

Free online availability substantially increases a paper's impact

Steve Lawrence

NEC Research Institute

4 Independence Way, Princeton, NJ 08540

lawrence@research.nj.nec.com

The volume of scientific literature far exceeds the ability of scientists to identify and use all relevant information. The ability to locate relevant research quickly will dramatically improve communication and scientific progress. Although availability varies greatly by discipline, more than a million research articles are now freely available on the web.

Here we investigate the impact of free online availability by analysing citation rates. Online availability of an article may not greatly improve access and impact without efficient and comprehensive search services; a substantial percentage of the literature needs to be indexed by these search services before scientists consider them useful. In computer science, a substantial percentage of the literature is online and available through search engines such as [Google](#), or specialized services such as [ResearchIndex](#) although the greatest impact of online availability is yet to come, because comprehensive search services and more powerful search methods have become available only recently.

We analysed 119,924 conference articles in computer science and related disciplines, obtained from DBLP (dblp.uni-trier.de). In these fields, conference articles are typically formal publications and are often more prestigious than journal articles, with acceptance rates at some conferences below 10%. We estimated citation counts and online availability using ResearchIndex, excluding self-citations.

Figure 1 shows the probability that an article is freely available online as a function of the number of citations to the article, and the year of publication of the article. The results are dramatic, showing a clear correlation between the number of times an article is cited and the probability that the article is online. More highly cited articles, and more recent articles, are significantly more likely to be online, in computer science. The mean number of citations to offline articles is 2.74, and the mean number of citations to online articles is 7.03, an increase of 157%.

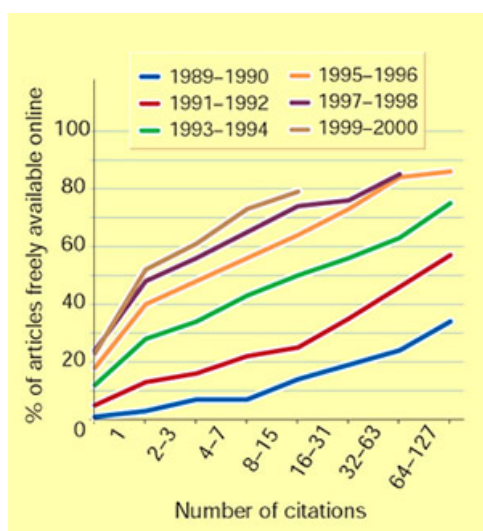


Figure 1

Analysis of 119,924 conference articles in computer science and related disciplines.

The actual percentage of articles available online is greater, owing to limitations in the extraction of article information from online documents and limitations in locating articles on the web. Only points with greater than 100 articles are computed.

We analysed differences within publication venues (the proceedings of a conference for a particular year, for example), looking at the percentage increase in citation rates for online articles. When offline articles were more highly cited, we used the negative of the percentage increase for offline articles: hence if the average number of citations for offline articles is two, and the average for online articles is four, the percentage increase would be 100%. For the opposite situation, the percentage increase would be \diamond 100%.

Averaging the percentage increase across 1,494 venues containing at least five offline and five online articles results in an average of 336% (median 158%) more citations to online computer-science articles compared with offline articles published in the same venue (see Figure 2).

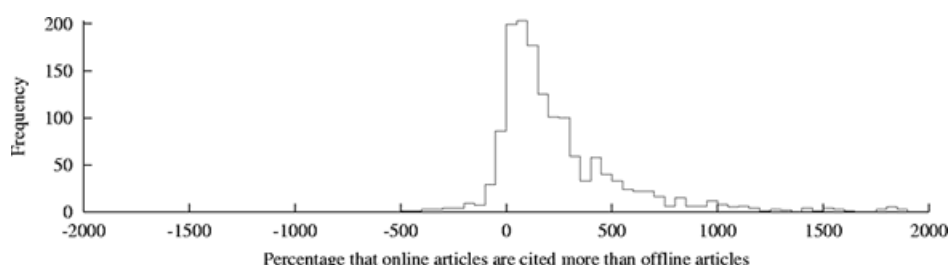


Figure 2

Analysis of citation rates within publication venues.

The graph shows the distribution of the percentage increase for the average number of citations to online articles compared to offline articles. The analysis covers 1,494 publication venues containing at least 5 online and 5 offline articles. For 90% of venues, online articles are more highly cited on average. On average there are 336% more citations to online articles compared to offline articles published in the same venue [the first, second (median), and third quartiles of the distribution are 58%, 158%, and 361%].

If we assume that articles published in the same venue are of similar quality, then the analysis by venue suggests that online articles are more highly cited because of their easier availability. This assumption is likely to be more valid for top-tier conferences with very high acceptance standards. Restricting our analysis to the top 20 publication venues by average citation rate gives an increase of 286% (median 284%) in the citation rate for online articles.

Free online availability facilitates access in many ways, including provision of online archives; direct connections among scientists or research groups; hassle-free links from e-mail, discussion groups and other services; indexing by web search engines; and the creation of third-party search services. Free online availability of scientific literature offers substantial benefits to science and society. To maximize impact, minimize redundancy and speed scientific progress, authors and publishers should aim to make research easy to access.