Webology, Volume 13, Number 2, December, 2016

arXiv popularity from a citation analysis point of view

Alireza Noruzi

Ph.D., Editor-in-Chief of Webology, Department of Knowledge and Information Science, Faculty of Management, University of Tehran, Iran. E-mail: noruzi@ut.ac.ir

---

Abstract

This study aims to provide an overview of the citation rate of arXiv.org since its launch in August 1991, based on the Scopus citation database. The total number of citations to arXiv in Scopus in the 26 year period was 135,782 of which the highest number of citations was 23,288 in 2016. It is also shown that arXiv-deposited papers are highly cited by physics and astronomy, mathematics, computer science, and engineering. It can be seen that researchers from the United States, Germany, China, United Kingdom, France, and Italy cite arXiv-deposited papers more than others. The analysis of document types indicate that articles rank first with 69% of all Scopus documents citing arXiv from 1991-2016, followed by conference papers (24.7%), reviews (3.2%), and book chapters (1.5%). It can be concluded that arXiv is cited increasingly by different subject areas, by different languages (especially English, Chinese and French), and by various countries.

Keywords

arXiv, e-prints, open access, citation analysis, Scopus

Introduction

arXiv is an e-print archive in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance and statistics. Started in August 1991, arXiv.org (formerly xxx.lanl.gov) is owned and operated by Cornell University. It is funded by Cornell University Library, the Simons Foundation and by the member institutions. This digital archive contains more than 1,240,000 open-access e-prints in physics, mathematics, computer science, quantitative biology, quantitative finance and statistics (arXiv, 2016).

This study aims to provide an overview of the citation rate of arXiv.org since its launch in August 1991. It is worth noting that since its inception arXiv —the international open-access multi-disciplinary repository— has been subject to criticism. Nevertheless, arXiv-deposited papers are increasingly cited by web resources (web pages, e-books, conference proceedings, journal papers, theses, ...) indexed by Google Scholar.

One way to test the popularity of a web resource or a digital archive is to use citation analysis method to predict to what extend it is cited by scientific documents. This research tries to identify and quantify the number of citations to arXiv in the documents indexed by the Scopus citation database, studying the citation behavior of authors who have published papers in journals indexed by Scopus.

Materials and Methods

An advanced search was conducted on December 24, 2016, in the Scopus citation database for total citations to arXiv-deposited papers, as follows:

REF (arXiv) AND PUBYEAR > 1990 AND PUBYEAR < 2017

Results

Table 1 reveals how many arXiv documents were cited in each year. The pattern shows the increasing trends of arXiv citations as used per annum. Yearly distribution of citations gives an idea about scattering of citations. The total number of citations to arXiv in Scopus in the 26 year period was 135,782 of which the highest number of citations was 23,288 in 2016 and the lowest number of citations was one in 1992.

Table 1. The total number of citations per year

Year	Number of citations
2016	23288
2015	21205
2014	17779
2013	14783
2012	12757
2011	11360
2010	9119
2009	7025
2008	5022
2007	3969
2006	3204
2005	2393
2004	1532
2003	959
2002	571
2001	404
2000	149
1999	74

1998	44
1997	23
1996	20
1995	37
1994	36
1993	11
1992	1
1991	17

Table 2 and Figure 1 show that arXiv-deposited papers are highly cited by physics and astronomy, mathematics, computer science, and engineering. Larivière et al. (2014) argued that arXiv has become central to the diffusion of scientific research, changing the scholarly communication patterns of physicists and mathematicians.

Table 2. Subject areas frequently citing arXiv

Rank	Subject Area	Number of citations
1	Physics and Astronomy	50607
2	Mathematics	49915
3	Computer Science	40606
4	Engineering	23850
5	Materials Science	6412
6	Social Sciences	5902
7	Earth and Planetary Sciences	4645
8	Biochemistry, Genetics and Molecular Biology	4135
9	Medicine	3903
10	Chemistry	3741
11	Decision Sciences	3658
12	Multidisciplinary	3129
13	Agricultural and Biological Sciences	1905
14	Arts and Humanities	1693
15	Environmental Science	1496
16	Chemical Engineering	1473
17	Neuroscience	1354
18	Economics, Econometrics and Finance	1180
19	Energy	1025
20	Business, Management and Accounting	983
21	Psychology	567
22	Immunology and Microbiology	487
23	Pharmacology, Toxicology and Pharmaceutics	239
24	Health Professions	236
25	Nursing	66
26	Veterinary	38
27	Dentistry	11

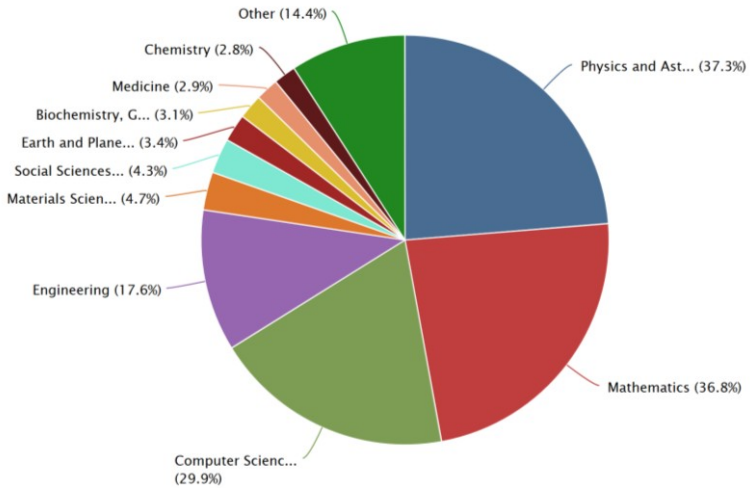

Figure 1. Subject areas frequently citing *arXiv* (Scopus, 2016)

Table 3 shows the distribution of citations by place of publication (affiliation). It can be seen that researchers from the United States, Germany, China, United Kingdom, France, and Italy cite papers available at arXiv more than others. In a citation analysis of arXiv-deposited papers, Amanl (2015) investigated the geographical distribution of authorship citing arXiv in their papers indexed in Web of Science (WOS) and showed that authors from the US, Germany, UK, France and Italy rely heavily on arXiv.

Table 3. Documents citing arXiv by country/territory

Rank	Country/Territory	No. of Citations
1	United States	43877
2	Germany	16619
3	China	15176
4	United Kingdom	13769
5	France	13547
6	Italy	10169
7	Canada	6854
8	Russian Federation	6554
9	Japan	6490
10	Spain	6305
11	Switzerland	5921
12	India	4953
13	Australia	4485
14	Netherlands	4009
15	Brazil	3399

Table 4 shows the language distribution of citations. The authors of citing documents cited arXiv in different languages. English language was found to be the most dominant language citing arXiv with 134,101 citations.

Table 4. Languages cited arXiv

Rank	Language	Number of citations
1	English	134,101
2	Chinese	879
3	French	361
4	Spanish	182
5	Russian	89
6	German	88
7	Portuguese	41
8	Japanese	30
9	Turkish	24
10	Italian	20
11	Ukrainian	20
12	Persian	17

Table 5 shows the number of documents by source, citing arXiv. It is worth noting that high-impact journals such as, *Journal of High Energy Physics*, *Nature*, and *Science* are citing arXiv.

Table 5. Source Documents

Rank	Source	No. of Documents
1	Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics	4266
2	Physics Letters Section B Nuclear Elementary Particle and High Energy Physics	3974
3	**Journal of High Energy Physics**	3638
4	Nuclear Physics B	1950
5	Communications in Mathematical Physics	1555
6	European Physical Journal C	1481
7	Physical Review D Particles Fields Gravitation and Cosmology	1468
8	IEEE International Symposium on Information Theory Proceedings	1417
9	Aip Conference Proceedings	1317
10	IEEE Transactions on Information Theory	1279
11	Advances in Mathematics	1005
12	Nuclear Physics A	1001
13	Nuclear Physics B Proceedings Supplements	950
14	Journal of Algebra	910
15	**Nature**	869
16	Proceedings of SPIE the International Society for Optical Engineering	860
17	Astronomy and Astrophysics	771
18	Journal of Mathematical Analysis and Applications	751
19	Journal of Physics Conference Series	706
20	Monthly Notices of the Royal Astronomical Society	681
21	International Journal of Modern Physics A	675
22	Nuclear Instruments and Methods in Physics Research Section A Accelerators Spectrometers Detectors and Associated Equipment	667
23	IEEE Transactions on Signal Processing	643
24	Nuclear and Particle Physics Proceedings	630
25	Nature Communications	600
26	Journal of Functional Analysis	597
27	**Science**	588
28	Nature Physics	572
29	Annals Of Physics	539
30	Proceedings of the IEEE Conference on Decision and Control	517

Table 6 shows the document type of citing resources. The analysis of document types indicate that *articles* rank first with 69% of all Scopus documents citing arXiv from 1991-2016, followed by *conference papers* (24.7%), *reviews* (3.2%), and *book chapters* (1.5%). *Books*, *short surveys*, *notes*, *letters*, *editorials*, and *erratum* also cite arXiv-deposited papers.

Table 6. Document type of citing resources

Rank	Document Type	Documents	%
1	Article	93708	69.0
2	Conference Paper	33472	24.7
3	Review	4279	3.2
4	Book Chapter	2000	1.5
5	Book	917	0.7
6	Short Survey	548	0.4
7	Note	341	0.3
8	Letter	242	0.2
9	Editorial	179	0.1
10	Erratum	96	0.1
	Total	**135782**	**100**

Discussion and Conclusion

This study shows that arXiv is cited 135,782 times in documents indexed in the Scopus citation database. The result showed that arXiv is cited increasingly by different subject areas (especially Physics and Astronomy, Mathematics, and Computer Science), by different languages (especially English, Chinese and French), and by various countries (especially by the United States, Germany, China, United Kingdom, France, and Italy).

The analysis of document types indicate that articles rank first with 69% of all Scopus documents citing arXiv from 1991-2016. Although, the credibility of some of arXiv content as an authoritative reference source was criticized, it is cited by international researchers. The significance of arXiv from a citation analysis point of view goes well beyond e-prints and open-access and enhanced opportunities for citation.

Previous studies show that e-prints, post-prints and open access advantage have an effect on how researchers communicate the research findings (Davis & Fromerth, 2007; Dietrich, 2008a, 2008b; Haque & Ginsparg, 2009; Kurtz et al., 2005; Moed, 2007; Larivière, et al., 2014).

References

- Aman, V. (2015). Citing e-prints on arXiv: A study of cited references in WoS-indexed journals from 1991- 2013.. In A. A. Salah, Y. Tonta, A. A. A. Salah, C. R. Sugimoto & U. Al (eds.), *Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference*, Istanbul, Turkey, 29 June to 3 July, 2015. Bogaziçi University Printhouse 2015.
- arXiv.org (2016). General information about arXiv. Retrieved December 20, 2016, from https://arxiv.org/help/general
- Bar-Ilan, J. (2014). Astrophysics publications on arXiv, Scopus and Mendeley: a case study. Scientometrics, 100(1), 217-225.
- Davis, P.M., & Fromerth, M.J. (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? *Scientometrics*, 71(2), 203-215.
- Dietrich, J.P. (2008a). The importance of being first: Position dependent citation rates on arxiv:astro-ph. *Publications of the Astronomical Society of the Pacific*, 120, 224-228.
- Dietrich, J.P. (2008b). Disentangling visibility and self-promotion bias in the arxiv:astro-ph positional citation effect. *Publications of the Astronomical Society of the Pacific*, 120, 801-804.
- Haque, A.-u., & Ginsparg, P. (2009). Positional effects on citation and readership in arXiv. *Journal of the American Society for Information Science & Technology*, 60(11), 2203-2218.
- Kurtz, M.J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E., & Murray, S.S. (2005). The effect of use and access on citations. *Information Processing & Management*, 41, 1395-1402.
- Moed, H.F. (2007). The effect of 'Open Access' on citation impact: An analysis of arXiv's condensed matter section. *Journal of the American Society for Information Science & Technology*, 58(13), 2047-2054.
- Larivière, V., Sugimoto, C.R., Macaluso, B., Milojević, S., Cronin, B., & Thelwall, M. (2014). arXiv e-prints and the journal of record: An analysis of roles and relationships. *Journal of the Association for Information Science and Technology*, 65(6), 1157-1169.

Bibliographic information of this paper for citing:

Noruzi, Alireza (2016). "arXiv popularity from a citation analysis point of view." *Webology*, **13**(2), editorial 22. Available at: http://www.webology.org/2016/v13n2/editorial22.pdf