Libri, 2006, vol. 56, pp. 145–156
Printed in Germany All rights reserved

Libri
ISSN 0024-2667

The Effect of Open Access on Citation Impact: A Comparison Study Based on Web Citation Analysis

YANJUN ZHANG
Faculty of Information and Media Studies, University of Western Ontario, London, Ontario, Canada

The academic impact advantage of Open Access (OA) is a prominent topic of debate in the library and publishing communities. Web citations have been proposed as comparable to, even replacements for, bibliographic citations in assessing the academic impact of journals. In our study, we compare Web citations to articles in an OA journal, the Journal of Computer-Mediated Communication (JCMC), and a traditional access journal, New Media & Society (NMS), in the communication discipline. Web citation counts for JCMC are significantly higher than those for NMS. Furthermore, JCMC receives significantly higher Web citations from the formal scholarly publications posted on the Web than NMS does. The types of Web citations for journal articles were also examined. In the Web context, the impact of a journal can be assessed using more than one type of source: citations from scholarly articles, teaching materials and non-authoritative documents. The OA journal has higher percentages of citations from the third type, which suggests that, in addition to the research community, the impact advantage of open access is also detectable among ordinary users participating in Web-based academic communication. Moreover, our study also proves that the OA journal has impact advantage in developing countries. Compared with NMS, JCMC has more Web citations from developing countries.

Introduction

On December 1–2, 2001, the Open Society Institute lunched the Budapest Open Access Initiative (BOAI), and gave an official definition of Open Access (OA):

> By 'open access' to literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself (BOAI 2006).

Since its nascence, open access has achieved great developments in scientific publication and is still quickly growing. In March 2005, there were 1,500 full text and quality controlled scholarly journals indexed in the Directory of Open Access Journals, covering various subject areas. In March 2006, this number became 2150 (DOAJ 2006). OA journals have also gradually gained acceptance by the research community. Of the 8,700 selected journals currently covered in Web of Science, 239 are OA journals (McVeigh 2004).

OA journals have established a new paradigm of scholarly communication. Consequently, their scholarly impact has been a topic of hot debate in the library and publishing communities. OA articles are expected to receive more citations than traditional access journals because the open access mode allows OA journals to be read by all or most of their intended audiences. In contrast, restrictive access policies cut readership of journal articles. Statistics indicate that the average annual price increased by 9.4% for all U.S. periodicals during the last 18-year period (1988–2005) while the inflation rate averaged 3.1% (Kean 2005). The budgets of libraries for literature subscriptions were unable to

Yanjun Zhang is Ph.D. Student in Library and Information Science, Faculty of Information and Media Studies, University of Western Ontario, London, Ontario, N6A 5B7, Canada, E-mail: yzhan266@uwo.ca

This paper has received *honourable mention* in the *2006 Libri Best Student Award Competition.*

keep up with such steep increases. A recent report from Australia described a 43.7% decline in total journal subscriptions for 38 university libraries between 1993 and 1998 (Create Change 2006).

OA journals' advantage in expanding potential readership in developing countries could be even more apparent. Blocked by expensive subscription fees, information resources available to researchers in developing countries have decreased significantly, as confirmed by Hu (2005): "the only choice of a new model in the academic exchange for our country [China] is the Open Access". In fact, OA journals did achieve faster growth in developing regions. Among all OA journals indexed in the ISI database, North America accounts for 24.3%, Western Europe accounts for 18.8%, South/Central America 13.8%, Middle East/Africa 2%, Eastern Europe 8%, and Asia/Pacific 33%; whereas among all non-OA journals, North America and Western Europe together account for 90.5% (McVeigh 2004).

In recent years, many studies (e.g. Lawrence 2001; Brody *et al.* 2004; Kurtz *et al.* 2004; McVeigh 2004) attempted to examine whether Open Access itself necessarily leads to more citations in OA journals or articles, finding both conclusive and inconclusive evidence. A common feature of these studies is that citation analysis was conducted based on the citations indexed in ISI Web of Science (referred to as ISI citation later). The limitation of these studies is, as indicated by Cronin *et al.* (1998, 1326), "While traditional citation analysis can tell us a lot about the formal bases of intellectual influence, it quite naturally, tells us nothing about the many other modalities of influence which comprise the total impact of an individual's ideas, thinking, and general professional presence." To date, there have been few studies on OA journal impact based on Web measures. However, we believe that since OA journals have moved knowledge dissemination to the Web, the OA advantage in academic communication and impact may possibly be more comprehensive and better measured in the Web context. In contrast, ISI citations only reflect the communication and use of cited articles for scholarly publications and only cover the authors indexed in ISI database as scholarly communication participants.

This study compares Web citations for an OA journal with those of a traditional access journal in the same discipline. We produce Web citations profiles for both. Our purpose is to explore whether the OA model may be a key explanatory variable in yielding more Web citations to OA journals, whether the intellectual impact of OA journals is higher than traditional access journals and whether journals can benefit from OA to expand their impact to a larger international scope.

Related studies

Previous studies on the impact of OA journals or articles were based on traditional bibliographic citations. Lawrence (2001) found that for computer science journals and proceedings, the citation impact of OA articles is 336% higher than the impact of non-OA articles. A similar study in physics (Brody *et al.* 2004) showed that the OA/non-OA citation ratio was even larger. Kurtz *et al.* (2004) reported similar estimates in astrophysics, and Odlyzko (2002) in mathematics. On the other hand, Thomson Scientific examined the coverage and citation performance of OA journals in the ISI citation databases based on 2002 and 2003 Journal Citation Reports (JCR), and found that OA journals and non-OA journals had similar citation impacts (McVeigh 2004). They concluded that "the evidence to date is inconclusive and the difference is not dramatic" (Pringle 2006).

The analogous relationship between bibliographic citations and hyperlinks has led to a great deal of study. Started by Larson (1996), Rousseau (1997) and Almind and Ingwersen (1997), citation analysis techniques were also applied to explore scholarly communication in the Web context. Ingwersen (1998) proposed "Web Impact Factor" as a Web counterpart of the ISI's Journal Impact Factor (JIF). Harter and Ford (2002) compared e-journal impact using the backlinks to the e-journals and their articles. Further, correlation between journal Web site inlink page counts and journal impact factors was found in the library and information science (LIS) field (Vaughan & Hysen 2002) and in law field (Vaughan & Thelwall 2002). Vaughan & Thelwall (2003) found that content level is one of the key factors affecting the Web links to a journal site – i.e. Web sites that have more content posted on the journal site attracted more inlinking pages relative to their JIF scores. This discovery implies the possibility that OA articles may have more inlinking (hyperlink) citations than traditional access articles.

In more recent studies of journal impact, the definition of Web citation was further articulated to include new elements: Web text citation and URL citation. Vaughan and Shaw (2003) used Web text citations to refer to text mentions of the title of published papers on Web pages. They studied 46 journals in LIS and found that, for most journals, Web text citations correlated significantly with both ISI bibliographic citations and the ISI's JIF. They also classified Web citation sources and found that "many of the Web citations represented intellectual impact, coming from other papers posted on the Web (30%) or from class readings lists (12%)"(Vaughan & Shaw 2003, 1313). Their finding suggests that Web citation counts may potentially supplement or even replace ISI citation counts as journal impact measures. In Vaughan and Shaw's later study (2005), they further recommended that Web citation counts might be conducted for "a more fine-grained assessment of an article's impact". Also, another cited advantage of the Web citations is that they reflect greater geographic diversity and therefore could provide a more comprehensive, global assessment of impact. Thus, we expect that Web citation analysis could be an effective approach for verifying the OA journals' advantage in gaining impact internationally.

Another type of Web citation, namely URL citation, was defined by Kousha and Thelwall (2005) as the presence of mentions of an URL in the text of a Web page. Fifteen LIS OA journals were involved in this study and motivations for creating URL citations were identified. In addition, a slight correlation was found between the average number of URL citations (Web impact factor) and the average number of ISI citations (ISI Impact factor).

Research questions

The association found between traditional impact measures and Web citation is a reassuring indicator that the study of Web citations for OA journals is likely to reveal information about the connection between academic communication and the OA format. The present study is an attempt to identify the open access model's effect on journal impact, following the supposition that the Web is an increasingly important platform for scholarly communication. Accordingly, the research tasks and questions of this study are:

- To examine whether the open access model itself equates to more Web citations in OA journals, compared with traditional access journals of similar academic quality and within same discipline. Do OA journal articles receive more Web citations?
- To classify Web citations by citation source type, and compare OA journals and traditional access journals so as to explore whether they have different proportions of Web citations that can be regarded as an indicator of the impact of cited articles.
- To test OA journals' advantage in gaining impact internationally, e.g. do OA journals have or tend to have higher impact in developing countries?

Research design

Journal selection

For this comparative study, we selected one OA journal and one traditional access journal from the same discipline – Communication, and from the same subject area – computer technology-based communication. Namely, these are the Journal of Computer-Mediated Communication (JCMC) and New Media & Society (NMS). We deemed these journals comparable, as both have the following common characteristics:

1. both journals are peer reviewed Communication journals with international reputation and similar length of publication history;
2. they have common and similar scope and topic coverage;
3. they have similar JIFs (the JIF has been shown to affect Web link counts in (Vaughan & Thelwall 2003); hence, we need to control JIFs' effect when examining the effect of access mode on Web citation counts);
4. both are quarterly journals and have a similar number of articles.

Journal of Computer-Mediated Communication (JCMC)

JCMC is an OA journal published in the United States, with the first issue published in 1995. Over time, there have been two JCMC Web sites: http://jcmc.indiana.edu and http://www.ascusc.org/jcmc. Although the latter one was discontinued and redirected to the former URL, a lot of citing documents still hyperlink to the URL on the old site (i.e. http://www.ascusc.org/jcmc/volX/issueY/filename.html). Thus in our study, we took Web documents citing either site as valid citations.

Table 1. Journals investigated in the study.

Journal	Articles published in 2001	Articles published in 2002	JIF 2003
JCMC	23	26	0.5
NMS	21	24	0.689

New Media & Society (NMS)

NMS is a traditional access journal published in the United Kingdom since 1999. It has two hosting sites: http://www.new-media-and-society.com and http://nms.sagepub.com. However, the electronic full texts of articles are not open to Web users except those having valid accounts and passwords for accessing Sage Publications (http://nms.sagepub.com). It is worth noting that NMS publishes in every issue not only articles but also editorial and book reviews, which are much different from normal journal articles in attracting citations. Indeed, editorials and book reviews were not included in the ISI JIF calculation. Thus in our study, we only considered citations to NMS articles.

As shown in Table 1, NMS's JIF in 2003 is 0.689 and JCMC's JIF in 2003 is 0.5. The former is given by ISI Journal Citation Reports, the latter is manually calculated according to the JIF definition (Thomson Scientific 2006) since JCMC has not been indexed in the ISI Web of Science database. 2001 and 2002 were chosen as the sample years of cited articles to allow 4 to 5 years for articles to be cited on the Web and in ISI journals.

Data collection

The data we collected contain two parts. The first part focused on ISI citations, which were used to form a comparison with Web citations in our study. Searching the ISI Web of Science database, we found journal articles published from 2001 to January 2006, which cite a JCMC or NMS article published in 2001 and 2002. Since JCMC articles were not indexed as citing items in the ISI database, we also manually checked the references of JCMC articles (2001 ~ January 2006) to collect self-citations. In total, there were 131 JCMC citations and 136 NMS citations.

The second part concerned Web citations. By Web citation, we meant to include and merge citations of the three different types reviewed above (i.e. Web text citation, URL citation and hyperlink citation) so that a complete set of Web citations can be guaranteed.

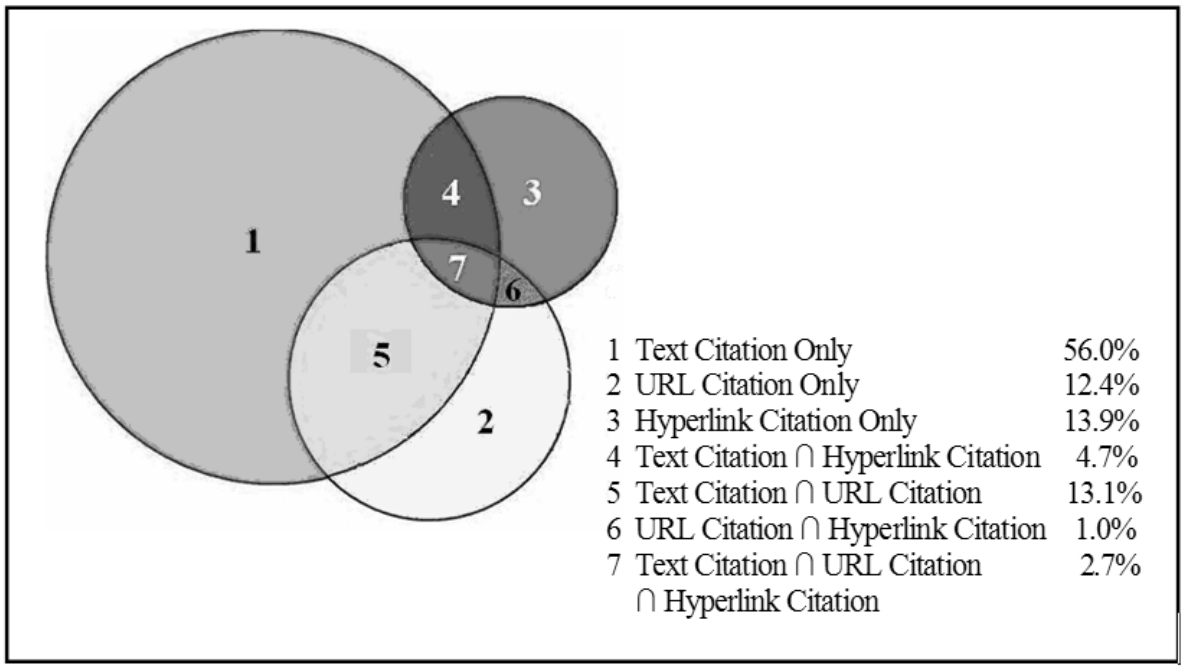
In our study, the collection of Web citation data was automated. We developed PHP programs to retrieve Web citations by invoking search engine APIs (Google API and Yahoo! API). The Google API was used to collect Web text citations and Web URL citations because results from previous studies showed that it provides the most comprehensive (Bar-Ilan 2004) and most stable search results over time (Vaughan & Shaw 2005). On the other hand, the Yahoo! API was used to retrieve hyperlink citations since Yahoo! outperforms other search engines in link searching (Vaughan & Zhang in press).

When searching for Web citations for a journal article, it was very important to create a combination of multiple queries which ensures that all qualified Web citations indexed in the search engine databases can be retrieved. With regard to searching Web text citations, our strategy was to enter the article's main title in a pair of double quotation marks to form a phrase search; if the main title was not sufficiently distinctive, we changed it into a two-term query, one being the main title phrase, the other the subtitle phrase or

Table 2. An example of Web citation search strategy

Article title	In Community We Trust: Online Security Communication at eBay
Author	Josh Boyd
URLs of the online article	http://jcmc.indiana.edu/vol7/issue3/boyd.html http://www.ascusc.org/jcmc/ vol7/issue3/boyd.html
Search Text citation	Query1 = [„In Community We Trust: Online Security Communication at eBay"] Query2 = ["In Community We Trust", "Online Security Communication at eBay"] Query3 = ["In Community We Trust", Boyd]
Search URL citation (for JCMC articles only)	Query1 = [jcmc.indiana.edu/vol7/issue3/boyd.html] Query2 = [www.ascusc.org/jcmc/vol7/issue3/boyd.html]
Search Hyperlink citation (for JCMC articles only)	Query1 = [link:http://jcmc.indiana.edu/vol7/issue3/boyd.html] Query2 = [link:http://www.ascusc.org/jcmc/vol7/issue3/boyd.html]

Figure 1. Proportion of different types of Web citations



author's last name. At the same time, we also entered the full title to form a phrase search because we found in practice that the results of a full-title search were not necessarily a subset of main title search results. Finally, the Web text citations of a specified article, as shown in Table 2, were generated by merging results of multiple search queries together. Similarly, this query-combination strategy was also applied for searching Web URL citations and hyperlink citations. Because a single JCMC article had two URLs, two different queries had to be generated. Since NMS articles were not published online, there was no need to conduct Web URL citations and hyperlink citations.

Undoubtedly, the results of the text citation search, URL citation search and hyperlink citation search overlapped somewhat. Therefore, we merged three groups of results together and deleted duplicate URLs. Finally, we identified 1860 Web citations for NMS and 3762 Web citations for JCMC. For JCMC, in particular, text citations, URL citations and hyperlink citations respectively made up 76.5%, 29.2% and 22.3% of the whole set (shown in Figure 1). This result showed that collection based on any single type of citation, like methods used in previous studies, would result in loss of raw data.

Data processing

Before data analysis, the collected Web citations were pre-processed as below:

1. For any given article, Web citations from the journal's hosting site were filtered out. Filtered citations included the online article itself and directory-like pages containing the journal's Table of Contents, article abstract, or indexes (e.g. http://nms.sagepub.com/content/vol3/issue4/, http://jcmc.indiana.edu/jcmcindex.html). We excluded these internal citations from our citation data set because overwhelming internal citations from either journal's (especially OA journal) hosting site could distort the citation analysis result. However, there was one exception for JCMC – if JCMC article A was cited by another JCMC article B instead of directory-like pages, this citation should be taken as self-citation and included in our data set.

2. We found that the collected Web citations were inflated by online database (e.g. DBLP - Digital Bibliography & Library Project) search results. From a same database site, Google and Yahoo! could retrieve up to hundreds of dynamic pages, which are actually database query results, regarding the queried article. Although these pages had distinct URLs, they had the same content – the bibliographic information about the queried article. Such types of query result pages were regarded as noise which can distort our citation analysis results. To eliminate this effect, we manually checked this type of query result citation and deleted them.

Comparison of Web citation counts

After data pre-processing, there were 2782 Web citations for JCMC (26.1% shrinkage) and 1436 Web Citations for NMS (22.8% shrinkage). Based on this data set, the Web citations for JCMC and NMS were compared. We counted Web citations for every JCMC and NMS article. On the individual article level, we examined whether there is significant difference between JCMC and NMS in the number of Web citations received. Meanwhile, a similar comparison based on the ISI citations was also conducted.

Classification of Web citations sources

To investigate Web citations in more detail, we classified citations by citing source type and country represented by citing items. According to the role of the citation source and the different modalities of the impact represented by citations, citing sources were classified into five categories and 14 sub-classes, as shown below:

Category I: Formal scholarly publications
Citation source belongs to formal scholarly publications such as journal papers. They are posted on the Web in various formats (e.g. HTML, PDF, etc.) and contain citations to JCMC or NMS articles published in 2001 and 2002. Basically, this type of Web citation is equal to the bibliographic citation in representing the intellectual impact of cited articles, although some of publications were not indexed in ISI Web of Science database. With this type of citation, the impact of cited items among research communities, particularly researchers involved in scholarly publication, is measurable based on citation counts received. These formal scholarly publications posted on the Web normally include:

a. Journal article;
b. Conference/workshop paper;
c. Technical report;
d. Book/book chapter;
e. Thesis/dissertation;

Category II: Course teaching materials
Citations from course syllabi/reading lists were also considered another type of citation reflecting the intellectual impact of cited articles (Vaughan & Shaw 2003). In this type of citation, the impact of cited items in educating the next generation could be reflected. In our study, articles listed in a bibliography/reading list for a course (includes continuing education as well as university-based courses) or citations from lecture notes were classified in this category.

f. Course reading list /Lecture Notes/Syllabus

Category III: Non-authoritative documents
Category I and II can claim to be authoritative information sources for research or study and hence such citations can be regarded as indicators of the impact of cited items. In addition, Web citations could also be found in a large number of useful information sources but whose credibility and authority are not necessarily guaranteed. For instance, academic essays or online tutorials (e.g. "Introduction to Web surveys") use journal articles (e.g. a JCMC article – Design of Web survey questionnaires) as references; Wikipedia uses a JCMC article as an external link for explaining "Internet Troll"; a NMS article is referred to within the discussion of "Gender and the use of the internet" on an online forum. Obviously, such types of citations represent the citing authors' acknowledgement of cited articles. These citing documents are widely used by Web users because of their considerable academic values and excellent accessibility supported by powerful search engines such as Google. Most often, authors of such documents are professors, students, researchers, and other knowledgeable people in a relevant field. We argue that citations from this type of Web document could also be considered indicators of the impact of cited articles, although they may be relatively weak indicators in relation to citations in category I and II. Generally, citation sources in category III were comprised of the following subclasses.

g. Academic essays/online encyclopedia article/online tutorial (e.g. Wikipedia article, guidelines for conducting Web surveys)
h. Online academic discussion (including mailing archive of academic discussion group, e.g. jESSE listerv®; debates posted on online academic forums or academic blogs)

i. Academic resources list – a selective reading list edited by individuals or organizations for study or research in a certain field. (In our study, many authors of such citing pages were found to be professors, researchers, graduate students or research institutions. However, we distinguished this type from category II because this type of citation source is not for teaching purposes and we assume that course teaching materials are posted after more careful preparation and strict examination, and therefore generally have higher credibility.)

j. Unpublished academic works (including research proposals, students' term papers, course project papers, etc.)

Category IV: Bibliography list

This type of citing source is solely functioning as bibliography. It cannot be regarded as the indicator of the impact of cited items.

k. Author's CV

l. Bibliography (different from reading list in category II, this type of citing pages list out publications by year, journal, publisher, or etc, rather than picked articles recommended by scholars)

m. Web directory (Web directories, search engine results, etc.) Basically, these citations are for navigation purposes.

Category V: Others

Any remaining citation sources were classified into this category. Mainly, they included Web pages which were inaccessible during the study and Web pages the required citation for which could not be found.

To reduce the classification task to manageable proportions, 468 samples were randomly selected from the whole set of Web citations. First, we divided Web citations into four groups by the journal title and the publication year of cited articles (i.e. JCMC2001, JCMC2002, NMS2001, NMS2002). Then from each group, we randomly selected one out of every nine Web citations. As a result, there were 159, 150, 82, 77 Web citation samples respectively for each group. Based on sampled data, we compared JCMC and NMS to determine if their citation sources have different distribution features.

Classification of Web citations by country

We also classified the sample's 468 Web citations by countries represented by citing sources. The purpose was to discover where these citations came from – developing countries or developed countries? The geographic locations of the citing sources were determined by the citing authors' affiliation locations; therefore, citing authors' countries of origin were not utilized for this purpose. The definition of developing and developed countries adopted for our study was the most widely used, official-source development data from the World Bank (World Bank 2005).

The global impacts (developing countries versus developed countries) of JCMC and NMS were compared from two aspects: (1) country coverage, i.e. how many developing and developed countries were covered by the citations for JCMC and NMS respectively? For any citing source, we can consider either the country represented by the first author only or countries represented by all co-authors. Both measures were adopted in our investigation; (2) how many cites were contributed by developing and developed countries respectively? Given a co-authored citing source (e.g. a US author and three Canadian authors), the number of cites from the US is then 0.25, and the number of cites from Canada is then 0.75. By aggregating all weighted cites by country, we obtained the number of cites from each individual country.

Comparison 1: Citation count to individual article

For both JCMC and NMS journals, we counted ISI and Web citations for individual articles. The results are shown in Table 3. For ISI citation counts at the individual article level, NMS is slightly higher than JCMC. The median ISI citation counts for JCMC and NMS are both 2 and their respective means are 2.6 and 2.9. A Mann-Whitney test shows that there is no statistically significant difference in the ISI citation count between JCMC and NMS.

However, the Web citation-based comparison shows different results. The median Web citation counts for JCMC and NMS are respectively 45 and 24. A Mann-Whitney test shows that the difference between JCMC and NMS is statistically significant ($p<0.001$). It should be noted that the Mann-Whitney test was used instead of the Independent T-test because the frequency distributions of citation counts for JCMC and NMS were both very skewed (Howell 2002, 707; Vaughan 2001, 122). In addition, the number of citations each journal re-

Table 3. Descriptive statistics of ISI & Web citation counts.

		Median	Mean	Mean 2001	Mean 2002
ISI Citation	JCMC	2	2.6	3.04	2.27
	NMS	2	2.9	3.14	2.75
Web Citation	JCMC	45	56.8	62.2	52
	NMS	24	31.9	35	29.2

Table 4. Classification of Web citations to JCMC and NMS published in 2001 and 2002.

	Formal Scholarly Publication	Teaching Materials	Non-authoritative Documents	Bibliography	Others	Total
JCMC	114 (36.9%)	18 (5.8%)	101 (32.7%)	58 (18.8%)	18 (5.8%)	309 (100%)
NMS	60 (37.7%)	14 (8.8%)	29 (18.2%)	43 (27%)	13 (8.2%)	159 (100%)
Total	174 (37.2%)	32 (6.8%)	130 (27.8%)	101 (21.6%)	31 (6.6%)	468 (100%)

ceived divided by the number of journal articles in each year (JCMC2001, JCMC2002, NMS2001 and NMS2002) is respectively 62.2, 52, 35, and 29.2. The ratios of Web-citation means (i.e. JCMC/ NMS) are stable and equal 1.78 in two continuous years, which implies that our test results are not affected by the publication year.

The preceding results indicate that the OA journal (i.e. JCMC) articles have more Web citations than the print journal (NMS) articles. A possible explanation for this phenomenon is that the access mode caused this difference.

We noticed that Web-citation sources contained not only formal scholarly publications which reflect the impact of cited articles, but also other pages such as author bibliographies which can not be considered as an indicator of the impact of cited articles. Thus, it is interesting to investigate in further depth whether JCMC articles receive more citations than NMS articles from formal scholarly publications posted on the Web, or whether JCMC's advantage of Web citation count over NMS comes from Web pages not representing any intellectual impact. Therefore, we did another Mann-Whitney test to compare JCMC and NMS's Web citations in category I only. The result shows that JCMC and NMS articles are still significantly different ($p=0.034$) in terms of the number of online formal scholarly citations per article. Almost half (23 out of 49) of the JCMC articles receive 2 citations; while half (22 out of 45) of the NMS articles receive 0 citations. This result clearly confirms that the OA journal JCMC received more bibliographic citations than NMS.

Comparison 2: Source of Web citations

Web citations were classified into four categories as described in detail in the "Data Processing" section above. Through classification, we further investigated whether these two journals are different in terms of Web citation distributions. Classification results are shown in Table 4.

A chi-square test shows a significant ($p < 0.01$) difference between two journals in terms of the category of the citation sources. Compared with NMS, the percentage of Web citations from formal scholarly publications to JCMC articles is 0.8% less; the percentage of Web citations from Teaching Materials to JCMC articles is 3% less; but with regard to the proportion of non-authoritative Web documents, JCMC is 14.5% higher than NMS. The difference between citations in category I, II and III is that they represent the impact of cited items from three different modalities of scholarly communication. Citations in formal scholarly publication reflect the use of cited articles in the process of producing formal scholarly works; the impact recipients are mainly publication (citing) authors. Citations in the course teaching materials reflect the use of cited articles in school education; the impact of cited items on teaching staff and students are represented. Citations in category III reflect the use of cited articles in relatively informal academic communication occasions; users can cite a journal article for more various purposes, such as online reviewing, debating, etc., rather than publishing papers only. Therefore, we can use citations in category III to investigate the impact

Table 5. Numbers of developing and developed countries contributing citations.

		Number of Developing countries	Number of Developed countries	#_Developing : #_Developed
ISI Citation	JCMC	2	17	1 : 8.5
	NMS	3	18	1 : 6
Web Citation	JCMC	8	23	1 : 2.875
	NMS	6	20	1 : 3.33

Table 6. Numbers of citations contributed by developing and developed countries

		Number of cites from developing countries	Number of cites from developed countries	Number of cites from country unknown	Percentage of cites from developing countries
ISI Citation	JCMC	2.45	128.55	0	1.87%
	NMS	9.7	126.3	0	7.13%
Web Citation	JCMC	14	214	80	4.55%
	NMS	8	134	18	5.0%

of cited articles permeating through everyday academic activities. Impact analysis can be extended to cover more impact recipients such as students and junior researchers, who represent the mass of the academic community, rather than cover only leading researchers such as publication authors. The different proportion of citations shown in Table 4, particularly in category III, reveals that, in the Web context, JCMC articles are more frequently cited/used in informal academic communication and JCMC articles have greater impact than NMS articles among grass-root members of the academic community.

Comparison 3: Web citation classification by country

First, we investigated the number of developing countries and developed countries from which JCMC and NMS have received all types of Web citations. Table 5 shows the analysis based on countries represented by first citing authors (the analysis based on countries represented by all citing authors also generated consistent results). ISI citations for NMS articles are dispersed in a slightly wider scope than JCMC in both developing and developed countries. But generally, JCMC and NMS are very similar in the number and the proportion of developing and developed countries covered by the ISI citations.

When Web citations are compared, the dissimilarity between JCMC and NMS is rendered obvious. Web citations for JCMC cover a wider scope in both developing and developed countries. More over, the JCMC's ratio of developing country counts over developed country counts is larger than the ratio for NMS. Such results indicate that OA articles, compared with NMS articles, tend to be cited on the Web by users distributed on a wider geographic scope. This pattern is particularly obvious in developing countries. A possible explanation of this phenomenon is that free online access causes OA articles to be disseminated more widely and it benefits developing country readership more than developed country readership. On the other hand, citing sources indexed in the ISI database are terribly predominated by journals and authors from developed countries – "North America and Western European publishers provide nearly 90% of the journal titles" (McVeigh 2004). That may be part of the reason why the OA journal advantage in developing countries was not shown in ISI citation samples.

Second, we compared JCMC and NMS in the number of Web citations contributed by developing and developed countries (shown in Table 6). The country distribution of ISI citations shows the two journals' significant ($p=0.039$) difference regarding their prevalence in developing and developed countries. That is, NMS has more citing authors from developing countries than JCMC. Interestingly, the country distribution of Web citations shows there is no significant ($p=0.89$) difference between the two journals. Such a result suggests that the open access mode is possibly an important factor that can affect the distribution of

Web citations in developing and developed countries. Since the two journals' impact (ISI-based) difference could also possibly affect the amount and proportion of citations from developing countries as shown by the ISI citation analysis, we need to control its effect when examining the OA mode's impact on increasing the proportion of Web citations from developing countries. This is achieved by developing a new measure, the ratio R defined as the percentage of Web citation from developing countries divided by the percentage of ISI citation from developing countries. It is astonishing that the R of JCMC was 3.47 times greater than the R of NMS. A possible explanation for this phenomenon is that the open access mode helps remove the barrier for users in developing countries to access and cite journal articles.

Results in Table 5 and 6 together show that on the Web, JCMC not only gets more citations from developing countries than NMS, but also receives Web citations from a wider scope of developing countries. Furthermore, this pattern was still identified when the same classification and analysis was conducted based on Web citations of category I, II and III only, which are indicators of the impact of cited articles. Such results suggest it is the access mode that makes JCMC disseminated in developing countries with greater width and depth than NMS.

Discussion and conclusion

This study has introduced a new methodology for collecting journal article Web citations. By including three different types of Web citations, the sample size was increased approximately 25%. Programming-based data collection makes it possible to collect Web citations on a very large scale in a short period of time. In our study, the JCMC Web citations and NMS Web citations were retrieved in the same day (February 14, 2006), thus the instability of sampled data caused by the update of search engine index databases could be decreased effectively.

JCMC and NMS represent respectively the OA journal and the traditional access journal, each having similar topic coverage, journal impact factors and international reputation. In our test, the effect of JIF on citation counts was controlled – there is no significant ($p=0.425$) difference between JCMC and NMS in the number of ISI citations received at the individual article level. However, we still found a statistically significant ($p<0.001$) difference between JCMC and NMS in Web citation counts per article. On average, JCMC articles receive Web citations twice as often as NMS articles. Such evidence suggests that the access mode may be the variable affecting the Web citation counts to journal articles, with OA articles receiving more Web citations. More importantly, when the citing source is restricted to formal scholarly publications posted on the Web, JCMC articles still have more citations than their traditional access counterparts – there is a significant ($p=0.034$) difference between JCMC articles and NMS articles in their Web citation counts from category I. This is actually to say that the OA mode can also affect the bibliographic citation counts of a journal, which is used to evaluate ISI journal impact.

Web citation classification outlined the citation source distribution of the two journals. JCMC has more Web citations than NMS in any category and subclass. But the advantages are not evenly distributed. The two journals have almost the same proportion of Web citations from formal scholarly publication. However, NMS is much (14.5%) lower than JCMC in the proportion of Web citations in category III, in which citations were created for various academic purposes other than producing formal publications and teaching courses, and citing authors ranged from top research scientists to the general public. Significant difference between JCMC and NMS actually attests that the impact advantage of the OA journal is visible not only among experts represented by formal publication authors and university professors, but even more remarkable among the general online public including students and beginners, creating citations for various academic reasons.

Through citation classification by country, we compared the two journals' impact in developing and developed countries. As the ISI database is overwhelmingly predominated by journals and authors in developed countries (McVeigh 2004; Vaughan & Shaw 2005), a large proportion of citations from developing countries could possibly be missed. Web citation analysis, on the other hand, can reduce the appearance of this kind of bias and reflect citations with greater geographic diversity. Comparisons show that Web citations for JCMC are more widely distributed than Web citations referring to NMS both in developing and

developed countries. Particularly with regard to developing countries, JCMC not only receives more Web citations than NMS, but also has citations distributed in a wider range of developing countries. Such evidence suggests that the open access mode could improve the dissemination and impact of journal articles globally. More importantly, the comparisons also demonstrate that applying the open access mode to a journal (e.g. JCMC) improved the proportion of Web citations from developing countries.

In conclusion, this experimental study generated the Web citation profiles of JCMC and NMS. The Web citation advantage of OA journal JCMC was demonstrated. Published online, OA articles are freely accessible to any user having Internet access so that they may potentially have a much larger size of readership than traditional access journal articles, and consequently receive far more citations. The classification of Web citation sources shows that traditional access journal articles have a significantly smaller proportion of citations from non-formally published academic materials than OA articles. This indicates that the OA articles' impact advantage over traditional access counterparts in informal academic communication is even more distinct than in formal communication, which is represented by formal publication and school education. The classification of Web citations by countries shows that JCMC articles receive a higher proportion of Web citations from developing countries and from a wider international scope. A convincing interpretation is that open access could effectively improve the articles' impact in developing countries and contribute to decreasing the academic gap between developing countries and developed countries.

One of the limitations in this study is that 6.5% Web citations have their citing source type unidentifiable and 21.5% Web citations have their country property unidentifiable. This is caused by the volatility of Web pages and the anonymity of Web page creators. Also, the comparison between OA journals and traditional access journals was conducted in a small scale – each type has only one journal as representative in this pilot study. In future work, we will include more OA journals and more print journals from a large discipline such as Library and Information Science so that the generality of our conclusion can be strengthened.

Acknowledgement

I am very grateful to my guide Prof. Liwen Vaughan for discussions and advice on this topic. I also thank Denver Vale Nixon and Nancy John for their very helpful comments and suggestions.

References

Almind, T.C. and P. Ingwersen. 1997. Info metric analysis on the World Wide Web: Methodological approaches to "Webometrics". *Journal of Documentation* 53(4): 404–426.

Bar-Ilan, J. 2004. The use of Web search engines in information science research. *Annual Review of Information Science and Technology* 38: 231–288.

BOAI. 2006. *Budapest Open Access Initiative: Frequently Asked Questions*. URL: http://www.earlham.edu/~peters/fos/boaifaq.htm [viewed March 30, 2006]

Brody, T., H. Stamerjohanns, S. Harnad, Y. Gingras, F. Vallieres, and C. Oppenheim. 2004. The effect of Open Access on Citation Impact. Presented at: *National Policies on Open Access (OA) Provision for University Research Output: an International meeting*. Southampton University, Southampton UK. February 19, 2004. URL: http://opcit.eprints.org/feb19prog.html [viewed March 30, 2006]

Create Change 2006. Scholars Under Siege: The Scholarly Communication Crisis. URL: http://www.createchange.org/librarians/issues/silent.html [viewed March 30, 2006]

Cronin, B., H. Snyder, R. Rosenbaum, A. Martinson, and E. Callahan. 1998. Invoked on the Web. Journal of the American Society for Information Science 49(14): 1319–1328.

DOAJ. 2006. Directory of Open Access Journals. URL: http://www.doaj.org [viewed March 30, 2006]

Harter, S. and C. Ford. 2000. Web-based analysis of E-journal impact: Approaches, problems, and issues. *Journal of the American Society for Information Science* 51(13): 1159–76.

Hu, Q. 2005. Open Access: A necessity for promoting capacity building in science and technology. URL: http://www.cae.cn/english/memberssuggestions/content.jsp?id=966 [viewed March 30, 2006]

Howell, D. 2002. *Statistical methods for psychology, 5th ed*. Pacific Grove, CA: Duxbury.

Ingwersen, P. 1998. The calculation of Web impact factors. *Journal of Documentation* 54(2): 236–243.

Kean, G. 2005. 18th Annual Study of Journal Prices for Scientific and Medical Society Journals. *The News Letter for Journal Publishers* 3, 2005. URL: http://www.allenpress.com/static/newsletters/pdf/JP-2005-03.pdf [viewed March 30, 2006]

Kousha, K. and M. Thelwall. 2005. Motivations for URL Citations to Open Access Library and Information Science Articles. In: *Proceedings* of *10th International Conference Society of Scientometrics and Informetrics (ISSI)*, Stockholm, July 24–28, 2005.

Kurtz, M.J. 2004. *Restrictive access policies cut readership of electronic research journal articles by a factor of two*, Harvard-Smithsonian Centre for Astrophysics, Cambridge, MA URL: http://opcit.eprints.org/feb19oa/kurtz.pdf [viewed March 30, 2006]

Larson, R.R. 1996. Bibliometrics and the World Wide Web: An exploratory analysis of the intellectual structure of cyberspace. *Proceedings of the 59th Annual Meeting of the American Society for Information Science* 71–83. URL: http://sherlock.berkeley.edu/asis96/asis96.html [viewed March 30, 2006]

Lawrence, S. 2001. Online or Invisible? *Nature* 41 (6837): 521 URL: http://www.neci.nec.com/~lawrence/papers/online-nature01/ [viewed March 30, 2006]

McVeigh, M. E. 2004. *Open Access Journals in the ISI Citation Databases: Analysis of Impact Factors and Citation Patterns*. URL: http://www.thomsonscientific.com/media/presentrep/essayspdf/openaccesscitations2.pdf [viewed March 30, 2006]

Odlyzko, A.M. 2002. The rapid evolution of scholarly communication. *Learned Publishing* 15: 7–19. URL: http://www.catchword.com/alpsp/09531513/v15n1/contp1-1.htm [viewed March 30, 2006]

Pringle, J. 2004. Do open access journals have impact? *Nature Web Focus*. URL: http://www.nature.com/nature/focus/accessdebate/19.html [viewed March 30, 2006]

Rousseau, R. 1997. Sitations: An exploratory study. *Cybermetrics* 1: 1-9. URL: http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html [viewed March 30, 2006]

Thomson Scientific. 2006. The ISI Impact Factor. URL: http://scientific.thomson.com/free/essays/journalcitationreports/impactfactor/ [viewed March 30, 2006]

Vaughan, L. 2001. *Statistical methods for the information professional: a practical, painless approach to understanding, using, and interpreting statistics*. (ASIST Monograph Series), Medford, New Jersey: Information Today, Inc.

Vaughan, L. and K. Hysen. 2002. Relationship between links to journal web sites and impact factors. *Aslib Proceedings: New Information Perspectives* 54(6): 356–361.

Vaughan, L. and M. Thelwall. 2002. Web link counts correlate with ISI Impact Factors: evidence from two disciplines, *Proceedings of the ASIST Annual Meeting* 39 (ASIST 2002): 436–443.

Vaughan, L. and D. Shaw. 2003. Bibliographic and web citations: What is the difference? *Journal of the American Society for Information Science and Technology* 54(14): 1313–1322.

Vaughan, L. and M. Thelwall. 2003. Scholarly use of the web: What are the key inducers of links to journal web sites? *Journal of the American Society for Information Science and Technology* 54(1): 29–38.

Vaughan, L. and D. Shaw. 2005. Web Citation Data for Impact Assessment: A Comparison of Four Science Disciplines. *Journal of the American Society for Information Science and Technology* 56(10):1075–1087.

Vaughan, L. and Y. Zhang. (in press). Equal Representation by Search Engines? A Comparison of Websites across Countries and Domains. *Journal of Computer-Mediated Communication*.

World Bank. 2005. Country Classification: Country Groups. URL: http://www.worldbank.org/data/countryclass/classgroups.htm [viewed March 30, 2006]

Editorial history:
paper received 30 May 2006;
accepted 14 August 2006.