conducting the process of disambiguation during the collection of data. The problem related to duplication of authors was encountered during the process of data collection on several occasions. Authors having the same name but different IDs were found as were the same authors with different IDs. On analysis, it was realized that the affiliation of an author during the time of being awarded the Nobel Prize was an important criterion that could help in the identification of the author. We/I are/am based on our/my search of identifying all authors having the same names. This was followed by checking their affiliation. Those authors whose affiliation matched the affiliation at the time of being considered for awarding the Nobel Prize were considered for the study. After completion of the process of de-duplication, we were left with 18 Nobel Laureates in Chemistry and 10 Nobel Laureates in Economic Sciences.

The required data were obtained from the profiles (Google Scholar and Scopus) by personal visit. Firstly, the Scopus website/database was accessed with the search query of name followed by matching and verifying their affiliation before final selection and the same criteria in case of Google Scholar as well. The data sets are downloaded till 30th November, 2021.

6. Results

During the period 2014 till 2021, 22 individuals have conferred the Nobel Prize in Chemistry while 15 economists were given the Nobel Prize in Economic Sciences. Their h-indexes and the numbers of citations have been tabulated in Table 1 and Table 2 (till 30th Nov., 2021). We have considered all works produced by the Nobel Laureates since the commencement of their productive life and till 2021 and have analyzed the same.

Table 1: h-Index of Chemistry Nobel Laureates

Year	Name	Scopus		Google Scholar	
		h-Index	CIT.	h-Index	CIT.
2014	Eric Betzig	61	25483	72	37848
	Stefan Walter Hell	99	46777	133	68949
	William Esco Moerner	78	29860	101	46099
2015	Tomas Robert Lindahl	99	38916	45	7277
	Paul Laurence Modrich	79	22816	101	36473
	Aziz Sancar	107	38982	135	59252
2016	Jean- Pierre Sauvage	99	37393	109	54363
	James Fraser Stoddart	139	103411	130	130864
	Bernard Lucas Feringa	120	63335	138	80381
2017	Jacques Dubochet	56	12557	49	14277
	Joachim Frank	91	30131	111	47268
	Richard Henderson	31	21336	64	29318
2018	Frances Hamilton Arnold	101	37177	136	61512
	George Pearson Smith	29	5412	52	37205

	Gregory Paul Winter	86	32293	103	58147
2019	John Bannister Goodenough	144	111407	20	1770
	Michael Stanely Whittingam	69	26712	44	5493
	Akiro Yashino	12	1284	2	426
2020	Emmanuelle Charpentier	45	20953	69	16891
	Jennifer Anne Doudna	107	54120	138	89621
2021	Benjamin List	88	32894	96	41823
	David MacMillan	100	43366	110	53163

Table 1 is a tabular representation of the h-Index and the number of citations that the works of Nobel Laureates in Chemistry have received. The figures have been extracted from Scopus and Google Scholar. The table shows that h-Index has a direct correlation with the number of citations. However, the h-Index calculated using data from Scopus deviates from that calculated using data from Google Scholar. The h-Index calculated by Google Scholar shows that despite having a lower citation number (80381 as against 130864), Bernard Lucas Feringa has a higher h-index (138) as against James Fraser Stoddart (h-Index = 130). Similarly Richard Henderson (h-Index = 64, citations = 29318) is higher than George Pearson Smith (h-Index = 52, citation = 37205).

Table 2: h-Index of Nobel Laureates in Economic Sciences

Year	Name	Scopus		Google Scholar	
		h-INDEX	CIT.	h-INDEX	CIT.
2014	Jean Tirole	72	27129	146	212779
2015	Angus Stewart Deaton	54	16381	112	98576
2016	Oliver Simon D’Arcy Hart	30	6519	16	1146
	Bengt Robert Holmstrom	24	7729	97	74745
2017	Richard H Thaler	51	30571	103	150730
2018	William Dawbney Nordhaus	41	10244	120	79588
	Paul Michael Romer	1	7	55	105010
2019	Abhijit Vinayak Banerjee	51	13768	98	66506
	Esther Duflo	50	17005	98	77527
	Michael Robert Kremer	48	10058	87	43598
2020	Paul Robert Milgrom	32	12592	83	108308
	Robert Butler Wilson	22	4642	54	30591
2021	David Card	50	14083	110	79862
	Joshua David Angrist	49	19669	84	82594
	Guido Wilhelmus Imbens	55	22806	85	75182

Table 2 tabulates the h-Index and number of citations in respect of Nobel Laureates in Economic Sciences as extracted from Scopus and Google Scholar. The data also shows the positive correlation among the two variables, viz. the h-Index, and the number of citations. However, the h-Index calculated by Google Scholar shows some discrepancies. While Richard H Thaler has an h-index of 103 with 150730 citations, the h-index of William Dawbney Nordhaus has been calculated at 120 though the works of the Nobel Laureate have been cited 79588 times. Further, Abhijit Vinayak Banerjee and Esther Duflo have an h-Index of 98 though they have a different number of citations.

A look into the data in Table 1 and 2 show the direct correlation that the number of citations has on the h-index. However, in most of the cases, the h-index calculated using data obtained from Google Scholar is found to be higher than that calculated using data from Scopus.

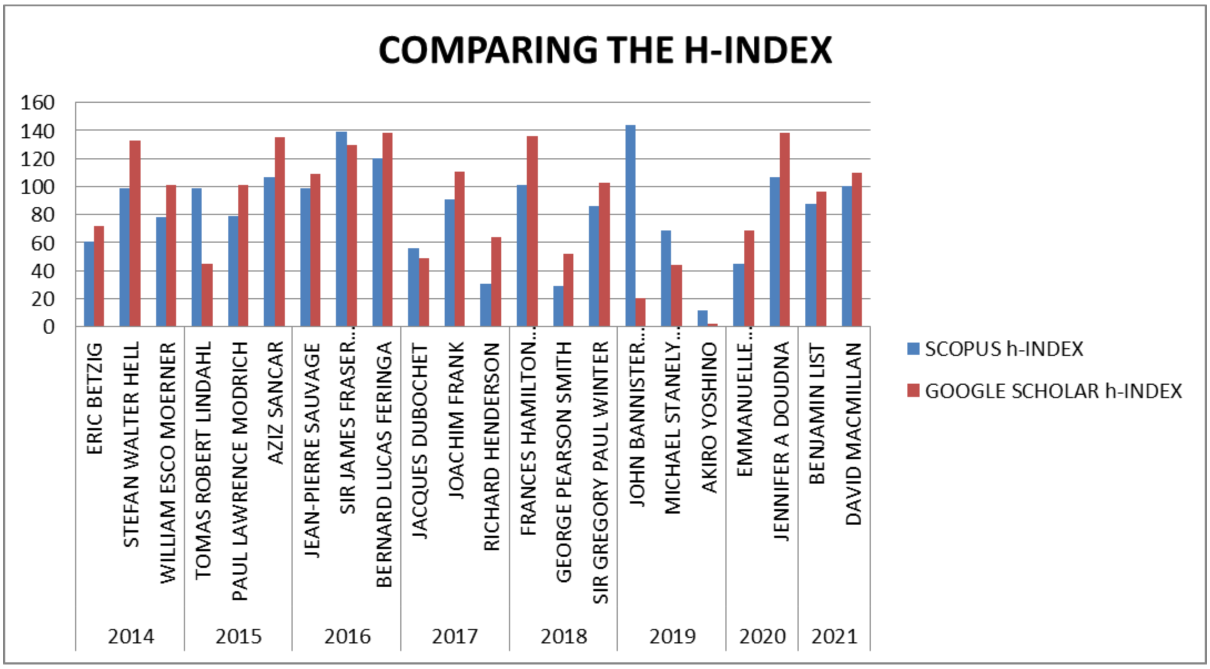

Fig 1: Comparing the h-index- Nobel Laureates in Chemistry

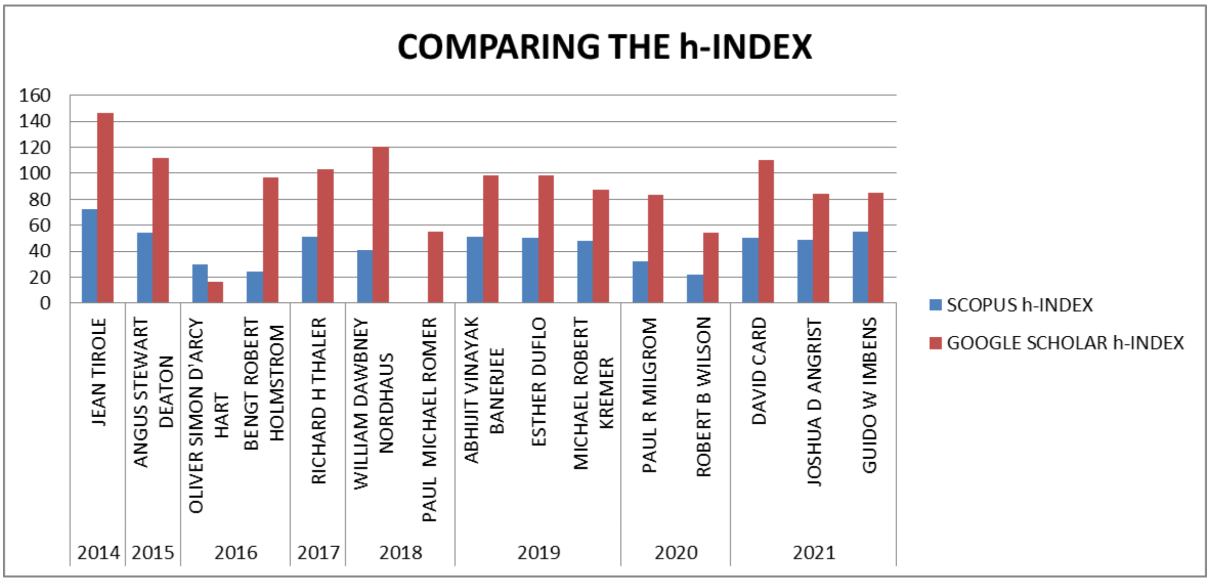

Fig 2: Comparing the h-index- Nobel Laureates in Economic Sciences

Figure 1 and 2 depicts the visual representation of the information provided in Table 1 and 2.

To correlate the number of citations with the h-index, both Pearson Coefficient and Spearman Rank Coefficient have been used. While the Pearson correlation coefficient measures the strength of the association between the h-Index and many citations, Spearman Rank Correlation measures the strength and direction of ranked h-Index and Ranked a number of citations.

The value of the Pearson Correlation Coefficient can vary from -1 to +1. While the value of +1 represents direct correlation, a value of -1 represents indirect correlation. A value of 0 indicates no correlation.

The values of the coefficients of both the correlations have been produced in Tables 2 to 5 below.

Table 2: Pearson Coefficient for Nobel Laureates in Chemistry using data from Scopus and Google Scholar

	Scopus	Google Scholar
Coefficient(r)	0.873105873	0.857106076
N Number	22	22
T Statistic	8.009031162	7.440880593
Degree of Freedom	20	20
P-Value	1.14552E- 07	3.50417E-07

The Pearson Correlation coefficient of the Nobel Laureates in Chemistry using data regarding h-Index and the number of citations extracted from Scopus and Google Scholar is tabulated in Table 2. Analysis of the data shows a high level of correlation of h-Index with the number of citations. However, the correlation of h-Index with the number of citations is higher for data extracted from Scopus compared to data collected from Google Scholar. However, the P-value less than the significant figure of 0.05 indicates a very high correlation.

Table 3: Pearson Coefficient for Nobel Laureates in Economic Sciences using data from Scopus and Google Scholar

	Scopus	Google Scholar
Coefficient(r)	0.83671053	0.713865926
N Number	15	15
T Statistic	5.50867897	3.675488913
Degree of Freedom	13	13
P-Value	0.000100653	0.002797834

A similar conclusion can be drawn for data regarding Nobel Laureates in Economic Sciences. Table 3 shows that the correlation of h-Index with the number of citations is higher for data extracted from Scopus as against data extracted from Google Scholar. The P-value lowers than 0.05 shows a positive correlation of h-Index with many citations.

Table 4: Spearman Rank Coefficient for Nobel Laureates in Chemistry using data from Scopus and Google Scholar

	Scopus	Google Scholar
Coefficient(r_S)	0.962398474	0.952542525
N Number	22	22
T Statistic	15.84431935	13.99415907
Degree of Freedom	20	20
P-Value	8.72992E- 13	8.5935E - 12

On analyzing the Spearman Rank coefficient for Nobel Laureates in Chemistry, Table 4 shows a high positive correlation of rank h-Index with citation numbers.

Table 5: Spearman Rank Coefficient for Nobel Laureates in Economic Sciences using data from Scopus and Google Scholar

	Scopus	Google Scholar
Coefficient(r_S)	0.8885511256	0.452189635
N Number	15	15
T Statistic	6.871788776	1.827954962
Degree of Freedom	13	13
P-Value	1.1324E-05	0.090589288

Table 5 which calculates the Spearman Rank coefficient for Nobel Laureates in Economic Sciences also shows similar results.

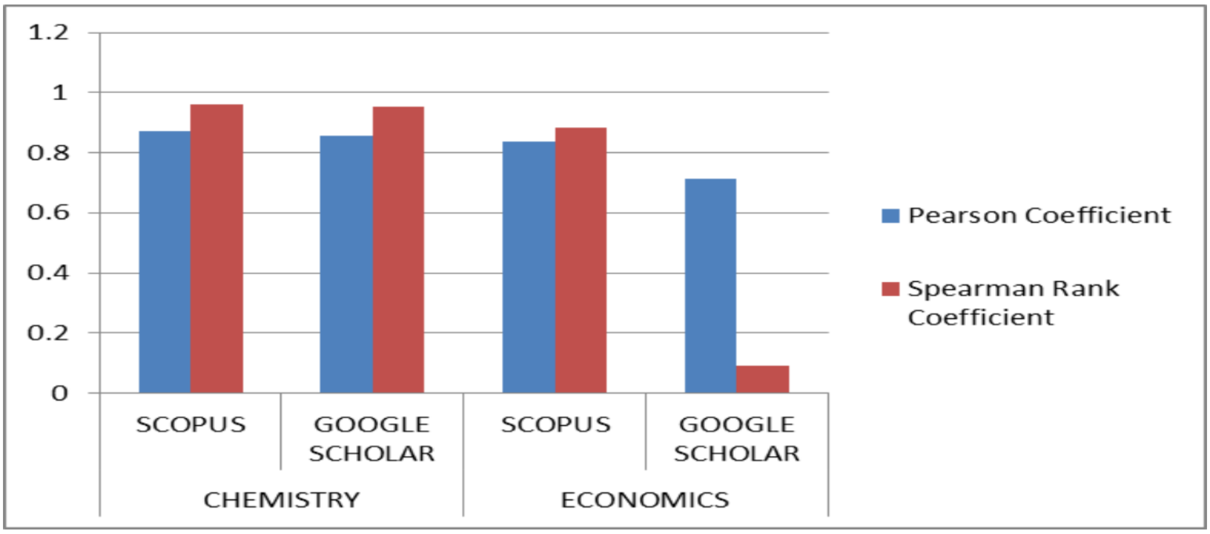

Fig 3: Correlation

7. Discussion and Conclusion

Analysis of the data regarding the h-Indices and the number of citations in respect of Nobel Laureates in Chemistry and Economic Sciences which have been extracted from Scopus and Google Scholar shows that the h-index is dependent upon the number of citations, a fact that has been mentioned by several scholars and detailed in the literature review. The data also show that the h-Index, for both Chemistry and Economic Sciences, extracted from Google Scholar is higher than that obtained from Scopus.

Results of the correlation among the h-Index and citation number are shown in Fig 3. While both Pearson correlation and Spearman correlation have been shown, for all Nobel Laureates, it has been observed that the Pearson correlation coefficient has lower values than Spearman ranked correlation coefficient. This observation implies that despite being highly correlated, the ranked lists obtained based on the h-Index and number of citations show a higher correlation. Pearson correlation is a linear relationship between two variables, while Spearman rank correlation is a measure of monotonic relationship which can be non-linear (J. Hauke & T. Kossowski, 2011).

There exists a substantial difference in the citation counts and the h-Index of the Nobel Laureates among the databases considered for the study. This study has extracted data from Scopus and Google Scholar. Scopus has a significantly higher number of indexed journals containing scientific literature. The contents in Scopus are authenticated by the Scopus Content Selection and Advisory Board, which is a conglomeration of international scientists, librarians, and researchers. Google Scholar, on the other hand, does not have any board for authentication. The h-index is calculated based on the number of publications and the number of times these publications have been cited by other researchers. As such, depending upon the area of focus, maturity, the h-Index of any researcher can vary across the databases. While Scopus has an independent reviewer, Google Scholar does not provide any information regarding the same. As has been previously mentioned, Google Scholar indexes publications in its database provided the researcher or the publisher is willing to provide free access to the abstract. This condition breeds dishonesty among the stakeholders to upload false scholarly papers to increase the citation counts. Further, the h-Index calculated by Google Scholar includes citations before the official publication of the article J. Bar-Ilan, 2007).

The analysis of the data relating to h-Index as extracted from Scopus and Google Scholar databases shows that both citation counts and the h-Index differ among the databases used in the study.

Considering the fact, that the authenticity, integrity, and reliability of Google Scholar, are a matter of intense debate, Scopus is a better database to calculate the h-Index.

REFERENCE

[1] J. Hirsch, "An index to quantify an individual's scientific research output", *Proceedings of the National Academy of Sciences*, vol. 102, no. 46, pp. 16569-16572, 2005. Available: 10.1073/pnas.0507655102.

[2] E. Garfield, "Science Citation Index", *Science Citation Index 1961*, vol. 1, p. v - xvi, 1963. Available: http://garfield.library.upenn.edu/papers/80.pdf.

[3] J. Bar-Ilan, "Which h-index? — A comparison of WoS, Scopus and Google Scholar", *Scientometrics*, vol. 74, no. 2, pp. 257-271, 2007. Available: 10.1007/s11192-008-0216-y.

[4] D. Payne, "Google Scholar welcomed", *Genome Biology*, vol. 5, pp. spotlight-20041123-01, 2004. Available: 10.1186/gb-spotlight-20041123-01.

[5] Elsevier, "About Scopus", 2021. [Online]. Available: https://www.elsevier.com/solutions/scopus.

[6] Google Scholar, "Google Scholar Support for Publishers", *Scholar.google.com*, 2021. [Online]. Available: http://scholar.google.com/intl/en/scholar/publishers.html.

[7] J. Giles, "Start your engines", *Nature*, vol. 438, no. 7068, pp. 554-555, 2005. Available: 10.1038/438554a.

[8] L. Mullen and K. Hartman, "Google Scholar and the Library Web Site: The Early Response by ARL Libraries", *College & Research Libraries*, vol. 67, no. 2, pp. 106-122, 2006. Available: 10.5860/crl.67.2.106.

[9] K. Bauer and N. Bakkalbasi, "An Examination of Citation Counts in a New Scholarly Communication Environment", *D-Lib Magazine*, vol. 11, no. 9, 2005.

[10] P. Jacso, "Deflated, inflated and phantom citation counts", *Online Information Review*, vol. 30, no. 3, pp. 297-309, 2006. Available: 10.1108/14684520610675816.

[11] J. Bar-Ilan, "Expectations versus reality–Search engine features needed for Web research at mid 2005.", *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics*, vol. 9, no. 9, 2005.

[12] D. Pauly and K. Stergiou, "Equivalence of results from two citation analyses: Thomson ISI’s Citation Index and Google’s Scholar service", *Ethics in Science and Environmental Politics*, vol. 9, pp. 33-35, 2005. Available: 10.3354/esep005033.

[13] E. Rahm and A. Thor, "Citation analysis of database publications", *SIGMOD Record*, vol. 34, no. 4, pp. 48-53, 2005.

[14] S. Gardner and S. Eng, "Gaga over Google? Scholar in the Social Sciences", *Library Hi Tech News*, vol. 22, no. 8, pp. 42-45, 2005. Available: 10.1108/07419050510633952.

[15] P. Jacso, "As we may search–Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases", *Current Science*, vol. 89, pp. 1537-1547, 2005.

[16] K. Kousha and M. Thelwall, "Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines", in *Book of Abstracts, 9th International Science & Technology Indicators Conference*, K. Kousha and M. Thelwall, Ed. Leuven, Belgium, 2006, pp. 72-73.

[17] R. Adler, J. Ewing and P. Taylor, "Citation statistics", *Statistical Science*, vol. 24, no. 1, pp. 1 - 14, 2009.

[18] A. Bletsas and J. Sahalos, "Hirsch Index Rankings Require Scaling and Higher Moment", *Journal of the American Society for Information Science and Technology*, vol. 60, no. 12, pp. 2577-2586, 2009.

[19] P. Batista, M. Campiteli, O. Kinouchi and A. Martinez, "Universal Behavior of a Research Productivity Index.", *ArXiv:physics/0510142*, 2005.

[20] J. Iglesias and C. Pecharromán, "Scaling the h-index for different scientific ISI fields", *Scientometrics*, vol. 73, no. 3, pp. 303-320, 2007. Available: 10.1007/s11192-007-1805-x.

[21] C. Malesios and S. Psarakis, "Comparison of the h-index for different fields of research using bootstrap methodology", *Quality & Quantity*, vol. 48, no. 1, pp. 521-545, 2012. Available: 10.1007/s11135-012-9785-1.

[22] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang and Z. Su, "ArnetMiner", *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08*, 2008. Available: 10.1145/1401890.1402008.

[23] J. Hauke and T. Kossowski, "Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data", *QUAGEO*, vol. 30, no. 2, pp. 87-93, 2011. Available: 10.2478/v10117-011-0021-1.

[24] J. Bar-Ilan, "Which h-index? — A comparison of WoS, Scopus and Google Scholar", *Scientometrics*, vol. 74, no. 2, pp. 257-271, 2008. Available: 10.1007/s11192-008-0216-y.