Do The Web Sites of Higher Rated Scholars Have Significantly More Online Impact?

Mike Thelwall[1]

School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1SB, UK. E-mail: *m.thelwall@wlv.ac.uk*

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

Gareth Harries

School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1SB, UK. E-mail: *g.harries@wlv.ac.uk*

Tel: +44 1902 321470 Fax: +44 1902 321485

The quality and impact of academic Web sites is of interest to many audiences, including the scholars who use them and Web educators who need to identify best practice. Several large-scale European Union research projects have been funded to build new indicators for online scientific activity, reflecting recognition of the importance of the Web for scholarly communication. In this paper we address the key question of whether higher rated scholars produce higher impact Web sites, using the UK as a case study and measuring scholars' quality in terms of university-wide average research ratings. Methodological issues concerning the measurement of the online impact are discussed, leading to the adoption of counts of links to a university's constituent single domain Web sites from an aggregated counting metric. The findings suggest that universities with higher rated scholars produce significantly more Web content but with a similar average online impact. Higher rated scholars therefore attract more total links from their peers, but only by being more prolific, refuting earlier suggestions. It can be surmised that general Web publications are very different from scholarly journal articles and conference papers, for which scholarly quality does associate with citation impact. This has important implications for the construction of new Web indicators, for example that online impact should not be used to assess the quality of small groups of scholars, even within a single discipline.

Keywords
Scholarly communication, Web, hyperlink analysis, cybermetrics, Web mining.

1. INTRODUCTION

In this paper we investigate whether the university Web sites created by or for higher rated scholars have more online impact, as measured by counts of links pointing to them. The focus is general academic Web sites rather than just online articles, and is therefore a complementary approach to the analysis of the increasingly ubiquitous academic digital libraries, including e-print archives (Brody *et al*., 2002; Harter & Ford, 2000). Links between university Web sites represent a wide range of types of predominantly informal scholarly communication (Wilkinson et al., 2003) and so the issue is whether highly rated scholars also have a high impact through informal communication channels, in particular the Web. Previous research has suggested that universities and departments with more and higher rated scholars attract more links from their peers (Chu et al., 2002; Li et al., 2003; Smith & Thelwall, 2002; Tang & Thelwall, 2003; Thelwall, 2001c, 2002a), but no firm conclusions have yet been reached as to why this should be the case, although it has been hypothesized that this could be because of a halo effect at the university level (Thelwall, 2001c). In this paper we will investigate this phenomena further by differentiating

[1] To whom all correspondence should be addressed

between the quality and quantity of Web sites. The overarching goal is to deepen understanding of how scholars interact with digital information, now a critical activity.

Historically, quantitative research into scholarly communication has been largely restricted to formal communications such as books and journal articles (Borgman, 1990). The need to understand informal communication has been hampered by its ephemeral nature, with groundbreaking findings coming from very small-scale constructivist studies such as that of Latour and Woolgar (1986), but the Web now gives scope for complementary large-scale approaches such as the investigation reported here. Researchers addressing problems related to quantitative analyses of the Web have developed the field of Webometrics, which is still in the process of developing and validating its methodologies. With the Web becoming central to an increasing proportion of scientific activity (Lawrence, 2001; Rees, 2002), a key question attracting widespread interest is whether the Web can be used as a data source with which to develop new indicators for aspects of the research process (Almind & Ingwersen, 1997; Aguillo, 1998; Ingwersen, 1998; Cronin, 2001, Thelwall, 2002ab; EICSTES, 2003; SIBIS, 2003; Vaughan & Thelwall, 2003; WISER, 2003). The issue of whether higher rated scholars produce higher impact Web sites is also a vital one for this because of the corresponding key belief in evaluative bibliometrics that better scholars produce higher impact publications.

2. BACKGROUND

2.1 Bibliometrics

The field of bibliometrics is concerned with conducting quantitative analyses of documents, traditionally printed ones such as books and journal articles. In bibliometrics, analyses of the impact of a body of work and relationships between publications or collections of publications are often based upon citation analysis. The basic objects counted are the citations between journal articles, conference articles and books, with the dual beliefs that counts of citations to a publication are an indicator of its impact within the academic community, and that citations broadly reflect the flow of use of knowledge between publications. These beliefs are controversial, being both empirically verified in statistical studies comparing results to other quality indicators (Bayer & Folger, 1966; Cole & Cole, 1967; Virgo, 1977; Smith, 1981; Lawani & Bayer, 1983; Cronin, 1984; Seng & Willett, 1995; Abt, 2000) and demonstrably false in many individual cases. Nevertheless, citations are being used both to assess research performance (Jimenez-Contreras *et al*., 2002) and to map the structure of science (Small, 1999). Bibliometric techniques have recently been applied to online journal articles (Harter & Ford, 2000), individual e-journals (Marek & Valauskas, 2002) and subject-based digital libraries including preprint archives (Brody *et al*., 2002). This may represent a long-term shift towards online analysis, even for formal scholarly communication (Rees, 2002).

Perhaps the most well-known publication impact metrics are the journal impact factors of the Institute for Scientific Information, although these are themselves problematic and controversial (Garfield, 1985; Moed, 2002). Bibliometric approaches are commonly applied to groups of researchers and are "particularly at the level of research groups (e.g. university departments and institutes) an indispensable element alongside peer review in the research evaluation process." (van Raan, 2000, p 101). The careful methods of van Raan and others (Moed *et al*., 1985; Nederhof & Noyons, 1992; van Raan, 1999; Ingwersen *et al*., 2000; Vinkler, 2000) show that great care should be taken in both the calculation and interpretation of bibliometric indicators. In particular, he recommends that a range of indicators should be used. For whole universities formal evaluation is normally not required, but quantitative approaches are still used at the national, university and lower organizational levels to track patterns of scholarly communication, both nationally and internationally, either using citations, cocitations or co-authorships (Zitt *et al*., 2000; Glänzel *et al*., 1999; Glänzel & Schubert, 2001).

2.2 Webometrics

Web content is studied from many perspectives in different disciplines. Computer scientists may be concerned with developing tools to enable users to retrieve information from the Web (Brin & Page, 1998; Kleinberg, 1999; Arasu *et al*., 2001) to organize their favorite pages (Amento *et al*., 1999; Terveen *et al*., 1999) or to design effective Web sites (Nel *et al*. 1999; Shneiderman, 1998; Furnas, 1997). Social scientists may be more concerned with what humans use the Web for (Rogers, 2000; Burnett & Marshall, 2003; Halavais & Garrido, 2003) or the effect that Web technologies can have on human behavior (Quan-Haas & Wellman, 2003). Information scientists span these concerns but many have a particular interest in how scholars use the Web (Almind & Ingwersen, 1997; Davenport & Cronin, 2000; Cronin, 2001). This stems, in part, from the tradition of bibliometrics (Borgman & Furner, 2002) and Scientometrics (Garg, 2003) discussed above in which metrics are developed to assess aspects of the behavior of researchers.

Webometrics is the application of informetric and other quantitative techniques to the study of the Web (Almind & Ingwersen, 1997; Björneborn & Ingwersen, 2001). The primary objects of study for one strand of Webometrics have been hyperlinks between pages, often through an analogy with citations between journal articles (Wilkinson et al., 2003b). Early studies explored possible techniques (Larson, 1996; Almind & Ingwersen, 1997; Rodriguez Gairín, 1997; Aguillo, 1998) and fitted mathematical models to Web data (Rousseau, 1997). Ingwersen (1998) then introduced a new metric that could be used to assess the impact of Web spaces such as university Web sites. His Web Impact Factor had many variants but the most successful was the ratio of the number of links pointing to the space in question from outside of that space (inlinks) divided by the total number of pages in the space. Subsequent authors attempted to validate link metrics by comparing them to other measures of research quality mainly at the national level (Smith, 1999; Thelwall, 1999; Thelwall, 2000; Thomas & Willett, 2000; Harter & Ford, 2000) without success. The link counts were obtained either through advanced searches in AltaVista or a personal Web crawler. Since these unpromising beginnings a series of positive results have shown that inlink count metrics can correlate significantly with measures of the research of universities (Thelwall, 2001; Smith & Thelwall, 2002; Thelwall, 2002a; Thelwall & Harries, 2003; Thelwall & Wilkinson, 2003a), and departments (Chu *et al*., 2002; Li *et al*., 2002; Tang & Thelwall, 2003) as well as correlating with journal impact factors (Vaughan & Huysen, 2002; Vaughan & Thelwall, 2003). This is important evidence that although links could be created at random or at the whim of Web authors, patterns may still be successfully mined from them.

An important question that has emerged from these studies is why university/department/journal Web sites with better/more research should attract more inlinks. A study of a random collection of 414 links between UK university Web sites found that although over 90% were related in some way to research, less than 1% targeted Web pages that were equivalent to refereed journal articles (Wilkinson *et al*., 2003a). Clearly even if higher rated researchers produce better formal publications, and place some on the Web, this is not the reason why their Web pages attract more links.

In addition to the studies above that have focused on the evaluation of online impact metrics, others have used hyperlinks to investigate patterns of communication between large areas of the Web or individual scientists. International links have been counted between the universities in whole countries, both in Europe (Thelwall *et al*., 2003) and in the Asia-Pacific region (Thelwall & Smith, 2002). These studies showed that the approach was feasible and that meaningful information could be extracted, but that care had to be taken when comparing the results for countries with significantly different scales of Web use. Björneborn (2001) has also investigated possibilities for information discovery through serendipity on the Web in a pilot study of online bookmark pages of a selection of researchers.

Similar questions have also been addressed by other Webometric studies that have used text rather than link analysis. Cronin *et al*. (1998) developed a typology of reasons for the online invocation of scholars, finding a diverse set. Vaughan and Shaw (2003) investigated Web citations, finding that online citations to journal articles correlated highly with citations from other journal articles. Goodrum *et al*., (2001) compared a Web based citation analysis of online PDF and postscript documents with a traditional approach. On the Web, published papers are more likely to be used (or better papers are more likely to be placed online) (Lawrence, 2001) but online articles in university Web sites are still relatively rarely linked to (Wilkinson *et al*., 2003a). Finally, Landes and Posner (2000) and Cronin and Shaw (2002) have both used counts of mentions of scholars in Web pages as partial indicators of public intellectual status, although this does not relate to the Web content produced by them.

In addition to applications in information science, Webometric results and approaches can also build understanding of online impact that can be of benefit to both the Web information retrieval and the hypertext design and evaluation specialisms. The reason is that hyperlink metrics are widely used as a quality indicator and for extracting structures from the Web. Perhaps the most visible example is the Google's PageRank algorithm, which attempts to rank Web pages by quality through the assumption that better pages will attract more links to them, particularly links from pages that are themselves high quality (Brin & Page, 1998). Kleinberg's (1999) HITS algorithm uses similar assumptions.

3. Measuring Online Impact

The measurement of the online impact of Web publications is a highly problematic task. There are many existing methods for evaluating aspects of the quality of hypertexts (e.g. Coulston & Vitolo, 2001) but these all focus on construction quality and usability rather than impact. University Web sites contain a wide range of types of information (Middleton *et al*., 1999) and it is clearly impractical to evaluate millions of Web pages, each by an appropriate human expert. We need, therefore, a measure that is capable of being calculated automatically over millions of Web pages, but which has some theoretical justification and experimental evidence to support its validity. A similar problem is faced by countries that seek to evaluate their scholars' work without resort to an expensive peer review process. The solution adopted in the latter case is often to use a form of citation counting as a proxy for quality in the belief that in general better articles attract more citations (Moed, 2002). This is a reasonable approach at a sufficiently large level of aggregation, e.g. for whole departments rather than individual researchers or papers. This approach logically suggests using links to a Web page as a proxy for its actual impact, however conceived. There are several problems with this.

3.1 Validity

The first and most fundamental problem for using Web link counts to evaluate online impact is the issue of validity. Citation counts have been subject to extensive research over the past 30 years to get to the position where experts can use them productively, if great care is taken (Moed, 2002). Web links need the same kind of research to give evidence of a consistent relationship with quality at any given level of aggregation. This process has begun, but is not complete. We contend, however, that link counts are the most appropriate objects available to use as a proxy for Web impact, especially when averaged over an entire site. Note also that links within a site should be excluded from the counts to improve the measurement (Ingwersen, 1998; Kleinberg, 1999; Thelwall 2000; Page, 2001; Arasu *et al*., 2001).

There are clearly methodological issues with the use of Web links. Undoubtedly there will be widely read Web pages that are influential but not linked to. Moreover, reasons for linking to pages can be tenuous or even apparently non-existent (Thelwall, 2003). Yet the statistical tests alluded to above provide some evidence that this should not be a major cause for concern at the university level of averaging, and the Wilkinson *et al*. (2003a) study shows that academic links typically bear some relationship to scholarly activities.

3.2 Reliability: Alternative Document Models

The second problem with link counting is reliability. Links would be most valuable if all were created individually by appropriate experts after careful consideration. This is undoubtedly the case for some, but at the other extreme many links are automatically generated by computer programs. The most common examples of the latter are probably credit links to software company home pages that are inserted by Web authoring software. A more insidious example is of repeated links in navigation bars, replicated to all pages of a site. This occurs in the sites of some joint research project sites in the UK to indicate shared multi-institution ownership of the project (Thelwall, 2002a).

One way of circumventing the problem of multiple links with 'low value' in navigation bars is to develop an algorithm to separate out the high value and low value links in a page by virtue of their position in the page. Some progress has been made in this direction (Miles-Board *et al*., 2002) but it is not sufficient for our purposes because there can also be repeated similar links in the more important 'content section' of pages. An example of this is tens of thousands of pages in the Warwick university site comprising an online biochemistry database, each one of which is specific to a single molecule and contains a link to a page describing the same molecule in similar online Web databases at Cambridge University and Imperial College. The solution that we will adopt to circumvent problems of this type is a conceptual shift implemented by a recently developed Webometric algorithm for counting.

The conceptual shift is to cluster pages together into multiple-page hypertexts and then count links between hypertexts instead of individual pages. Ideally, the clustering would be done by a program automatically and on the basis of identifying the genre that the collection of pages fits into (e.g. 'scholar's personal site' or 'departmental Web site') but automatic genre identification is currently a very distant goal (Rehm, 2001). Instead, we make a simplification and regard all pages with the same domain name as being a single hypertext. This idea, has been used before to improve counting results and is known as the domain alternative document model (ADM), (Thelwall, 2002a). It has given improved results compared to the page model in some situations, but the area where it works better includes the UK academic domain. In this paper the term *domain name based Web site*, or just *domain*, will be used to refer to the collection of pages that share a common domain name in their URL. To illustrate this, in the example below the pages 1.htm, 3.htm and 5.htm are all part of the same domain name based Web site, that of a.ac.uk, but 7.htm is part of a different domain name based Web site, that of d.a.ac.uk.

We give an example to illustrate counting with the domain ADM. Consider the following set of links from university A to university B.

```
a.ac.uk/1.htm → b.ac.uk/2.htm
a.ac.uk/3.htm → b.ac.uk/4.htm
a.ac.uk/5.htm → c.b.ac.uk/6.htm
d.a.ac.uk/7.htm → b.ac.uk/8.htm
```

These would count as three domain ADM links from university A to university B since the first two links connect the same hypertexts, and so the second one would be ignored in counting. Note that links between different hypertexts within the same university are still ignored, including the following example.

```
a.ac.uk/9.htm → e.a.ac.uk/10.htm
```

The disadvantage of the domain ADM is that in addition to eliminating replicated links, it will conflate links that were created for two completely different purposes between the same pair of domains. Perhaps in the above example, pages 1.htm and 3.htm were created by different authors but happened to link to a common target domain by coincidence. Statistically, however, the advantages have been shown to outweigh the disadvantages (Thelwall, 2002a).

3.3 Site Coverage: The Publicly Indexable Set

The only practical method for extracting links from a collection of large Web sites is to use a crawler. In a crawl of a large multi-domain site, such as a university Web site, pages and entire domains will be missed because they were not linked to by other pages crawled on the same site, or because the links were hidden in JavaScript, databases or similar techniques. As a result it is impossible to guarantee complete coverage. This is an unavoidable problem (Thelwall, 2002b; Lawrence & Giles, 1999), but has not stopped the significant results reported above and so does not seem to be an obstacle in practice. For counting links between Web sites, therefore, the pages indexed are those that can be found by a crawler by recursively following links from the home page. This will be referred to as the publicly indexable (PI) set, adapting terminology from Lawrence & Giles (1999).

4. THE RESEARCH QUESTION

The question of whether higher rated scholars produce higher online impact Web sites is clearly subject to a wide a range of interpretations that make it ambiguous. Having already operationalized Web sites on the basis of previous research as domain name based Web sites, we now operationalize 'higher rated scholars' and 'higher online impact'.

For the measurement of the quality of scholars we will use the figures provided for UK universities by the official 2001 Research Assessment Exercise (2003). This is a peer-review, subject-based process that is used to direct government research funding and hence is taken very seriously by all concerned. Our notion of 'higher rated' for scholars is therefore a nationally recognized highly important one, based upon the judgments of subject experts. We believe that there is not a more reliable set of research quality figures for any other country in the world. In fact the very expense of providing this level of quality is itself controversial (Holmes & Oppenheim, 2001; Jaffe, 2002).

We will use inlink counts obtained using the domain ADM discussed above to evaluate online impact. The concrete statement of the research question is therefore as follows.

Do the domain ADM inlink counts of the domains of UK university Web sites correlate significantly with the RAE research quality of the owning institutions?

We will also investigate whether higher rated scholars are more productive on the Web, in terms of generating more publications. We will use a count of domain name-based Web sites as an indicator of the total Web productivity for a university.

5. METHODS

5.1 Crawling the UK Academic Web

The Web sites of all UK universities were crawled in June 2002 by a specialist information science Web crawler (Thelwall, 2001ab). The crawler is fed with a starting page, normally the university home page, and then iteratively downloads all pages that it can find by extracting links from the anchor tags, client side image maps, meta tag and HTTP server redirection instructions. It obeys the robot.txt convention, avoiding areas of sites declared as out of bounds to crawlers by the site or page owners. It is not able to extract links from non-HTML files, JavaScript in HTML or objects embedded in HTML, such as Java or Flash. Also, it does not follow links that appear to be database queries because they contain a question mark in the URL. A few universities have ASP interfaces on their sites with all main pages delivered by ASP queries, and these would not be indexed. For this type of site an alternative starting page was found at a list of departmental Web sites in a standard (static) Web page. Each university was identified through its domain names. For example, for Wolverhampton University, all pages with domain names ending in .wlv.ac.uk or .wolverhampton.ac.uk were included, in addition to the root domains wlv.ac.uk and wolverhampton.ac.uk. Unfortunately, universities also use a range of other domain names for publishing in addition to their official ones. These can include, for example, a .com domain for an individual research group or project. Probably there are at least a thousand of these in the UK but we do not have an effective method for identifying them. As a result the crawl data is restricted to domain names derived from official university root domain names.

The crawler was designed to test exhaustively for duplicate pages and to discard all copies. University Web sites normally contain a variety of information that was not created by the university's staff, and this had to be manually identified and excluded from the study. Examples of this include copies of manufacturers' computer software documentation (e.g. versions of Java) and the SunSITE, but also e-journals since the papers are typically written by academics from a variety of institutions. The approach adopted was to combine URL path name heuristics (mainly for Java documentation) with a manual inspection of areas of a site that either attracted many links, or were very large and standardized in structure. This is clearly an error-prone process designed only to identify the most blatant offenders. Random samples of individual pages (e.g. Wilkinson *et al*., 2003a; Thelwall, 2001; Thelwall & Harries, 2003) have shown that it is highly effective, however.

The link data collected has been placed online and open access for other researchers to freely use (cybermetrics.wlv.ac.uk/database/) in the belief that since Web crawling is a process that uses the resources of the servers crawled without charge, there is a moral obligation on researchers to allow others to use all data obtained in this way.

5.2 Compiling the Link Data

A program was used on the link structure databases created by the crawler to convert them into domain ADM inlink counts. After this process, there were a number of domains that had links to them but had not been crawled. This is possible because the crawls of university Web sites are conducted independently. In other words the discovery of a link in one university to a page in another university would not result in the link target being crawled. We therefore had to decide what to do with the counts of links to uncrawled domains.

The most complete data set available after the crawls was the union of the set of domains crawled with the set of domains linked to, and this would have been a possible choice of final data set. It would not be a *representative* sample, however, since it would contain too small a proportion of domains that were not linked to. This would be a consequence of only including the uncrawled domains that were linked to, and not the uncrawled domains that were not linked to. Of course we could not add the latter set simply because we did not know their identity or size. We therefore only included links to crawled domains (i.e. the publicly indexable set) and links to uncrawled domains were discarded. Note that the restriction to the publicly indexable set also solves another problem: that of mistyped URLs. Many URLs have typographical errors but these will automatically be excluded since only valid crawled domains are used.

5.3 Calculating the Research Quality Figures

The research quality figures provided by the RAE are a set of scores for up to 68 sets of subjects. The scores are 1, 2, 3a, 3b, 4, 5, 5* where 1 is the lowest (almost none of the assessed research is of national excellence) and 5* the highest (over half of the assessed research is of international excellence). We mapped this onto a numerical scale 1 – 7 and for each university multiplied the score for each set of academics by the number of academics in the set, totaling the products over all sets in the university. This is making the assumption of a linear scale, which is the standard interpretation (Mayfield University Consultants, 2000; Education Guardian, 2001) that has been used successfully before in academic research (Thelwall, 2001a, 2002a). In the following formula the *total research productivity R(A)* of university *A* is defined, where $R_{u(A)}$ is the RAE score for the unit of research *u(A)* submitted by university *A* from the set *U(A)* of all units submitted by university *A*, and $S_{u(A)}$ is the number of staff (faculty) submitted by university *A* for unit *u(A)*.

$$R(A) = \sum_{u(A)\in U(A)} S_{u(A)} R_{u(A)}$$

When the research productivity is divided by the total full-time equivalent academic staff S_A of the university A (www.rae.ac.uk) the result RAE_A will be called its *average research productivity*, or its *RAE research quality*.

$$RAE_A = \frac{R(A)}{S_A}$$

Full-time equivalent academic staff counts S_A will be used as a measure of university size in the analysis below. Note that $S_A > \sum_{u(A)\in U(A)} S_{u(A)}$ for all universities *A* since not all faculty in a university are actually submitted for evaluation in the RAE.

There is a difficulty in the interpretation of quality and productivity in the above constructions. This problem is also present in the original RAE, with individuals normally being assessed on their best four publications over the time period of the study. It would be reasonable to expect individuals to be rated based upon the average quality of the assessed publications but there is no uniformity in this, with the History panel now assessing only the best one of the four publications (McNay, 2003). This latter example is almost a pure quality assessment, but an averaging process necessarily incorporates an element of quantity: a good rating is based upon producing *enough* (four per submitted faculty) high quality publications. The standard interpretation of the RAE values, however, is as purely indicators of quality (Mayfield University Consultants, 2000; Education Guardian, 2001). Our notion of research productivity is 'quality times quantity': how much good research an institution produces. This is a simplification of a complex issue: for example an institution would be judged to have the same research productivity if it doubled the number of academics submitted to the RAE but their average score was numerically halved. Nevertheless this seems to be a logical approach to estimating research productivity and receives support from significant correlations with Web data (Thelwall, 2002a; Thelwall & Harries, 2003; Thelwall & Wilkinson, 2003a).

6. RESULTS

6.1 Inlinks and Online Impact

The total number of Web pages retrieved in the crawl was 5,438,941, containing a total of 34,041,179 links of all kinds. After restricting the data to just the publicly indexable set, and applying the domain ADM, there were 7,493 domains with a total of 82,672 domain ADM links between them, approximately 11 inlinks each. The excluded links included all links to domains not crawled (because they were not to a UK university site or were outside the publicly indexable set) as well as links between different domains owned by the same university.

In order to assess whether (domain based) sites created by or on behalf of higher RAE-scoring researchers attracted more domain ADM site inlinks we correlated inlinks with university RAE research quality and found a small but highly significant correlation, a Spearman's rho value of 0.082, significant at the 0.1% level (n=7,493). This shows that there is a significant positive relationship but its effect on inlink counts is extremely small. We confirmed that the effect was small by correlating the median inlink count for domains in each university against the university RAE, getting a Spearman's rho value of 0.180, which was not significant even at the 5% level (n=111). Spearman tests were used instead of Pearson in both cases because all data sets failed Kolmogorov-Smirnov normality tests. The median distribution is illustrated in Figure 1. The outliers are all universities for which only one or two domains were crawled, so the median is dominated by the count of inlinks to the home domain, which typically has a much larger count than the rest of the domains in a site. This shows how the statistics can be inappropriate for individual universities due to site design considerations. Removing outliers *can* be a permissible step in a statistical analysis, and in our case the correlation after these have been removed is 0.232 (n=108) which is significant at the 5% level, but is still a numerically small value that does not change the conclusions.

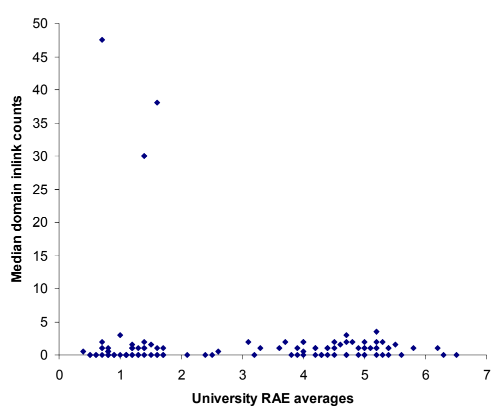
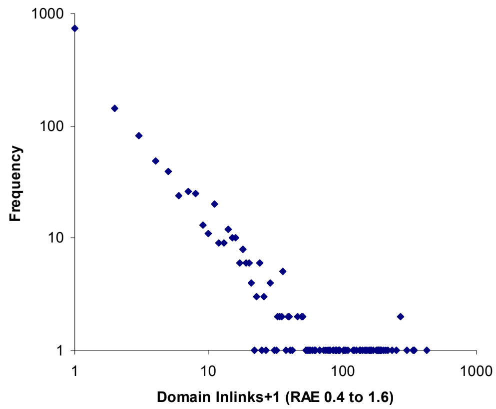


Figure 1. Median domain inlink counts show little relationship to RAE averages for the 111 UK universities.

Figures 2 and 3 give the frequency distribution of inlinks for the domains with the lowest RAE scores and those with the highest respectively. The inlinks have 1 added to them in the graph to allow them to be plotted on a logarithmic scale. We chose the RAE ranges to give similar numbers of domains for each graph. As can be seen, the basic shapes are similar in both cases, confirming that the difference in inlink distributions between the two is minor. The graphs displaying the approximate power laws (Lotka, 1926) that are commonly found on the Web (Rousseau, 1997; Barabási & Adamic, 1999; Thelwall & Wilkinson, 2003b). We tested for genuine power laws for the non-zero inlink counts using Rousseau's (2000) program implementing Nicholls' (1986) test and both failed. As a result we were unable to further statistically compare the two distributions.



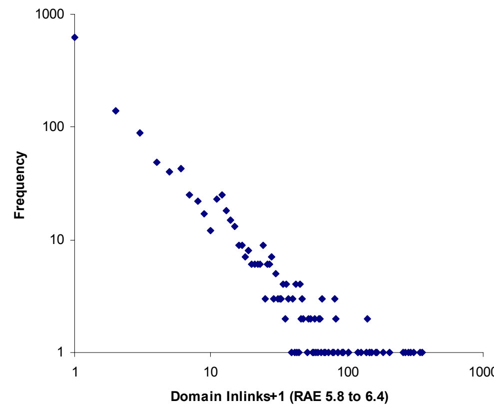


Figure 2. A log-log scale frequency diagram for inlinks to domains from universities with RAE averages between 0.4 and 1.6. Total domains = 1353.

Figure 3. A log-log scale frequency diagram for inlinks to domains from universities with RAE averages between 5.8 and 6.4. Total domains = 1348.

The small correlations found so far would not be enough to explain the strong association between links to a university and its research productivity found in a previous study of the 2001 UK academic Web (Thelwall, 2002a). To confirm this, we calculated the Spearman correlation coefficient for our data set (i.e. domain ADM inlink count totals for each university) with RAE averages. The resulting correlation was a very high 0.914 (n=111, significant at the 0.1% level). After factoring out university sizes by dividing both data sets by university staff numbers, the result was a still highly significant 0.784 (n=111, significant at the 0.1% level). The strong trend can clearly be seen in Figure 4. Note that this second calculation is effectively a comparison of Ingwersen's (1998) Web Impact Factor (in a subsequently modified version (Thelwall, 2001c)) with what we have termed RAE research quality, but there are two differences: the domain ADM is being used and the inlinks are restricted to the publicly indexable set.

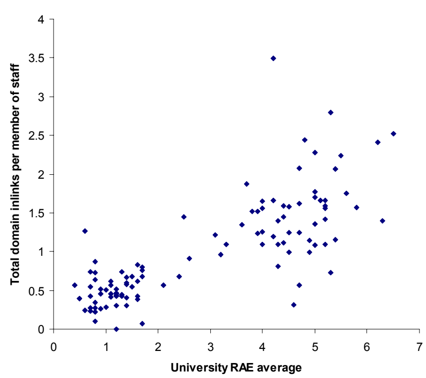


Figure 4. Domain inlinks per member of staff against RAE averages for 111 UK universities.

So why are total domain ADM inlink counts for universities with higher RAE scores much higher than those for universities with lower scores, with the difference being striking rather than just marginal? The second research question is an attempt to answer this: do higher rated scholars simply produce more Web content?

6.2 Research Quality and Web Productivity

For the second question we correlated total research productivity with total PI domains for each university. This time the correlation value was much higher, a Spearman value of 0.762, significant at the 0.1% level (n=111). Universities with higher total research productivity therefore produce many more domains.

Part of the correlation may again be due to size issues: larger universities would be expected to have, in general, both larger Web sites and higher total research productivity. In order to factor out size we repeated the test with both variables divided by institutional academic staff numbers. In other words we were comparing university RAE scores with total PI domains per member of staff. The results were still significant, but with a lower correlation of 0.509, significant at the 0.1% level (n=111). See Figure 5 for a graph of the data. Universities with higher average research quality have many more domains per member of staff.

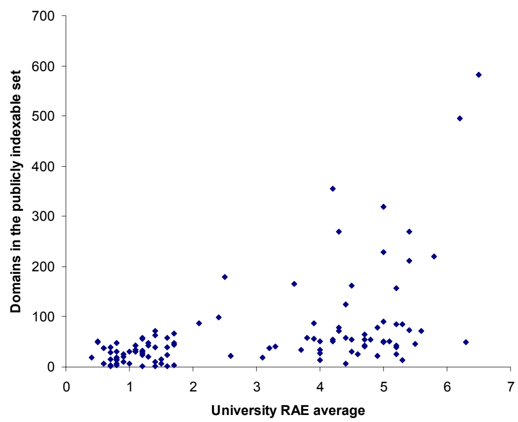


Figure 5. Domains per member of staff against RAE averages for 111 UK universities.

6.3 Triangulation with the Page ADM

Although the domain ADM is the best model for reliable link analysis, we also conducted a similar analysis for the page ADM as part of a triangulation approach, and specifically to address the obvious concern that the domain ADM results will be influenced for individual universities by their university domain naming strategy. The page ADM is the standard counting of links by individual pages rather than aggregating into whole domains. Trivially, median inlinks per page do not correlate with university RAE averages, because these are zero for all universities: most pages are not linked to from other universities. As a second-best choice, we also correlated mean inlinks per page with RAE averages and found the result of 0.182 to be insignificant at the 5% level (n=111). By contrast, the number of pages per member of staff correlated highly significantly with university RAE averages, with a value of 0.641 (n=111, significant at the 0.1% level). In summary, high inlink counts to universities with higher rated scholars are better explained by the quantity than the impact of the pages produced.

7. DISCUSSION

Our findings can be summarized as evidence to suggest that, on average, higher rated scholars produce much more Web content that is of slightly higher quality.

There are a number of limitations of the study, many of which have already been discussed in the choice of methods above. We list here those that we see as the key ones.

- The crawler will have missed some domains, perhaps even many domains in individual sites.
- Universities have different policies for domain name use (Thelwall & Harries, 2003) and so our domain based sites are not a homogeneous construction, even though statistically, as discussed above, they are more reliable than page counts. The triangulation with the page ADM significantly reduces this concern.

- Sites take time to attract inlinks (Barabási & Albert, 1999; Vaughan & Thelwall, 2003; Pennock *et al*., 2001) and so universities with predominantly newer domains would be disadvantaged.
- Web use and linking is discipline-dependent (Tang & Thelwall, 2003), giving an advantage to universities with large excellent departments in high Web using specialisms.
- The study only covers the UK and although the Web is international, it has geographic patterns of use (Thelwall, 2002c) and so it is possible that cultures of use in other countries would be different, so the results would not necessarily extend to all other countries.

As a result of the first four concerns, we feel that the small correlation found between domain inlinks and university RAE averages may be attributed to another cause than actual site impact, however defined. For example lower RAE scoring universities being late starters in the Web could also be an explanation. We have more confidence in the much higher correlation for domain counts, which seems to be much more robust. Unfortunately, in the nature of the unknowns involved it is impossible to give valid statistical estimates of reliability, a problem that is inherent to any similar Web site counting study. In the final analysis our interpretation of the results has to be mediated by qualitative judgments, a mixed model (pragmatic) research paradigm (Tashakkori & Teddlie, 1998).

8. CONCLUSION

Universities with higher RAE quality staff produce significantly more (domain name based) Web sites, but those sites are only marginally higher in online impact, as measured in (domain ADM) inlink counts, if at all. The answer to the main research question is therefore negative: the study has not found evidence that higher rated scholars produce significantly higher impact Web content, only that they produce more. This was corroborated by the page-based counting triangulation.

There is an important implication for those wishing to use links to assess the quality of small groups of researchers by counting links to their Web output: it should not be done. Although inlink counts to pages do seem to reflect the usefulness of the page's or site's information to some extent, and on average, (Brin & Page, 1998; Kleinberg, 1999), and inlink counts to whole universities do associate with their research productivity (Thelwall, 2002a), it is not appropriate to interpolate from this to individual scholars or even small groups, because the key factor is content production. Since inlink counts are by their nature relatively unreliable, a better measure for small groups of researchers could be just total content produced, the factor that most explains the university research – inlink count association. Clearly, such a measure would only be suitable for *academic* studies of scholarly communication and productivity, but not for *evaluating* the researchers concerned since Web publications are not quality controlled and it would be easy to make a mockery of the results by artificially putting pages online and creating new domains. This is in contrast to traditional bibliometrics, with van Raan (2000) describing the number of international refereed journal publications produced by a research institute as being a "first but good" indication of output size. It should also be mentioned that at the level of small groups of researchers there will be big differences due to disciplinary and personal factors (Kling & McKim, 2000). For example scientists seem to publish much more on the Web than others (Tang & Thelwall, 2003) and individual scholars can sometimes create enormous numbers of Web pages (Thelwall & Harries, 2003).

It seems that scholars that are highly rated by peers for their research, and presumably attract more citations to their formal publications, only attract a similarly higher level of hyperlinks to their Web sites by being more prolific, and not as a result of any halo effect or higher intrinsic quality. This is an important step in developing a sense of how this critical medium is used for informal scholarly communication and should be taken into consideration when constructing new Web indicators. Previous Webometric results will also need to be reinterpreted in the light of these new findings.

9. ACKNOWLEDGMENTS

This work was supported by a grant from the Common Basis for Science, Technology and Innovation Indicators part of the Improving Human Research Potential specific programme of the Fifth Framework for Research and

Technological Development of the European Commission. It is part of the WISER project (Web indicators for scientific, technological and innovation research) (Contract HPV2-CT-2002-00015) (www.webindicators.org). The referees are thanked for their helpful advice.

10. REFERENCES

Abt, H. A. (2000). Do important papers produce high citation counts? Scientometrics, 48(1), 65-70.

Aguillo, I. F. (1998). STM information on the Web and the development of new Internet R&D databases and indicators. In: Online Information 98: Proceedings. Learned Information, 239-243.

Almind, T. C. & Ingwersen, P. (1997). Informetric analyses on the world wide Web: methodological approaches to 'Webometrics'. Journal of Documentation, 53(4), 404-426.

Amento, B., Hill, W., Terveen, L., Hix, D. & Ju, P. (1999). An empirical evaluation of User Interfaces for Topic Management of Web Sites. CHI 99 Conference Proceedings, pp. 552-559, Addison Wesley, New York.

Arasu, A., Cho, J., Garcia-Molina, H., Paepcke, A. & Raghavan, S. (2001). Searching the Web. ACM Transactions on Internet Technology, 1(1), 2-43.

Barabási, A. L. & Albert, R. (1999). The emergence of scaling in random networks. Science, 286, 509-512.

Bayer, A. E. & Folger, J. (1966). Some correlates of a citation measure of productivity in science, Sociology Education, 39, 381-390.

Björneborn, L. (2001). Small-world linkage and co-linkage. In: Proceedings of the 12th ACM Conference on Hypertext and Hypermedia (pp. 133-134). New York: ACM Press.

Björneborn, L. & Ingwersen, P. (2001). Perspectives of webometrics. Scientometrics, 50(1), 65-82.

Borgman, C. (1990). Scholarly communication and bibliometrics, California: Sage.

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., Carr, L. & Harnad, S. (2002). Evidence of hypertext in the scholarly archive. In Proceedings of ACM Hypertext 2002, 74-75.

Burnett, R. & Marshall, P. D. (2003). Web theory. London: Routledge.

Chu, H., He, S. & Thelwall, M. (2002). Library and information science schools in Canada and USA: A Webometric perspective. Journal of Education for Library and Information Science, 43(2), 110-125.

Cole, J. R. & Cole, S (1967). Scientific output and recognition: a study in the operation of the reward system in science. American Sociology Review, 32, 377-390.

Coulston, C. & Vitolo, T. M. (2001). A hypertext metric based on huffman coding. In: Proceedings of ACM Hypertext 2001, 243-244.

Cronin, B. & Shaw, D. (2002). Banking (on) different forms of symbolic capital. Journal of the American Society for Information Science and Technology, 53(14), 1267-1270.

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

Cronin, B. (1984). The citation process. London: Taylor Graham.

Cronin, B. (2001). Bibliometrics and beyond: Some thoughts on Web-based citation analysis. Journal of Information Science, 27(1), 1-7.

Davenport, E. & Cronin, B. (2000). The citation network as a prototype for representing trust in virtual environments. In: Cronin, B. and Atkins, H. B. (Eds.). The Web of knowledge: a festschrift in honor of Eugene Garfield. Metford, NJ: Information Today Inc. ASIS Monograph Series, 517-534.

Education Guardian (2001). About the tables, http://education.guardian.co.uk.

EICSTES (2003). http://www.eicstes.org/

Furnas, G. W. (1997). Effective View Navigation. In: Pemberton, S., CHI 97 Conference Proceedings, USA:ACM Press, 367-374.

Garfield, E. (1979). Citation indexing: its theory and applications in science, technology and the humanities. New York: Wiley Interscience.

Garg, K. C. (2003). An overview of cross-national, national, and institutional assessment as reflected in the international journal Scientometrics. Scientometrics, 56(2), 169-199.

Glänzel, W. Schubert, A. & Czerwon, H. –J. (1999). A bibliometric analysis of international scientific cooperation of the European Union (1985-1995). Scientometrics, 45(2), 185-202.

Glänzel, W. & Schubert, A. (2001). Double effort = Double impact? A critical view at international co-authorship in chemistry. Scientometrics, 50(2), 199-214.

Goodrum, A. A., McCain, K. W., Lawrence, S. & Giles, C. L. (2001). Scholarly publishing in the Internet age: a citation analysis of computer science literature. Information Processing & Management, 37(5), 661-676.

Halavais, A. & Garrido, M. (2003). Mapping networks of support for the Zapatista movement. In McCaughy, M, and Ayers, M. (Eds.), Cyberactivism: Online Activism in Theory and Practice. London: Routledge, pp. 165-184.

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

Holmes, A. & Oppenheim, C. (2001). Use of citation analysis to predict the outcome of the 2001 Research Assessment Exercise for Unit of Assessment (UoA) 61: Library and Information Management. Information Research, 6(2). http://informationr.net/ir/6-2/paper103.html

Ingwersen, P., Larsen, B. & Wormell, I. (2000). Applying diachronic citation analysis to research program evaluations. In: Cronin, B. and Atkins, H. B. (Eds.). The Web of knowledge: a festschrift in honor of Eugene Garfield. Metford, NJ: Information Today Inc. ASIS Monograph Series, 373-387.

Ingwersen, P. (1998). The calculation of Web Impact Factors. Journal of Documentation, 54(2), 236-243.

Jaffe, S. (2002). The next Research Assessment Exercise will probably include citation analysis. The Scientist. Available: http://www.the-scientist.com/yr2002/nov/prof1_021111.html, accessed 11 November, 2002.

Jimenez-Contreras, E., Delgado Lopez-Cozar, E., Ruiz-Perez, R., & Fernandez, V. M. (2002). Impact-factor rewards affect Spanish research. Nature, 417(6892), 898.

Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment, Journal of the ACM, 46(5), 604-632.

Kling, R. & McKim, G. (1999). Scholarly communication and the continuum of electronic publishing. Journal of the American Society for Information Science, 50(10), 890-906.

Landes, W. M. & Posner, R. A. (2000). Citations, age, fame, and the Web. Journal of Legal Studies, 29(319), 329-341

Larson, R. R. (1996). Bibliometrics of the World Wide Web: an exploratory analysis of the intellectual structure of cyberspace. Proceedings of the AISS 59th annual meeting, 71-78.

Latour, B. & Woolgar, S. (1986). Laboratory Life: the construction of scientific facts. (2 Ed). Princeton, NJ: Princeton University Press.

Lawani, S. M. & Bayer, A. E. (1983). Validity of citation criteria for assessing the influence of scientific publications: new evidence with peer assessment. Journal of the American Society of Information Science, 34(1), 59-66.

Lawrence, S., and Giles, C. L. (1999). Accessibility and distribution of information on the Web. Nature, 400, 107-110.

Lawrence, S. L. (2001). Online or invisible? Nature, 411(6837) 521.

Li, X., Thelwall, M., Musgrove, P. & Wilkinson, D. (2003). The relationship between the links/Web Impact Factors of computer science departments in UK and their RAE (Research Assessment Exercise) ranking in 2001. Scientometrics, 57(2), 239-255.

Lotka, A.J. (1926). The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences, 16, 317-323.

Marek, K., & Valauskas, E. J. (2002). Web logs as indices of electronic journal use: Tools for identifying a 'classic' article. Libri, 52(4), 220-230.

Mayfield University Consultants, (2000). League Tables 2000. The Times Higher Education Supplement, April 14, II-III.

McNay, I. (2003). Assessing the assessment: an analysis of the UK Research Assessment Exercise, 2001, and its outcomes, with special reference to research in education. Science and Public Policy, 30(1), 47-54.

Middleton, I., McConnell, M. & Davidson, G. (1999). Presenting a model for the structure and content of a university World Wide Web site. Journal of Information Science, 25(3), 219-227.

Miles-Board, T. Carr, L. & Hall, W. (2002). Looking for linking: associative links on the Web. In Proceedings of ACM Hypertext 2002, 76-77.

Moed, H., Burger, W. J. M., Frankfort, J. G. & van Raan, A. F. J (1985). The use of bibliometric data for the measurement of university research performance. Research Policy, 14, 131-149.

Moed, H. (2002). The impact-factors debate: the ISI's uses and limits. Nature, 415, 731-732.

Nel, D., van Niekerk, R., Berthon, J. & Davies, T. (1999). Going with the flow: Web sites and customer involvement. Internet Research, 9(2), 109-116.

Nederhof, A. J. & Noyons, E. C. M. (1992). Assessment of the international standing of university departments' research: A comparison of bibliometric methods. Scientometrics, 24(3), 393-404.

Nicholls, P.T. (1986). Empirical validation of Lotka's law. Information Processing & Management, 22, 417-419.

Page, L. (2001). Method for node ranking in a linked database, United States Patent 6,285,999.

Pennock, D., Flake, G., Lawrence, S., Glover, E. & Giles, C. L. (2002). Winners don't take all: Characterizing the competition for links on the Web. Proceedings of the National Academy of Sciences, 99(8), 5207-5211.

Quan-Haas, A. & Wellman, B. (2003, to appear). How does the Internet affect social capital? In: Huysman, M. & Wulf, V. (Eds). IT and Social Capital.

Rees, M. (2002). Not worth the paper. New Scientist, 2370, 27.

Rehm, G. (2002). Towards automatic Web genre identification. In: Proceedings of the 35th Hawaii International Conference on System Sciences. (http://www.computer.org, electronic restricted access document).

Research Assessment Exercise (2003). Available: www.rae.ac.uk.

Rodríguez Gairín, J. M. (1997). Valorando el impacto de la informacion en Internet: AltaVista, el "Citation Index" de la Red. Revista Espanola de Documentacion Cientifica, 20, 175-181. http://www.kronosdoc.com/publicacions/altavis.htm

Rogers, R., & Marres, N. (2000). Landscaping climate change: A mapping technique for understanding science and technology debates on the world wide Web. Public Understanding of Science, 9, 141-163.

Rousseau, R. (1997). Sitations: An exploratory study. Cybermetrics, 1(1), Retrieved July 30, 2002, from http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html

Rousseau, R. (2000). LOTKA: A program to fit a power law distribution to observed frequency data. Cybermetrics, 4, Retrieved January 31, 2003, from http://www.cindoc.csic.es/cybermetrics/articles/v4i1p4.html

Seng, L.B. & Willett, P. (1995). The citedness of publications by United Kingdom library schools. Journal of Information Science, 21, 68-71.

Shneidermann, B. (1998) Designing the User Interface. Reading: Addison-Wesley.

SIBIS (2003). Statistical Indicators Benchmarking the Information Society. http://www.sibis-eu.org/sibis/

Small, H. (1999). Visualising science through citation mapping. Journal of the American Society for Information Science, 50(9), 799-812.

Smith, A. G. & Thelwall, M. (2002). Web Impact Factors for Australasian universities. Scientometrics, 54(3), 363-380.

Smith, A. G. (1999). A tale of two Web spaces: comparing sites using Web Impact Factors. Journal of Documentation, 55(5), 577-592.

Smith, L. C. (1981). Citation analysis. Library Trends, 30, 83-106.

Tang, R. & Thelwall, M. (2003, to appear). Disciplinary differences in US academic departmental Web site interlinking. Library and Information Science Research.

Tashakkori, A. & Teddlie, C. (1998). Mixed methodology. London: Sage.

Terveen, L., Hill, W. & Amento, B. (1999). Constructing, organizing, and visualizing collections of topically related Web resources. ACM Transactions on Computer-Human Interaction, 6(1), 67-94.

Thelwall, M. & Harries, G. (2003). The connection between the research of a university and counts of links to its Web pages: An investigation based upon a classification of the relationships of pages to the research of the host university. Journal of the American Society for Information Science and Technology, 54(7), 594-602.

Thelwall, M. & Smith, A. G. (2002). A study of the interlinking between Asia-Pacific university Web sites. Scientometrics, 55(3), 363-376.

Thelwall, M., Tang, R. & Price, E. (2003). Linguistic patterns of academic Web use in Western Europe. Scientometrics, 56(3), 417-432.

Thelwall, M. & Wilkinson, D. (2003a). Three target document range metrics for university Web sites. Journal of the American Society for Information Science and Technology, 54(6), 489-496.

Thelwall, M. & Wilkinson, D. (2003b). Graph structure in three national academic Webs: Power laws with anomalies. Journal of the American Society for Information Science and Technology, 54(8), 706-712.

Thelwall, M. (2000). Web Impact Factors and search engine coverage. Journal of Documentation, 56(2), 185-189.

Thelwall, M. (2001a). Extracting macroscopic information from Web links. Journal of the American Society for Information Science and Technology, 52 (13), 1157-1168.

Thelwall, M. (2001a). A Web crawler design for data mining. Journal of Information Science, 27(5), 319-325.

Thelwall, M. (2001b). A publicly accessible database of UK university Web site links and a discussion of the need for human intervention in Web crawling. http://www.scit.wlv.ac.uk/~cm1993/papers/a_publicly_accessible_database.pdf

Thelwall, M. (2001c). Extracting macroscopic information from Web links. Journal of the American Society for Information Science and Technology, 52 (13), 1157-1168.

Thelwall, M. (2002a). Conceptualizing documentation on the Web: an evaluation of different heuristic-based models for counting links between university Web sites. Journal of the American Society for Information Science and Technology, 53(12), 995-1005.

Thelwall, M. (2002b). Research dissemination and invocation on the Web. Online Information Review, 26(6), 413-420.

Thelwall, M. (2002c). Evidence for the existence of geographic trends in university Web site interlinking. Journal of Documentation, 58(5), 563-574.

Thelwall, M. (2003). What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation, Internet Research, 8(3), paper no. 151. Available: http://informationr.net/ir/8-3/paper151.html.

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. (1999). Advanced bibliometric methods for the evaluation of universities. Scientometrics, 45(3), 417-423.

van Raan, A. F. J. (2000). The Pandora's box of citation analysis: Measuring scientific excellence – the last evil? In: Cronin, B. and Atkins, H. B. (Eds.). The Web of knowledge: a festschrift in honor of Eugene Garfield. Metford, NJ: Information Today Inc. ASIS Monograph Series, 301-319.

Vaughan, L. Q. & Hysen, K. (2002). Do Web link counts resemble citation counts: An empirical examination. ASLIB Proceedings, 54(6), 356-361.

Vaughan, L. & Shaw, D. (2003, to appear). Bibliographic and Web citations: What is the difference? Journal of the American Society for Information Science and Technology.

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

Vinkler, P. (2000). Evaluation of the publication activity of research teams by means of scientometric indicators. Current Science, 79(5), 602-612

Virgo, J. A. (1977). A statistical procedure for evaluating the importance of scientific papers. Library Quarterly, 47, 415-430.

Wilkinson, D., Harries, G., Thelwall, M. & Price, E. (2003a). Motivations for academic Web site interlinking: Evidence for the Web as a novel source of information on informal scholarly communication. Journal of Information Science, 29(1), 59-66.

Wilkinson, D., Thelwall, M. & Li, X. (2003b, to appear). Exploiting hyperlinks to study academic Web use. Social Science Computer Review, 21(3).

WISER (2003). http://www.webindicators.org.

Zitt, M., Bassecoulard, E. & Okubo, Y. (2000). Shadows of the past in international cooperation: Collaboration profiles of the top five producers of science. Scientometrics, 47(3), 627-657.