Pergamon

Library & Information Science Research
25 (2003) 437–458

Library & Information Science Research

U.S. academic departmental Web-site interlinking in the United States Disciplinary differences

Rong Tang[a,*], Mike Thelwall[b]

[a] School of Library and Information Science, Catholic University of America, Marist Hall 2nd floor, Washington D.C. 20064, USA. E-mail address: tangr@cua.edu (R. Tang).
[b] School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1SB, United Kingdom

Abstract

This article explores disciplinary differences in academic Web-site interlinking using the university departments of chemistry, psychology, and history. Research has suggested that Web-link counts are related to research productivity and geographic distance between source and target, but no previous Webometric studies have comparatively analyzed academic departments from different disciplines. This study shows large differences in Web use by discipline for both Web-site size and the extent of interlinking, with the history department making little use of the Web and the chemistry department the most. There are significant correlations between in-links and research impact for the psychology and chemistry departments, with a stronger association for the psychology department. There was little evidence, however, of a geographic trend in interlinking.

1. Introduction

Many information science researchers have asserted the value of citation ranking and have suggested that it reflects research quality (e.g., Clark, 1957; Cronin & Overfelt, 1994; Garfield, 1970, 1998; Koenig, 1982, 1983; Lawani, 1977; Narin & Hamilton, 1996; Virgo, 1977). Other researchers, however, have questioned the validity of applying bibliometric indicators blindly (e.g., MacRoberts & MacRoberts, 1986, 1989; Seglen, 1992), irrespective of types of citations or citers' motivations (Bavelas, 1978; Brooks, 1986; Garfield, 1965) and

* Corresponding author.

0740-8188/03/$ – see front matter
doi:10.1016/S0740-8188(03)00053-7

without consideration of domain dependency or country specificity (Aksnes, Olsen, & Seglen, 2000; Garfield, 1979; Moed, Burger, Frankfort, & van Ranan, 1985; Winclawska, 1996). Citation analysis also has been used to track the relationships among scholars, research groups, departments, institutions, or countries, through their formal scholarly output, journal articles, and, in some cases, conference papers (Borgman & Furner, 2002).

The development of the Internet and Web technology has made possible a new kind of analysis, called "Webometric" analysis. Although bibliometrics is the quantitative study of patterns in formal scholarly communication, such as in books, journals, and other printed materials, Webometrics is the quantitative study of Web-based information resources, including patterns of use. It is proposed that, parallel with journal impact factors, Web impact factors (Ingwersen, 1998) based on counts of links to a page or site, or "sitations" (Rousseau, 1997), may be used to model informal communication between different academic institutions.

Although Web data are not proven to be reliable enough to be used as a primary means of assessing the influence of a body of research, there are still several important reasons to pursue this line of enquiry. First, informal scholarly communication is a key part of the research process, but one that has been difficult to capture in the past because it is typically conducted through communication channels that are temporal, private, or hidden. Examples are phone conversations, face-to-face meetings, and e-mail messages. The Web, which is largely publicly accessible, makes data collection possible. Second, the Web is an important medium for scientists to communicate their findings and activities to others. It is useful to be able to assess how effective researchers are in taking advantage of the Web to disseminate their scholarly work, with the objective of identifying areas of good practice and situations requiring remedial attention (e.g., Thelwall, Binns, Harries, Page-Kennedy, Price, & Wilkinson, 2002). In bibliometrics, numerous studies have investigated the relationship between interlinking counts of a selected set of universities and the research productivity of those universities. Positive correlations were found for universities in the United Kingdom (Thelwall, 2001a), Australian universities (Smith & Thelwall, 2002), and universities in mainland China (Tang & Thelwall, 2002) and Taiwan (Thelwall & Tang, 2002). In addition, the number of links between universities may be dependent on geographic distance (Thelwall, 2002a), and counts of links to departments may be associated with the research reputation of these departments (Chu, He, & Thelwall, 2002; Li, Thelwall, Musgrove, & Wilkinson, 2003). All of these findings are so far limited to specific localities—the countries (and discipline in the last case) that were studied—but they showed that patterns can be extracted from links and that aggregations of links bear some relationship to research.

Although disciplinary differences in citation characteristics have been discussed in several bibliometric studies, they have not yet been subjected to a major investigation in Webometric analysis. Vaughan and Thelwall (2003) found statistically significant disciplinary differences for links to academic journal Web sites. This study, however, investigated links to academic departments in universities in the United States and associated them with the estimated citation impact of each department (defined later) to explore and suggest disciplinary characteristics for "sitation" patterns.

2. Literature review

2.1. Disciplinary differences in citation patterns

Two major applications of citation analysis, as reviewed by Smith (1981), are to identify scholarly communication patterns and to conduct evaluative bibliometrics. Although the former uses citations as indicators of scientific communication patterns, the latter uses citations as a measure of scientific influence and productivity. In evaluative bibliometrics, scholars long ago issued a caution concerning improper comparisons between citation counts generated in different fields because "citation potential can vary significantly from one field to another" (Garfield, 1979, p. 248). In summarizing research work on citation analysis, Garfield (1979) pointed out that variations were found in several citation characteristics, such as the size of the field, the degree of integration, and the maturity of the research. The citation potential of a field is also characterized by "how quickly a paper will be cited, how long the citation rate will take to peak and how long the paper will continue being cited" (p. 248).

Moed, Burger, Frankfort, and van Ranan (1985), at the University of Leiden, conducted a study on field-dependent citation practices and concluded that "rankings of publications from different fields, based on citation counts, can be affected seriously by differences between citation characteristics in those fields" (p. 177). Lange (1985), who investigated country differences on citation habits, found evidence that national factors affected scientists' citation behaviors. Winclawska (1996), on the other hand, addressed the distinction between hard and soft science research. She suggested that using the Social Science Citation Index (SSCI) and the Arts and Humanities Citation Index (AHCI) for bibliometric analysis would likely produce unreliable results because research productivity in social science and the humanities is highly culturally oriented, which is different from the relative universality of the hard sciences. By focusing on journal articles, the SSCI and AHCI do not capture the publication nature of soft science research. In those sciences, "the basic mean[s] of communication is still a monograph, in contrast to hard core sciences where scholars communicate by means of a journal paper or report" (Winclawska, 1996, p. 388). Winclawska further investigated the inefficiency of the SSCI in covering non–English-language materials and non-American authors. On the distinction between monograph citation and journal citation in social science, Line (1979) found that the reference pattern of monographs differs from journals in terms of the date distributions, forms of material cited, subject distributions, and countries of publication.

2.2. Bibliometric analyses of departments in a single country

One of the general approaches identified by the bibliometricians Nederhof and Noyons (1992) for the assessment of the research standing of university departments is comparison with similar departments nationally. Most of the citation studies that concerned the data within a country covered departments within the same subject area. Several of these studies have concentrated on the correlation between the ratings by the peer-review–based Research Assessment Exercises (RAEs) in the United Kingdom and the citation counts of departments in one discipline. Several studies have looked at the performance by university library and information science (LIS) departments in the United Kingdom. Seng and Willett (1995) and Oppenheim (1995) revealed the correlation between citation counts and department ratings in the 1992 RAEs. Holmes and Oppenheim (2001) further used citation data to predict the 2001 RAE ranking of LIS schools in the United Kingdom. Similar studies on other disciplines include a study by Zhu, Meadows, and Mason (1991) on chemistry departments and a study by Oppenheim (1997) on genetics, anatomy, and archeology departments in British universities. Correlations between citation counts and RAE scores were found in both studies, and results of the first study demonstrated the key role of the most productive researchers in departments. Within a discipline, citation analyses are able to give valuable information, although it is disputed whether they could be used to substantially replace peer review for the assessment of research quality (Holmes & Oppenheim, 2001; MacRoberts & MacRoberts, 1989; Moed, 2002; Warner, 2000).

Although the Institute for Scientific Information (ISI) citation data have been considered biased toward U.S. scholars, seldom has any study investigated the departmental performance in U.S. universities, including multiple departments across various disciplines. Most previous citation research that focused on the United States was oriented to a particular field, such as a study on psychology by Clark (1957). For the purpose of studying academic historians' use of electronic reference resources, Graham (2001) randomly selected 30 history departments in the continental United States as her sample and found little evidence of their use in citations.

2.3. General properties of Web links

Web links have been studied by scholars from several disciplines using very different perspectives. In theoretical physics, the emphasis has been on developing mathematical models for the underlying properties of the network created by pages and links, in addition to other topologically similar types of networks (Adamic & Huberman, 2000; Watts & Strogatz, 1998). Computer science contributed a revised model to this debate (Pennock, Flake, Lawrence, Glover, & Giles, 2002), which showed that when a new link is created, pages that already have many links pointing to them are statistically more likely to be selected as the target of the link than are poorly linked pages (the "rich get richer" or "Matthew" effect). This phenomenon is balanced by a countervailing tendency, which varies by type of page, for even poorly linked pages to have a small chance of being targeted by a proportion of the newly created links. A commonsense explanation for this finding is that pages can be linked to not merely because they are well known but also because they are useful or their owner has a good reputation. Web links also have been the subject of information retrieval research by computer and information scientists. Such studies have used Web-link structures essentially on their own (Brin & Page, 1998; Flake, Lawrence, Giles, & Coetzee, 2002), or in conjunction with page text analysis (Gao, Walker, Robertson, Cao, He, Zhang, & Nie, 2001; Kleinberg, 1999). Even though empirical evidence for the efficacy of using links in this context is lacking, the widespread belief in their efficacy is fuelled by qualitative analyses and case studies (Brin & Page, 1998; Kleinberg, 1999), as well as the effectiveness and popularity of Google (Sullivan, 2002; Thelwall, 2002b) with its link-based PageRank algorithm trumpeted as the reason for success: "The heart of our software is PageRank™" (Google, 2002).

2.4. Web-link analyses of departments

Research on scholarly communication from the information science perspective has taken a different approach, using techniques derived from bibliometrics to study conceptually coherent networks of pages or sites: for example, a topic-specific collection (Larson, 1996), groups of countries (Ingwersen, 1998), and groups of universities (Thelwall, 2001a). Social network analysts also have used links to study collections of sites (Garrido & Halavais, 2000; Park, Barnett, & Nam, 2002). The most developed area of study is that of the whole set of universities within a single country (Smith & Thelwall, 2002; Thelwall, 2001a, 2002e; Thelwall & Tang, 2003) from which positive results about relationships of links to research productivity have emerged after early negative results (e.g., Thelwall, 2000). Departments also have been the targets of Webometric analysis, but early research did not produce statistically significant findings (Thomas & Willett, 2000). One recent study found that the counts of links to LIS departments in the United States correlated with their *U.S. News and World Report* ranking (Chu et al., 2002). A later study found a similar connection between research and in-link counts for U.K. computer science departments (Li et al., 2003). Nonetheless, no causative connection is claimed between research and link targeting, although such claims have been controversially put forward for the more formal arena of journal article citations (Borgman & Furner, 2002; Cronin, 1984).

Most academic Web-link studies have investigated the reason for the statistical association between links and research to some extent, but one recent article focused on this single issue (Wilkinson, Harries, Thelwall, & Price, 2003). Although the vast majority of links between U.K. universities were connected in some way with scholarly activities (including teaching), less than 1% were formally referencing journal-quality publications. Counts of links, therefore, represent predominantly an amalgamation of many different causes loosely related to research and various forms of informal scholarly communication. One consequence of this finding is that their use to evaluate research impact fails the validity test.

3. Research questions

Despite the empirical evidence for the efficacy of studying links as an aspect of informal scholarly communication among departments, at the departmental level thus far there have been no comparisons of disciplines and no geographic studies. The current study fills these gaps using data collected from the United States and addresses the following research questions:

1. Are there clearly identifiable disciplinary differences in basic Web-site use and interlinking?

2. Do departments that conduct more and better research attract more links from their peers? Does the answer to this question vary by discipline?
3. Is Web-site interlinking affected by geographic factors for departments within a single discipline?

4. Methods

The researchers explored the previously described research questions with the Web sites of a selection of departments in each of three disciplines: history, psychology, and chemistry. These disciplines were chosen in an attempt to select from three different scholarly domains—science, the humanities, and behavioral science—and to avoid an overlap with previous single-discipline studies.

4.1. Variables, data source, and methodological specifications

4.1.1. Departmental estimated research productivity

For the three targeted disciplines, the authors gathered data to estimate the citation-based impact of each department. The raw data concerned (1) the top 30 departments with regard to the number of publications in the year 2000 in each discipline and (2) the ISI citation impact factor arranged by fields and by universities.

4.1.1.1. Top 30 departments ranked based on year 2000 publications. For the three disciplines, several ISI citation databases were searched. By examining the dominating number of publications and overlaps, it was determined by the first author (R.T.) that the best database for chemistry is the Science Citation Index (SCI); for psychology, SSCI; and for history, AHCI. In addition, it was determined that the year 2000 publications would be considered. At the time of the data collection, the 2002 data were incomplete and the year 2001 was not considered because of concern about the September 11th event influencing the normal publication pattern. Searches were conducted via dialog link in file 34 for chemistry, file 7 for psychology, and file 439 for history. The exact command procedure was as follows, with separate search commands for each discipline included in the brackets:

S SC = chemistry [SC = psychology, SC = history]

S CS = (dept? (3N) chem?) [CS = (dept? (3N) psychol?), CS = (dept? (3N) hist?)]

S PY = 2000 and GL = USA

S S1 AND S2 AND S3

This sequence allowed the search to be limited to a particular subject category with publications from a particular department, publication year as 2000, and the location as the United States. After this procedure was completed, a rank command was issued: RANK CS.

Upon issuing the rank command, the publication count in chemistry was more than 10,000, which exceeds the limit of ranking set. Two rank commands were then issued to perform the ranking for the first 10,000 publications and then another for the remaining items. The top 100−ranked departments from both lists were merged and numbers were added to produce the entire ranking for chemistry.

The CS ranking function in dialog may produce errors in data because of the same department being indexed with different names, and thereby given a different entry in the ranking. To compensate for this, the first author selected the top 100 CS for each discipline, double-checked for duplicate entries and made corrections for those that were in fact the same department or institute but were counted separately. For instance, the University of Illinois' Department of Psychology was listed four times within the top 100 most published departments as ranks 2, 14, 61, and 69. Having verified that all four were the same department and were indexed differently (e.g., Rank 2 UNIV ILLINOIS, DEPT PSYCHOL, 603 E DANIEL ST/CH; and Rank 14 UNIV ILLINOIS, DEPT PSYCHOL/URBANA//IL/61801), the item numbers from each rank were added together to make the total departmental publication count. In the end, the top 33 departments were selected for each discipline. The first author (R.T.) deliberately collected a few more departments from each discipline as a precaution, and ended up with a sample of 89 departments for Webometric analysis. There were fewer than 25 departments in history because of the relatively scarce publication count and missing data, and a little more than 30 departments in psychology and chemistry.

4.1.1.2. ISI citation impact factor. Because the number of publications by a department is not the same as the number of citations to publications by that department, and because the authors were more interested in the "sitation" counterpart of the bibliometric indicator, a decision was made to get specific citation counts for each department in the selection. The first author (R.T.) therefore contacted ISI Contract Research Services and obtained the citation impact factor (i.e., average citations per article) for the years 1997 to 2001 for universities (not necessarily from only the particular department that was selected) by fields of chemistry, psychology, and history.

4.1.1.3. Estimated citation impact of departments. Given that the ISI citation impact data were oriented to a university and not necessarily the particular department in that university, the authors multiplied the number of publications from a department with the citation impact factor of a given field from the university of that department. This creates a count that was operationally defined as the estimated citation impact (ECI) of a department. For instance, suppose the publication counts in the year 2000 for the Department of Psychology at the University of Illinois at Urbana-Champaign is X, and the ISI citation impact factor in the field of psychology for the University of Illinois at Urbana-Champaign is Y. The ECI for the Department of Psychology at the University of Illinois at Urbana-Champaign is as follows:

$$\mathrm{ECI} = X * Y$$

Anew ranking was thus generated for each discipline based on the ECI of the top 24 to 33 departments in that discipline.

4.1.2. Counts of links between and to departments

To count accurately the number of links between each pair of departments in all of the sets, the associated Web sites first were identified and then all pages were visited and checked for links. The Web sites were identified by searching universities' Web sites and locating relevant departments. Each site was then visited by a specialized crawler (Thelwall, 2001b), which is designed for accurate and complete crawling of specified sites, including the elimination of duplicate pages. Coverage was limited to only those pages that were findable by following links from the home page and not banned to crawlers (Thelwall, 2002c). This process created a database of the link structures of each site, which was then used to calculate the total number of links between each pair of departments and the total links to each department from all the other departments.

Previous research has found that counting links between pages is not always the best approach; counting between domains and directories is also useful when pages are grouped together into an alternative domain or directory document (Thelwall 2002e). The four alternative document models are described in the following list. These models are used to give three more link-counting methods.

1. Page: Each separate HTML file is treated as a document for the purposes of extracting links. Each unique link uniform resource locater (URL) is treated as pointing to a separate document for the purposes of finding link targets. URLs are truncated before any internal target marker "#" character found, however, to avoid multiple references to different parts of the same page.
2. Directory: All HTML files in the same directory are treated as a document. All target URLs are automatically shortened to the position of the last slash, and links from multiple pages in the same directory are combined and duplicates eliminated.
3. Domain name: Same procedure as above, except all HTML files with the same domain name are treated as a single document for both link sources and link targets. In particular, this clusters together all pages hosted by a single subdomain of a university site.
4. University: Same procedure as above, except that all pages belonging to a university are treated as a single document for both link sources and link targets (Thelwall, 2002e).

The authors also wished to investigate links to the departments from two other arenas: all U.S. universities and the rest of the Web. It was not possible to crawl either of these spaces because of their size; therefore, AltaVista advanced searches were used. Commercial search engines have been criticized for a lack of reliability and for results and coverage being outside of the control of the researcher (Bar-Ilan, 2001; Björneborn & Ingwersen, 2001; Snyder & Rosenbaum, 1999). Consequently, their results should be treated with caution.

For U.S. university links, the .edu domain was used as a proxy. Some universities not based in the United States and non-university entities also have an .edu ending, but for the purpose of counting links to departments, the researchers assumed that this would not be a significant limitation. The queries used excluded all links from the host university, as the focus was on external impact. A sample query is as follows:

link : chem.umich.edu AND
domain : edu AND NOThost : umich.edu

For all Web links, the host university was again excluded.

link : chem.umich.edu AND NOT host : umich.edu

The focus on external impact (links originating outside the host institution) follows established practice (e.g., Chu et al., 2002; Li et al., 2003), but there is a difference with the citation indicator used. Although the value of self-citations has long been disputed (Garfield, 1974, Snyder & Bonzi, 1998), and even citations to authors within the same department may be a biased source of evidence (Cronin & Shaw, 2002), journal articles are subject to peer reviews that would limit excessive self-citations. It did not seem necessary to make a complex attempt to compensate for this with the scale of the current study. Link counts are different, however, in that links are typically used freely for navigation within a departmental site and sparingly to point to other sites, and are not subject to peer review for appropriateness. Other factors being equal, the inclusion of links internal to a site tends to give results that dwarf counts of links between sites and therefore measures only the size of the target site. The restriction to intersite links hence gives a data source that is more likely to be related to the concept of online impact.

4.1.3. Faculty numbers

To collect information about the number of faculty members of all the departments in the year 2000, two steps were taken. First, the first author (R.T.) contacted the American Historical Association, the American Chemical Society, and the American Psychological Association, in an attempt to gather most of the statistics for departments. The three organizations provided very helpful information; however, there were several departments' data that were unavailable. Second, all the individual departments that had missing statistics were contacted and asked to provide the information. In the end, faculty statistics on each of the 89 departments were obtained. Faculty numbers were used to normalize link counts as detailed in later sections.

4.1.4. Geographic distances between departments

Postal zip codes for all 89 departments were collected from departments' Web sites. The zip codes were converted into coordinates in meters using multimap.com. In one case, the chemistry department at the University of Houston, the zip code generated an error and so a neighboring zip code from the city of Houston was used instead. The coordinates were converted to distances between every pair of departments in each set using the standard straight-line measure. No correction for the earth's curvature was used. These distances are physical and do not really represent geographical distance in the human sense. For instance, two universities at opposite ends of California are far apart physically but perhaps conceptually closer together in the minds of researchers (e.g., because of a common name, shared state interests) than closer institutions in neighboring states. In addition, in a densely populated area, a university might have to be within 10 miles to be considered a neighbor, but in the Midwest, a neighboring university might be a thousand miles away, yet still be the closest institution. Physical distance is therefore only one form of distance. The concept of distance occurs in a variety of settings with different connotations (Bourgeois & Friedkin, 2001; Ghemawat, 2001; Redmond, 2000).

4.2. Data analysis

During the investigations, the researchers attempted to use multidimensional scaling on the data (as in Thelwall, 2002d), but the results did not appear to follow any pattern and were dramatically dependent on the exact variant used. As a result, this option was not pursued. The authors conjectured that the matrix of data was too sparse and the low individual counts too unreliable for this technique to work. The results and exact details of this approach were therefore omitted.

4.2.1. Basic Web-site use and interlinking by discipline

For each discipline, the total number of pages across all sites was recorded, in addition to the total number of interdepartmental links. This latter count was based on identifying all links originating in a page within one department that targeted any page within another from the same data sets. Self-links (i.e., links between different pages in the same department) were excluded because this study focused on interdepartmental communication. The two figures used are essentially measures of the size of Web and Web-link use in each disciplinary set of departments.

The authors also measured the degree of interlinking between departments by calculating the ratio of the total departments interlinked to the maximum possible interlinking. This was done by dividing the total number of ordered pairs of departments such that there is at least one link from any page in the first to any page in the second by the total number of ordered pairs in the data set (excluding same department pairs). The resulting figure is the fraction of ordered pairs of departments that are linked. For example, if there are X departments, then there would be $X^2 - X$ ordered pairs used and if there were Y of these pairs in which the first department linked to the second, then the statistic produced would be $Y/(X^2 - X)$.

4.2.2. Research impact and link counts

For each discipline, the extent to which the total links to and from each department correlate with its ECI can be assessed using the Spearman correlation coefficient. A failed Kolmogorov-Smirnov normality test precluded the use of the Pearson test. It would be expected that larger departments tend to conduct more research, to have larger Web sites, and to be more visible on the Web. It is logical, therefore, to attempt to cancel out this size variable by dividing both links to and from each department by total faculty numbers. Significant correlations between the resultant data sets would be more convincing as indicators of an association between citation impact and Web links. Results per faculty member are important to avoid results being skewed by individual large departments getting high values on both links to and from each department simply because of their size.

4.2.3. Geographic trends

To identify evidence of a geographic trend in Web-site interlinking (for whole universities), a previous study used a graphical technique (Thelwall, 2002a). A statistical hypothesis test was also added in this study. Even though it would be possible to directly correlate the distance data and the link data matrices, using the Spearman test, this is not appropriate for matrix data. Nonetheless, a calculation named the Quadratic Assignment Procedure (QAP) correlation is designed specifically for assessing how well the matrices match compared with matching them after random permutations of the rows and columns (Hubert & Schultz, 1976). The QAP figure is affected by size considerations if the entities associated with the nodes are of different sizes. As shown later, however, this factor does not seem to be a significant. Furthermore, Krackhardt (1992) reported that QAP is a flawed methodology because, for certain data distributions, a positive result would need to be verified by the construction of an explicit linking model. This turned out to be unnecessary for this data set, so the authors have not included a detailed description of the technique. In line with the previous study, the data were analyzed, after attempting to factor out a connection between research and links, by dividing each link count by the product of the estimated research productivity of the source and target universities.

5. Results

5.1. Basic Web-site use and interlinking by discipline

Table 1 presents differential use of the Web and Web interlinking by discipline. Because of the skewness of the page count and in-link count data, the authors used the medians for comparison purposes. An examination of the Table 1 shows a greatly varying amount of Web use and Web interlinking by departments, with linking differences too large to be explained by faculty numbers or department counts. Those using the Web more and linking more are also more comprehensively interlinked with their sibling departments, as the ratio statistic shows. For chemistry, the average department links to over a third of its peers, more than double the interlinking in psychology and history departments. This difference can be only partly accounted for by faculty numbers: larger departments could be expected to create more links, although there was not any evidence for such a trend in this data set (see later). The ratio calculation is explained in the Methods section.

The pattern of links to and from each department is described separately by discipline and then the results are compared at the end.

Table 1
Descriptive statistics for each of the disciplines

Variable	History	Psychology	Chemistry
Departments	24	32	33
Total pages	10,840	63,459	152,825
Median pages per department	113	603.5	1,786
Total in-links	3	275	1,268
Median in-links per department	0	6	31
Total departments interlinked/maximum possible interlinking	3/552 (1%)	139/992 (14%)	380/1056 (36%)
Total faculty	780	821	1,119.5

5.2. Research impact and link counts

5.2.1. History

There were only three links between history sites, rendering statistical tests inappropriate. AltaVista was still used to get counts of links to the departments, but, from Table 2, it can be seen that no significant relationship was found, with neither statistic being significant at the p = .05 level. The three queries used were as follows:

1. hnet.uci.edu/history/faculty/block/ TO princeton.edu/~history/
2. fas.harvard.edu/~hsdept/faculty/index.html TO sscnet.ucla.edu/history/hammonds/
3. history.ucla.edu/undergrad/pat/links.php TO orpheus.ucsd.edu/history/ancienthistory/links.htm

5.2.2. Psychology

Table 3 gives the results for comparing links with the ECI. It can be seen that size plays only a small factor in the correlations and that in-links to a department (its online impact) appear to be more strongly related to research. This finding is in line with a university-level study, where a similar conclusion was found (Thelwall, 2003). A broadly linear trend is a reasonable interpretation of Figure 1. The greater scatter for out-links is visible in Figure 2. The results from AltaVista were less emphatic than for the crawler data. Page models were graphed for clarity, because results were similar for each model. There was very little correlation between ECI and faculty numbers (p = .156, which is an insignificant value).

5.2.3. Chemistry

The results were similar to those for psychology in terms of the relative sizes of figures in rows and columns, but the association between links and research was weaker (Table 4;

Table 2
Spearman correlations for links versus estimated citation impact for history departments

Counting model	Links to department
AltaVista all in-links	– 0.040
AltaVista edu in-links	– 0.011

Table 3
Spearman correlations for links versus estimated citation impact for psychology departments

Counting model	Total links to department against ECI	Total links from department against ECI	Links to department per faculty member against ECI per faculty	Links from department per faculty member against ECI per faculty
File	0.613*	0.312	0.577*	0.381**
Directory	0.590*	0.327	0.563*	0.405**
Domain	0.567*	0.372**	0.539*	0.428**
University	0.571*	0.348	0.545*	0.430**
AltaVista all in-links	0.359**	—	0.346	—
AltaVista edu in-links	0.387**	—	0.419**	—

* Significant at p = .01.
** Significant at p = .05.

Figures 3 and 4). This phenomenon was also evident in Figures 5 and 6. Moreover, the AltaVista link data show no evidence of a relationship, and out-link correlations all fall below the 5% significance level. The ECI again did not correlate significantly with faculty numbers. The Spearman value was -0.059.

5.3. Geographic trends

For psychology and chemistry, the interlinking counts were large enough to be assessed for geographic trends. The QAP correlations produced were insignificant, however. The data were further examined by graphing (see Figures 5 and 6), and a slight trend can be seen for the chemistry departments, but not for psychology. Each psychology bar represents at least 30 department pairs except the last, which is for only 12. Each chemistry bar represents at least 44 department pairs except the third from last, which is for only 10. The spike in the psychology data at 2,000- to 2,500-km distance apart was mainly from a total

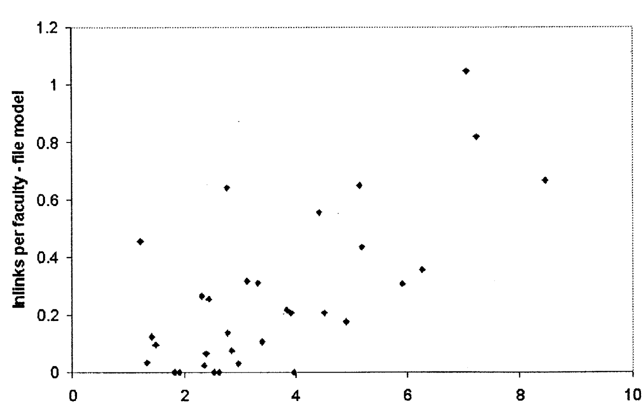

Fig. 1. Links per faculty member to each psychology department from the other psychology departments.

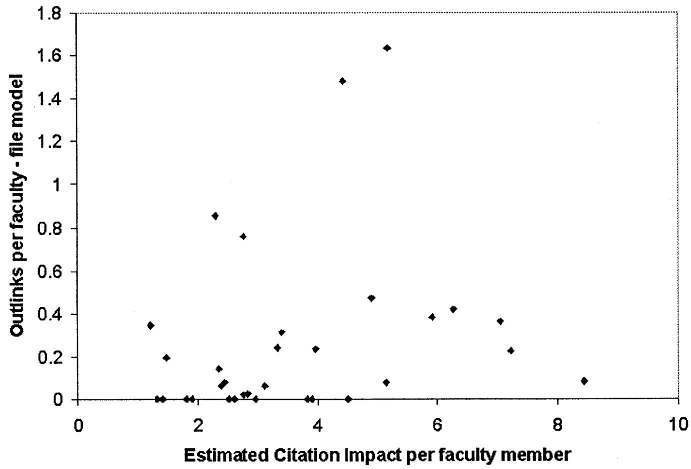

Fig. 2. Links per faculty member from each psychology department to other psychology departments.

of 12 links exchanged between the University of Texas and the University of California–Santa Barbara.

6. Discussion

In answer to the first research question, the Web is being used to a substantially different extent along disciplinary lines. This is evident from the statistics of Web-site size and interlinking in Table 1. The particular finding is an empirical validation of the Kling and McKim (2000) theories of differential Web use. The authors have claimed that this is an important finding—the first concrete evidence of its kind concerning sites from different disciplines crawled using a consistent method.

Even though there was no evidence of a significant association with research impact for history department links, such evidence was found for psychology and chemistry depart-

Table 4
Spearman correlations for links versus estimated citation impact for chemistry departments

Counting model	Total links to department against ECI	Total links from department against ECI	Links to department per faculty member against ECI per faculty	Links from department per faculty member against ECI per faculty
File	0.427*	0.267	0.356*	0.159
Directory	0.433*	0.287	0.389*	0.190
Domain	0.397*	0.336	0.353*	0.303
University	0.357*	0.337	0.357*	0.325
AltaVista, all in-links	– 0.029	—	0.021	—
AltaVista, edu links	– 0.004	—	0.073	—

* Significant at $p = .05$.

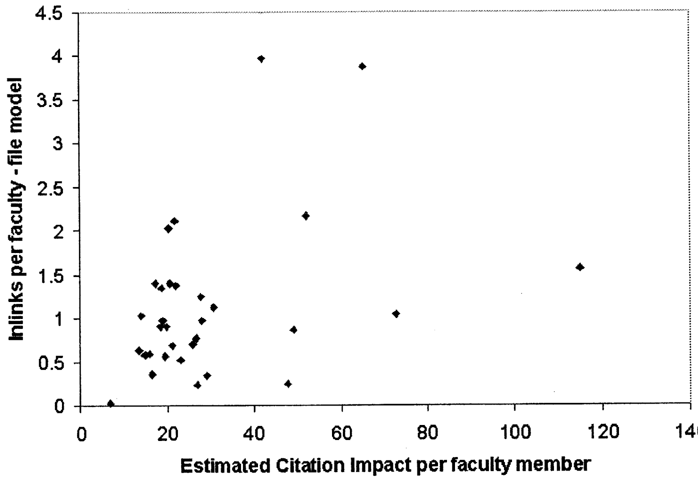

Fig. 3. Links per faculty member to each chemistry department from other chemistry departments.

ments. Given the range of departments now covered by all Webometric research, the current authors believe that link counts associate with research productivity or impact in most hard and soft science disciplines from economically advanced nations. It seems likely that this does not extend to areas that, because of economic factors or being a nonscientific discipline, have low overall use.

The AltaVista correlations between links and total ECI in both chemistry and psychology were similar but much lower and less significant than the disciplinary values obtained from the crawler. A possible explanation is reduced reliability of commercial search engine results, although this has not been found in a previous comparison of data from AltaVista and from the specialized crawler used here (Thelwall, 2001a). A more likely explanation is that Web

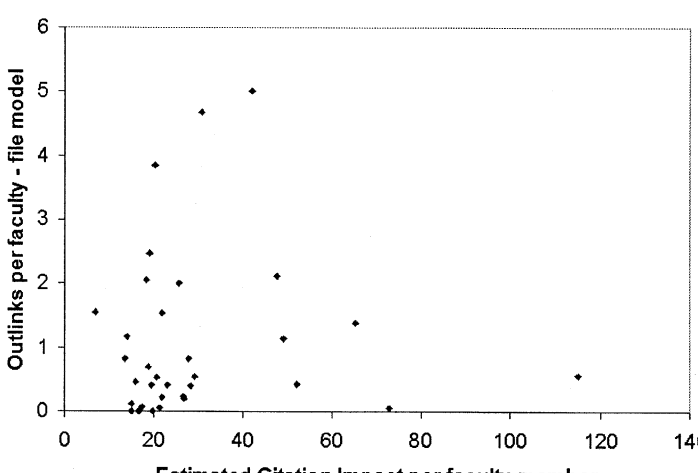

Fig. 4. Links per faculty member from each chemistry department to other chemistry departments.

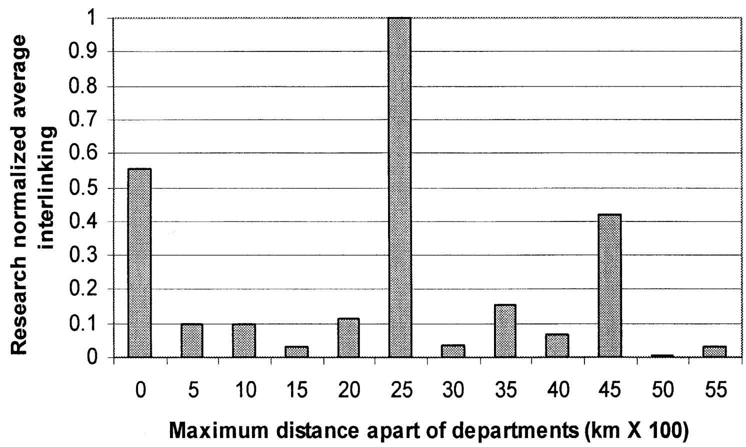

Fig. 5. Normalized average linking between psychology departments.

linking is more related to citation impact within a discipline than outside of it. This interpretation would be consistent with the logical possibility that scholars within a discipline are more influenced by publication quality in their own field than in others.

This study shows that different alternative document models did not make a significant contribution to mining the link data. The original rationale for using these models was to reduce anomalies. It seems, however, that outliers were not significantly present in this data set. Future research is needed to decide whether the application of document models is better confined to certain disciplines or countries.

The results failed to show geographic relationships with links in any discipline. There are three possible interpretations: (1) straight-line distance is not a meaningful measure in a geographically diverse country, such as the United States; (2) the department level is too small to generate statistically significant results; and (3) hyperlinking within a discipline is

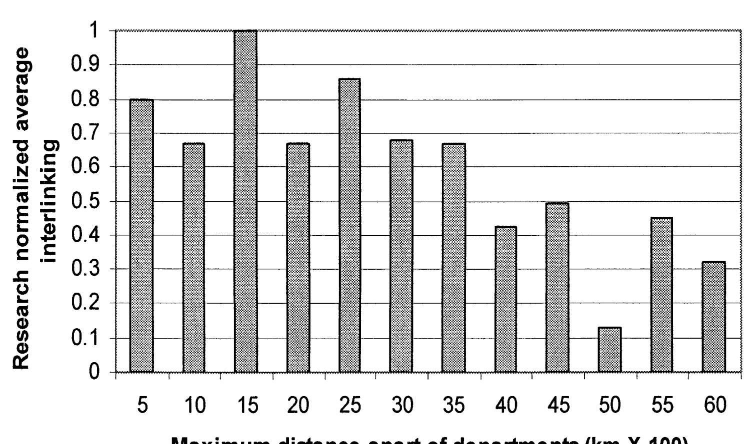

Fig. 6. Normalized average linking between chemistry departments.

not constrained by distance in the way that interdisciplinary connections are (Thelwall, 2002a). For psychology, the spike in Figure 5 illustrates how the low numbers involved in the calculations allow small, unusual factors to dominate the results. The chemistry results are more robust because of the larger numbers involved. It is possible that there is a trend, but this trend is subtler than that of research productivity association, and more data are required to produce the necessary averaging. Nevertheless, the lack of a clear and significant association with geographic distance is important new evidence. Such evidence demonstrates that distance does not make a great difference in same-discipline interlinking within the United States. This finding contrasts with the significant geographic factors found for university interlinking in the United Kingdom (Thelwall, 2002a).

7. Interpretation of results

The previous analysis of hyperlinks has been of a quantitative nature, and such an analysis is not connected to the actions of individual departments. Little is known in concrete terms about who creates hyperlinks in academic Web sites and why. One recent article has suggested that creation motivations can be much more trivial than for journal citations (Thelwall, 2003), even though most academic Web links seem related to scholarly concerns (Wilkinson et al., 2003).

Clearly, different groups, such as faculty, students, and administrators, can create Web pages as well as Web links. These links may be created as a result of an aspect of the person's role in the university or as a part of his or her private life. Links may be created singly with great deliberation, in large quantities as a part of an integrated site, or even at random using a computerized program. A considerable body of research is needed to establish a full understanding of academic link-creation processes and motives for link creation. Caution therefore should be exercised when interpreting link counts and interlinking patterns. Any meaningful interpretation should be augmented by statistical comparisons with other valid measures.

The authors found clear disciplinary differences in hyperlinking patterns, and it is tempting to conclude that the variations reflect differing patterns of Web use, or even Internet use. Alternative explanations are probable, however. For instance, it is possible that the three disciplines actually use the Web at a similar level of intensity, and that the difference found is only in Internet publishing. It is also possible that disciplines use different forms of computer-mediated communication, with humanities scholars preferring e-mail and discussion lists over the Web presentation of their work. The third possibility is that the departmental Web-publishing differences are attributable to higher employment of information technology (IT) staff in the sciences. This line of reasoning is undermined by the correlations found with ECI, unless the higher ECI is associated with the greater employment of supporting staff.

Based on the academic Web pages that the authors have visited in this and other research, they believe that pages with links to other universities have been predominantly created by academics rather than by the IT staff. Nevertheless, because of the diverse contexts of links

(Wilkinson et al., 2003), it is not possible to ascribe a unified motivation more specific than "informal scholarly communication." It also seems reasonable to speculate that Web linking does reflect, to some extent, academics' Web use, yet disciplinary differences would also be modified by differences in the technical knowledge to create pages. In addition, for individual departments, local policies may have an impact on Web publishing. In extreme cases, there may be controls to restrict Web publication or, conversely, encouragement to publish and link widely. The correlations found alleviate the concern that these phenomena may be widespread, but they can still affect individual departments.

The degree of interlinking of disciplines is closely related to the degree of Web publishing. If one assumes that faculty members are creating the pages, interdisciplinary links would be a natural thing for scholars to include in their Web pages, whether to link to important authors, colleagues, or information for the use of students or other researchers. The correlations found in this study suggest that, within a discipline, scholars who are producing more or better academic publications are also either using the Web more or simply publishing more pages and links.

8. Conclusion

The findings in this study have implications for applications of Web links in comparing departments and groups of departments. First, geographic location is not an influential factor for departmental interlinking in the United States. Second, Webometric comparisons of departments within the same discipline in economically advanced nations have some validity in assessing online impact from the perspective of identifying trends and areas of good practice and poor performance. Third, it would be unfair to compare departments from different disciplines for the purpose of identifying individual units that are examples of good or bad practice. For example, comparing all the departments of a single university with each other could give an advantage to the sciences and would therefore not be a fruitful exercise.

More than 10 years after the creation of the Web, the very low Web publishing and linking by history departments in the United States, as shown in this study, is a particular cause for concern (see also Graham, 2001). Historians might very well be missing an opportunity to disseminate their research findings and activities and to use the Web for other scholarly purposes. The findings of this study show that, even though hyperlinks may not necessarily serve as a direct valid data source for evidence of research impact, academic departmental links and their interlinking patterns can still be valuable for comparative analyses that help to establish a better understanding of disciplinary-specific Web behaviors.

Acknowledgments

The ISI Contract Research Services provided citation impact data. The authors also thank the American Historical Association, the American Chemistry Society, the American

Psychological Association, and relevant academic departments for providing information about the number of full-time faculty members in the departments.

References

Adamic, L. A., & Huberman, B. A. (2000). Power-law distribution of the World-Wide Web. *Science*, *287*, 2115a.

Aksnes, D. W., Olsen, T. B., & Seglen, P. O. (2000). Validation of bibliometric indicators in the field of microbiology: A Norwegian case study. *Scientometrics*, *49*, 7–22.

Bar-Ilan, J. (2001). Data collection methods on the Web for informetric purposes—a review and analysis. *Scientometrics*, *50*, 7–32.

Bavelas, J. (1978). The social psychology of citations. *Canadian Psychological Review*, *19*, 158–163.

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

Borgman, C., & Furner, J. (2002). Scholarly communication and bibliometrics. In B. Cronin (Ed.), *Annual Review of Information Science and Technology* (vol. 36, pp. 3–72). Medford, NJ: Information Today.

Bourgeois, M., & Friedkin, N. E. (2001). The distant core: Social solidarity, social distance and interpersonal ties in core-periphery structures. *Social Networks*, *23*, 245–260.

Brin, S., & Page, L. (1998). The anatomy of a large scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, *30* (Retrieved February 5, 2003, from http://citeseer.nj.nec.com/brin98anatomy.html).

Brooks, T. A. (1986). Evidence of complex citer motivation. *Journal of the American Society for Information Science*, *37*, 34–36.

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*, 110–125.

Clark, K. E. (1957). *America's psychologists: A survey of a growing profession*. Washington, DC: American Psychological Association.

Cronin, B. (1984). *The citation process: The role and significance of citations in scientific communication*. London, UK: Taylor Graham.

Cronin, B., & Overfelt, K. (1994). Citation-based auditing of academic performance. *Journal of the American Society for Information Science*, *45*, 61–72.

Cronin, B., & Shaw, D. (2002). Identity-creators and image makers: Using citation analysis and thick descriptions to put authors in their place. *Scientometrics*, *54*, 31–49.

Flake, G. W., Lawrence, S., Giles, C. L., & Coetzee, F. M. (2002). Self-organization and identification of web communities. *IEEE Computer*, *35*, 66–71.

Gao, J., Walker, S., Robertson, S., Cao, G., He, H., Zhang, M., & Nie, J.-Y. (2001). *TREC-10 Web track experiments at MSRA 384–392* (Retrieved February 5, 2003, from http://trec.nist.gov/pubs/trec10/t10_proceedings.html).

Garfield, E. (1965). Can citation indexing be automated? In M. E. Stevens (Ed.), *Statistical association methods for mechanized documentation* (pp. 189–192). Washington, DC: National Bureau of Standards.

Garfield, E. (1970). Citation indexing for studying science. *Nature*, *227*, 669–671.

Garfield, E. (1974). Journal citation studies XVII. Journal self-citation rates—there's a difference. *Current Contents*, *52* (Retrieved February 5, 2003, from http://www.garfield.library.upenn.edu/essays/v2p192y1974-76.pdf).

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

Garfield, E. (1998). From citation indexes to informetrics: Is the tail now wagging the dog? *Libri*, *48*, 67–80.

Garrido, M., & Halavais, A. (2003). Mapping networks of support for the Zapatista movement: Applying social network analysis to study contemporary social movements. In M. McCaughey, & M. Ayers (Eds.), *Cyberactivism: Online activism in theory and practice* (pp.165–184). London, UK: Routledge.

Ghemawat, P. (2001). Distance still matters—the hard reality of global expansion. *Harvard Business Review*, *79*, 137–147.

Google. (2002). *Our search: Google technology* (Retrieved February 5, 2003, from http://www.google.com/technology/index.html).

Graham, S. (2001). Historians and electronic resources: A second citation analysis. *Journal of the Association for History and Computing*, *5* (Retrieved February 5, 2003, from http://mcel.pacificu.edu/jahc/jahciv2/articles/graham/graham.html).

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* (Retrieved February 5, 2003, from http://informationr.net/ir/6-2/paper103.html).

Hubert, L., & Schultz, L. (1976). Quadratic assignment as a general data analysis strategy. *British Journal of Mathematical and Statistical Psychology*, *29*, 190–241.

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

Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the Association for Computing Machinery*, *46*, 604–632.

Kling, R., & McKim, G. (2000). Not just a matter of time: Field differences in the shaping of electronic media in supporting scientific communication. *Journal of the American Society for Information Science*, *51*, 1306–1320.

Koenig, M. E. D. (1982). Determinants of expert judgment of research performance. *Scientometrics*, *4*, 361–378.

Koenig, M. E. D. (1983). Bibliometric indicators versus expert opinion in assessing research performance. *Journal of the American Society for Information Science*, *34*, 136–145.

Krackhardt, D. (1992). A caveat on the use of the quadratic assignment procedure. *Journal of Quantitative Anthropology*, *3*, 279–296.

Lange, L. (1985). Effects of disciplines and countries on citation habits. *Scientometrics*, *8*, 205–215.

Larson, R. R. (1996). Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of cyberspace. In C. S. Schwartz (Ed.), *Proceedings of the ASIS 59th Annual Meeting* (pp. 71–78). Medford, NJ: Information Today, Inc.

Lawani, S. M. (1977). Citation analysis and the quality of scientific productivity. *BioScience*, *27*, 26–31.

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.

Line, M. B. (1979). The influence of the type of sources used on the results of citation analysis. *Journal of Documentation*, *35*, 265–284.

MacRoberts, M. H., & MacRoberts, B. R. (1986). Quantitative measures of communication in science: A study of the formal level. *Social Studies of Science*, *16*, 151–187.

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

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

Moed, H. F., Burger, W. J. M., Frankfort, J. G., & van Ranan, A. F. J. (1985). The application of bibliometric indicators: Important field- and time-dependent factors to be considered. *Scientometrics*, *8*, 177–203.

Narin, F., & Hamilton, K. S. (1996). Bibliometric performance measures. *Scientometrics*, *36*, 293–310.

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

Oppenheim, C. (1995). The correlation between citation counts and the 1992 Research Assessment Exercise ratings for British library and information science university departments. *Journal of Documentation*, *51*, 18–27.

Oppenheim, C. (1997). The correlation between citation counts and the 1992 Research Assessment Exercise ratings for British research in genetics, anatomy and archaeology. *Journal of Documentation*, *53*, 477–487.

Park, H. W., Barnett, G. A., & Nam, I. (2002). Hyperlink-affiliation network structure of top Web sites: Examining affiliates with hyperlink in Korea. *Journal of the American Society for Information Science and Technology*, *53*, 592–601.

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*, 5207–5211.

Redmond, M. V. (2000). Cultural distance as a mediating factor between stress and intercultural communication competence. *International Journal of Intercultural Relations*, *24*, 151–159.

Rousseau, R. (1997). Sitations: An exploratory study. *Cybermetrics*, *1* (Retrieved February 5, 2003, from http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html).

Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science*, *43*, 628–638.

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

Smith, A., & Thelwall, M. (2002). Web impact factors for Australasian universities. *Scientometrics*, *54*, 363–380.

Smith, L. (1981). Citation analysis. *Library Trends*, *30*, 83–106.

Snyder, H., & Bonzi, S. (1998). Patterns of self-citation across disciplines. *Journal of Information Science*, *24*, 431–435.

Snyder, H., & Rosenbaum, H. (1999). Can search engines be used for Web-link analysis? A critical review. *Journal of Documentation*, *55*, 375–384.

Sullivan, D. (2002). Google tops in "search hours" ratings. *Search Engine Watch* (Retrieved February 5, 2003, from http://searchenginewatch.com/sereport/02/05-ratings.html).

Tang, R., & Thelwall, M. (2002). Exploring the pattern of links between Chinese university Web sites. In E. G. Toms (Ed.), *ASIST 2002: Proceedings of the 65th ASIST Annual Meeting*, *39*, 417–424.

Thelwall, M. (2000). Web impact factors and search engine coverage. *Journal of Documentation*, *56*, 185–189.

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

Thelwall, M. (2001b). A Web crawler design for data mining. *Journal of Information Science*, *27*, 319–325.

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

Thelwall, M. (2002b). In praise of Google: Finding law journal Web sites. *Online Information Review*, *26*, 271–272.

Thelwall, M. (2002c). Methodologies for crawler based Web surveys. *Internet Research: Electronic Networking and Applications*, *12*, 124–138.

Thelwall, M. (2002d). An initial exploration of the link relationship between UK university Web sites. *ASLIB Proceedings*, *54*, 118–126.

Thelwall, M. (2002e). 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*, 995–1005.

Thelwall, M. (2003). Web use and peer interconnectivity metrics for academic Web sites. *Journal of Information Science*, *29*, 11–20.

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

Thelwall, M., Binns, R., Harries, G., Page-Kennedy, T., Price, E., & Wilkinson, D. (2002). European Union associated university Websites. *Scientometrics*, *53*, 95–111.

Thelwall, M., & Tang, R. (2003). Disciplinary and linguistic considerations for academic Web linking: An exploratory hyperlink mediated study with Mainland China and Taiwan. *Scientometrics*, 58 (1), 153–179.

Thomas, O., & Willett, P. (2000). Webometric analysis of departments of librarianship and information science. *Journal of Information Science*, *26*, 421–428.

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*, 29–38.

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

Warner, J. (2000). A critical review of the application of citation studies to the Research Assessment Exercises. *Journal of Information Science*, *26*, 453–460.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature*, *393*, 440–442.

Wilkinson, D., Harries, G., Thelwall, M., & Price, E. (2003). 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*, 59–66.

Winclawska, B. A. (1996). Polish sociology citation index: Principles for creation and the first results. *Scientometrics*, *35*, 387–391.

Zhu, J., Meadows, A. J., & Mason, G. (1991). Citations and departmental research ratings. *Scientometrics*, *21*, 171–179.