

Is Sci-Hub Increasing Visibility of Indian Research Papers? An Analytical Evaluation

Vivek Kumar Singh^{1,*}, Satya Swarup Srichandan¹, Sujit Bhattacharya²

¹Department of Computer Science, Banaras Hindu University, Varanasi, Uttar Pradesh, INDIA.

²CSIR-National Institute of Science Technology and Development Studies, New Delhi, INDIA.

ABSTRACT

Sci-Hub, founded by Alexandra Elbakyan in 2011 in Kazakhstan has, over the years, emerged as a very popular source for researchers to download scientific papers. It is believed that Sci-Hub contains more than 76 million academic articles. However, recently three foreign academic publishers (Elsevier, Wiley and American Chemical Society) have filed a lawsuit against Sci-Hub and LibGen before the Delhi High Court and prayed for complete blocking these websites in India. It is in this context, that this paper attempts to find out how many Indian research papers are available in Sci-Hub and who downloads them. The citation advantage of Indian research papers available on Sci-Hub is analysed, with results confirming that such an advantage do exist.

Keywords: Indian Research, Indian Science, Black Open Access, Open Access, Sci-Hub.

Correspondence

Vivek Kumar Singh

Department of Computer Science,
Banaras Hindu University,
Varanasi-221005, INDIA.

Email id: vivekks12@gmail.com

Received: 16-03-2021

Revised: 29-03-2021

Accepted: 25-04-2021

DOI: 10.5530/jscores.10.1.16

INTRODUCTION

Responsible Research and Innovation (RRI) has become one of the essential characteristics of 'good' research. This has led to various new normative principles that researchers, funding bodies, and knowledge users are expected to adhere. Various bodies such as the European Commission has spelt out the specific aspects that they would like to embed in member states under the RRI.^[1] One key attribute among the different parameters they underscore is the need to make sustained efforts for knowledge to be Open Access so that it promotes accessibility of science and research to public.^[2] The skewedness in availability of research papers shows large divergence between developed and developing economies and even within a country that needs to be addressed.

One of the key issues is the increasing subscription cost of journals from large publishing houses that is becoming a serious deterrent in universal open access of knowledge. Not only institutions in developing countries are finding it difficult to pay these costs but several institutions in the developed world are also cancelling subscriptions of many journals. Many well-known Universities in the world are now entering into different kinds of agreements with publishers for open

access publishing of their research output, and at the same time encouraging their researchers to publish in openly accessible forms. Universities and Institutions are also creating Institutional Repositories (IRs) to facilitate researchers to keep local copies of papers published by them in different journals. In fact, some of the top research funding agencies of the world have now made it mandatory that any research supported by them should be submitted as pre- or post-print in different repositories, either institutional or disciplinary. However, despite all these efforts barriers to universal access of knowledge still exist.

Sci-Hub, founded by Alexandra Elbakyan in 2011 in Kazakhstan has, over the years, emerged as a very popular source for researchers to download scientific papers. It is believed that Sci-Hub contains more than 76 million academic articles. Bohannon^[3] worked with Alexandara Elbakyan, and analysed access log of Sci-Hub searching for an answer to his question "Who's downloading pirated papers from Sci-Hub", and answered it as "Everyone". While, Alexndra Elbakyan,^[4,5] terms Sci-Hub a "true solution to open access" few others^[6,7] argue that Sci-Hub is at least a signal of failure of current open access models if not a solution. Some others, on the other hand, have criticized Sci-hub for various reasons ranging from copyright violation to threatening commercial viability of publishers.

Due to its controversial nature, Sci-Hub has faced many legal suites, resulting to being blocked in several countries such as United States, Sweden, Russia, France etc. Sci-Hub, however, has managed to stay alive by changing its URLs and location

Copyright

© The Author(s). 2021 This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

of its servers. Recently three foreign academic publishers (Elsevier, Wiley and American Chemical Society) have filed a lawsuit against Sci-Hub and LibGen before the Delhi High Court and prayed for complete blocking these websites in India through a so-called dynamic injunction.^[8] The matter is being heard by the Court and if the petition succeeds, these websites may face similar action to what happened in United States in 2017.^[9,10] Many argue that the outcome of the case against Sci-Hub and LibGen may have long-term consequence to research and education in India^[11] and that blocking Sci-Hub may actually hurt national interest.^[12] This has motivated us to undertake a two-part study on Sci-Hub in the Indian context. This paper, the first part of the study, attempts to find out how many Indian research papers are available in Sci-Hub and who downloads them, and the second part of the study^[13] analyses the quantum of papers downloaded by Indian researchers from Sci-Hub. The question that we intend to answer through the two studies is the influence of Sci-Hub on Indian research visibility, and access. The evidence-based study would help inform the current debate i.e., the consequence of blocking Sci-Hub. At the same time, it would also show the need to create mechanisms that allows researchers to have access to papers that are not in Open Access so that they are not compelled to download from sites like Sci-Hub.

DATA AND METHODOLOGY

The complete publication record for India for the year 2016 as indexed in Web of Science is obtained and analysed. This comprised of 76,530 publication records, out of which 67,857 publication records had a DOI. Since Sci-Hub mainly provides an efficient retrieval for DOI-based queries, we used the publication records with DOI for the analysis. The publication records for the year 2016 were used as analytical data due to the fact that last available access log of Sci-Hub is only for the year 2017. Therefore, it would not have been possible to find out who is downloading Indian papers if data for later publication year would have been used.

The *first* point of analysis was to find out how many Indian research papers are available in Sci-Hub for download. For this purpose, a computational query system was designed to automatically query the Sci-Hub website for each DOI in the data set. Each DOI in the data was searched for presence of full-text in Sci-Hub. The hits and misses were recorded. *Secondly*, the availability of Indian research papers was seen in the category of different document types, publishers and subject areas. The subject area grouping proposed by Rupika *et al.*^[14] was used for showing subject area-wise availability. *Thirdly*, the Sci-Hub access log^[15] for 2017 was obtained and analysed to find out the quantum of download of Indian research papers for the year 2016 and to identify locations from which they are downloaded.

Indian research papers available in Sci-Hub

The automated lookup for Indian papers in Sci-Hub resulted in hit for 61,243 unique papers out of total 67,857 papers. This constitutes about 90.25% of the total research output for India in the year 2016 as indexed in Web of Science (Table 1). It would be relevant here to mention a previous study,^[16] which showed that only 24% research papers for the year 2016 indexed in Web of Science are available in legal gold-green forms of open access. When this is compared to availability in Sci-Hub, there is a huge contrast observed. This is perhaps also an indication of the fact that the legal open access models are not working.

Table 2 shows the availability in Sci-Hub for the papers of different document types. It is observed that the three major

Table 1: Dataset Description.

No. of papers in WoS	Papers with DOI	No. of papers that are on Sci-Hub	Percentage ¹
76,530	67,857	61,243	90.25%

¹Percentage is calculated with respect to number of articles in WoS having a DOI, i.e. column 2.

Table 2: Sci-Hub availability by Document Type.

Document Type	2016			Percentage ¹
	No. of papers in WoS	No. of papers with DOI	No. of papers that are on Sci-Hub	
Article	63259	59369	54739	92.20
Review	3372	3256	2855	87.68
Meeting Abstract	5404	1458	1368	93.83
Editorial Material	1832	1555	1130	72.67
Letter	1899	1593	1056	66.29
Proceedings Paper	1039	1025	967	94.34
Correction	328	318	301	94.65
Book Review	256	214	195	91.12
Book Chapter	60	58	57	98.28
Retraction	34	32	31	96.88
Retracted Publication	18	18	18	100.00
News Item	86	37	15	40.54
Biographical-Item	52	22	13	59.09
Software Review	2	2	2	100.00
Data Paper	1	1	1	100.00
Reprint	1	1	1	100.00

¹Percentage is calculated with respect to number of articles in WoS having a DOI, i.e. column 3.

document types that contain useful material, namely Article, Proceedings Paper and Book Chapter, the availability is more than 90%. For the document type Review, it's a bit less (~88%). Thus, all the major document types have significantly high proportion of papers available in Sci-Hub. Table 3 shows the publisher-wise paper availability in Sci-Hub. It is observed that for the majority of the 20 publishers analysed, the availability percentage of Indian papers is more than 90%. Thus, Sci-Hub hosts papers from almost all major publishers in same proportions.

The next analysis of availability percentage was done with respect to subject area. Table 4 shows the subject area-wise availability for the following 14 major subject areas: Agriculture (AGR), Art and Humanities (AH), Biology (BIO), Chemistry (CHE), Engineering (ENG), Environment Science (ENV), Geology (GEO), Information Sciences (INF), Material Science (MAR), Mathematics (MAT), Medical Science (MED), Multidisciplinary (MUL), Physics (PHY) and Social Science (SS). It is seen that the availability percentage varies

Table 3: Sci-Hub availability by Publisher (Top 20).

Article Type	2016			
	No. of papers in WoS	No. of papers with DOI	No. of papers that are on Sci-Hub	Percentage ¹
Elsevier Science BV	6954	6949	6850	98.50
Springer	5895	5581	5536	93.91
Royal Society of Chemistry	3936	3936	3929	99.82
Pergamon-Elsevier Science Ltd	3246	3234	3194	98.40
Elsevier Sci Ltd	3433	3164	3141	91.49
Taylor and Francis Ltd	2562	2525	2515	98.17
Taylor and Francis Inc	1828	1811	1807	98.85
American Chemical Society	1966	1811	1774	90.23
Wiley-Blackwell	2458	1727	1701	69.20
Elsevier Science Inc	1772	1717	1691	95.43
Wiley	1665	1533	1519	91.23
Elsevier Science Sa	1506	1506	1489	98.87
Springer Heidelberg	1284	1283	1272	99.07
Wiley-V C H Verlag Gmbh	1170	1170	1157	98.89
IEEE	1158	1150	1125	97.15
IOP Publishing Ltd	1052	1044	997	94.77
Springer India	1074	907	900	83.80
Nature Publishing Group	972	907	900	92.59
American Physical Society	919	918	899	97.82

¹ Percentage is calculated with respect to number of articles in WoS having a DOI, i.e. column 3

Table 4: Subject area-wise availability of papers in Sci-Hub.

Subject Area	2016				
	No. of papers in WoS	No. of papers with DOI	No. of papers that are on Sci-Hub	Percentage of papers available on Sci-Hub ¹	Proportionate Contribution to the total articles on Sci-Hub ²
MED	15307	11884	9422	79.28	13.89
CHE	14092	13496	13092	97.01	19.29
ENG	6708	6389	6120	95.79	9.02
ENV	3563	3164	2869	90.68	4.23
MAR	5818	5562	5356	96.30	7.89
PHY	7838	7598	7286	95.89	10.74
GEO	3912	3038	2628	86.50	3.87
BIO	7094	6394	5871	91.82	8.65
INF	2857	2728	2593	95.05	3.82
MUL	2754	2351	1853	78.82	2.73
AGR	4582	3516	2803	79.72	4.13
MAT	1648	1487	1240	83.39	1.83
SS	1759	1478	1424	96.35	2.10
AH	184	98	89	90.82	0.13

¹ Percentage is calculated with respect to number of articles in WoS having a DOI, i.e. column 3.

² Percentage is calculated with respect to number of articles in WoS having a DOI, i.e. 67857

from 97% for Chemistry to 78% for multidisciplinary field. Further, 9 out of 14 subject areas have more than 90% of their research papers available in Sci-Hub. When these results are seen with findings of a previous study,^[17] it is seen that the subject areas which have less percentage of articles available as legal gold-green open access have better coverage in Sci-Hub (for example Information Science or Engineering). Thus, Sci-Hub, in a sense, is also seen complementing the open access availability in legal gold-green models.

Downloads of Indian research papers available in Sci-Hub

The second major analysis we performed was to find out how many times the Indian research papers available in Sci-Hub are downloaded. For this purpose, the Sci-Hub access log for 2017 was processed. The log entries were analysed to find out how many of the download entries were for the Indian research papers (for the year 2016) available in Sci-Hub. It was found that out of 61,243 unique research papers (for year 2016) available in Sci-Hub, 36,063 research papers were actually downloaded i.e., only 58.8% of the research papers, for the year 2016, available in Sci-Hub were downloaded in 2017. The total number of downloads for these papers in this period is 476,694. The average download per paper for the Indian papers available in Sci-Hub is thus

7.78. Figure 1 shows the download activity of these papers. The plot is arranged in descending order of downloads of papers. It is observed that the download pattern has a long tail, indicating that only few papers are downloaded frequently, and most others are downloaded lesser.

Figure 2 shows the daily download statistics for the year 2017, indicating both the total downloads and the unique downloads per day. The analysis of access log shows that on an average 2,500 Indian research papers are downloaded every day. Next, we wanted to find out who is downloading the Indian research papers available in Sci-Hub, i.e., whether most of the downloads are coming from India or downloads are distributed to other regions of the world as well. For this purpose, the latitude and longitude information in the access log was processed and they were geo-tagged on the world map. Figure 3 shows the distribution of download requests for Indian research papers in Sci-Hub. It can be observed that majority of the download activity is located in the European region, followed by United States and Japan. Interestingly, institutions in these regions are known to have the richest subscription to journals and digital libraries. Thus, the access to Indian research papers in Sci-Hub is not localized to developing countries but is apparently coming from across the world.

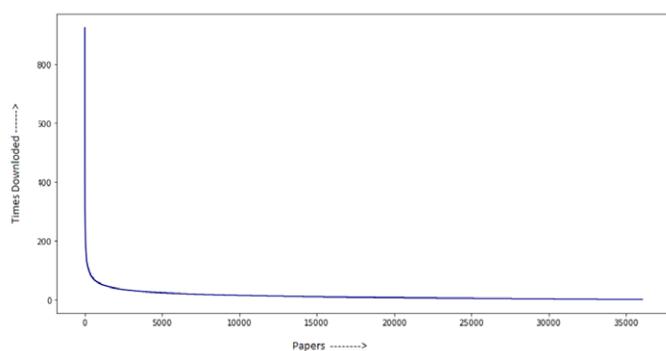


Figure 1: Download frequency of Indian papers in Sci-Hub as per 2017 access log.

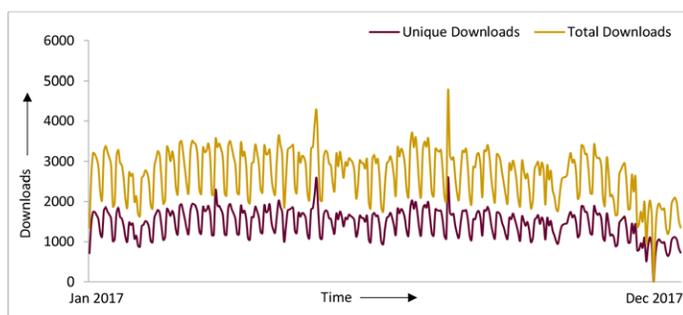


Figure 2: Number of Indian Papers downloaded from Sci-hub (1.1.2017 to 31.12.2017).



Figure 3: Geographical distribution of download activity for Indian Papers.

Table 5: Average Citation Per Paper (ACPP) for papers present and not present in Sci-Hub.

No. of articles in WoS	Articles with DOI	ACPP articles that are on Sci-Hub (61,243)	ACPP articles that are not on Sci-Hub (6,151)
76,530	67,857	11.71	3.93

Impact of presence in Sci-Hub on citations

We also tried to find out what impact the presence of paper in Sci-Hub may have on its citations. For this purpose, we computed the average citations per paper for the papers present in Sci-Hub as well as the papers not present in Sci-Hub. Table 5 shows the average citations per paper for papers that are present on Sci-Hub and those that are not. The average citation per paper for papers present on Sci-Hub is 11.71 as compared to 3.93 for the papers that are not on Sci-Hub. Thus, papers that are present on Sci-Hub apparently get higher visibility/ downloads and higher citations, as also observed by Correa *et al.*^[18] The Open access leading to higher impact in terms of increasing citations has been shown in some recent studies. Recent study of the journal Nature Communications undertaken by Research Information Network (RIN) has found that frequency of viewing of open access articles are on an average three times more than subscription articles. They also found higher citation of open access articles as compared to subscription articles. In an earlier study by SPARC Europe (2016), in their OpCit project,^[19] the question 'whether or not there is citation advantage for open access articles' was examined. They found 65% of such articles, identifying citation advantage. However, a much-dated study which still has high contemporary relevance is by Craig *et al.*^[20] They underscore that paper influence comes from various contributing factors and it can lead to fallacy of generalisation to draw conclusion from relationship that are shown between open access articles and high citations. Studies across diverse disciplines, removing selection bias, developing robust methodology can draw more informed understanding of the effect of Open Access articles on citations. Thus, drawing from this, we would also say that higher citation impact of

papers from Sci-Hub vis-a-vis papers not available therein may be due to other factors and broad generalisation from this would be fallacious. However, one can be tempted to postulate that easy availability of a paper allows an author to consult it and this increases the chances of the paper attracting higher citations.

CONCLUSION

The study analysed how many Indian research papers are available on Sci-Hub and who is downloading those papers. The results show that about 90.25% of Indian research papers (for the year 2016) are available in Sci-Hub for free download. Out of these, about 58.8% of papers have been downloaded at least once in the year 2017. These downloads are distributed to different parts of the world, with major download activity coming from the European region, US, Japan and South East Asia. Further, the paper available in Sci-Hub get an advantage in citations per paper which our study has shown. This however, needs more diverse study and across disciplines as we have discussed earlier to make an assertion.

Previous studies have shown that Sci-Hub provides access to all scientific literature for the world. The analysis by Greshake^[21] have shown that Sci-Hub is a huge repository of research papers and is being used heavily by people across the world, including significant number of accesses coming from the developed countries. The analysis of Sci-Hub access log shows that a good number of downloads come from developed countries, as also shown by Bohannon.^[3] Further, Travis^[22] points out that users who use Sci-Hub for access to research papers do really appreciate its existence, in form of ‘thumbs up’ to the site. This suggests that researchers are attracted by the simple, easy to use interface of Sci-Hub, and possibly see it as a more certain way of getting access to research papers. It appears that researchers downloading papers from Sci-Hub seem totally unconcerned with legal issues associated with the Sci-Hub’s model. On the contrary what matters to them is the free and easy access to research papers. This may be seen as the desire of researchers for universal access to knowledge without any barriers. As Green^[23] points out, existence of Sci-Hub is an indication that we must change our approach to access of scientific literature. Sci-Hub is thus not only a signal of a broken open access system but also an indication that the entire science dissemination model needs to be simplified. This has a much wider implication as this also is an important attribute of Responsible Research and Innovation.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. European Commission. Directorate-General for Research and Innovation. Responsible Research and Innovation: Europe’s Ability to Respond to Societal Challenges. Publications Office of the European Union. 2014. Retrieved from: <https://op.europa.eu/en/publication-detail/-/publication/2be36f74-b490-409e-bb60-12fd438100fe>
2. Das AK, Dutta B. Open Science in Addressing Research and innovation: evidence from India and other countries. *Journal of Scientometric Research*. 2020;9(2s):s14-23. DOI: 10.5530/jscires.9.2s.33
3. Bohannon J. Who’s downloading pirated papers? *Everyone*. Science. 2016.
4. Elbakyan A. Sci-Hub is a goal, changing the system is a method. *Engineering*. 2016.
5. Elbakyan A. Why Sci-Hub is the true solution for open access: Reply to criticism. *Engineering*. 2016.
6. Faust JS. Sci-Hub: A solution to the problem of Paywalls, or merely a diagnosis of a broken system?. *Annals of Emergency Medicine*. 2016;68(1):15-7.
7. Priego E. Signal, not solution: Notes on why Sci-Hub is not opening access. *The Winnower*. 2016;3:e145624.
8. Scaria AJ. Sci-Hub Case: The Court Should Protect Science From Greedy Academic Publishers, *The Wire*. 2020. Retrieved from <https://thewire.in/law/sci-hub-elsevier-delhi-high-court-access-medical-literature-scientific-publishing-access-inequity>
9. Schiermeier Q. US court grants Elsevier millions in damages from Sci-Hub. *Nature News*. 2017. Retrieved from: <https://www.nature.com/news/us-court-grants-elsevier-millions-in-damages-from-sci-hub-1.22196>
10. Schiermeier Q. Pirate paper website Sci-Hub dealt another blow by US courts. *Nature News*. 2017. Retrieved from: <https://www.nature.com/news/pirate-paper-website-sci-hub-dealt-another-blow-by-us-courts-1.22971>
11. Trivedi D. Cases against Sci-Hub and Libgen imply long-term consequences to research and education in India. *Frontline*. 2021. Retrieved from: <https://frontline.thehindu.com/the-nation/locking-up-research-cases-against-sci-hub-and-libgen-imply-long-term-consequences-to-research-and-education-in-india/article33641506.ece>
12. Pai N. Why blocking Sci-Hub will actually hurt national interest. *The Print*. 2020. Retrieved from: <https://theprint.in/opinion/why-blocking-sci-hub-will-hurt-national-interest/575577/>
13. Singh VK, Srichandan SS, Bhattacharya S. What do Indian researchers download from Sci-Hub?. *arXiv preprint arXiv:2103.16783*. 2021. Pre-print available at: <https://arxiv.org/abs/2103.16783>
14. Rupika UA, Singh VK. Measuring the university–industry–government collaboration in Indian research output. *Current Science*. 2016;110(10):1904-9.
15. Tzovaras BG. Sci-Hub download log of 2017. *Zenodo*. 2018. DOI: 10.5281/zenodo.1158301.
16. Piryani R, Dua J, Singh VK. Open Access Levels and Patterns in Scholarly Articles from India. *Current Science*. 2019;117(9):1435-40.
17. Singh VK, Piryani R, Srichandan SS. The Case of Significant Variations in Gold-Green and Black Open Access: Evidence from Indian Research Output. *Scientometrics*. 2020;124(1):515-31.
18. Correa JC, Laverde-Rojas H, Tejada J, Marmolejo-Ramos F. The Sci-Hub effect on papers’ citations. *Scientometrics*. 2021;1-28.
19. The Open Access Citation Advantage Service (OACA) of The Open Citation Project - Reference Linking and Citation Analysis for Open Archives. 2021. retrieved from: <http://opcit.eprints.org/oacait-biblio.html>
20. Craig ID, Plume AM, McVeigh ME, Pringle J, Amin M. Do open access articles have greater citation impact?: A critical review of the literature. *Journal of Informetrics*. 2007;1(3):239-48.
21. Greshake B. Looking into Pandora’s Box: The Content of Sci-Hub and its Usage. *F1000Research*. 2017;6: 541. DOI: 10.12688/f1000research.11366.1
22. Travis J. In survey, most give thumbs-up to pirated papers. *Science*. 2016.
23. Green T. We’ve failed: Pirate black open access is trumping green and gold and we must change our approach. *Learned Publishing*. 2017;30:325-9.