

Web Logs as Indices of Electronic Journal Use: Tools for Identifying a “Classic” Article

KATE MAREK AND EDWARD J. VALAUSKAS

Graduate School of Library and Information Science, Dominican University, River Forest, IL USA

First Monday (<http://firstmonday.org>)

Because of the importance to the academic community of shared research through publication, information specialists have developed mechanisms to identify important articles. One accepted method is citation analysis, where the use of an author's work is tracked by counting subsequent citations to that work. The development of Web-based scholarly

journals offers new tools, such as server logs, for tracking an article's use and distribution. This paper explores the potential of server logs as sources of information to reveal distribution levels of individual articles. A new model is proposed for designating a work of significant scholarly impact, the “classic” article.

Introduction

The academic community is known for its culture of sharing information, research, and new ideas. Sharing new information through publication is important for the development of scholarship and scientific discovery. Because of the importance of shared scholarship, academic reward systems consider research publications by faculty to be vital. With this in mind, information professionals have worked to develop tools for identification of successful research publications.

Citation analysis is one tool which is used extensively to identify active use of published information. Measures of individuals' academic and scientific contributions frequently include studies of how their work has been used by other scholars in subsequent publications. Eugene Garfield, via The Institute for Scientific Information (ISI) and its various citation indexes, has been a leader in promoting citation analysis over the last 50 years as a tool for evaluating scientific contribution. “By recognizing that the value of infor-

mation is determined by those who use it, what better way to measure the quality of work than by measuring the impact it makes on the community at large ... ” (Isinet.com 2002).

In an early article describing citation analysis developed in conjunction with the *Science Citation Index*, Garfield explains the emerging opportunities offered by computer manipulation of large data files. “To date, studies of the network and of the interrelation of its components have been limited in the number of journals, the areas of scientific study, and the periods of time their authors were able to consider. Such shortcomings ... [have been due] to the practical difficulty of compiling and manipulating manually the enormous amount of necessary data” (Garfield 1972). Garfield's project was one of several in the 1960s and early 1970s in which he began to build and mine databases. This study focused on *Science Citation Index* data for a three-month period in 1969. Garfield generated three specific listings for that time period: cumulation of all citations of the same titles with the number of times each title was cited;

Kate Marek is Assistant Professor, Graduate School of Library and Information Science, Dominican University, River Forest, IL, USA. E-mail: kmarek@dom.edu

Edward J. Valauskas is Chief Editor, *First Monday*, Post Office Box 87636, Chicago IL 60680 USA, E-mail: ejv@uic.edu

a detailed citation history of each cited title; and, citations arranged by citing journal rather than by cited article title. Garfield laid the foundation with this paper for future applications of citation analysis, including designations of "citation classics," "high impact journals," and "hot papers."

A *citation classic* as defined by Garfield is an individual work which is cited frequently and consistently over time. Citation frequency is the first criteria for selection as a classic (Garfield 1984). To be designated a "classic", an individual work must be cited at least 400 times. The formula, however, is designed to include articles published in small journals from narrower areas of specialization. Thus articles in specialty journals may be cited as few as 50 times to qualify as a "classic". A "*high impact journal*" is defined as a journal judged by the "accumulated total of citations to all articles published by a journal" (Harter 1996). "*Fast Breaking Hot Papers*" (available at <http://www.esi-topics.com> [viewed November 15, 2002]) are designated by *The Scientist* through citation analysis of the most recent three months, based on the realization that "citation of significant breakthroughs usually [is] quite rapid" (Garfield 2002).

This discussion of citation analysis and its background provides a basis for considering current developments in academic publishing. The mechanisms for analysis over the last fifty years have been electronic databases built to store, organize, and analyze bibliographic and statistical information. Just as the early development of computers enabled the first efforts at large-scale citation analysis, today's Internet-based mechanisms offer a new generation of analysis. Web-based journals, for example, have the added advantage of server logs, which can identify users from a variety of statistical data points. "Citation-based measures are not the only, or even necessarily the best, measures of [scholarly] impact. ... Publishers [of e-journals] can gain some information regarding use by noting the number of subscriptions to their journals or by counting the number of times articles are accessed or downloaded by host servers" (Harter 1996). Study of this electronic subscription and usage data gives journal editors new opportunities for timely analysis, and offers new value to citation analysis concepts such as "high impact" and "citation classic."

Server logs recording the use of Web-based journals reveal patterns of long-term use over time, including indications of use within the journal that is generated specifically from user access to highly cited works in that same journal. Thus, frequently cited articles, or citation classics, in online journals generate more visits to the Web-based journal site, in turn giving all articles within that journal higher impact. This demonstrates a new facet of peer review, not unlike the model used by the successful Web search engine Google. In Google, rankings of Web pages are based in part on hyperlinks from other Web sites, with higher rankings given to sites with greater numbers of links. This process, termed "PageRank" by Google, is based on counting links as "votes" in their ranking structure (Google 2002).

The cycle of increased use is continually stimulated by active readership, and ultimately speaks positively to the heightened value of quality Web journals. This paper examines the annual logs of the Internet journal *First Monday* in an attempt to demonstrate the ways in which a "classic article" and "high impact journal" might be defined in this Internet environment for scholarly work. It also examines the overall effects of the availability of an Internet journal's back issues and the use of this archival content over time.

Methodology

First Monday (<http://www.firstmonday.org>) is a monthly, peer-reviewed journal focusing on topics pertinent to the Internet and its impact. It began publication in May of 1996, and since then has published nearly 400 articles (*First Monday* 2002). Web-based articles are stored on a server and made available through the World Wide Web, where appropriate analysis software monitors and collects data about visitors to the *First Monday* Web site and their use of individual articles.

Journal usage logs

Logs recording the use of *First Monday* via the World Wide Web were examined on a monthly basis during the period January 1999 through December 2001 in order to understand the use of *First Monday's* content over time. These logs were analyzed by *Analog* (<http://www.analog.cx/>), ver-

sion 4.16, which extracts information from the logs and provides it in several categories, including yearly, monthly, daily, and hourly summaries; domain, organization, directory, file type, status code, and request reports; and, failure, referrer, referring site, and search word reports.

How should these logs be used? Certainly, individual contributors to a peer-reviewed Internet journal such as *First Monday* should have some sense of the circulation of their work, once published. *First Monday* provides, on request by a contributor, statistical information on the circulation of a given paper. This information may be used in turn for tenure review or for sheer personal insight into the use of a paper. The editors of *First Monday* also use information from the logs for editorial work. Log information provides clues to the utility of papers on certain specific subjects. Ever since the publication of David Noble's classic paper "Digital Diploma Mills" in the January 1998 issue of *First Monday*, the journal has made an effort to publish papers on distance education and especially papers on the impact of the Internet on higher education. In part, this has been stimulated by the editorial knowledge that readers turn to *First Monday* for information on these topics, based on log records for not only Noble's paper but contributions commenting on Noble's ideas. *First Monday* seeks papers on open source software for similar reasons, given the utility of papers published in its virtual pages on the topic, notably by Eric Raymond.

Efforts are made, through the posting of an official Web site privacy statement (<http://www.firstmonday.org/privacystatment.html> [viewed November 15, 2002]) to make information available to readers and authors about *First Monday's* usage logs. An Opt-Out option is offered to persons who prefer to not be included in the statistical analysis of article downloads.

Data collection and analysis

For the purpose of this study, the specific yearly reports for 1999, 2000, and 2001 were studied, specifically looking at the request reports for each year. Each request report lists files with at least twenty requests per year, sorted by the number of requests. Each report provides details on the most requested files, indicating the number of requests; percentage of requests for a given file compared

to the number of all files on the server; number of gigabytes transferred; percentage of bytes transferred for a specific file compared to the total number of files transferred; and, the last time a specific file was requested and transferred. Tables 1–7 in this paper represent a summary of the logs, as based on reports generated by *Analog* for the most requested files for the years 1999, 2000 and 2001.

Discussion

"... the more widely scientists make their intellectual property freely available to others, the more securely it becomes identified as their property. For science is public not private knowledge." (Merton 1983)

Ideas, expressed in scholarly communication, need to be widely distributed to have any effect on a research community or the world at large. Paradoxically, traditional, peer-reviewed, print journals have been the medium for scholarship for the past three centuries, and these journals have been anything but free or widely available in many cases. For the most part, and with the exception of a few notable titles, peer-reviewed, scholarly journals have had limited circulation, usually to well-defined academic communities. Because of the nature of these journals – that is printed and sold by subscription or by membership in a particular society or professional organization – the exact circulation of an idea, as expressed in a published paper, has never been known. Indirect indices, such as citation records, provide one imperfect measure of the distribution and utility of an idea as expressed in a scholarly contribution printed in a peer-reviewed journal.

With the development and maturation of peer-reviewed Web journals over the past decade, it has become possible to more closely measure the circulation of published papers. Thanks to server logs, it is now possible to extract more quickly exact values of the use of a specific contribution over time. Server logs record the use of files stored on a specific World Wide Web server, providing detailed snapshots of the number of times a given paper is downloaded from the server to a local computer via the Internet (Thelwall 2001). As such, an analysis of the information collected in server logs would provide more immediate pictures of the utility of scholarship as printed in Internet journals.

Table 1: Defining a Classic Paper
Most Read Papers, based on Cumulative Logs of *First Monday*, 1999–2001
(based on appearance in Top 10 requested papers for two or more years)

rank	paper	Total Number of requests	Number of months in top 10 requested	Average requests per month
1	Eytan Adar and Bernardo A. Huberman, "Free Riding on Gnutella", <i>First Monday</i> , volume 5, number 10 (October 2000), http://firstmonday.org/issues/issue5_10/adar/	82,396	15	5,493
2	David F. Noble, "Digital Diploma Mills: The Automation of Higher Education," <i>First Monday</i> , volume 3, number 1 (January 1998), http://firstmonday.org/issues/issue3_1/noble/	58,122	36	1,614
3	Craig H. Rowland, "Covert Channels in the TCP/IP Protocol Suite", <i>First Monday</i> , volume 2, number 5 (May 1997), http://firstmonday.org/issues/issue2_5/rowland/	55,680	24	2,320
4	Karen L. Murphy and Mauri P. Collins, "Communication Conventions in Instructional Electronic Chats," <i>First Monday</i> , volume 2, number 11 (November 1997), http://firstmonday.org/issues/issue2_11/murphy/	31,446	24	1,310
5	Noriko Hara and Rob Kling, "Students' Frustrations with a Web-Based Distance Education Course," <i>First Monday</i> , volume 4, number 12 (December 1999), http://firstmonday.org/issues/issue4_12/hara/	30,167	24	1,257
6	John Kelsey and Bruce Schneier, "The Street Performer Protocol and Digital Copyrights," <i>First Monday</i> , volume 4, number 6 (June 1999), http://firstmonday.org/issues/issue4_6/kelsey/	27,610	18	1,534
7	Mary Minow, "Filters and the Public Library: A Legal and Policy Analysis," <i>First Monday</i> , volume 2, number 12 (December 1997), http://firstmonday.org/issues/issue2_12/minow/	21,013	24	876

In this study, we examined how papers were used in *First Monday*. By "used," we mean a paper that actually was downloaded from the *First Monday* server to a local computer via screen display, printing, or downloading to disk. Statistical analysis of *First Monday*'s logs allows us to collect all of the hyperlinks to a given paper as one value, as the Uniform Resource Locator (URL) for a given paper. Challenges in the analysis arise when various types of data are presented independently in one article, and thus it is important for the logging software to compile the various parts of one HTML document as one journal article, or one statistical value. For example, the most popular paper for the last three years in *First Monday* is "Free Riding on Gnutella" written by Eytan Adar and Bernardo A. Huberman of the Xerox Palo Alto Research Center in Palo Alto, Calif. It appeared in the October 2000 issue of *First Monday* along with seven other papers and book reviews. The paper contains text, graphics as part of the *First Monday* formatting, and graphics (graphs) that illustrate points made in the paper. There are, in this particular case, ten graphics that are part of the

paper for stylistic reasons, and a separate seven graphics that are graphs as figures in the paper. Each illustration – stylistic or scientific – generates a "hit" in the server log, as well as a "hit" for the text of the paper. Log analysis software used in the course of this project synthesizes all of those "hits" into a single value for the paper, as summarized in statistical reports under the URL, http://firstmonday.org/issues/issue5_10/adar/ [viewed November 15, 2002].

For this analysis, once a paper has been downloaded, no efforts were made to track subsequent use of that article. However, with the logs, we have an exact value that we can attribute to readers of *First Monday* who examined specific papers from the Web site locally on their computers. It is important to note that we assume someone does not look at a paper without reading and/or printing the entire document or some portion of it. In future studies on the use of *First Monday*, we will need to survey *First Monday*'s readers more exactly to see what some sample of the readership actually does with the contents of each issue (read it online only, print, save to a local storage device,

Table 2: Most Requested Papers, *First Monday*, 1999 (based on an analysis of monthly logs, January–December 1999)

rank	Paper	Number of requests	Number of months in 1999 paper available	Average requests per month in 1999
1	Kerry Coffman and Andrew Odlyzko, "The Size and Growth Rate of the Internet", <i>First Monday</i> , volume 3, number 10 (October 1998), http://firstmonday.org/issues/issue3_10/coffman/	28,447	12	2,371
2	Nikolai Bezroukov, "Open Source Software Development as a Special Type of Academic Research (Critique of Vulgar Raymondism)", <i>First Monday</i> , volume 4, number 10 (October 1999), http://firstmonday.org/issues/issue4_10/bezroukov/	21,135	3	7,045
3	David F. Noble, "Digital Diploma Mills: The Automation of Higher Education," <i>First Monday</i> , volume 3, number 1 (January 1998), http://firstmonday.org/issues/issue3_1/noble/	14,850	12	1,237
4	Gisle Hannemyr, "Technology and Pleasure: Considering Hacking Constructive," <i>First Monday</i> , volume 4, number 2 (February 1999), http://firstmonday.org/issues/issue4_2/gisle/	12,997	11	1,182
5	John Kelsey and Bruce Schneier, "The Street Performer Protocol and Digital Copyrights," <i>First Monday</i> , volume 4, number 6 (June 1999), http://firstmonday.org/issues/issue4_6/kelsey/	10,175	7	1,454
6	Eben Moglen, "Anarchism Triumphant: Free Software and the Death of Copyright," <i>First Monday</i> , volume 4, number 8 (August 1999), http://firstmonday.org/issues/issue4_8/moglen/	7,631	5	1,526
7	Jeannette Allis Bastian, "Filtering the Internet in American Public Libraries: Sliding Down the Slippery Slope," <i>First Monday</i> , volume 2, number 10 (October 1997), http://firstmonday.org/issues/issue2_10/bastian/	7,584	12	632
8	Mary Minow, "Filters and the Public Library: A Legal and Policy Analysis," <i>First Monday</i> , volume 2, number 12 (December 1997), http://firstmonday.org/issues/issue2_12/minow/	7,401	12	617
9	Michael H. Goldhaber, "The Attention Economy and the Net," <i>First Monday</i> , volume 2, number 4 (April 1997), http://firstmonday.org/issues/issue2_4/goldhaber/	5,788	12	482
10	Hal R. Varian, "Differential Pricing and Efficiency," <i>First Monday</i> , volume 1, number 2 (August 1996), http://firstmonday.org/issues/issue2/different/	5,371	12	448

Table 3: Analysis of Most Frequently Requested Papers, *First Monday*, 1999

Top 10 papers requested, number published in 1999: 4

Top 10 papers requested, number published in 1998: 2

Top 10 papers requested, number published in 1997: 3

Top 10 papers requested, number published in 1996: 1

Total number of requests for Top 10 requested papers published in 1999: 121,379

Total number of requests for 1999 papers in Top 10 for 1999: 51,938

Percentage of total: 42.8

Total number of requests for 1998 papers in Top 10 for 1999: 43,297

Percentage of total: 35.7

Total number of requests for 1997 papers in Top 10 for 1999: 20,773

Percentage of total: 17.1

Total number of requests for 1996 papers in Top 10 for 1999: 5,371

Percentage of total: 4.4

Total number of requests for Top 10 requested papers published in 1999: 121,379

Total number of requests for 1999 papers in Top 10 for 1999: 51,938

Percentage of total: 42.8

Total number of requests for 1996–1998 papers in Top 10 for 1999: 69,441

Percentage of total: 57.2

Table 4: Most Requested Papers, *First Monday*, 2000 (based on an analysis of monthly logs, January–December 2000)

rank	Paper	Number of requests	Number of months in 2000 paper available	Average requests per month in 2000
1	Craig H. Rowland, "Covert Channels in the TCP/IP Protocol Suite", <i>First Monday</i> , volume 2, number 5 (May 1997), http://firstmonday.org/issues/issue2_5/rowland/	33,885	12	2,824
2	Eytan Adar and Bernardo A. Huberman, "Free Riding on Gnutella", <i>First Monday</i> , volume 5, number 10 (October 2000), http://firstmonday.org/issues/issue5_10/adar/	23,161	3	7,720
3	Steve Cisler, "Letter from Cambridge: Digital Nations and eDevelopment meetings," <i>First Monday</i> , volume 5, number 11 (November 2000), http://firstmonday.org/issues/issue5_11/cisler/	22,205	2	11,102
4	David F. Noble, "Digital Diploma Mills: The Automation of Higher Education," <i>First Monday</i> , volume 3, number 1 (January 1998), http://firstmonday.org/issues/issue3_1/noble/	19,924	12	1,660
5	John Kelsey and Bruce Schneier, "The Street Performer Protocol and Digital Copyrights," <i>First Monday</i> , volume 4, number 6 (June 1999), http://firstmonday.org/issues/issue4_6/kelsey/	17,435	12	1,453
6	Noriko Hara and Rob Kling, "Students' Frustrations with a Web-Based Distance Education Course," <i>First Monday</i> , volume 4, number 12 (December 1999), http://firstmonday.org/issues/issue4_12/hara/	14,560	12	1,213
7	Karen L. Murphy and Mauri P. Collins, "Communication Conventions in Instructional Electronic Chats," <i>First Monday</i> , volume 2, number 11 (November 1997), http://firstmonday.org/issues/issue2_11/murphy/	13,903	12	1,159
8	Mary Minow, "Filters and the Public Library: A Legal and Policy Analysis," <i>First Monday</i> , volume 2, number 12 (December 1997), http://firstmonday.org/issues/issue2_12/minow/	13,612	12	1,134
9	Chun Wei Choo, Brian Detlor, and Don Turnbull, "Information Seeking on the Web: An Integrated Model of Browsing and Searching," <i>First Monday</i> , volume 5, number 2 (February 2000), http://firstmonday.org/issues/issue5_2/choo/	13,390	11	1,217
10	Paul Bambury, "A Taxonomy of Internet Commerce," <i>First Monday</i> , volume 3, number 10 (October 1998), http://firstmonday.org/issues/issue3_10/bambury/	11,715	12	976

Table 5: Analysis of Most Frequently Requested Papers, *First Monday*, 2000

Top 10 papers requested, number published in 2000: 3

Top 10 papers requested, number published in 1999: 2

Top 10 papers requested, number published in 1998: 2

Top 10 papers requested, number published in 1997: 3

Top 10 papers requested, number published in 1996: 0

Total number of requests for Top 10 requested papers published in 2000: 183,790

Total number of requests for 2000 papers in Top 10 for 2000: 58,756

Percentage of total: 32.0

Total number of requests for 1999 papers in Top 10 for 2000: 31,995

Percentage of total: 17.4

Total number of requests for 1998 papers in Top 10 for 2000: 31,639

Percentage of total: 17.2

Total number of requests for 1997 papers in Top 10 for 2000: 61,400

Percentage of total: 33.4

Total number of requests for Top 10 requested papers published in 2000: 183,790

Total number of requests for 2000 papers in Top 10 for 2000: 58,756

Percentage of total: 32.0

Total number of requests for 1997–1999 papers in Top 10 for 2000: 125,034

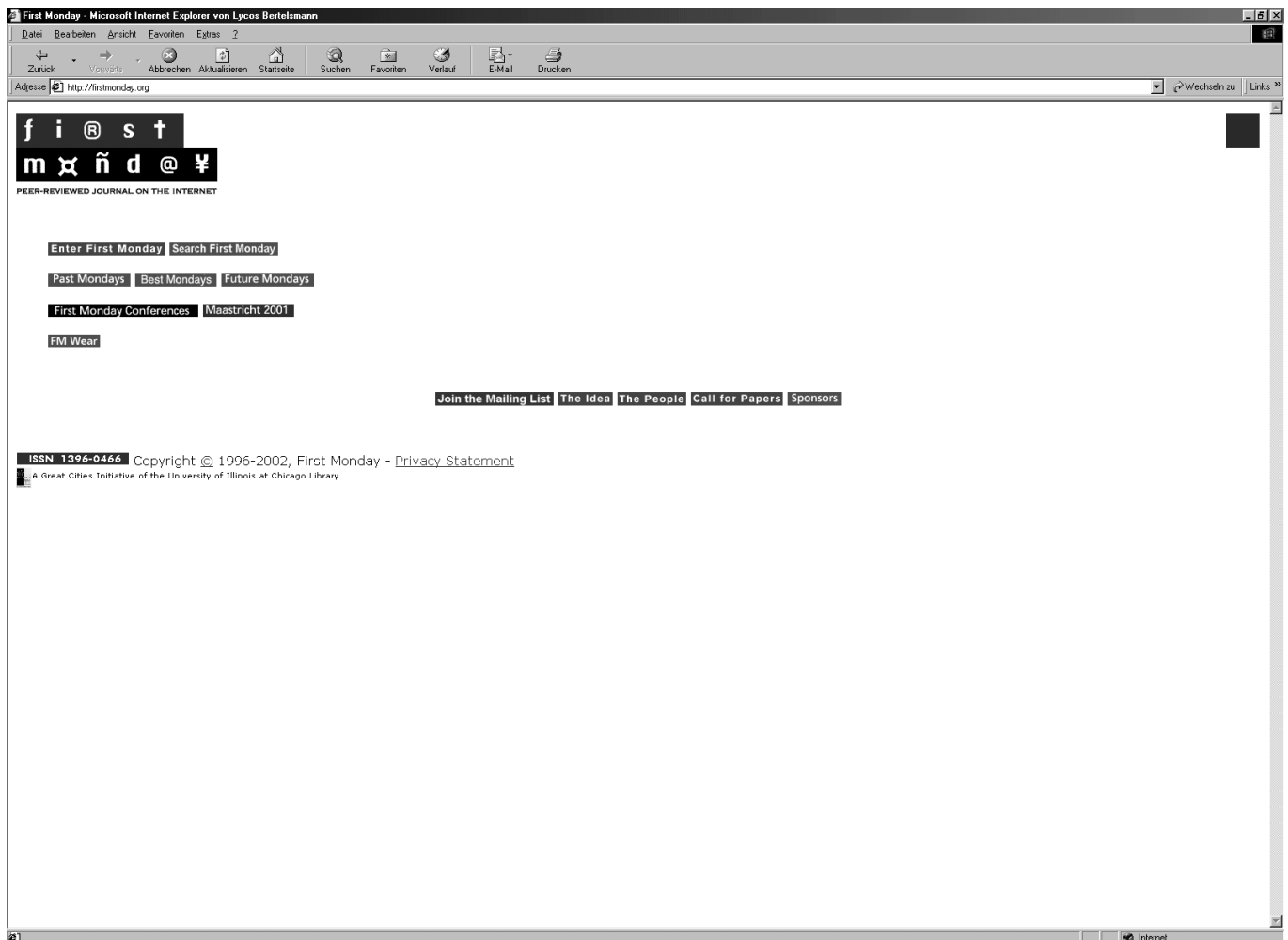
Percentage of total: 68.0

Table 6: Most Requested Papers, *First Monday*, 2001 (based on an analysis of monthly logs, January–December 2001)

rank	Paper	Number of requests	Number of months in 2001 paper available	Average requests per month in 2001
1	Eytan Adar and Bernardo A. Huberman, "Free Riding on Gnutella", <i>First Monday</i> , volume 5, number 10 (October 2000), http://firstmonday.org/issues/issue5_10/adar/	59,235	12	4,936
2	Richard W. Wiggins, "The Effects of September 11 on the Leading Search Engine", <i>First Monday</i> , volume 6, number 10 (October 2001), http://firstmonday.org/issues/issue6_10/wiggins/	31,078	3	10,359
3	Andrew Odlyzko, "Content is Not King," <i>First Monday</i> , volume 6, number 2 (February 2001), http://firstmonday.org/issues/issue6_2/odlyzko/	30,161	11	2,742
4	Clifford Lynch, "The Battle to Define the Future of the Book in the Digital World," <i>First Monday</i> , volume 6, number 6 (June 2001), http://firstmonday.org/issues/issue6_6/lynch/	29,295	7	4,185
5	Roberto Zamparelli, "Copyright and Global Libraries: Going with the Flow of Technology," <i>First Monday</i> , volume 2, number 11 (November 1997), http://firstmonday.org/issues/issue2_11/zamparelli/	23,919	12	1,993
6	David F. Noble, "Digital Diploma Mills: The Automation of Higher Education," <i>First Monday</i> , volume 3, number 1 (January 1998), http://firstmonday.org/issues/issue3_1/noble/	23,348	12	1,947
7	Craig H. Rowland, "Covert Channels in the TCP/IP Protocol Suite", <i>First Monday</i> , volume 2, number 5 (May 1997), http://firstmonday.org/issues/issue2_5/rowland/	21,795	12	1,816
8	Karen L. Murphy and Mauri P. Collins, "Communication Conventions in Instructional Electronic Chats," <i>First Monday</i> , volume 2, number 11 (November 1997), http://firstmonday.org/issues/issue2_11/murphy/	17,543	12	1,462
9	George N. Dafermos, "Management and Virtual Decentralised Networks: The Linux Project," <i>First Monday</i> , volume 6, number 11 (November 2001), http://firstmonday.org/issues/issue6_11/dafermos/	16,462	2	8,231
10	Noriko Hara and Rob Kling, "Students' Frustrations with a Web-Based Distance Education Course," <i>First Monday</i> , volume 4, number 12 (December 1999), http://firstmonday.org/issues/issue4_12/hara/	15,607	12	1,301

Table 7: Analysis of Most Frequently Requested Papers, *First Monday*, 2001

Top 10 papers requested, number published in 2001: 4
 Top 10 papers requested, number published in 2000: 1
 Top 10 papers requested, number published in 1999: 1
 Top 10 papers requested, number published in 1998: 1
 Top 10 papers requested, number published in 1997: 3
 Top 10 papers requested, number published in 1996: 0
 Total number of requests for Top 10 requested papers published in 2001: 268,443
 Total number of requests for 2001 papers in Top 10 for 2001: 106,996
 Percentage of total: 40.0
 Total number of requests for 2000 papers in Top 10 for 2001: 59,235
 Percentage of total: 22.1
 Total number of requests for 1999 papers in Top 10 for 2001: 15,607
 Percentage of total: 5.6
 Total number of requests for 1998 papers in Top 10 for 2001: 23,348
 Percentage of total: 8.7
 Total number of requests for 1997 papers in Top 10 for 2001: 63,257
 Percentage of total: 23.6
 Total number of requests for Top 10 requested papers published in 2001: 268,443
 Total number of requests for 2001 papers in Top 10 for 2001: 106,996
 Percentage of total: 40.0
 Total number of requests for 1996–2000 papers in Top 10 for 2001: 161,447
 Percentage of total: 60.0

Figure 1. *First Monday's* home page (<http://firstmonday.org> – November 28, 2002)

or some combination of all three options). This caveat notwithstanding, we believe that each individual access of a journal article is a valid indication of the circulation of the ideas presented in that article.

Current issues and back files

Tables 2–7 examine the use of *First Monday's* papers in the years 1999, 2000 and 2001. Table 2 provides a view of the ten most downloaded papers in 1999, Table 4 lists the ten most downloaded papers in 2000 and Table 6 lists the ten most downloaded papers in 2001. Keep in mind that all papers published in *First Monday*, from its first issue in May 1996 to the most recent issue, are equally accessible from the *First Monday* server. Of immediate importance to *First Monday* and its readership is access to papers in its archives. For any given year examined in this study, papers in

First Monday's archives attracted more traffic than current content (that is, published within the most recent calendar and volume year). In 1999, 42.8 percent of all traffic to issues was aimed at the current year, while 57.2 percent was directed to papers published in 1996–1998, or the back file (see Table 3). In 2000, for volume 5, only 32% of all traffic was for the current year, while 68 percent was directed to content published earlier (see Table 5). In 2001, for volume 6, 40 percent of the total requests for papers in the “Top 10” (or most read) were for papers published in 2001; 60 percent of the traffic was for papers published earlier (see Table 7).

It is important to remember this highly distributed nature of *First Monday's* Web site traffic. The design of a Web journal's opening page can invite immediate use of the journal's entire back file of articles. Because all of the previous issues of *First Monday* are as accessible as the most current issue (see Figure 1), it should not be a great

surprise that readers of *First Monday* frequently turn to papers in the archives. Indeed, past publications are often accessed more frequently than papers in the current issue.

Frequency of access

Server logs indicate this flexibility of access; all papers published in *First Monday* and subsequently available on the server are examined and downloaded. Some papers, however, are more frequently downloaded than others. With print journals, a citation pattern over time emerges as some papers are found to be more utilitarian than others. With *First Monday*, a more immediate measure of utility is made possible through the server logs' data of downloading rates.

The definition of a "classic" article

In traditional print journals, a "citation classic" is defined as a highly cited publication, with "highly cited" defined by ISI's Science Citation Index (SCI) statistical measure. A "classic" may vary from discipline to discipline, as citation patterns vary; what may be a classic in terms of sheer citation numbers may be some large value (approximately 400) in one discipline, and a much lower value (50) in another, emerging, or more tightly defined field of study.

We propose that with Web-based journals, aggregate quantitative data from the server logs provides opportunity for a redefinition of the classic scholarly article: a paper that is frequently downloaded over time. In this new definition, statistical analysis of access through Web logs replaces citation indexing as the numerical basis for evaluation of an article's usage in the academic community. As previously stated, we posit that circulation of an article's ideas through each electronic access is a valid contemporary measure of that idea's scholarly impact.

"Frequency of use over time" as a designator of value can be further refined by examining the impact level of the Web journal where the article was published. For example, analysis of Tables 2, 4, and 6 reveal that in *First Monday*, the number of hits required for an individual article to perform in the Top 10 for the year triples from 1999–2001. In 1999, the tenth most frequently cited article received 5,371 hits over the course of the

year. In 2000, the tenth most frequently cited article received 11,715 hits, and in 2001, to make the Top 10 an article had to receive 15,607 hits. This dramatic growth in overall journal readership demonstrates the success of the journal, but also reveals the difficulty of assigning any standard numeric quotient in our formula toward "classic" designation. Therefore, an additional criterion may be added: the overall Web usage level of the journal itself.

Garfield uses the term "high impact journal" to identify a publication that consistently produces a high rate of cited articles. On the World Wide Web, an indicator of the level of use of any given Web journal is the number of times that journal's URL appears in other pages; i.e., the number of hyperlinks to that journal. A standard Web search tool such as Google can be used to identify the number of hyperlinks to a specific Web journal. With Google (<http://www.google.com>) a request for "link:firstmonday.org" provides an immediate measure of the number of active hyperlinks to *First Monday*. *First Monday*, as of July 2002, was linked from 2,800 Web sites, according to Google. Hence further research will be needed to determine if that value makes *First Monday* a "high impact journal" or not, by comparing it to other journals and magazines on the Web.

Further illustration and discussion of the model

Using *First Monday* server records, we can illustrate the designation of a classic article by segregating, through statistical analysis, papers selected from the *First Monday* server repeatedly over the course of two or more years to the extent that they appear in the Top 10 most downloaded papers year after year. Table 1 shows *First Monday*'s most downloaded papers for the years 1999–2001. The most popular paper, "Free Riding on Gnutella," by Eytan Adar and Bernardo Huberman, was requested on 82,396 occasions, or 5,493/month, since its first appearance in October 2000.

The average monthly values for each year – as recorded in Tables 2 (for 1999), Table 4 (2000), and Table 6 (2001) – demonstrate that some papers indeed experience a brief stint of popularity (thrusting them into the "Hot Paper" category of immediate high interest in the readership), only to drop off the subsequent year's list of most read papers. There are a variety of factors leading a

paper to have a rush of readers, including publicity on discussion lists, notices in print, word-of-mouth, and hyperlinks to a specific paper from other Web sites and journals. Lasting value – as we may define a “classic” – only appears in the constant use of a paper over time; in this era of rapid information dissemination, we define “frequent use over time” to indicate a period of more than twelve months. Table 1 is a list of the seven most requested papers from the *First Monday* server for the period 1999–2001, and hence provides an illustration of the model of analysis for defining a classic scholarly article in a Web-based environment.

See again for example Adar and Huberman’s article, published in October of 2000 and therefore only available for three of the 12 statistical months for that year’s annual hit rate data. However, the article was so popular that it not only made the Top 10 for the year, but also indeed placed first among 2000’s downloads. That performance would easily qualify the article as a Hot Paper. This paper was subsequently transformed from Hot Paper to Classic through the “over time” nature of its download record, where as Table 1 indicates, the paper continues to be the most popular *First Monday* download two years after its original publication.

Questions for further study

It is important in defining the use of an idea presented in an academic article to determine the effects of that idea on subsequent research. By defining a classic article as one that is frequently accessed over time, and with the assumption that the circulation of an idea is a valid measure of its academic impact, gaps still exist in our knowledge about how that information is manifested in subsequent academic discourse. In addition, what quantitative impact does the listing of an important article within an electronic journal’s monthly table of contents have on the readership rate of other articles in that same issue, and thus on the circulation of those ideas? What impact is there on the readership of other articles in the entire journal, based on the existence of an efficient and clearly visible internal search tool? What patterns emerge in usage over time, and how can those patterns inform the study of search behaviors? Finally, further research needs to follow regard-

ing the change in academic reward structures for print vs. electronic faculty publication records. Qualitative as well as quantitative study can help explore these questions and are tools for further investigation. These studies can best take place with the full participation of the academic community, including the use of statistical data from Web journals and their logs. Open sharing of these logs is necessary to further test the definitions and models proposed here.

Conclusion

As larger numbers of scholars move to Internet-only publications to distribute their work, new measures will be needed to test the use of these papers over time. In traditional peer-reviewed publications, citation analysis has been used to evaluate scholarship, providing a measure of productivity for academic research (Adam 2002; Jimenez-Contreras et al. 2002). The World Wide Web provides rich opportunities to measure the use of scholarship in ways that simply are not possible with traditional print publications, and examples of consistently high performance articles from *First Monday* illustrate a new model for designation of a classic article. Still needed for a more transferable model is a standard numeric indicator of a paper’s success based on comparative Web journal usage data across disciplines. Opportunities for further data analysis should be shared in the academic community and should be used to the best advantage of scholars as we seek to understand trends in research and scholarly publication.

References

- Adam, D. 2002. The Counting House. *Nature* 415: 726–729.
- First Monday* 2002. *First Monday* Basics. URL: <http://www.firstmonday.org/idea.html> [viewed November 15, 2002].
- Garfield, E. 1972. Citation Analysis as a Tool in Journal Evaluation. *Science*, New Series, 178: 471–479.
- Garfield, E. 1984. The 100 Most-Cited Papers Ever and How We Select *Citation Classics*. *Current Comments*, 23: 3–9.
- Garfield, E. 2002. Fast-Breaking Hot Papers. *The Scientist*. Available at URL: http://www.the-scientist.com/yr2002/apr/comm_020415.html [viewed November 15, 2002].

- Google 2002. Our Search: Google Technology. Available at URL: <http://www.google.com/technology/index.html> [viewed November 15, 2002].
- Harter, S. P. 1996. The Impact of Electronic Journals on Scholarly Communication: A Citation Analysis. *Public-Access Computer Systems Review* 7. Available at <http://info.lib.uh.edu/pr/v7/n5/hart7n5.html> [viewed November 15, 2002].
- Isinet.com 2002. History of Citation Indexing. Available at <http://www.isinet.com/isi/hot/essays/citationindexing/21.html> [viewed November 15, 2002].
- Jimenez-Contreras, E., Lopez-Cozar, E. D., Ruiz-Perez, R. and Fernandez, V. M. 2002. Impact-factor rewards affect Spanish research. *Nature* 417: 898.
- Merton, R. K. 1983. Foreword. In: Eugene Garfield. *Citation Indexing – Its Theory and Application in Science, Technology, and Humanities*. Philadelphia: ISI Press.
- Thelwall, M. 2001. Web Log File Analysis: Backlinks and Queries. *ASLIB Proceedings* 53: 217–223.

Editorial history:

Paper received 24 July 2002;

Final version received 3 September 2002;

Accepted 5 September 2002.