Measuring Citation Advantages of Open Accessibility

Samson C. Soong
The Hong Kong University of Science and Technology
<soong@ust.hk>

Introduction

Over the past decade, a good number of studies have attempted to demonstrate that open access journal articles have higher citation rates than traditionally published ones. These open access articles are either published in open access journals or are made publicly available through open access repositories. More methodologically sound studies, however, are needed to pinpoint more exactly the impact, if any, of open accessibility on citation rates and to look at the issue of impact (or non-impact) more closely. For instance, of those journal articles that have been made openly accessible, about what percent of them benefit from open online availability? What percent of them do not?

This article describes a study, involving a set of articles published in scholarly journals by faculty members of the Hong Kong University of Science and Technology (HKUST) that have also been deposited in the HKUST Institutional Repository. The study was conducted to measure the actual effect of their open accessibility on citation rates. More importantly, the study suggests another quantitative method, and it is hoped that the study will inspire more effective methods of measuring the net impact of open access.

Level of research at HKUST

Established in 1991, HKUST was ranked 35th among the top universities in the world by the Times Higher Education Supplement in 2009. The University, by a whole range of measures, is considered one of the leaders both in education and research. Although a relatively young university, it has already gained an international reputation for innovation and cutting-edge research in a number of fields, including nanoscience and nanotechnology, biological sciences,
wireless and information technology, environment and sustainable development, and management science.

**HKUST Institutional Repository**

In 2003 the HKUST Library established a digital institutional repository (IR) to host, preserve and facilitate access to the scholarly output of the University community. This repository uses DSpace as its platform, which complies with the Open Archives Initiative (OAI), thus allowing archived documents to be easily discovered by web search engines and services, such as Google Scholar, OAIster, and Scirus. There are currently over 5,600 documents in the Repository.

Documents deposited in the IR include conference papers, research and technical reports, book chapters, theses and dissertations, patents, and research datasets, as well as published journal articles. By submitting articles to the Repository after they are published by a journal, researchers at HKUST make their work more readily available to the global scholarly community. Depositing published articles in an open access repository usefully complements their publication by scholarly journals. Thus, the IR serves as a secure central repository of the scholarly output of the institution and offers greater visibility for the research carried out by the University community.

**Methodology and data sources**

Since methodological soundness is critical to assessing the results of any study, a full account of the methodology and data sources used to obtain the citation rates is given here.

A total of 50 archived journal articles that already have 10 or more citation counts in Scopus were randomly selected for inclusion in this study. This could be done quickly, as the system grabs in real time the current total citation count from Scopus and displays that piece of information as part of each record in the Repository, e.g., [http://repository.ust.hk/dspace/handle/1783.1/191](http://repository.ust.hk/dspace/handle/1783.1/191).

Each "full item record" in the Repository includes fields on "date.accessioned" and "date.available" which help to ascertain when a full-text journal article was actually added to the Repository and made openly accessible. Elsevier's Scopus database was then searched to get the citation counts of these 50 articles in each of the years since they were published, including the years after these articles were deposited in the HKUST IR.

For example, the Scopus record on the journal article "Computer supported augmentation and collaborative decision making: the HERMES system" by Karacapilidis and Papadias provides the following citation accounts: 2008(10

The absolute citation counts were used to derive two average-per-year citation counts, before and after the articles are made openly accessible through the IR. The Karacapilidis and Papadias article was published in 2001 and deposited in the IR in 2003, according to the "date.accessioned" field and the "date.available" field of the full item record for this archived journal article. Thus, the average-per-year citation counts before and after this article becomes openly accessible are 1.5 and 7 respectively. The number of citations in 2009 is not included in the calculation, as the year is not over yet. The number of citations during the year of publication (most of the time during part of the year) is also excluded in the computation.

The following table shows the results of this study.

<table>
<thead>
<tr>
<th>IR record #</th>
<th>Pub Year</th>
<th>IRD epYear</th>
<th>Number of Citations</th>
<th>Before</th>
<th>After</th>
<th>% of Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Average</td>
<td>Average</td>
<td></td>
</tr>
<tr>
<td>Math - 1775</td>
<td>200</td>
<td>200</td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>2.2%</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Math - 1787</td>
<td>199</td>
<td>199</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>bich - 1605</td>
<td>199</td>
<td>200</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>Econ - 2149</td>
<td>200</td>
<td>200</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>bich - 2466</td>
<td>200</td>
<td>200</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>bich - 2467</td>
<td>200</td>
<td>200</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>200</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0%</td>
</tr>
</tbody>
</table>

The table shows the number of citations before and after the articles become openly accessible.
<table>
<thead>
<tr>
<th></th>
<th>200</th>
<th>200</th>
<th>200</th>
<th>200</th>
<th>200</th>
<th>200</th>
<th>200</th>
<th>200</th>
<th>200</th>
<th>eas e</th>
<th>increas e</th>
</tr>
</thead>
<tbody>
<tr>
<td>bich - 1356</td>
<td>2148</td>
<td>199</td>
<td>199</td>
<td>199</td>
<td>199</td>
<td>199</td>
<td>199</td>
<td>199</td>
<td>199</td>
<td>No incr ease</td>
<td>1220</td>
</tr>
<tr>
<td>bich - 1608</td>
<td>1218</td>
<td>1218</td>
<td>1218</td>
<td>1218</td>
<td>1218</td>
<td>1218</td>
<td>1218</td>
<td>1218</td>
<td>1218</td>
<td>No incr ease</td>
<td>1219</td>
</tr>
<tr>
<td>ECE - 1251</td>
<td>1251</td>
<td>1251</td>
<td>1251</td>
<td>1251</td>
<td>1251</td>
<td>1251</td>
<td>1251</td>
<td>1251</td>
<td>1251</td>
<td>No incr ease</td>
<td>1262</td>
</tr>
<tr>
<td>biol - 639</td>
<td>639</td>
<td>639</td>
<td>639</td>
<td>639</td>
<td>639</td>
<td>639</td>
<td>639</td>
<td>639</td>
<td>639</td>
<td>No incr ease</td>
<td>640</td>
</tr>
<tr>
<td>biol - 1605</td>
<td>1605</td>
<td>1605</td>
<td>1605</td>
<td>1605</td>
<td>1605</td>
<td>1605</td>
<td>1605</td>
<td>1605</td>
<td>1605</td>
<td>No incr ease</td>
<td>1431</td>
</tr>
<tr>
<td>biol - 2262</td>
<td>2262</td>
<td>2262</td>
<td>2262</td>
<td>2262</td>
<td>2262</td>
<td>2262</td>
<td>2262</td>
<td>2262</td>
<td>2262</td>
<td>No incr ease</td>
<td>2263</td>
</tr>
<tr>
<td>biol - 1262</td>
<td>1262</td>
<td>1262</td>
<td>1262</td>
<td>1262</td>
<td>1262</td>
<td>1262</td>
<td>1262</td>
<td>1262</td>
<td>1262</td>
<td>No incr ease</td>
<td>1263</td>
</tr>
<tr>
<td>biol - 1431</td>
<td>1431</td>
<td>1431</td>
<td>1431</td>
<td>1431</td>
<td>1431</td>
<td>1431</td>
<td>1431</td>
<td>1431</td>
<td>1431</td>
<td>No incr ease</td>
<td>1431</td>
</tr>
<tr>
<td>biol - 640</td>
<td>640</td>
<td>640</td>
<td>640</td>
<td>640</td>
<td>640</td>
<td>640</td>
<td>640</td>
<td>640</td>
<td>640</td>
<td>No incr ease</td>
<td>640</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>-----------</td>
<td>---------------</td>
<td>-----------</td>
<td>--------------</td>
<td>-----------------</td>
<td>--------------------------</td>
<td>-----------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>biol - 2255</td>
<td>2003</td>
<td>2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.00</td>
<td></td>
<td>No increase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>biol - 2256</td>
<td>1999</td>
<td>2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.83</td>
<td></td>
<td>38%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>biol - 1606</td>
<td>2000</td>
<td>2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4.25</td>
<td></td>
<td>6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>biol - 2260</td>
<td>1997</td>
<td>2005</td>
<td></td>
<td>0 4 6 4 1 5 8 1 8 4 6</td>
<td></td>
<td></td>
<td>4.13</td>
<td></td>
<td>45%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECE - 2160</td>
<td>2000</td>
<td>2005</td>
<td></td>
<td>0 3 4 3 5 1 4 0 2 2</td>
<td></td>
<td></td>
<td>3.20</td>
<td></td>
<td>No increase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>biol - 1432</td>
<td>2001</td>
<td>2004</td>
<td>1 7 4 1 1 2 0</td>
<td></td>
<td>5.50</td>
<td></td>
<td>No increase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>biol - 2259</td>
<td>2000</td>
<td>2005</td>
<td></td>
<td>0 2 5 2 6</td>
<td></td>
<td></td>
<td>2.00</td>
<td></td>
<td>117%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>biol - 1433</td>
<td>1999</td>
<td>2004</td>
<td></td>
<td>5 9 5 1 2 0 8 5 3 7 3</td>
<td></td>
<td></td>
<td>8.80</td>
<td></td>
<td>No increase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>biol - 1224</td>
<td>1999</td>
<td>2004</td>
<td></td>
<td>0 0 0 0 0 1 4 1 1 2 3 4 1</td>
<td></td>
<td></td>
<td>0.88</td>
<td></td>
<td>186%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finan - 362</td>
<td>1998</td>
<td>2003</td>
<td></td>
<td>0 2 0 3 5 2 1 7 4 5 1 0</td>
<td></td>
<td></td>
<td>2.40</td>
<td></td>
<td>125%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>biol - 1430</td>
<td>2001</td>
<td>2004</td>
<td></td>
<td>0 1 5 3 7 1 8 5 3.00</td>
<td></td>
<td></td>
<td>8.25</td>
<td></td>
<td>175%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>biol - 2261</td>
<td>1999</td>
<td>2005</td>
<td></td>
<td>0 3 3 3 6 3 3 3 3 1 3</td>
<td></td>
<td></td>
<td>3.50</td>
<td></td>
<td>No increase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>biol - 2256</td>
<td>1999</td>
<td>2005</td>
<td></td>
<td>3 5 5 7 4 3 4 7 7 3 6</td>
<td></td>
<td></td>
<td>4.67</td>
<td></td>
<td>43%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>biol - 1431</td>
<td>1999</td>
<td>2004</td>
<td></td>
<td>0 8 5 1 6 6 2 7 4 3 7.00</td>
<td></td>
<td></td>
<td>4.00</td>
<td></td>
<td>No increase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBME - 645</td>
<td>1998</td>
<td>2003</td>
<td></td>
<td>0 0 3 5 5 3 1 7</td>
<td></td>
<td></td>
<td>4.00</td>
<td></td>
<td>600%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Key Findings

Of the 50 articles included in the study, 29 (or 58%) have had a higher average-citation-rate after they have been deposited in the IR and made openly accessible than they had
prior to being available in the IR. The rest of the open access articles, or 42%, have not yet experienced a similar increase in the average-citation-rate. Although we cannot simply contrast these two percentages to derive an overall rate or a potential rate of increase, these results suggest that there is a citation advantage for a larger percent of articles that are made publicly accessible.

While there is an increase in the average citation rate for 58% of the articles, there is no obvious year-by-year incremental increase. In other words, we have not seen that the number of years an article becomes openly available has any particular effect. Much bigger increases for some articles suggest that the subject areas of these articles matter to a large extent. For instance, a number of articles in Chemical and Biomolecular Engineering in this study stand out in particular.

Conclusions

The citation advantage of open access needs to be studied more in depth to determine their impact in different subject areas or disciplines. An easy-to-follow framework for citation impact analysis of open accessibility is presented here. This framework allows for direct measurement and comparison of citation rates before and after journal articles are made openly available. This approach also allows for easy contrast between open access articles that have produced citation impact or advantage and those that have not. Any impact study would be undermined, however, if such a study does not take difference in disciplines into consideration. Although the value of time intervals to the study framework was not particularly highlighted, it is totally feasible to conduct the study longitudinally to see if the impact of open access changes over time.

The rapid growth of institutional repositories has helped the open access movement take hold with many more scholars and researchers in the world. Additional methodologically sound studies should be encouraged to explore more effective ways to measure and ascertain citation advantage of open accessibility of journal articles. The entire scholarly community will significantly benefit when open access to research output is compellingly advocated and maximized.

Notes and references


2 "Scirus is the most comprehensive scientific research tool on the web. With over 450 million scientific items indexed at last count, it allows researchers to search for not only journal content but also scientists' homepages, courseware, pre-print server material, patents and institutional repository and website information." Scirus website at
3 "Scopus is the largest abstract and citation database of peer-reviewed literature and quality web sources with smart tools to track, analyze and visualize research. Scopus is designed and developed with over 500 users and librarians internationally. Updated daily, it includes the abstracts and references of 15,000 peer-reviewed journals from more than 4,000 international publishers, ensuring broad interdisciplinary coverage." Scopus website at <http://www.elsevier.com/wps/find/electronicproductdescription.cws_home/704746/description#description>.

4 Karacapilidis N., Papadias, D., "Computer supported argumentation and collaborative decision making: The HERMES system", 2001, Information Systems, 26 (4), pp.259-277. Professor Papadias is a faculty member of the Department of Computer Science and Engineering at HKUST.