WILEY InterScience®
DISCOVER SOMETHING GREAT

Usage Impact Factor: The Effects of Sample Characteristics on Usage-Based Impact Metrics

Johan Bollen and Herbert Van de Sompel
Digital Library Research and Prototyping Team, Los Alamos National Laboratory, Los Alamos, New Mexico 87545. E-mail: {jbollen, herbertv}@lanl.gov

There exist ample demonstrations that indicators of scholarly impact analogous to the citation-based ISI Impact Factor can be derived from usage data; however, so far, usage can practically be recorded only at the level of distinct information services. This leads to community-specific assessments of scholarly impact that are difficult to generalize to the global scholarly community. In contrast, the ISI Impact Factor is based on citation data and thereby represents the global community of scholarly authors. The objective of this study is to examine the effects of community characteristics on assessments of scholarly impact from usage. We define a journal Usage Impact Factor that mimics the definition of the Thomson Scientific ISI Impact Factor. Usage Impact Factor rankings are calculated on the basis of a large-scale usage dataset recorded by the linking servers of the California State University system from 2003 to 2005. The resulting journal rankings are then compared to the Thomson Scientific ISI Impact Factor that is used as a reference indicator of general impact. Our results indicate that the particular scientific and demographic characteristics of a discipline have a strong effect on resulting usage-based assessments of scholarly impact. In particular, we observed that as the number of graduate students and faculty increases in a particular discipline, Usage Impact Factor rankings will converge more strongly with the ISI Impact Factor.

Introduction

Numerous publications [Medicine (Eds.) PLoS, 2006; Monastersky, 2005; Nature, 2005; Opthof, 1997; Reedijk, 1998; Weingart, 2005] have lamented the growing and often indiscriminate use of the Thomson Scientific ISI Impact Factor (ISI IF) as an indicator of journal impact. Despite these objections, the ISI IF and the citation data on which it is based continue to enjoy widespread acceptance. One could attribute this to two factors. First, the intuitive definition of the ISI IF (i.e., as a mean citation rate per article) facilitates its general acceptance among the various technical and nontechnical communities involved in the assessment of scholarly impact. Second, the fact that citation data are derived from the body of peer-reviewed literature may create the perception that it is officially sanctioned, accurate, and representative.

The ISI IF's basis in citation statistics, however, introduces a number of significant problems regarding its accuracy and representativeness. First, publication delays cause citation data to lag scholarly developments by a significant period of time (Luwel & Moed, 1998; Rinia, Leeuwen, Bruins, Vuren, & Raan, 2001). Second, citation data are largely limited to scholarly journal articles, which excludes the growing body of scholarly artifacts emerging from e-science and electronic publishing. Third, the absence of any machine-readable representation of author motivations makes it difficult to reconstruct the context in which a citation was put in place (e.g., disagreement, agreement, or the desire to reference background knowledge). Fourth, since citations are mostly recorded for journal articles, they pertain to a community consisting of those who author journal articles. This is a small subset of the scholarly community that may exclude (nonpublishing) practitioners, students, and developers.

Efforts therefore have been made to expand the community and the set of scholarly communication items to which assessments of scholarly impact can be applied. Usage data have in that context often been pointed to as a promising alternative to citation data because data can in principle be recorded for any member of the scholarly community and any type of communication item. However, due to technological limitations, the recording of usage data has in the past been limited to onsite library usage of printed matter. For example, Scales (1976), Galvin and Kent (1977), and more recently King, Tenopir, and Clarke (2006) operationalized usage as reshelving and circulation statistics. Similarly, Tsay (1998) determined journal usage from reshelving rates and found statistically significant correlations between journal usage and citation impact rankings.

The various operationalizations of usage in these studies were based on circulation data collected for aggregated,

Received November 19, 2006; revised April 29, 2007, August 2, 2007; accepted August 2, 2007

 • Published online 10 December 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20746

printed matter (i.e., serials and books) within the framework of a particular library and its users. This approach has the following disadvantages. First, given the tremendous growth of online usage, physical library visitors may no longer be representative of the larger community of library users. Second, users can take journals and books from a shelf for any number of reasons (e.g., by mistake and serendipitous curiosity). Reshelving and circulation statistics record no information on user interests and motivations. Finally, usage of individual articles cannot be recorded from reshelving and circulation statistics. The mentioned studies thus had to focus on aggregate usage statistics for serials and books.

As usage of digital information services increasingly has come to replace the use of printed materials (Groote & Dorsch, 2001), studies of usage data started to shift their focus to online usage. This brought the following opportunities. First, online usage data can be recorded at a very large scale for a community of users that extends well beyond those who physically visit a library. Second, most online information services record particular user-request types, thereby allowing different aspects of user interest to be disambiguated. For example, a user can explicitly select a "View abstract" or "Download Full-Text Article" option from an online information service. Both types of requests thus can be separately logged. Third, online usage data can be recorded for items at nearly any level of granularity (i.e., books, journals, individual articles, and even their contents). Fourth, online usage can be recorded and analyzed immediately after articles are published online, thereby greatly reducing the effects of publication delays in conjunction with library acquisition and shelving delays.

Usage of scholarly resources as recorded by digital information services has therefore been gaining acceptance as a means to study the scholarly community. Usage data have been used to study trends in science (Bollen, Luce, Vemulapalli, & Xu, 2003) as well as to map the interests of certain subsets of the scholarly community (Bollen & Van de Sompel, 2006a). In addition, usage data have been shown to be a promising alternative to citation data in the assessment of scholarly impact. As early as 2002, Darmoni, Roussel, Benichou, Thirion, and Pinhas (2002) proposed a reading factor to rank journals according to their impact derived from a library's access statistics. Bollen and Luce (2002) and Bollen, Van de Sompel, Smith, and Luce (2005) proposed the use of social network metrics calculated for journal networks derived from usage sequences in a library's access log. Kurtz et al. (2004a, 2004b) discussed the potential of usage data for impact ranking. Brody, Harnad, and Carr (2006) later explored how early article usage statistics can predict citation rates. Shepherd (2007) investigated the feasibility of "journal usage factors." In addition to these research developments, the COUNTER[1] project (Shepherd, 2004) and the SUSHI[2] project (Chandler & Jewell, 2006) are proposing practical standards for reporting and transmitting usage statistics recorded by scholarly publishers. Thomson Scientific has recently started to include usage statistics in its ISI *Web of Knowledge* product.[3]

However, a number of significant problems must be juxtaposed to the many advantages offered by online usage data. First, online usage can be biased by a myriad of interface characteristics (Davis & Price, 2006). This problem is compounded by the fact that digital information services can record any variety of usage types that may each indicate different levels and modalities of user interest. Fortunately, most modern digital information services exist to accommodate similar user needs and are therefore designed to respond to similar types of user requests; therefore, it may in principle be possible to standardize a common set of request types that are deemed to most strongly and consistently express user interest (e.g., full-text article downloads). The U.S. National Information Standards Organization (NISO) has initiated several efforts that will explore this issue in more detail and that may lead to the development of a community standard regarding which request types best express certain aspects of usage. The second and arguably most illusive issue is that of sample characteristics. The community for which usage has been recorded by a digital information service may have a significant effect on the outcomes of a subsequent analysis. This problem reduces the generalizability of usage-based assessments of scholarly impact and hampers their acceptance as a valid complement to citation analysis. No feasible solutions have thus far been proposed for this problem, and furthermore, it is poorly understood.

The issue of sample characteristics cannot be avoided when dealing with usage data acquired by digital information services. When Bollen and Luce (2002) ranked journals according to their usage, this was done on the basis of usage data recorded by the Los Alamos National Laboratory (LANL) Research Library servers and therefore reflects the preferences of the LANL community. In a similar manner, the results reported by Brody et al. (2006) apply to the user community of the UK arXiv mirror.[4] A similar argument can be made for the "citation-download correlation tool" of the University of Southampton's CiteBase system,[5] which uses download information from the U.K. arXiv mirror. In all cases, the community for which usage is recorded is limited to those who accessed (or were allowed to access) that particular information service. Additionally, in nearly all cases, the characteristics of the sample of the scholarly community for which usage data have been recorded is unknown in terms of its diversity and its span. The resulting usage data and its subsequent analysis therefore could be shaped by a set of sample characteristics that are not well understood (Franklin & Plum, 2004; Luther, 2006). For example, The CiteBase user community could in terms of its diversity be an unknown and highly biased mix of undergraduate students, professors, university staff, laypersons, and scholars. Its span

[1]http://www.projectcounter.org/
[2]http://www.niso.org/committees/SUSHI/SUSHI_comm.html
[3]ISI *Web of Knowledge* Usage Reporting System (WURS).
[4]http://uk.arxiv.org/
[5]http://www.citebase.org/

DOI: 10.1002/asi

may or may not be limited to the United Kingdom and may include many other nations. In fact, when considering usage statistics as a population statistic, the question emerges for which sample of the scholarly community usage has been recorded, and how the characteristics of that particular sample will influence the outcomes of an assessment of scholarly impact based on these statistics (Winship & Mare, 1992).

The issue of sampling permeates the field of scholarly impact assessments, even where citation data are used. Thomson Scientific's ISI IF (Garfield, 1999) is calculated from citation data (Garfield, 1979) recorded for a set of journals selected by Thomson Scientific. The corresponding sample of the scholarly community consequently has the following characteristics:

- *Span:* Extends to the *global* set of scholarly authors who published in the set of journals selected by Thomson Scientific.
- *Diversity:* Limited to scholarly *authors*, and articles published in the set of journals selected by Thomson Scientific.

Despite the latter limitation, the general acceptance of the ISI IF as an indicator of scholarly impact derives from the perception that the sample on which it is based is representative and respected.

In comparison to the ISI IF, usage-based assessments of scholarly impact are generally based on samples of the scholarly community with the following characteristics:

- *Span:* Delimited by the *local* boundaries of a particular information service.
- *Diversity:* Extends to *all user types* who can request *a variety of services* for *any type of scholarly communication unit*, including but not limited to journal articles, that is made accessible by the information service in question.

To realize impact measures derived from usage data that could achieve the same level of acceptance as the ISI IF, explorations need to take place along both of the aforementioned dimensions. The first dimension, *span*, entails the aggregation of usage data across a wide range of information services to create a more global, representative sample of the scholarly community (i.e., increase the sample's *span*). In fact, Bollen and Van de Sompel (2006b) proposed an architecture for the large-scale aggregation of usage data which could be employed to achieve such global samples; however, this architecture addresses only the technical issues involved in aggregating such samples; it does not address the issue of what constitutes a representative global sample or which digital information services that usage should be aggregated for. The second dimension, *diversity*, entails efforts to understand and control how community characteristics affect usage-based impact assessments, regardless of whether the sampled community is representative of the global scholarly community.

Whereas Bollen and Van de Sompel (2006b) focused on aspects of the first dimension (i.e., sample span), this article addresses the second dimension (i.e., sample diversity): How do the characteristics of the community for which usage was recorded affect usage-based assessments of impact? Usage of scholarly resources for all 23 California State University (CSU) campuses, comprising about 405,000 students and 44,000 faculty and staff, was recorded throughout the entire period of October 2003 to August 2005 by the CSU linking servers (Van de Sompel & Beit-Arie, 2001). The resulting large-scale, high-granularity usage dataset covers one of the world's largest and most diverse scholarly communities. A simple Usage Impact Factor (UIF) was defined to mimic the definition of the ISI IF and was then used to rank journals on the basis of the recorded CSU usage data. Correlations between the resulting CSU UIF and ISI IF rankings are determined for a set of scholarly disciplines demarcated by ISI journal classification codes. These correlations are then matched to the demographic features of the CSU community to yield insights into how the latter affect usage-based assessment of impact.

IFs and Data Collection

Citation IF

The IF of a particular journal in a given year (Garfield, 1979) is defined as the ratio of two quantities; namely, the number of citations received by the citable[6] articles published in the journal during the previous 2 years divided by the total number of articles published by the journal in that 2-year period. As such, the IF is the mean citation rate of articles published in a particular journal over a 2-year period.

More formally, the IF can be defined as follows. We denote the Set A of articles published in Journal j in Year y as A^y_j so that $A^y_j = \{a_1, a_2, \ldots a_n\}$, where $a_i \in A^y{}_j$ represents an article published in Journal j in Year y. We introduce the citation function C^y that maps a set of articles to the number of times these articles were cited by articles published in year y i.e., $C^y(A) \rightarrow N$ where N denotes the set of natural numbers $\{0,1,2,3,\ldots\}$. It follows that $C^y(A^k_j)$ returns the number of citations recorded in Year y that points to the set of articles published in Journal j in Year k.

The IF of a Journal j in Year y, denoted IF^y_j, is defined as the ratio of two quantities:

$$IF^y_j = \frac{C^y(A^{y-1}_j \cup A^{y-2}_j)}{|A^{y-1}_j \cup A^{y-2}_j|} \quad (1)$$

where $C^y(A^{y-1}_j \cup A^{y-2}_j)$ is the number of citations in Year y to all citable articles published in Journal j in the proceeding 2 years $y-1$ and $y-2$, and $|A^{y-1}_j \cup A^{y-2}_j|$ is the number of citable articles published by Journal j in the proceeding 2 years $y-1$ and $y-2$.

[6]Thomson Scientific, formerly ISI, who publishes the journal *Impact Factor*, makes a distinction between citable and noncitable articles (Monastersky, 2005). For reasons of comparability, any further mention of articles in the context of determining the rate at which they were cited will refer to the set of citable articles as defined by Thomson Scientific.

DOI: 10.1002/asi

UIF

A similar reasoning can be applied to the definition of a UIF, which can be framed in terms of the mean rate by which an article published in a particular journal over a 2-year period is used rather than cited. Analogous to the IF, we define the UIF of Journal j in Year y, denoted UIF_j^y, as follows. We replace the citation function $C^y(A_j^k)$ with the usage function $R^y(A_j^k)$ that returns the number of times the articles in A_j^k are used in Year y. The UIF then can be defined as the ratio between two quantities:

$$UIF_j^y = \frac{R^y(A_j^{y-1} \cup A_j^{y-2})}{|A_j^{y-1} \cup A_j^{y-2}|} \tag{2}$$

where $R^y(A_j^{y-1} \cup A_j^{y-2})$ is the number of uses recorded in Year y of articles published in Journal j in the proceeding 2 years $y-1$ and $y-2$, and $|A_j^{y-1} \cup A_j^{y-2}|$ is the number of articles published by Journal j in the proceeding 2 years $y-1$ and $y-2$.

The UIF is the mean rate by which an article published in a journal within a 2-year period is used in a particular year, much like the IF is the mean rate by which an article published in a journal within a 2-year period is cited in a particular year. The similarities between the IF and the UIF are clarified in Figure 1.

To ensure that the IF and the UIF for a particular journal are determined on the basis of similar samples, the UIF denominator can be that of the IF; namely, the number of citable items published by Journal j in Years $y-1$ and $y-2$. In other words, the number of citable or "usable" articles in a journal is then considered the same quantity for a particular year.

In this work, we use the full-text downloads of an article as an approximation of article usage. A similar problem of approximation exists in citation analysis, where author motivations to cite a particular article can vary strongly (MacRoberts & MacRoberts, 1989) and a citation can express any modality of agreement, disagreement, or interest. Contrary to citation data, which lack any formal indication of author motivation, online usage data typically do specify the type of requests issued by the user, thereby allowing a careful selection of which to consider for a particular analysis. Although still finer distinctions can be made between different types of usage (e.g., surveys to determine actual reading rates; King et al., 2006), such an investigation was beyond the scope of this study; full-text downloads were considered to be the strongest, if somewhat partial, indicator of user interest.

Data Acquisition

Sample considerations. The significance of sample span and diversity was outlined in the Introduction. Therefore, when discussing usage- or citation-based metrics of impact, two orthogonal factors need to be taken into account:

- the characteristics of the sample that the specific metric has been calculated for (i.e., sample span and diversity), and
- the formal definition of a metric as an indicator of scholarly impact.

This perspective is represented in Figure 2. The IF, as defined in Equation 1, can be calculated for any set of journal citation data; however, the most common instantiation of the IF is the one published by Thomson Scientific's ISI. This ISI IF is calculated on the basis of citation data for a core set of about 8,000 ISI-selected journals. Regarding the span of its sample, the ISI IF places no restrictions on the origin or affiliation of authors. It therefore represents a global sample of the scholarly community, albeit one whose diversity is limited by the focus on authors who published journal articles in the set of ISI-selected journals.

The IF can be calculated for local citation samples. For example, McDonald (2006) extracted citation data pertaining only to California Institute of Technology authors to determine a local perspective of citation impact. This approach results in a Local Impact Factor (LIF), as indicated in Figure 2.

The UIF, as defined in Equation 2, can in principle be calculated for any usage dataset, but the nature of usage data is such that they are generally recorded for the local user communities of a specific information service. This article

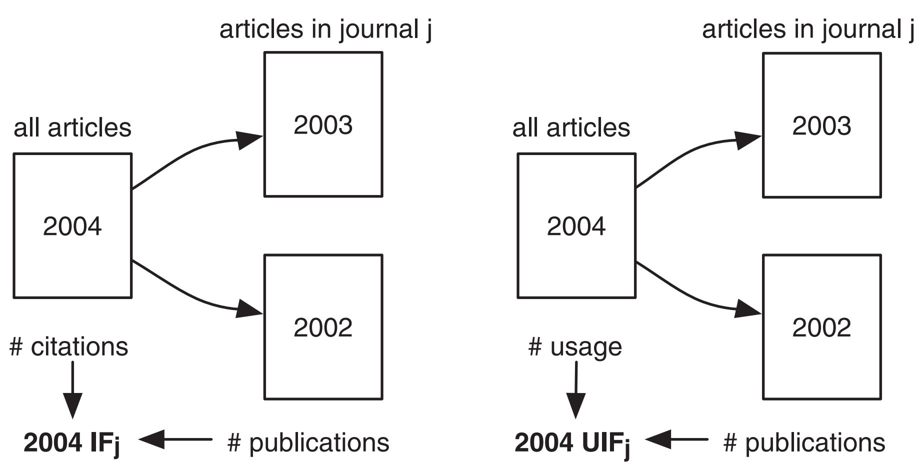

FIG. 1. Usage Impact Factor (UIF) defined in analogy to the ISI Impact Factor (ISI IF).

DOI: 10.1002/asi

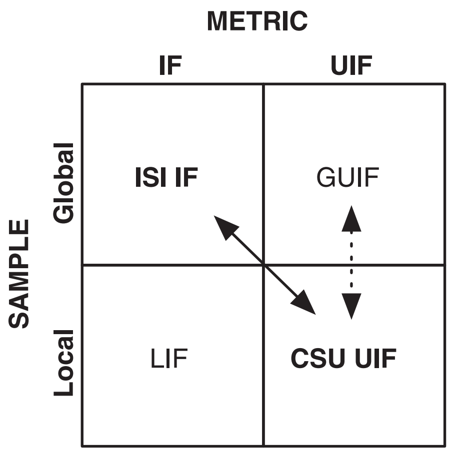

FIG. 2. Two orthogonal factors: formal metric definition and the sample to which the metric has been applied.

reports on UIF values calculated on the basis of usage data recorded for the CSU system, which correspond to a local, CSU-specific sample of the scholarly community. We therefore label the consequent UIF values "CSU UIF" to indicate the fact that they apply to local CSU usage.

The aggregation of usage datasets across different information services and institutions may in the future yield increasingly global samples of the scholarly community. The resulting UIF rankings would then reflect a more global rather than a local, institutional sample of the scholarly community. Such metrics are labeled Global Usage Impact Factor (GUIF) in Figure 2.

This article outlines a comparison of the globally oriented ISI IF, which is used as a baseline indicator of general impact, versus the CSU UIF, which represents a local, CSU-specific facet of scholarly impact. However, once aggregated usage data become available, a comparison between CSU UIF and the GUIF, the latter used as a global baseline, could be equally informative.

ISI IF citation data. ISI IF values were extracted from the 2004 *Journal Citation Reports* (*JCR*), which are published on a yearly basis by Thomson Scientific's ISI. Combined, the Science and Social Science edition of the 2004 *JCR* contained impact factors for 7,356 scholarly journals.

CSU UIF usage data. A large-scale usage log was created by aggregating usage data recorded by the linking servers (Van de Sompel, 1999a, 1999b; Van de Sompel & Beit-Arie, 2001) of the entire CSU system from November 2003 to August 2005. Recording started November 11, 2003 (10:44 a.m.), and continued uninterrupted until August 8, 2005 (11:43 p.m.). Linking server logs aggregate usage across all OpenURL-enabled information services, and they thereby contain records of all types of user requests, including requests for the abstract or full-text version of an article. They may additionally provide extensive usage, document, and user metadata, which allows, for example, requester and referent type to be taken into account when considering usage-based indicators of scholarly impact. As linking servers become increasingly prevalent, their importance grows among the tools by which usage of OpenURL-enabled information services can be recorded (Gallagher, Bauer, & Dollar, 2005, McDonald, 2006).

Usage for nine major institutions (i.e., Chancellor, California Polytechnic State University, CSU Los Angeles, CSU Northridge, CSU Sacramento, San Jose State University, CSU San Marcos, San Diego State University, and San Francisco State University) was retained since they had recorded usage data most consistently and reliably, and accounted for the majority of CSU linking server data. A total of 3,679,325 unique usage events was thus recorded in the resulting master log for a total of 176,575 users (identified by their IP addresses[7]), requesting services for 1,657,312 unique documents. A majority (i.e., 73%) of the requests pertained to journal articles. A range of service request types was recorded including, but not limited to, users requesting full-text downloads, requests for library holding information, requests for journal citation data, and requests for the abstract of an article.

The resulting master log was then filtered to only include events conforming to the following requirements:

- Request type is full-text article download.
- Year of download was 2004.
- Download concerned articles published in 2002 and 2003.

A total of only 140,675 usage requests remained after this filtering, which is not surprising since usage has been shown to decay rapidly with publication date (Nicholas et al., 2005) and we are examining usage pertaining only to articles published in a 2-year window. These events pertained to articles published in 6,423 unique journals. The number of full-text article downloads was tallied for each of these journals. The resulting download frequency table was then merged with the 2004 ISI IF data, resulting in a list of 3,146 journals for which download data as well as nonzero ISI IF were available. Following Equation 2, the journal download frequency values were then divided by the same number of citable articles as was used to calculate the 2004 ISI IF, resulting in a 2004 CSU UIF value in conjunction with every 2004 ISI IF value for each journal.

Results

CSU UIF Journal Rankings

Table 1 lists the 10 journals with highest 2004 CSU UIF as well as their 2004 ISI IF values. The list reveals a strong social science focus in the CSU community. The journals *Topics in Early Childhood Special Education* (*TOP EARLY CHILD SPEC*), *Hispanic Journal of Behavior Sciences* (*HISPANIC J BEHAV SCI*), *Intervention in School and Clinic* (*INTERV SCH CLIN*), and *Monographs of the Society for Research in*

[7]It is acknowledged that IP addresses do not uniquely identify individual users; however, the presented analysis relies on overall article-download frequencies and does not require unique user identification.

DOI: 10.1002/asi

TABLE 1. Journals ranked by 2004 CSU UIF and 2004 ISI IF values.

	Ordered by 2004 CSU UIF			Ordered by 2004 ISI IF		
Rank	Title	UIF04	IF04	Title	UIF04	IF04
1	*TOP EARLY CHILD SPEC*	6.759	0.862	*ANNU REV IMMUNOL*	0.059	52.431
2	*HISPANIC J BEHAV SCI*	6.720	0.500	*CA-CANCER J CLIN*	0.667	44.515
3	*INTERV SCH CLIN*	6.017	0.172	*NEW ENGL J MED*	0.262	38.570
4	*MONOGR SOC RES CHILD*	5.571	7.286	*PHYSIOL REV*	0.164	33.918
5	*J SCHOOL PSYCHOL*	5.000	1.750	*NATURE*	0.277	32.182
6	*J FAM VIOLENCE*	4.964	0.491	*SCIENCE*	0.288	31.853
7	*SEX ROLES*	4.804	0.639	*ANNU REV BIOCHEM*	0.077	31.538
8	*J YOUTH ADOLESCENCE*	4.723	0.855	*CELL*	0.002	28.389
9	*EDUC URBAN SOC*	4.653	0.224	*JAMA-J AM MED ASSOC*	1.196	24.831
10	*J AUTISM DEV DISORD*	4.513	2.128	*ANNU REV NEUROSCI*	0.048	23.143

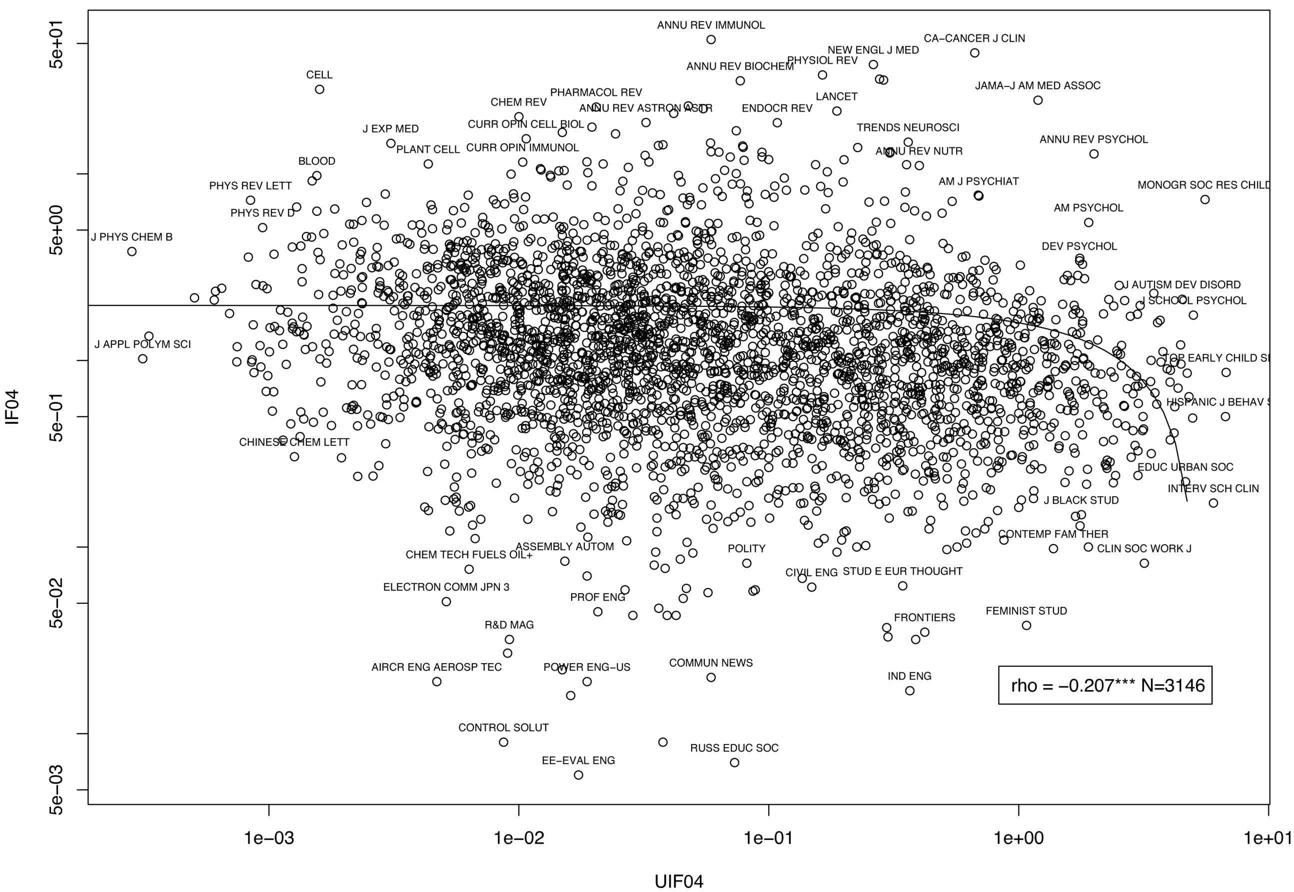

FIG. 3. CSU UIF and ISI IF values for 3,146 journals.

Child Development (*MONOGR SOC RES CHILD*) are found at the top of the list. The low 2004 ISI IF values of these journals indicate a strong discrepancy between the degree by which journals are used by the CSU community and their overall scholarly impact as indicated by the 2004 ISI IF.

The 10 journals with highest 2004 ISI IF values are listed on the right-hand side of Table 1 along with their CSU UIF values. This ISI IF ranked list contains journals with high IF rankings such as *Nature*, *Science*, *New England Journal of Medicine* (*NEW ENGL J MED*), *Cell*, and the *Journal of the American Medical Association* (*JAMA*). The corresponding 2004 CSU UIF values are relatively low for these journals despite their high 2004 ISI IF rankings.

Correlating CSU UIF and the ISI IF

The Spearman rank-order correlation coefficient, denoted ρ, between 2004 CSU UIF and 2004 ISI IF values was found to be $\rho = -0.207$ $N = 3{,}146$, $p < .001$, indicating a modest negative correlation between usage and the ISI IF for the CSU community. This negative relationship is confirmed by the log–log scaled scatterplot in Figure 3. Some of the journals

DOI: 10.1002/asi

on the extremities of the scatterplot are labeled. Note that the journals with a high ISI IF value (top of plot), regardless of their 2004 CSU UIF values, mostly correspond to medicine. In addition, a significant number of prominent physics journals (*Physical Review B* and *Physical Review Letters*) are located in the quadrant of the plot that corresponds to high ISI IF and low CSU UIF values. In other words, they are considered high impact in the general scholarly community, but their articles are used relatively infrequently in the CSU community.

This comparison of 2004 CSU UIF and 2004 ISI IF values fails to take into account variations among the different disciplines in the CSU system. A set of discipline-specific comparisons of the correlation between the 2004 CSU UIF and 2004 ISI IF is therefore provided in the following sections.

Discipline-Specific Comparisons

The scatterplot in Figure 3 suggests that the relationship between the 2004 CSU UIF and 2004 ISI IF values differ for particular disciplines (e.g., among the set of journals with high ISI IF values and low CSU UIF values, we find a preponderance of physics journals). It is therefore warranted to assess the CSU UIF and ISI IF correlations within, rather than between, individual scholarly disciplines.

The disciplines used by the CSU to tally enrollment and faculty numbers in its Statistical Abstracts (Analytic Studies Division, 2004) are the starting point of the discipline-specific comparisons of 2004 CSU UIF and 2004 ISI IF values in this article. These disciplines are listed in Table 2 (reproduced from Analytic Studies Division, 2004, p. 125, Table 81).

To separate the group of examined journals in discipline-related sets, we manually matched each of the listed CSU disciplines with a set of ISI journal classification codes.[8] These classification codes were then used to demarcate discipline-related sets of journals within which a comparison of CSU UIF and ISI IF could be conducted. The ISI journal classification codes for the CSU disciplines listed in Table 2 are provided in Table 7 (Appendix). The 2004 CSU UIF and 2004 ISI IF correlations calculated for each of the thus demarcated CSU disciplines are listed in Table 3. Statistically significant correlations, marked in bold font, were found for only 3 of the 17 disciplines; namely, Interdisciplinary Studies ($\rho = -0.470$, $N = 89$, $p < .001$), Education ($\rho = 0.228$, $N = 127$, $p = .010$), and Engineering ($\rho = -0.147$, $N = 259$, $p = .018$). Physical Sciences was found to have a marginally significant, negative correlation ($\rho = -0.225$, $N = 56$, $p = .096$). Log–log scaled scatterplots of the 2004 CSU UIF versus 2004 ISI IF values for the mentioned four disciplines are shown in Figure 4 and confirm the reported correlations.

It is of particular interest that three of the four mentioned disciplines exhibit a negative correlation between 2004 CSU UIF and 2004 ISI IF values. Whereas a zero correlation would have indicated the absence of a relationship, in this case the two metrics are inversely correlated, indicating that members of the communities interested in the particular discipline specifically do not frequently use articles published in high-impact journals, and vice versa. However, for Education, a significant, positive correlation was found between the 2004 CSU UIF and the 2004 ISI IF, indicating that for this particular CSU discipline, journal usage is moderately related to scholarly impact as indicated by the 2004 ISI IF.

The size of a discipline in terms of the number of journals that it comprises may affect ISI IF values. A marginally significant correlation was found between the CSU UIF and ISI IF correlation versus the number of journals in that particular discipline ($\rho = -0.459$, $N = 17$, $p = .065$). However, the correlation between CSU UIF and ISI IF values was not affected by the total number of students enrolled in a particular discipline. No statistically significant correlation was found between total student enrollment numbers and the correlation

TABLE 2. California State University disciplines used to tally enrollment and faculty numbers.

Disciplines
Agriculture and Natural Resources, Architecture and Environmental Design, Area Studies, Biological Sciences, Business and Management, Communications, Computer and Information Sciences, Education, Engineering, Fine and Applied Arts, Foreign Languages, Health Professions, Home Economics, Interdisciplinary Studies, Letters, Library Science, Mathematics, Physical Sciences, Psychology, Public Affairs, Social Sciences

TABLE 3. Discipline-specific 2004 CSU UIF and 2004 ISI IF Spearman rank-order correlations.

	2004 CSU UIF vs. 2004 ISI IF		
Discipline	ρ	N	p
Interdisciplinary Studies	−0.470	89	<.001
Education	+0.228	127	.010
Engineering	−0.147	259	.018
Physical Sciences	−0.225	56	.096
Agriculture and Natural Resources	+0.238	40	.138
Business and Management	+0.132	115	.160
Computer and Information Sciences	+0.077	155	.338
Area Studies	+0.169	27	.397
Public Affairs	−0.073	106	.455
Library	+0.126	25	.546
Psychology	+0.033	316	.556
Architecture & Environmental Design	+0.041	188	.572
Mathematics	+0.077	44	.617
Biological Sciences	−0.024	331	.669
Communications	+0.049	58	.712
Social Sciences	+0.026	59	.843
Health Professions	−0.012	126	.890

[8]This is a subjective matter; however, specific care was taken to match ISI Journal Classification Codes as literally as possible to the specific CSU disciplines.

DOI: 10.1002/asi

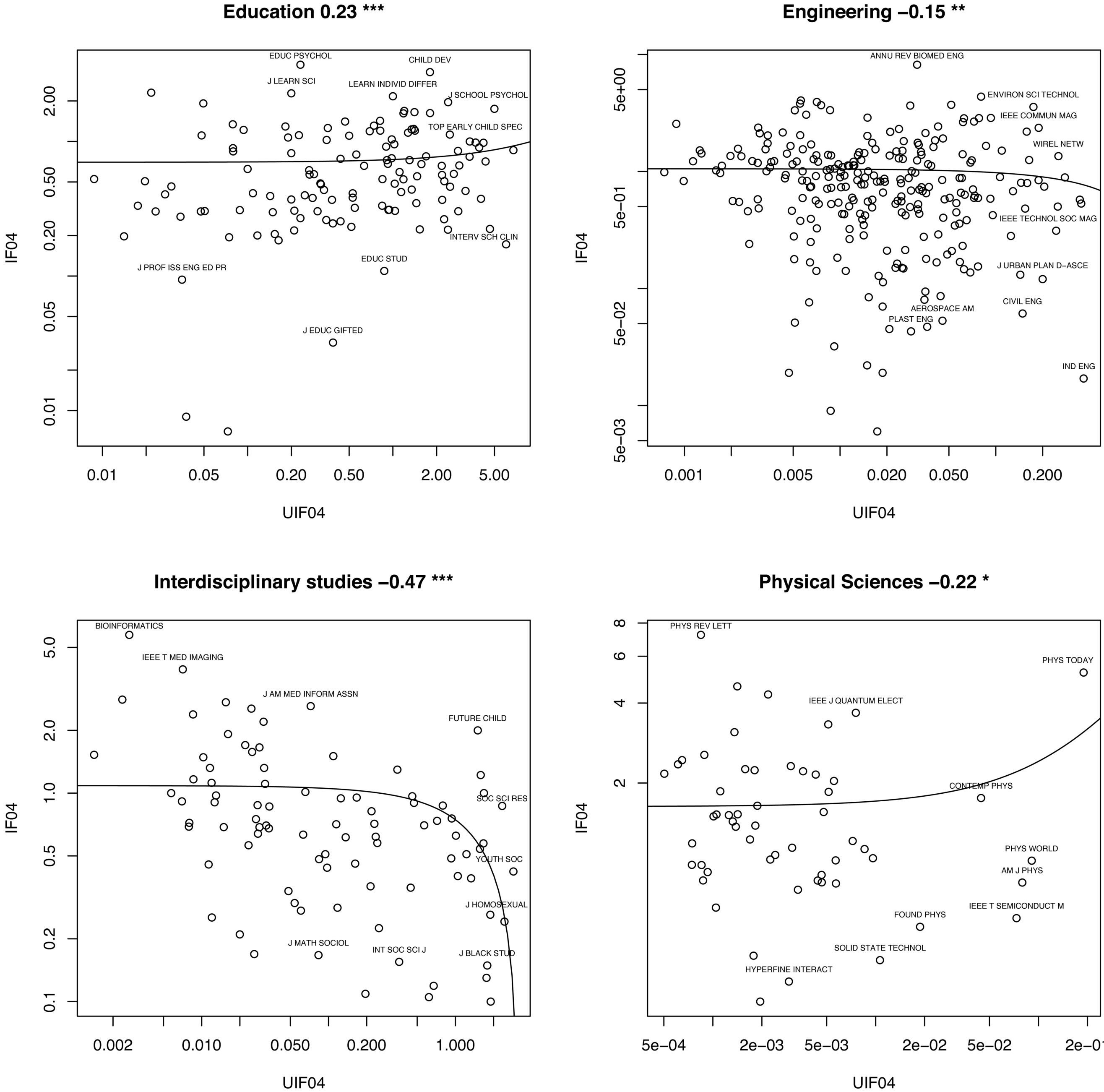

FIG. 4. CSU UIF and ISI IF comparisons for four disciplines with highest and lowest correlation values.

between CSU UIF and ISI IF correlations ($\rho = -0.262$, $N = 17$, $p = .308$).

Community Demographics

On the basis of the hypothesis that the observed correlations between CSU UIF and ISI IF values for these disciplines may be related to the academic demographics of the CSU communities corresponding to the investigated disciplines, 2004 undergraduate and graduate enrollment and faculty numbers were matched to the observed correlations. Faculty numbers are estimated in terms of Full-Time-Equivalent Faculty (FTEF) (i.e., the total number of hours taught in a particular division divided by the assumed 15 hr required for full-time faculty status). The particular number of FTEF and students, respectively, teaching or enrolled at the undergraduate or graduate level are listed in Table 4. Note that undergraduate FTEF numbers are split into low and high divisions, which need to be summed to determine total undergraduate FTEFs.

Three ratios of the size of the undergraduate versus the size of the graduate community were defined as follows. We denote the entire CSU community as the set of n individuals $V = \{v_1, v_2, \ldots, v_n\}$. We distinguish four subsets of the CSU community along two dimensions: (a) undergraduate versus

TABLE 4. California State University student enrollment and Full-Time Equivalent Faculty (FTEF) numbers (undergraduate & graduate) for 2004.

	Students		FTEF		
Discipline	U.Grad	Grad.	Low	High	Grad.
Agriculture & Natural Resources	5,381	302	62.7	127.5	21.0
Architecture & Environmental Design	2,902	358	33.9	72.1	19.2
Area Studies	319	148	12.9	25.1	4.3
Biological Sciences	13,642	1,052	243.3	264.7	89.1
Business & Management	60,069	5,242	143.3	914.4	161.3
Communications	14,252	674	139.5	299.5	31.4
Computer & Information Science	16,415	2,322	119.7	223.8	68.3
Education	16,084	15,452	49.6	750.7	836.6
Engineering	22,877	4,146	191.8	483.6	123.9
Fine & Applied Arts	19,418	1,321	425.3	712.1	102.0
Foreign Languages	2,252	486	226.2	138.5	21.2
Health Professions	13,386	3,984	31.2	142.9	143.1
Home Economics	3,261	738	29.4	93.0	16.4
Interdisciplinary Studies	29,780	948	146.6	225.5	24.8
Letters	13,594	3,413	729.6	691.6	170.9
Library	–	561	6.6	2.0	17.3
Mathematics	3,325	816	488.6	189.8	48.5
Physical Sciences	3,310	741	425.6	320.2	75.3
Psychology	16,944	1,380	84.6	332.9	108.9
Public Affairs	14,250	4,643	47.4	287.0	216.8
Social Sciences	24,597	2,956	570.4	1,081.9	162.8

graduate and (b) students versus faculty. The set of undergraduate students is denoted $V_{us} \subset V$. The set of graduate students is denoted $V_{gs} \subset V$. The set of undergraduate faculty is denoted $V_{uf} \subset V$, and finally, the set of graduate faculty is denoted $V_{gf} \subset V$. Faculty numbers are approximated from the FTEF data in Table 4.

The three defined ratios across these subsets are then formalized as follows:

$$R_a = \frac{|V_{gs} \cup V_{gf}|}{|V_{us} \cup V_{uf}|} \tag{3}$$

$$R_s = \frac{|V_{gs}|}{|V_{us}|} \tag{4}$$

$$R_f = \frac{|V_{gs}|}{|V_{us}|} \tag{5}$$

where R_a denotes the ratio of total graduate student enrollment plus graduate FTEF numbers over the total number of undergraduate student enrollment plus undergraduate (high and low divisions combined) FTEF numbers, R_s denotes the ratio of graduate over undergraduate student enrollment, and R_f denotes the ratio of graduate FTEF numbers over undergraduate FTEF numbers.

The thus defined ratios were then compared to the observed CSU UIF versus ISI IF correlations in Table 3. It must be stressed that this comparison was restricted to the mentioned four disciplines for which significant or marginally significant CSU UIF versus ISI IF correlations were observed. The results are listed in Table 5, and suggest the possibility of a relationship between the ratio of the graduate to undergraduate community within a discipline and the observed CSU UIF versus ISI IF correlations.

In particular, the discipline of Interdisciplinary Studies is characterized by an approximate ratio of 1 graduate to 15 undergraduate students (R_s = 0.067), and an approximate ratio of 1 graduate to 30 undergraduate faculty (R_f = 0.032). A highly significant, negative CSU UIF versus ISI IF correlation was observed for this discipline.

Conversely, Education is characterized by an approximate ratio of 1 graduate student to 1 undergraduate student (R_s = 1.045). The ratio of graduate faculty to undergraduate

TABLE 5. The 2004 CSU UIF and ISI IF correlations compared to ratios of faculty and student numbers.

				Graduate vs. undergraduate ratio		
Discipline	ρ(UIF, IF)	n	p	R_s	R_f	R_a
Interdisciplinary Studies	20.470	89	.000	0.067	0.032	0.032
Physical Sciences	−0.225	56	.096	0.101	0.224	0.202
Engineering	−0.147	259	.018	0.183	0.180	0.180
Education	+0.228	127	.010	1.045	0.881	0.888

DOI: 10.1002/asi

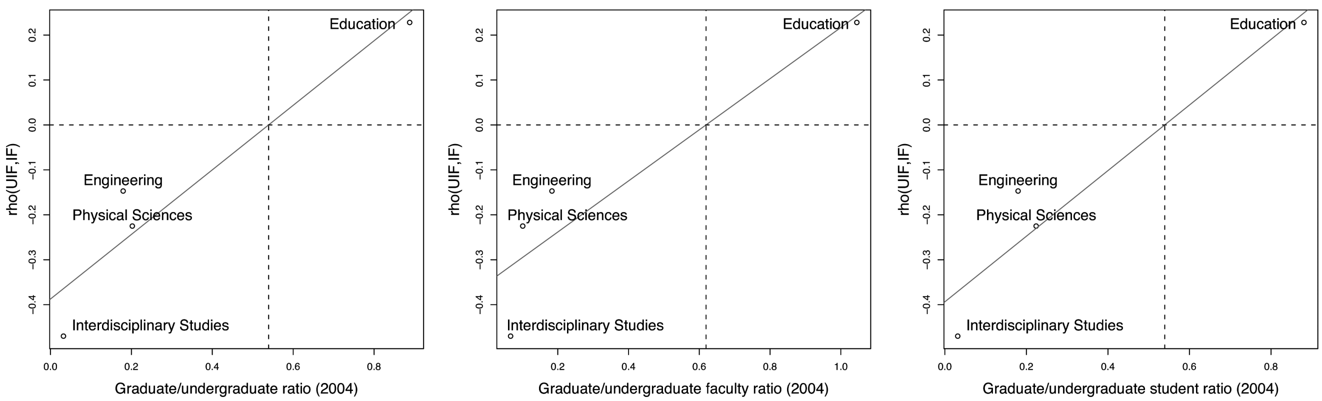

FIG. 5. Comparisons of Fall 2004 student and faculty populations versus 2004 CSU UIF versus ISI IF correlation.

TABLE 6. Spearman rank-order correlation values between 2004 Usage Impact Factor and 1997–2004 ISI ISI Impact Factors.

		ISI IF year							
		1997	1998	1999	2000	2001	2002	2003	2004
2004 CSU UIF	ρ	−0.186	−0.159	−0.170	−0.171	−0.197	−0.203	−0.204	−0.207
	n	2,636	2,750	2,819	2,892	2,960	3,050	3,096	3,146
	p	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001

faculty is of the same order ($R_f = 0.881$). A significant, positive correlation was observed between journal CSU UIF versus ISI IF values within this discipline.

This pattern is further confirmed by the undergraduate versus graduate ratios for Engineering and Physical Sciences. Engineering has a ratio of approximately 1 graduate student to 5.5 undergraduate students ($R_s = 0.183$) and a ratio of approximately 1 graduate faculty to 5.5 undergraduate faculty ($R_f = 0.180$). Physical sciences has a ratio of 1 graduate student to 10 undergraduate students ($R_s = 0.101$) and 1 graduate faculty to 4.5 graduate faculty ($R_f = 0.224$). Moderate negative CSU UIF versus ISI IF correlations were observed for both Engineering and Physical Sciences.

A linear regression model was generated for the relation between the ratio of graduate to undergraduate numbers versus the observed 2004 UIF and 2004 ISI IF correlations on the basis of the 4 data points listed in Table 5. Since similar results were obtained for all three demographic ratios ("All", "Student," and "Faculty"), only the linear regression model for the combined student and faculty ratios ("All") is discussed. Figure 5 shows a scatterplot of the mentioned values and the corresponding linear regression model. The linear regression model was found to have an intercept of -0.3873 and a slope of 0.7183 ($r^2 = 0.9029$).

From this, it could be predicted that CSU UIF versus ISI IF correlations become positive as soon as the graduate community becomes half as large as the undergraduate community in a particular discipline. Note that the overall ratio of undergraduate to graduate enrollment for the entire CSU system is approximately 6 to 1 (326,483/51,694 = 6.315), which together with the observed UIF versus ISI IF correlations of $\rho = -0.207$ ($p < .001$, $N = 3{,}146$) supports the aforementioned pattern.

Baseline Assessment

The 2004 ISI IF is used as a baseline assessment of scholarly impact against which 2004 CSU UIF values can be compared. Although CSU UIF and ISI IF are deliberately compared for the same years in which usage, citation, and publication samples were recorded, questions arise regarding the sensitivity of this comparison to longitudinal changes in the ISI IF over time.

For this reason, we investigated the degree of correlation between the 2004 CSU UIF versus past ISI IF values (i.e., ISI IF values that were published in 1997–2004).[9] The results are listed Table 6. These correlations indicate a stable, negative correlation between 2004 CSU UIF values and past ISI IF values over the mentioned period of 8 years. The absence of a particular trend in CSU UIF versus ISI IF correlations is supported by Table 6 and Figure 6. The scatterplots of CSU UIF versus ISI IF values for each specific year are shown in Figure 7.

Results Summary

The picture that emerges from these results can be summarized as follows:

[9]At the time this analysis was conducted, 2005 ISI IF values were not yet available.

DOI: 10.1002/asi

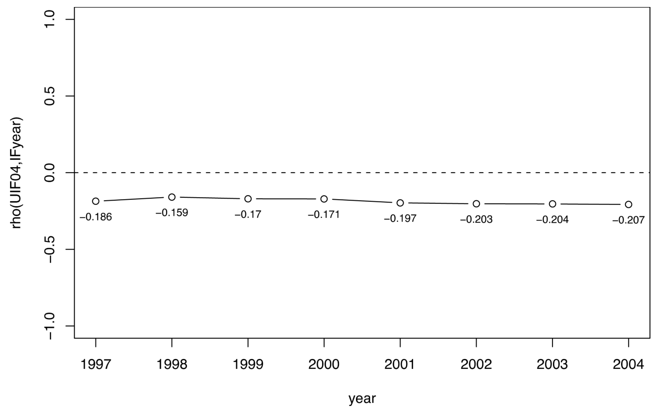

FIG. 6. Spearman rank-order correlation values between 2004 UIF and 1997–2004 ISI ISI IFs as listed in Table 6.

- A moderate negative correlation between 2004 CSU UIF and 2004 ISI IF values was found when not taking into account CSU disciplines.
- This negative correlation persists over a period of 8 years when 2004 UIF values are compared to 2004–1997 ISI IF values.
- Some CSU disciplines exhibit negative correlations between CSU UIF and ISI IF values whereas others exhibit positive correlations; however, most disciplines exhibit zero or insignificant correlations.
- CSU UIF versus ISI IF correlations seemed to be related to the ratio between the sizes of the undergraduate and graduate community in a discipline.

Discussion

Usage-based metrics of scholarly impact are gradually gaining acceptance in the domain of bibliometrics; however, little attention has been paid to how usage-based impact assessments are influenced by the demographic and scholarly characteristics of particular communities. The discussed analysis of CSU usage data indicates significant, community-based deviations between local usage impact and global citation impact as indicated by the generated CSU UIF and ISI IF rankings, respectively. In particular, we found a general negative correlation between the CSU IF and the ISI IF, which indicates usage over the entire CSU community is inversely related to general citation impact.

The observed negative correlation between the CSU UIF and the ISI IF runs counter to previous findings. Darmoni et al. (2002), Perneger (2004), and Moed (2005) all reported positive correlations between either article downloads and citations or between journal downloads and the ISI IF. Brody et al. (2006) and Bollen et al. (2005) also reported positive correlations between usage and citation rates. However, the LANL Research Library information services mostly accommodate a community of scholars in computer science and physics. In light of our results, the restricted sample may explain the observed positive correlations between usage and the ISI IF. A similar effect may occur for the U.K. arXiv mirror. Conversely, the CSU community for which usage was recorded is composed of a diverse mix of students, faculty, staff, and practitioners who are focused on a variety of science and social science domains. Related to this observation, Saha and Saint (2003) reported that the subjective journal ratings of medical researchers correlated more strongly with the ISI IF than those of medical practitioners. Both the nature of the CSU library collection and the CSU community that uses it may therefore have contributed to the negative correlation between CSU UIF and ISI IF values.

Within different scholarly disciplines, we observed both positive as well as negative CSU UIF versus ISI IF

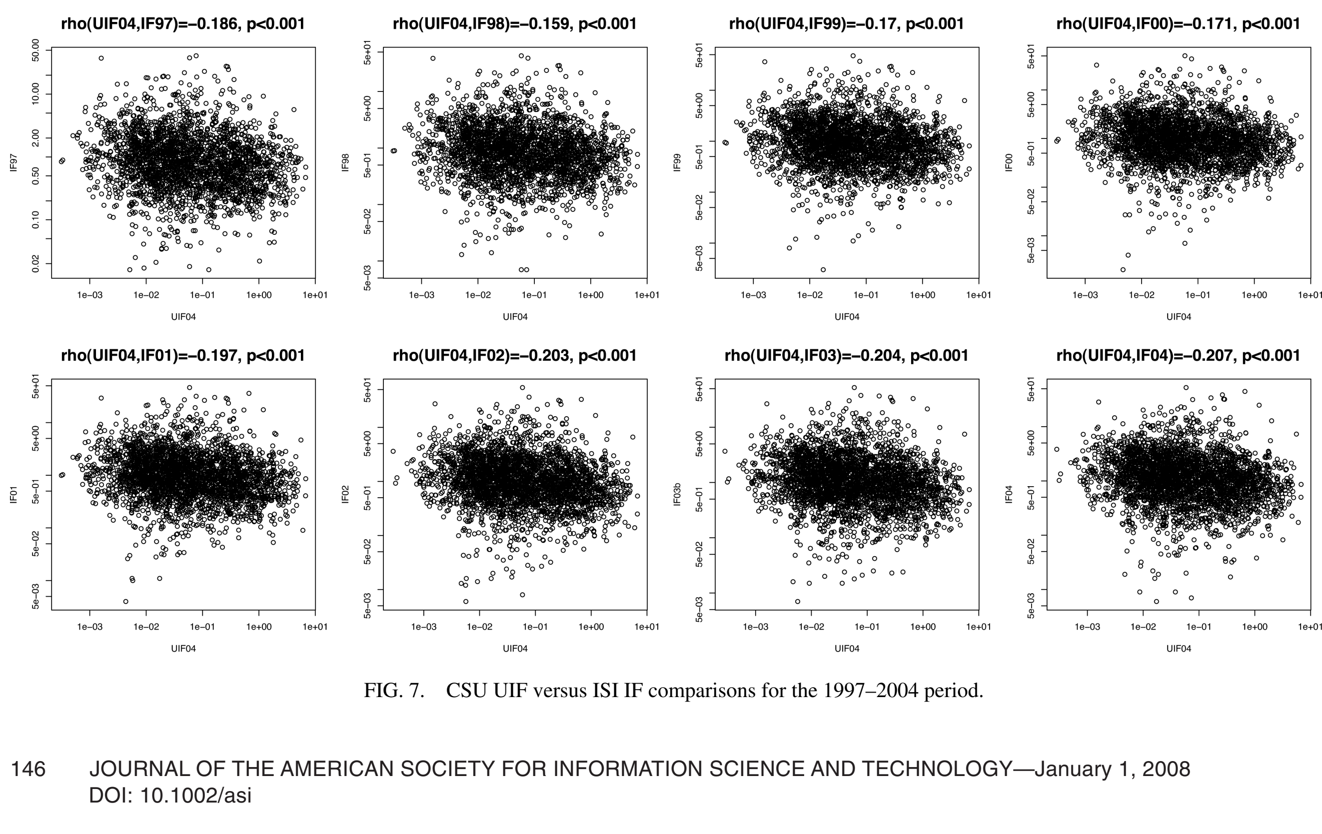

FIG. 7. CSU UIF versus ISI IF comparisons for the 1997–2004 period.

DOI: 10.1002/asi

correlations. In addition, a comparison of the relative sizes of the undergraduate and graduate communities at CSU to the correlations of CSU UIF versus ISI IF values within specific disciplines suggested that the size of the graduate community (students and faculty) relative to that of the undergraduate community within a discipline could be related to the magnitude of the observed CSU UIF versus ISI IF correlations. The tentative linear relationship that was observed between the ratio of graduate to undergraduate enrollment and CSU UIF versus ISI IF correlations raises the possibility that applications of usage data can take into account demographic data to extract different facets of impact; however, we must caution that the latter observations are based on only those four disciplines for which significant or marginally significant CSU UIF versus ISI IF correlations were observed. Future research could focus on validating these tentative results for a larger number of disciplines.

Earlier in this article, we distinguished two factors that shape metric-based assessments of scholarly impact: (a) the formal definition of a metric and (b) the sample that it has been applied to. Although the UIF has been defined to mimic the IF, the CSU UIF and ISI IF rankings in this article have been generated for very different samples of the scholarly community. The ISI IF rests on citation data collected for a set of ISI-selected journals; its rankings therefore express the global community of all scholarly authors publishing in those journals. The CSU usage data, on the other hand, reflect the characteristics of the local CSU academic community that comprises a mix of students and faculty, among others. It can therefore be considered at the same time more diverse than the ISI-defined sample in terms of its composition, yet more limited in terms of its span since it applies to CSU users only.

We envision three future paths along which usage-based metrics such as the UIF can be developed. These paths are not mutually exclusive and are related to the issues mentioned in the Introduction.

The first path is one in which attempts are undertaken to mimic the properties of the ISI IF on the basis of usage data. This requires the aggregation of a meaningful, representative sample of the scholarly community, similar in span to the ISI IF sample, and efforts to compensate for the increased diversity of the usage data sample (e.g., excluding all agents that are not scholarly authors and taking into account particular discipline-specific demographics and preferences). This article has provided an initial exploration of the second issue whereas the architecture described by Bollen and Van de Sompel (2006b) may offer at least a technical solution to the first issue, combined with efforts to standardize the various types of online usage. Questions remain as to how one can create a truly representative usage sample of the global scholarly community.

The second path along which usage-based metrics of scholarly status can be developed is focused on leveraging the greater diversity (in terms of agents and community characteristics) that usage data generally engender. This path may still require the aggregation of a meaningful, representative sample of the scholarly community, but its assessment of scholarly impact specifically leverages sample diversity to assess the many different facets of impact as they exist in the scholarly community. Indeed, one could argue that an article that is often read by a majority of students yet seldom cited by scholars in this field still can have considerable scholarly impact. In fact, on the basis of sufficiently detailed usage data, impact could be separately assessed for any subset of the scholarly community including undergraduate and graduate students, research faculty, lecturers, and the public at large.

Finally, where only local usage data are collected, there is still particular value in being able to determine local impact rankings which correspond to the preferences and characteristics of specific communities such as CSU. The CSU UIF generated in this article may not be globally applicable, but offers CSU administrators an interesting perspective on what is valued in their community. Our analysis demonstrates that considerable, yet locally meaningful, deviations can occur between impact as it is perceived by particular scholarly disciplines and the ISI IF. Such deviations are not problematic, but offer considerable possibilities to optimize local information services and adopt policies to accommodate the preferences of local communities.

Many issues remain to be addressed in future research on this topic. The Andrew W. Mellon Foundation has awarded a grant to our team to investigate a range of issues related to the definition of usage-based metrics of scholarly impact. The funded project, named MESUR,[10] aims to construct a large-scale model of the scholarly community which merges usage and bibliographic data to support the definition and validation of a range of usage-based metrics of scholarly status. This article describes our first explorations in this research area.

Acknowledgments

We thank the Andrew W. Mellon Foundation for supporting this research. We also thank Marko A. Rodriguez for proofreading the earlier versions of this article, and Joan Smith at the Department of Computer Science at Old Dominion University for producing the raw citation data on which parts of this analysis are based. Los Alamos Unlimited Release (LAUR) No: LA-UR-06–7626.

References

Analytic Studies Division. (2004). Statistical abstract 2004–2005 (Tech. Rep.). California State University. http://www.calstate.edu/as/stat_abstract/stat0405/

Bollen, J., & Luce, R. (2002). Evaluation of digital library impact and user communities by analysis of usage patterns. D-Lib Magazine, 8(6). http://www.dlib.org.

Bollen, J., Luce, R., Vemulapalli, S., & Xu, W. (2003). Detecting research trends in digital library readership. In Proceedings of the 7th European Conference on Digital Libraries (LNCS 2769) (pp. 24–28). Trondheim, Norway: Springer-Verlag.

Bollen, J., & Van de Sompel, H. (2006a). Mapping the structure of science through usage. Scientometrics, 69(2), 227–258.

[10]http://www.mesur.org/

DOI: 10.1002/asi

Bollen, J., & Van de Sompel, H. (2006b). An architecture for the aggregation and analysis of scholarly usage data. In Joint Conference on Digital Libraries (JCDL2006) (pp. 298–307). Chapel Hill, NC.

Bollen, J., Van de Sompel, H., Smith, J., & Luce, R. (2005). Toward alternative metrics of journal impact: A comparison of download and citation data. Information Processing and Management, 41(6), 1419–1440.

Brody, T., Harnad, S., & Carr, L. (2006). Earlier web usage statistics as predictors of later citation impact. Journal of the American Society for Information Science and Technology, 57(8), 1060–1072.

Chandler, A., & Jewell, T. (2006). Standards—Libraries, data providers and SUSHI: The Standardized Usage Statistics Harvesting Initiative. Against the Grain, 18(2), 82–83.

Darmoni, S.J., Roussel, F., Benichou, J., Thirion, B., & Pinhas, N. (2002). Reading factor: A new bibliometric criterion for managing digital libraries. Journal of the Medical Library Association, 90(3), 323–327.

Davis, P.M., & Price, J.S. (2006). eJournal interface can influence usage statistics: Implications for libraries, publishers, and project COUNTER. Journal of the American Society for Information Science and Technology, 57, 1243–1248.

Franklin, B., & Plum, T. (2004). Library usage patterns in the electronic information environment. Information Research, 9(4), Paper No. 187.

Gallagher, J., Bauer, K., & Dollar, D.M. (2005). Evidence-based librarianship: Utilizing data from all available sources to make judicious print cancellation decisions. Library Collections, Acquisitions and Technical Services, 29(2), 169–179.

Galvin, T.J., & Kent, A. (1977). Use of a university library collection:A progress report on a Pittsburgh study. Library Journal, 102(20), 2317–2320.

Garfield, E. (1979). Citation indexing: Its theory and application in science, technology, and humanities. New York: Wiley.

Garfield, E. (1999). Journal impact factor: A brief review. Canadian Medical Association Journal, 161, 979–980.

Groote, S.L.D., & Dorsch, J.L. (2001). Online journals: Impact on print journal usage. Bulletin of the Medical Library Association, 89(4), 372–378.

King, D.W., Tenopir, C., & Clarke, M. (2006). Measuring total reading of journal articles. D-Lib Magazine, 12(10).

Kurtz, M.J., Eichhorn, G., Accomazzi, A., Grant, C.S., Demleitner, M., & Murray, S.S. (2004a). The bibliometric properties of article readership information. Journal of the American Society for Information Science and Technology, 56(2), 111–128.

Kurtz, M.J., Eichhorn, G., Accomazzi, A., Grant, C.S., Demleitner, M., & Murray, S.S. (2004b). Worldwide use and impact of the NASA Astrophysics Data System digital library. Journal of the American Society for Information Science and Technology, 56(1), 36–45.

Luther, J. (2006). White paper on electronic journal usage statistics. The Journal of Electronic Publishing, 6(3). http://www.press.umich.edu/jep/06-03/luther.html

Luwel, M., & Moed, H.F. (1998). Publication delays in the science field and their relationship to the ageing of scientific literature. Scientometrics, 41(1–2), 29–40.

MacRoberts, M.H., & MacRoberts, B.R. (1989). Problems of citation analysis: A critical review. Journal of the American Society for Information Science, 40(5), 342–349.

McDonald, J.D. (2006). Understanding online journal usage: A statistical analysis of citation and use. Journal of the American Society for Information Science and Technology, 57(13), 39–50.

Medicine (Eds.) PLoS. (2006). The impact factor game. PLoS Med, 3(6), DOI: 10.1371/journal.pmed.003029.

Moed, H.F. (2005). Statistical relationships between downloads and citations at the level of individual documents within a single journal. Journal of the American Society for Information Science and Technology, 56(10), 1088–1097.

Monastersky, R. (2005). The number that's devouring science. Chronicle of Higher Education, 52, A12.

Nature. (2005). Editorial: Not-so-deep impact. Nature, 435, 1003–1004.

Nicholas, D., Huntington, P., Dobrowolski, T., Rowlands, I., Jamali, M.H.R., & Polydoratou, P. (2005). Revisiting obsolescence and journal article decay through usage data: An analysis of digital journal use by year of publication. Information Processing and Management, 41(6), 1441–1461.

Opthof, T. (1997). Sense and nonsense about the impact factor. Cardiovascular Research, 33, 1–7.

Perneger, T.V. (2004). Relation between online hit counts and subsequent citations: Prospective study of research papers in the BMJ. BMJ, 329(7465), 546–547.

Reedijk, J. (1998). Sense and nonsense of science citation analyses: Comments on the monopoly position of ISI and citation inaccuracies. Risks of possible misuse and biased citation and impact data. New Journal of Chemistry, 767–770.

Rinia, E.J., Leeuwen, T.N. van, Bruins, E.E.W., Vuren, H.G. van, & Raan, A.F.J. van. (2001). Citation delay in interdisciplinary knowledge exchange. Scientometrics, 51(1), 293–309.

Saha, S., & Saint, S. (2003). Impact factor: A valid measure of journal quality? Journal of the Medical Library Association, 91(1), 42–46.

Scales, P.A. (1976). Citation analysis as indicators of use of serials—Comparison of ranked title lists produced by citation counting and from use data. Journal of Documentation, 32(1), 17–25.

Shepherd, P.T. (2004). Project COUNTER—Setting international standards for online usage statistics. Journal of Information Processing and Management, 47(4), 245–257.

Shepherd, P.T. (2007). The feasibility of developing and implementing journal usage factors: A research project sponsored by UKSG. Journal for the Serials Community, 20(2), 117–123. DOI: 10.1629/20117.

Tsay, M.Y. (1998). The relationship between journal use in a medical library and citation use. Bulletin of the Medical Library Association, 86(1), 31–39.

Van de Sompel, H. (1999a). Reference linking in a hybrid library environment (Part 1). D-Lib Magazine, 5(4).

Van de Sompel, H. (1999b). Reference linking in a hybrid library environment (Part 2). D-Lib Magazine, 5(4).

Van de Sompel, H., & Beit-Arie, O. (2001). Open linking in the scholarly information environment using the OpenURL framework. D-Lib Magazine, 7(3).

Weingart, P. (2005). Impact of bibliometrics upon the science system: Inadvertent consequences? Scientometrics, 62(1), 117–131.

Winship, C., & Mare, R.D. (1992). Models of sample selection bias. Annual Review of Sociology, 18, 237–350.

Appendix

TABLE 7. ISI journal classification codes for CSU disciplines listed in Table 2.

Agriculture and Natural Resources
AD (AGRICULTURE, DAIRY ANIMAL SCIENCE), AE (AGRICULTURAL ENGINEERING), AF (AGRICULTURAL ECONOMICS POLICY), AH (AGRICULTURE, MULTIDISCIPLINARY), XE (AGRICULTURE, SOIL SCIENCE)
Architecture and Environmental Design
IH (ENGINEERING, ENVIRONMENTAL), JA (ENVIRONMENTAL SCIENCES), NE (PUBLIC, ENVIRONMENTAL OCCUPATIONAL HEALTH), JB (ENVIRONMENTAL STUDIES)
Area Studies
BM (AREA STUDIES)

Appendix (*continued*)

Biological Sciences

CQ (BIOCHEMISTRY MOLECULAR BIOLOGY, CU (BIOLOGY), DB (BIOTECHNOLOGY APPLIED MICROBIOLOGY), DR (CELL BIOLOGY), HT (EVOLUTIONARY BIOLOGY), HY (DEVELOPMENTAL BIOLOGY), PI (MARINE FRESHWATER BIOLOGY), QU (MICROBIOLOGY), WF (REPRODUCTIVE BIOLOGY), BV (PSYCHOLOGY, BIOLOGICAL)

Business and Management

DI (BUSINESS), DK (BUSINESS, FINANCE), PE (OPERATIONS RESEARCH MANAGEMENT SCIENCE), PC (MANAGEMENT)

Communications

YE (TELECOMMUNICATIONS), EU (COMMUNICATION)

Computer and Information Sciences

EP (COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE), ER (COMPUTER SCIENCE, CYBERNETICS), ES (COMPUTER SCIENCE, HARDWARE ARCHITECTURE), ET (COMPUTER SCIENCE, INFORMATION SYSTEMS), EV (COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS), EW (COMPUTER SCIENCE, SOFTWARE ENGINEERING), EX (COMPUTER SCIENCE, THEORY METHODS), ET (COMPUTER SCIENCE, INFORMATION SYSTEMS), PT (MEDICAL INFORMATICS), NU (INFORMATION SCIENCE LIBRARY SCIENCE)

Education

HB (EDUCATION, SCIENTIFIC DISCIPLINES), HA (EDUCATION EDUCATIONAL RESEARCH), HE (EDUCATION, SPECIAL), HI (PSYCHOLOGY, EDUCATIONAL)

Engineering

AE (AGRICULTURAL ENGINEERING), AI (ENGINEERING, AEROSPACE), EW (COMPUTER SCIENCE, SOFTWARE ENGINEERING), IF (ENGINEERING, MULTIDISCIPLINARY), IG (ENGINEERING, BIOMEDICAL), IH (ENGINEERING, ENVIRONMENTAL), II (ENGINEERING, CHEMICAL), IJ (ENGINEERING, INDUSTRIAL), IK (ENGINEERING, MANUFACTURING), IL (ENGINEERING, MARINE), IM (ENGINEERING, CIVIL), IO (ENGINEERING, OCEAN), IP (ENGINEERING, PETROLEUM), IQ (ENGINEERING, ELECTRICAL ELECTRONIC), IU (ENGINEERING, MECHANICAL), IX (ENGINEERING, GEOLOGICAL), PZ (METALLURGY METALLURGICAL ENGINEERING)

Fine and Applied Arts

No results

Foreign Languages

No results

Health Professions

HL (HEALTH CARE SCIENCES SERVICES), NE (PUBLIC, ENVIRONMENTAL OCCUPATIONAL HEALTH), LQ (HEALTH POLICY AND SERVICES)

Home Economics

No results

Interdisciplinary Studies

EV (COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS), PO (MATHEMATICS, INTERDISCIPLINARY APPLICATIONS), WU (SOCIAL SCIENCES, INTERDISCIPLINARY)

Letters

No results

Library

NU (INFORMATION SCIENCE LIBRARY SCIENCE)

Mathematics

PN (MATHEMATICS, APPLIED), PO (MATHEMATICS, INTERDISCIPLINARY APPLICATIONS), PQ (MATHEMATICS)

Physical Sciences

UB (PHYSICS, APPLIED), UF (PHYSICS, FLUIDS PLASMAS), UH (PHYSICS, ATOMIC, MOLECULAR CHEMICAL), UI (PHYSICS, MULTIDISCIPLINARY), UK (PHYSICS, CONDENSED MATTER)

Psychology

VI (PSYCHOLOGY), BV (PSYCHOLOGY, BIOLOGICAL), EQ (PSYCHOLOGY, CLINICAL), HI (PSYCHOLOGY, EDUCATIONAL), MY (PSYCHOLOGY, DEVELOPMENTAL), NQ (PSYCHOLOGY, APPLIED), VJ (PSYCHOLOGY, MULTIDISCIPLINARY), VP (PSYCHOLOGY, PSYCHOANALYSIS), VS (PSYCHOLOGY, MATHEMATICAL), VX (PSYCHOLOGY, EXPERIMENTAL), WQ (PSYCHOLOGY, SOCIAL)

Public Affairs

NE (PUBLIC, ENVIRONMENTAL, and OCCUPATIONAL HEALTH), VM (PUBLIC ADMINISTRATION)

Social Sciences

PS (SOCIAL SCIENCES, MATHEMATICAL METHODS), WU (SOCIAL SCIENCES, INTERDISCIPLINARY), WV (SOCIAL SCIENCES, BIOMEDICAL)

DOI: 10.1002/asi