

Journal Impact Evaluation: A Webometric Perspective¹

Mike Thelwall

Statistical Cybermetrics Research Group, School of Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, UK.

E-mail: m.thelwall@wlv.ac.uk

Tel: +44 1902 321470, Fax: +44 1902 321478

Abstract

In theory, the web has the potential to provide information about the wider impact of academic research, beyond traditional scholarly impact. This is because the web can reflect non-scholarly uses of research, such as in online government documents, press coverage or public discussions. Nevertheless, there are practical problems with creating metrics for journals based on web data: principally that most such metrics should be easy for journal editors or publishers to manipulate. Nevertheless, two alternatives seem to have both promise and value: citations derived from digitised books and download counts for journals within specific delivery platforms.

Introduction

The widely used Journal Impact Factors (JIFs) are indicators of intellectual impact based upon the average number of citations to the recently published articles of any given journal. More precisely, the JIF of a journal for a given year is the number of citations given in that year to articles published in the previous two years – divided by the number of “citable items” in the journal (as judged by Thomson Reuters) published in the previous two years. Citations only count if they are in journals, serials or other items indexed by Thomson Reuters and are identified as pointing to an item in the given journal in the previous two years. Despite the significant technical and theoretical limitations of the JIF (Garfield, 2005; Moed, 2010; Vanclay, in press), many of which seem to be widely-known and to have been recognised in scientometrics for a long time (Cole & Cole, 1967; Garfield, 1972), it continues to be used as a convenient proxy for journal quality. This article focuses not on technical limitations, however, but on the wider limitation of the JIF due to it being based upon published scholarly work alone. In particular, this article evaluates the potential to replace or supplement the JIF with information derived from web-based information. This is timely due to the recent emergence of the altmetrics movement, which aims to develop a range of new indicators for scientific research that exploit the potential of information that is openly provided on the web (Priem & Hemminger, 2010; Priem, Taraborelli, Groth, & Neylon, 2011).

The use of citation counts as indicators of scientific impact draws upon Merton’s (1973) belief in science as a normative and cumulative enterprise. When new contributions are made to science, they are typically formalised in published journal articles. The cumulative nature of science is recognised by scientists citing the work that they used to develop their new knowledge. The cumulative nature of science is acknowledged in the phrase written by Newton, “If I have seen a little further it is by standing on ye shoulders of Giants.” (Newton, 1676; 1992 republication). Merton’s normative theory was built upon by his some of his students explicitly using citation analysis. Early work was very promising, finding empirical evidence that citation counts correlated with expert judgment better than any other readily available quantitative measure (Cole & Cole, 1967). Nevertheless, there are many criticisms that could be made of this citation theory. First, authors may choose citations for reasons other than scientific quality and relevance (Bornmann & Daniel, 2008; Brooks, 1986; Cronin, 1984). More fundamentally, the value of science is in its ability to allow us to understand or control the physical environment and so it is possible to conceive of scientific advances that get rarely cited despite fulfilling this goal, for example by showing an apparently important area of research, such as cold fusion, to be false or useless (Huizenga, 1992). Alternatively, a major scientific issue might be solved so thoroughly that the problem itself is no longer urgent and the

¹ This is a pre-final version of: Thelwall, M. (2012). Journal impact evaluation: A webometric perspective, *Scientometrics*, 92(2), 429-441.

research field concerned may stagnate. A simple example of this is the discovery that anti-septic hand washing stops the spread of puerperal fever, although in this case the finding took a while to be accepted (De Costa, 2002).

More generally, published articles may make valuable contributions to science as a whole in ways that are unlikely to be reflected by citations because they will be used predominantly for things other than supporting further research. For example, the area of patent citation analysis has emerged to recognise and quantify the value of some scientific research that leads to commercial exploitation (Meyer, 2003; Oppenheim, 2000). This approach only works for areas of applied research where patenting is common, however. There are also many non-commercial applications of research, such as improving education and training, or aiding various kinds of policy making, such as economic, social or medical. In recognition of this, UK government funding for research will be given to those that can demonstrate “academic impact” or “economic or societal impact” as part of its periodic research evaluation exercise. The economic or societal impact component includes “enhancing cultural enrichment, quality of life, health and well-being” as well as “contributing towards evidence based policy-making and influencing public policies and legislation at a local, regional, national and international level”, amongst other things (RCUK, 2011). Research may also be oriented towards a profession, such as librarianship, social work, nursing or law, and thus contribute primarily to professional practice. Scholars may therefore have an impact on the world that is insufficiently represented by the citations that their publications receive. This may also apply to entire journals, such as those focusing on policy-relevant or professional issues, or education. A case in point is the *Journal of Education for Library and Information Science* (JELIS), which was dropped from the Science Citation Index due to consistently low JIFs, despite having a good reputation within the library and information science community (Coleman, 2007).

Another fundamental issue with the Mertonian perspective is that arguably social sciences and humanities research may not primarily be cumulative but may make contributions that are much vaguer than providing facts, such as providing interesting interpretations or perspectives. This seems to be typical of fields in which the phenomena investigated are hard to control because knowledge may be more fragile and less universally agreed as a result. This may be described as high “technical task uncertainty” in the language of Whitley (2000) or the disciplines may be described as “soft” (Becher & Trowler, 2001).

Despite all of the limitations of citation analysis, it is still a useful tool for evaluating researchers, if used cautiously with safeguards appropriate to the level of aggregation employed. Essentially, this is because some of the factors discussed above tend to average out over large groups of researchers. At the level of individual journals, however, there may be clear and systematic biases that would lead to misleading JIFs, even if correctly calculated. In particular, any journals focussing on professional, educational or commercial applications can expect to have their value poorly represented in comparison to theoretical journals that might not have much impact outside of the academic literature. Reliance upon the JIF seems likely therefore to lead to a systematic bias in reputation in favour of journals with a theoretical orientation. In simple terms, in a mixed rank order list of journals from the same subject area, such as the *Journal Citation Reports* categories, the journals with a professional, applied or educational orientation may be lower down than their value would justify. Of course, other factors may also influence JIFs that may exacerbate or mitigate this, such as field norms for the amount of citations and any cross-disciplinary influences in the applied and educational research. These factors can be alleviated to some extent by introducing a much more complex calculation for new journal IFs (Moed, 2010) but this article focuses instead on new sources of evidence to address these problems.

The web and the digitalisation of journals have made possible the calculation of new metrics for journals and articles, and this may have the potential to remedy some of the shortfalls of the JIF. This can happen in two main ways. First, because the web contains not just scholarly articles but a wider variety of public content, such as online newspapers and university class reading lists, it may be possible to gather citations from more types of document and hence capture wider types of impact, including some not captured well by the JIF. Second, it is possible to calculate indicators for how often an article is read based upon access statistics for electronic versions. This gives the potential to calculate impact in the form of readership, which could include reading to inform new research, as well as educational, commercial, education and other purposes. This article reviews studies that have

investigated online metrics of this kind, whether for individual articles or for entire journals, and then discusses the potential to use this kind of evidence to help evaluate academic journals.

Online Readership Indicators

Perhaps the most obvious source of new information about journal articles in the electronic era is usage information: how often they are read. Publishers of journals stored online in digital libraries will be able to record how often each article is accessed or downloaded. This information is sometimes given to editors or editorial boards and may be used to produce lists of the most accessed/downloaded articles or may be displayed on information pages associated with each article. Journals that use at least one of these approaches (as of January, 2012) include PLoS ONE (e.g., “Most Viewed” on the home page), the Journal of Informetrics (e.g., Elsevier access statistics reported annually to the editorial board) and the Journal of Medical Information Research (e.g., “Top Articles” on the home page). Publishers may already calculate download-based Impact Factors (IFs) for journals in their collection and use these statistics to evaluate them. This makes sense in the era of big deals for journals: selling groups of titles rather than individual titles. Without big deals, the bottom line for journals might be simple sales or profitability. Download information can also be used to demonstrate the value of a publication to the purchaser, especially if the download counts can be broken down by institution. In principle, it would be simple to calculate the average number of downloads per article for each journal, a download IF. In practice, there are three important problems, however.

The first problem is that access counts may not reflect reader numbers. Articles may be downloaded but not read; they may be printed and distributed to many people, such as a college class; they may be read in the print version of the journal; versions may be also read via the author’s institutional repository or copy stored on their web site.

The second problem is that if the figures are used to compare journals produced by different publishers then a reliable means must be agreed to make the figures comparable. Without this, each publisher would be free to inflate their figures as much as they liked. Hence an industry standard would be needed to make download IFs feasible as a general supplement for JIFs. Two such are the Standardized Usage Statistics Harvesting Initiative (SUSHI), which is formalised standard ANSI/NISO Z39.93 – 2007 (NISO, 2011), and the Counting Online Usage of Networked Electronic Resources (COUNTER) initiative (COUNTER, 2011a). The latter certifies specific report types as being “COUNTER-compliant” if they meet its standards. COUNTER-compliant vendors are also subject to periodic audits. The list of compliant vendors for journal download statistics at the time of writing included many large publishers like Elsevier, John Wiley and Sons and Springer Verlag and so this seems like a promising initiative (COUNTER, 2011b). In principle, usage IFs for any of the listed publishers would be comparable. A follow-on project has begun to investigate the potential for global usage IFs for journals (Shepherd, 2011). It recommended using the median rather than mean of article downloads for a journal because of skewed download distributions, amongst other things. Publisher acceptance is still a significant hurdle to overcome, however.

Finally, all downloads may not have equal value. To give an extreme example, an article mentioned in a newspaper may be frequently accessed by casual readers that never read the full text after accessing it because it is too difficult for the uninitiated to understand. In contrast, another article may be accessed by a person that uses the information to produce something of value, such as follow-up research, a commercial application, or professional guidelines. More seriously, the figures would be fairly easy to manipulate unless effective safeguards could be devised. An author may download their own articles repeatedly or make it compulsory for their students to access them, even though they were not central to any course.

A number of studies have investigated the use of access counts for scientometrics by assessing whether they can be used to predict future citations. These have shown that download counts can predict future citations, so download counts could be used as early evidence of the likely impact of articles (Brody, Harnad, & Carr, 2006; Moed, 2005) and, subject to appropriate safeguards, perhaps also for journals. Significant correlations between downloads and citation counts also provide evidence that downloads tend to be academic-related in some way, rather than just random. This goes some way to justifying their use for scientometric indicators. There do not seem to have been

systematic studies of articles that attract relatively many downloads compared to citations, however. These would need to give evidence that there were positive reasons why articles attracted relatively more citations than downloads to justify the claim that download indicators would be useful to supplement citation based indicators for individual articles or entire journals.

Despite the significant correlations found between citations and accesses for individual articles, at the journal level, this correlation does not seem to exist. Using download statistics from nine California State University institutions, usage IFs for data collected in 2004 were found to correlate negatively with JIFs, overall (Bollen & Van de Sompel, 2008). At the level of individual disciplines, only one out of 17 studied (education) exhibited a significant positive correlation between usage IFs and JIFs. A possible explanation for the lack of significant correlations is that the specialist interests of the community represented in this example do not fully reflect global interests, but positive correlations were found in disciplines with relatively large *graduate* populations and negative correlations in disciplines with relatively large *undergraduate* populations (Bollen & Van de Sompel, 2008). A potential conclusion from this is that journals with low IFs may tend to be *more* useful for undergraduate teaching than journals with high IFs. To illustrate other possible explanations, the journals Science and Nature both had low usage IFs (0.3) compared to high JIFs (31.9 and 32.2 respectively) but both market themselves extensively for individual user print subscriptions and so their readership may be underrepresented by usage IFs. For comparison, the top usage IF journal scored over twenty times higher: *Topics in Early Childhood Special Education* (usage IF 6.8; JIF 0.9).

A novel method to calculate usage factors is based upon the accesses of journals within a particular community, such as a group of universities (Bollen & Van de Sompel, 2008). This seems to be practical for large communities that share a common library service giving them access to the journals. If software could be developed to record accesses along with publication information about the number of articles per journal then this would allow localised JIFs to be calculated. These would have the advantage of being adapted to local conditions but would not serve the marketing needs of publishers in the way that the JIF does. Presumably localised usage IFs would be less tempting to manipulate as a result, but publishers may still need to reach agreements to provide necessary information to allow article accesses to be counted in a uniform way.

Social Bookmarking Indicators

An alternative to online readership indicators is to count how many people have added a given article to their online reference management software or archive. Such people seem likely to be those that have read an article and appreciate it enough to want to record it, as well as those that intend to read the article in the future – presumably because they have judged it relevant from the article title or abstract - and are recording it for this purpose. Online reference managers such as Zotero, CiteULike (Bogers & Bosch, 2008), Mendeley (Henning & Reichelt, 2008), BibSonomy and Connotea are therefore logical places to seek public evidence of the readership of articles. These offer a free online article bookmarking service with the added value of facilities for social interaction (Maxmen, 2010). This source of evidence is likely to suffer from similar biases to online readership figures, however. In particular, bookmarks may be predominantly created by students and reflect course reading lists. Moreover, articles related to social bookmarking and the social web seem likely to attract disproportionately many social bookmarks.

There is some evidence of the value of bookmark counts as an indicator of scholarly value because one study found citation counts to correlate with Mendeley bookmark counts for a set of 1,163 articles from Science and Nature in 2007 (Li, Thelwall, & Giustini, in press), although the study from 2011 found social bookmarking systems to be too rarely used to give generally useful data for individual article impact evaluation purposes. This may not be true for entire journals, however. A study of 45 physics journals with CiteULike, BibSonomy and Connotea compared IFs to a range of metrics derived from these services, such as the number of users per journal and the number of bookmarked articles per journal, finding significant correlations with most of them (Haustein & Siebenlist, 2011). The range of indicators used suggests that social bookmarking services can be used for indicators for a range of different types of journal impact. These and other studies also found significant practical issues with inaccurate and duplicate data (Bar-Ilan, 2011). Nevertheless, there is some evidence of the value of social bookmark counts.

Despite correlations between bookmarking indicators and citation counts, some articles are highly cited but not highly bookmarked and vice versa (Bar-Ilan, 2011), suggesting that bookmarks reflect a different aspect of impact to that of citations.

One final advantage of social bookmarking systems is that they seem to be open and transparent to that any manipulation of them might be easy to find. Nevertheless, because there is little or no quality control over them it is easy to conceive of legitimate ways of using them that would manipulate the results of any statistics generated from them. For instance, almost any lecturer could set their class the task of learning how to use a social bookmarking system and set them the assignment of bookmarking a set of articles of the lecturer's choice. Thus the results would be quite easy to legitimately manipulate.

Link Analysis

Hyperlinks in web pages are a technical device that allows a person to move from the page that they are reading to another page by clicking on the link. Information scientists identified in the early days of the web that, like citations, hyperlinks were inter-document connections (Larson, 1996; Rousseau, 1997) and could potentially be used for citation analysis purposes. When a commercial search engine introduced a link search capability then this turned it into a citation index and the web itself into a potential source of impact evidence (Ingwersen, 1998; Rodríguez i Gairín, 1997). In response to this, a number of researchers began to investigate the validity of using hyperlink counts for citation-like metrics for journals, individuals, research groups and institutions, as described below.

Outside of scientometrics, hyperlinks are widely used as evidence of the importance of documents or web sites. For instance, Google's key PageRank Algorithm (Brin & Page, 1998) is explicitly motivated by citation analysis and is designed to help rank web pages matching a search. Other hyperlink-based ranking algorithms have also been developed (e.g., Kleinberg, 1999). Hyperlinks are also used as evidence of the topics of web pages (Chakrabarti et al., 1998; Chakrabarti, Joshi, Punera, & Pennock, 2002; Chakrabarti, VanDen Berg, & Dom, 1999) amongst other applications. Hence the value of hyperlinks is widely exploited. A consequence of the importance of hyperlinks for search engine ranking and the commercial value of search engine positioning is that there are widespread attempts to create spam links that have the sole purpose of improving the ranking of the target web sites. This has, in turn, led to the on-going development of spam-detection algorithms that use various hyperlink properties as indicators of spam or explicitly aim to detect "link spam" (e.g., Han, Ahn, Moon, & Jeong, 2006). For example, TrustRank (Gyongyi, Garcia-Molina, & Pedersen, 2004) uses the link structure of the web to assign web sites a trust value: essentially the most trustworthy pages are those that are linked to from other trustworthy pages. TrustRank assumes that academic web sites are relatively trustworthy and so uses them to initialise the recursive algorithm.

Within Information science, hyperlinks have been evaluated as impact indicators for universities within a country (Kousha & Horri, 2004; Qiu, Chen, & Wang, 2004; Smith, 1999; Thelwall, 2001; Vaughan & Thelwall, 2005), departments within a country (Li, Thelwall, Wilkinson, & Musgrove, 2005; Thomas & Willett, 2000) and individual research groups (Barjak & Thelwall, 2008). These studies found significant correlations between counts of links to universities or departments and measures of their research productivity. Nevertheless, the correlations were stronger for larger units – i.e., strong for entire universities, weak for departments within a field and insignificant for research groups within a field. This suggests that the aggregation level is important and, since journals are relatively small, hyperlink counts seem likely to work poorly for them (Smith, 1999). This line of research also revealed that the key attractor of academic hyperlinks was the quantity of web publishing rather than its quality – with researchers producing more and better research also producing more web pages and attracting more hyperlinks (Thelwall & Harries, 2004). It also found that it is better to count inlinking web sites rather than individual hyperlinks (Thelwall, 2002).

There are two logical ways to use hyperlinks to replace JIFs: to calculate the total number of hyperlinks to a journal web site, divided by a quantity representing the size of the journal or its web site, or to calculate the average number of hyperlinks received by each published article. The former is much easier but it is not clear what the best figure would be to divide the link counts by and so all

research seems to have just counted total links to journal web sites. Early studies demonstrated that, within a single subject area and amongst a relatively homogeneous collection of journals, the number of hyperlinks to a journal web site correlates with its JIF (Vaughan & Hysen, 2002). This may not occur for non-homogeneous sets of journals, for example with some open access and across different subjects (Smith, 1999). Web site age is another important factor (Vaughan & Thelwall, 2003).

Despite the positive results discussed above, it seems clear that early predictions that citations will not be replaced by hyperlinks (e.g., van Raan, 2001) have been confirmed in the sense that there is still no hyperlink-based IF for journals. The main reason why no serious attempt has been made to construct this is probably the ease with which hyperlinks can be manipulated and the presumption that this manipulation will occur as soon as links are used tangibly to help indicate the reputation of journals. They are used for university reputations now to some extent, not just in search engine ranking but also in the Webometrics ranking of world universities (Aguillo, 2009). Nevertheless, practical problems with counting links to journals have increased due to the disappearance of many journals into publishers' web sites and their accessing online from within subscription-based services. As a result of this, there seems to be little need for independent web sites for journals and for scholars to link to such web sites.

Web Citations

One way in which to construct an IF with a wide range of coverage of types of impact would be to count citations not just from academic journals but from the entire web. Since the web contains commercial, academic, governmental and other information, such an IF would encompass types of impact ignored by the academic-based JIF. Some research has attempted to mimic the JIF using web data in this way, leading to promising results. The term web citation has been coined to mean a reference to a specific academic article from within a web page. Web citations can typically be searched for by constructing appropriate queries and submitting them to search engines. For example, such a query might contain the article title as a phrase search, together with the publication year and first author's name.

Web citations for articles in individual journals, as described above, have been shown to correlate significantly with traditional Web of Science citations for many disciplines (Kousha & Thelwall, 2007; Vaughan & Shaw, 2003, 2005). A web citation – hyperlink hybrid measure, URL citations, has also been shown to correlate significantly with citations (Kousha & Thelwall, 2007). Some studies have gone further and only counted web citations from specific types of documents, such as online presentations (Thelwall & Kousha, 2008), online syllabuses (Kousha & Thelwall, 2008), digitised books (Kousha & Thelwall, 2009), and blogs (Kousha, Thelwall, & Rezaie, 2010), all showing significant correlations with citation counts. The significant correlations give evidence that the different types of web citation counts are related to academic impact, but the correlations are typically not high and so it is also reasonable to hypothesise that the different sources might give indicators for somewhat different types of impact. For example, the study of online citations to journal articles from digitalised books via Google Book Search found that some articles had a high book-based impact, despite a moderate journal-based impact (Kousha & Thelwall, 2009). This may be particularly true for humanities-oriented articles due to the importance of books in the humanities.

Many researchers have employed an implicit type of web citation by using Google Scholar to calculate citation counts to articles (Jascó, 2005; Mayr & Walter, 2007; Meho & Yang, 2007; Vaughan & Shaw, 2008). The results correlate significantly with Web of Science citations (Kousha & Thelwall, 2007) and incorporate a range of source documents, including, but not limited to, publishers' digital libraries. This is effectively a hybrid web/journal database source, but seems to be heavily focused on academic sources and does not permit automatic searching in a way that would be necessary for automatic IF calculations for journals, unless provided by Google.

On the surface, the different kinds of web-based citation counts have great potential to be used as alternatives to the JIF. Nevertheless, there are two problems. First, with the exception of the Google Book Search citation counts, all the above sources can easily be manipulated, if adopted for widely disseminated IFs. Second, large-scale web citation counting would be time-consuming to do well because it does not seem possible to automatically construct the search engine queries that would be necessary. For example, it might not be possible to construct an effective query for a journal article

with a short, common title (e.g., Atoms) and a common author name (e.g., J. Smith), especially taking into account that any citation could be written in a number of different standard citation formats.

Twitter

Twitter seems to be a logical place to identify scholarly impact because it is used to recommend articles that they have read and some scholars believe that such citations reflect scholarly impact (Priem & Costello, 2010). Moreover, there is empirical evidence that citations in twitter correlate with later citation impact for one journal (Eysenbach, 2011). There are two practical problems with using Twitter for impact metrics, however. First, as with most web data, it has the potential to be manipulated and may be affected by spam. Second, and more specific to Twitter, the restriction on text message lengths to 140 characters is insufficient for complete references in most cases and even hyperlink identification is not straightforward because of the use of short encoded bit.ly and other URLs.

It is possible to circumvent the practical problems with Twitter citations in some ways, but some of them are not scalable. For instance, the problems of spam and identifying relevant tweets can be avoided by manual content analysis of the tweets of selected individual scholars (Priem & Costello, 2010). Scholarly-relevant and presumably spam-free tweets can also be gathered by monitoring specific academic hashtags, such as those for conferences (Weller, Dröge, & Puschmann, 2011). More generally, datasets of scholarly tweets can, in theory, be collected via queries for lists of scientific twitter users, scientific hashtags or general search terms (Weller & Puschmann, 2011). Of these, the first two seem to be the most practical but both would rely upon large lists of users or hashtags, which seem impossible to create automatically (Weller & Puschmann, 2011) and may need too much human time to create.

Assuming that relevant tweets have been identified, citations would need to be extracted. A recommendation for a specific article seems likely to be in the form of an URL or an indirect short description, such as “the article by X in the current Nature” or “the bioflavonoids article in the 12(4) issue of Food Hygiene”, or even “library searching article! #jasist”. The simplest solution would be to collect only tweets that mention a specific journal’s hashtag and ignore all other Twitter citations. This would give a clear advantage to journals with simple and well-used hashtags however, and also would not stop the spam problem. Alternatively, only tweets containing URLs of published articles could be counted. Since URLs are typically encoded, this approach would need to be combined with a large scale harvesting of Twitter. This is necessary because the URLs could only be found by decoding the short URLs in all tweets to find the ones relevant to journal articles. This approach has been successfully used for one online journal (Eysenbach, 2011) but for a large-scale analysis, significant manual work would be needed to start with to identify the URL structures of all journals to be assessed. This method would presumably favour online journals and journals with simple URL structures that readers would feel comfortable to forward. For other journals, readers may be more likely to reference an article in other ways than including an URL.

Twitter citation counting seems to be practical, given a source of funding for identifying journals’ URL structures, but the metric produced is likely to be particularly easy to manipulate, for example through false Twitter accounts, and its likely bias towards online journals seems unavoidable.

Discussion and Conclusions

As mentioned above, an important principal with indicator development is that when an indicator becomes widely recognised, there are likely to be attempts to manipulate it. When manipulation occurs for JIFs by journal editors promoting journal self-citations, it can be caught by calculating journal self-citation rates and then sanctions can be instigated (Davis, 2011). In contrast, any web based data seems likely to be easy to manipulate and this could even be cheap, as has been shown by the search engine optimization industry. This industry is also experienced at avoiding detection and so it seems that detecting manipulation attempts could not be easily automated and would therefore be expensive to attempt. This seems to rule out all IFs using open web data, including those based upon hyperlink counts and various types of web citation counts. The two possible exceptions are download IFs and IFs using book citations in Google Book Search.

Any IF relying upon Google Book Search would presumably need support from Google to implement and would be based upon a large number of queries to check for book citations of articles – one per published article. These queries would also need to be supported by human checking for cases when appropriate queries cannot be constructed. The results seem likely to be of particular value to humanities-oriented journals in subject areas combining humanities research and social science research, such as the library and information science discipline. This is because they would help to offset the JIF advantage of social science journals compared to humanities-oriented journals. A limitation of book-based IFs, however, is that books can be slow to write and publish compared to journal articles and so book-based journal IFs would need to cover a longer publication period – perhaps five years by default instead of two – and would be less responsive to changes over time as a result. If this approach was adopted then checks against manipulation would also be needed. In particular, citations from vanity publishers and university in-house publishing may need to be examined particularly closely or excluded. This idea would need the backing of a likely beneficiary to succeed – perhaps Google, benefitting from the publicity, or a national humanities funding council seeking to protect its research area.

Download IFs can also be manipulated to some extent. This manipulation could be direct in the form of authors or editors repeatedly downloading their own articles, or instructing students or others to do the same for them. It could also be indirect with a trick used by viruses (Provos, Mavrommatis, Rajab, & Monrose, 2008): using iFrames in unrelated web pages to get browsers to download specified articles without the user being aware of it. One protection for download statistics is that they are more numerous than citations and hence more need to be faked to have an impact. On the other hand, faking a download is much easier and less traceable than creating a citation in a published journal article. Another practical challenge for the creation of a general download IF for science is the need for publisher and distributor co-operation to ensure reasonably accurate and comparable download statistics. Projects like COUNTER and SUSHI, described above, seem to have made this a distinct possibility for the future. Intuitively, usage IFs seem to be problematic, however, because of the commercial competition involved and the need for publisher co-operation to generate comparable statistics. Another possible avenue is the development of localised download-based IFs for journals, calculated for specific digital library systems. This seems possible because the localised nature would make them less tempting to manipulate and the existence of such tools could be a marketing aid for the digital library, with the publishers hence supporting the software creation necessary to produce them. Such regionalised IFs may even be combined across multiple similar sites (e.g., universities) to create national or even international aggregated versions – although limited by only being comparable between journals within a specific supplier.

Finally, an alternative approach would be to employ a reasonably large set of altmetrics with the belief that it would be harder to manipulate multiple metrics if there were enough of them. There seems to be a wide potential to create different altmetrics and so this is a possibility but seems undesirable for several reasons. Creating a large number seems likely to increase the overall creation effort whilst reducing the time spent on each individual metric. Each individual metric may therefore be weaker and easier to manipulate as a result, unless the creation effort could be distributed in some way.

In conclusion, whilst there is a clear case for the value of web-based metrics in creating IFs for journals that would capture wider types of impact than that of the current JIFs, there are practical problems that seem to rule out most initiatives. The two approaches that seem possible, albeit with a limited scope of application for both and with financial backing needed to implement them, are book citations via Google Book Search, and localised download IFs.

References

- Aguillo, I. F. (2009). Measuring the institution's footprint in the web. *Library Hi Tech*, 27(4), 540-556.
- Bar-Ilan, J. (2011). *Articles tagged by 'bibliometrics' on Mendeley and CiteULike*. Paper presented at the Metrics 2011 Symposium on Informetric and Scientometric Research.

- Barjak, F., & Thelwall, M. (2008). A statistical analysis of the web presences of European life sciences research teams. *Journal of the American Society for Information Science and Technology*, 59(4), 628-643.
- Becher, T., & Trowler, P. (2001). *Academic tribes and territories (2ed)*. Milton Keynes, UK: Open University Press.
- Bogers, T., & Bosch, A. v. d. (2008). Recommending scientific articles using citeulike. In *Proceedings of the 2008 ACM conference on Recommender Systems (RecSys '08)* (pp. 287-290). New York, NY: ACM.
- Bollen, J., & Van de Sompel, H. (2008). Usage impact factor: The effects of sample characteristics on usage-based impact metrics. *Journal of the American Society for Information Science and Technology*, 59(1), 136-149.
- Bornmann, L., & Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45-80.
- Brin, S., & Page, L. (1998). The anatomy of a large scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1-7), 107-117.
- 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.
- Brooks, T. A. (1986). Evidence of complex citer motivations. *Journal of the American Society for Information Science*, 37, 34-36.
- Chakrabarti, S., Dom, B., Raghavan, P., Rajagonpalan, S., Gibson, D., & Kleinberg, J. M. (1998, April 1998). *Automatic resource compilation by analyzing hyperlink structure and associated text*. Paper presented at the the 7th International World Wide Web Conference.
- Chakrabarti, S., Joshi, M. M., Punera, K., & Pennock, D. M. (2002). The structure of broad topics on the Web. from <http://www2002.org/CDROM/refereed/338>
- Chakrabarti, S., VanDen Berg, M., & Dom, B. (1999, May 1999). *Focused crawling: A new approach to topic-specific Web resource discovery*. Paper presented at the the 8th International World Wide Web Conference.
- Cole, S., & Cole, J. R. (1967). Scientific output and recognition: A study in the operation of the reward system in science. *American Sociological Review*, 32(3), 377-390.
- Coleman, A. (2007). Assessing the value of a journal beyond the impact factor. *Journal of the American Society for Information Science & Technology*, 58(8), 1148-1161.
- COUNTER. (2011a). Counting Online Usage of Networked Electronic Resources. Retrieved December 14, 2011 from: <http://www.projectcounter.org/>.
- COUNTER. (2011b). Register of vendors providing usage reports compliant with Release 3 of the Code of Practice for Journals and Databases. Retrieved December 14, 2011 from: http://www.projectcounter.org/r2013/R2013CV_Dec2011.pdf.
- Cronin, B. (1984). *The citation process: The role and significance of citations in scientific communication*. London: Taylor Graham.
- Davis, P. (2011). Gaming the Impact Factor puts journal in time-out. Retrieved December 14, 2011 from: <http://scholarlykitchen.sspnet.org/2011/2010/2017/gaming-the-impact-factor-puts-journal-in-time-out/>.
- De Costa, C. M. (2002). "The contagiousness of childbed fever": A short history of puerperal sepsis and its treatment *The Medical Journal of Australia*, 177(11/12), 668-671.
- Eysenbach, G. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research*, 13(4), e123.
- Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, 178(4060), 471-479.
- Garfield, E. (2005). The agony and the ecstasy: The history and the meaning of the journal impact factor. *Fifth International Congress on Peer Review in Biomedical Publication, in Chicago, USA*, Retrieved September 27, 2007: <http://garfield.library.upenn.edu/papers/jifchicago2005.pdf>.
- Gyongyi, Z., Garcia-Molina, H., & Pedersen, J. (2004). Combating web spam with TrustRank. *Proceedings of the thirtieth international conference on Very large data bases*, 30, 576-587.

- Han, S., Ahn, Y.-y., Moon, S., & Jeong, H. (2006). Collaborative blog spam filtering using adaptive percolation search. *WWW2006 Workshop*, Retrieved May 5, 2006 from: <http://www.blogpulse.com/www2006-workshop/papers/collaborative-blogspam-filtering.pdf>.
- Haustein, S., & Siebenlist, T. (2011). Applying social bookmarking data to evaluate journal usage. *Journal of Informetrics*, 5(3), 446–457.
- Henning, V., & Reichelt, J. (2008). Mendeley - A last.fm for research? In *IEEE Fourth International Conference on eScience (eScience '08)* (pp. 327-328). Los Alamitos: IEEE.
- Huizenga, J. R. (1992). *Cold fusion: The scientific fiasco of the century*. Rochester, NY: University of Rochester Press.
- Ingwersen, P. (1998). The calculation of Web Impact Factors. *Journal of Documentation*, 54(2), 236-243.
- Jascó, P. (2005). Google Scholar: the pros and the cons. *Online Information Review*, 29(2), 208-214.
- Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5), 604-632.
- Kousha, K., & Horri, A. (2004). The relationship between scholarly publishing and the counts of academic inlinks to Iranian university web sites: Exploring academic link creation motivations. *Journal of Information Management and Scientometrics*, 1(2), 13-22.
- Kousha, K., & Thelwall, M. (2007). Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology*, 58(7), 1055 -1065.
- Kousha, K., & Thelwall, M. (2008). Assessing the impact of disciplinary research on teaching: An automatic analysis of online syllabuses. *Journal of the American Society for Information Science and Technology*, 59(13), 2060-2069.
- Kousha, K., & Thelwall, M. (2009). Google Book Search: Citation analysis for social science and the humanities. *Journal of the American Society for Information Science and Technology*, 60(8), 1537-1549.
- Kousha, K., Thelwall, M., & Rezaie, S. (2010). Using the Web for research evaluation: The Integrated Online Impact indicator. *Journal of Informetrics*, 4(1), 124-135.
- Larson, R. R. (1996). Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of cyberspace, *ASIS 59th annual meeting*. Baltimore, MD.
- Li, X., Thelwall, M., & Giustini, D. (in press). Validating online reference managers for scholarly impact measurement. *Scientometrics*.
- Li, X., Thelwall, M., Wilkinson, D., & Musgrove, P. B. (2005). National and international university departmental web site interlinking, part 2: Link patterns. *Scientometrics*, 64(2), 187-208.
- Maxmen, A. (2010). Science networking gets serious. *Cell*, 141(3), 387-389.
- Mayr, P., & Walter, A. K. (2007). An exploratory study of Google Scholar. *Online Information Review*, 31(6), 814-830.
- Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science vs. Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58(13), 2105-2125.
- Merton, R. K. (1973). *The sociology of science. Theoretical and empirical investigations*. Chicago: University of Chicago Press.
- Meyer, M. (2003). Academic patents as an indicator of useful research? A new approach to measure academic inventiveness. *Research Evaluation*, 12(1), 17-27.
- 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 & Technology*, 56(10), 1088-1097.
- Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, 4(3), 265-277.
- Newton, I. (1676; 1992 republication). Letter from Isaac Newton to Robert Hooke, 5 February 1676. In J.-P. Maury (Ed.), *Newton: Understanding the Cosmos*. New York: New Horizons.
- NISO. (2011). ANSI/NISO Z39.93 - 2007 The Standardized Usage Statistics Harvesting Initiative (SUSHI) protocol. Retrieved December 14, 2011 from: http://www.niso.org/apps/group_public/project/details.php?project_id=2021.

- Oppenheim, C. (2000). Do patent citations count? In B. Cronin & H. B. Atkins (Eds.), *The web of knowledge: a festschrift in honor of Eugene Garfield* (pp. 405-432). Metford, NJ: Information Today Inc. ASIS Monograph Series.
- Priem, J., & Costello, K. L. (2010). How and why scholars cite on Twitter. In *Proceedings of the American Society for Information Science and Technology (ASIST 2010)* (Vol. 47, pp. 1-4).
- Priem, J., & Hemminger, B. M. (2010). Scientometrics 2.0: Toward new metrics of scholarly impact on the social Web. *First Monday*, 15(7), Retrieved December 7, 2011 from: <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2874/2570>.
- Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2011). altmetrics: a manifesto. Retrieved December 14, 2011 from: <http://altmetrics.org/manifesto/>.
- Provos, N., Mavrommatis, P., Rajab, M. A., & Monrose, F. (2008). All your iFRAMEs point to us. Retrieved December 13, 2011 from: http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en/archive/provos-2008a.pdf.
- Qiu, J. P., Chen, J. Q., & Wang, Z. (2004). An analysis of backlink counts and Web Impact Factors for Chinese university websites. *Scientometrics*, 60(3), 463-473.
- RCUK. (2011). Types of Impact. Retrieved December 12, 2011 from: <http://www.rcuk.ac.uk/documents/impacts/TypologyofResearchImpacts.pdf>.
- Rodríguez i Gairín, J. M. (1997). Valorando el impacto de la información en Internet: AltaVista, el "Citation Index" de la Red (Evaluating the impact of Internet information: Altavista, the "Citation Index" of the Web). *Revista Española de Documentación Científica*, 20(2), 175-181.
- Rousseau, R. (1997). Sitations: an exploratory study. *Cybermetrics*, 1(1), Retrieved July 25, 2006 from: <http://www.cindoc.csic.es/cybermetrics/articles/v2001i2001p2001.html>.
- Shepherd, P. (2011). Journal Usage Factor: Results, recommendations and next steps. Retrieved December 14, 2011 from: <http://www.uksg.org/sites/uksg.org/files/JournalUsageFactorReport080711.pdf>.
- Smith, A. G. (1999). A tale of two web spaces; comparing sites using Web Impact Factors. *Journal of Documentation*, 55(5), 577-592.
- Thelwall, M. (2001). Extracting macroscopic information from web links. *Journal of American Society for Information Science and Technology*, 52(13), 1157-1168.
- Thelwall, M. (2002). Conceptualizing documentation on the Web: An evaluation of different heuristic-based models for counting links between university web sites. *Journal of American Society for Information Science and Technology*, 53(12), 995-1005.
- Thelwall, M., & Harries, G. (2004). Do the Web sites of higher rated scholars have significantly more online impact? *Journal of the American Society for Information Science and Technology*, 55(2), 149-159.
- Thelwall, M., & Kousha, K. (2008). Online presentations as a source of scientific impact?: An analysis of PowerPoint files citing academic journals. *Journal of the American Society for Information Science and Technology*, 59(5), 805-815.
- Thomas, O., & Willett, P. (2000). Webometric analysis of departments of librarianship and information science. *Journal of Information Science*, 26(6), 421-428.
- van Raan, A. F. J. (2001). Bibliometrics and Internet: Some observations and expectations. *Scientometrics*, 50(1), 59-63.
- Vanclay, J. K. (in press). Impact Factor: Outdated artefact or stepping-stone to journal certification? *Scientometrics*.
- Vaughan, L., & Hysen, K. (2002). Relationship between links to journal Web sites and impact factors. *ASLIB Proceedings*, 54(6), 356-361.
- Vaughan, L., & Shaw, D. (2003). Bibliographic and Web citations: What is the difference? *Journal of the American Society for Information Science and Technology*, 54(14), 1313-1322.
- Vaughan, L., & Shaw, D. (2005). Web citation data for impact assessment: A comparison of four science disciplines. *Journal of the American Society for Information Science & Technology*, 56(10), 1075-1087.
- Vaughan, L., & Shaw, D. (2008). A new look at evidence of scholarly citation in citation indexes and from web sources. *Scientometrics*, 74(2), 317-330.

- Vaughan, L., & Thelwall, M. (2003). Scholarly use of the Web: What are the key inducers of links to journal Web sites? *Journal of American Society for Information Science and Technology*, 54(1), 29-38.
- Vaughan, L., & Thelwall, M. (2005). A modeling approach to uncover hyperlink patterns: The case of Canadian universities. *Information Processing & Management*, 41(2), 347-359.
- Weller, K., Dröge, E., & Puschmann, C. (2011). Citation analysis in Twitter: Approaches for defining and measuring information flows within tweets during scientific conferences. *1st Workshop on Making Sense of Microposts*, Retrieved June 16, 2011 from: http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-2718/paper_2004.pdf.
- Weller, K., & Puschmann, C. (2011). Twitter for scientific communication: How can citations/references be identified and measured? In *Proceedings of the Poster Session at the Web Science Conference 2011* (pp. Retrieved October 21, 2011 from: <http://journal.webscience.org/2500/>). Koblenz, Germany.
- Whitley, R. (2000). *The intellectual and social organization of the sciences* (2 ed.). Oxford: Oxford University Press.