

A Research and Institutional Size Based Model for National University Web Site Interlinking

Mike Thelwall¹

School of Computing and Information Technology, University of Wolverhampton,
35/49 Lichfield Street, Wolverhampton, WV1 1EQ, UK.

Email: cm1993@wlv.ac.uk

Web links are a phenomenon of interest to bibliometricians by analogy with citations, and to others because of their use in web navigation and search engines. It is known that very few links on university web sites are targeted at scholarly expositions and yet, at least in the UK and Australia, a correlation has been established between link count metrics for universities and measures of institutional research. This paper operates on a finer-grained level of detail, focussing on counts of links between *pairs* of universities. It provides evidence of an underlying linear relationship with the quadruple product of the size and research quality of both source and target institution. This simple model is proposed as applying generally to national university systems, subject to a series of constraints to identify cases where it is unlikely to be applicable. It is hoped that the model, if confirmed by studies of other countries, will open the door to deeper mining of academic web link data.

Introduction

Web links are a fascinating object of study because of the implicit information that they contain: the premise that a page with many inlinks is likely to be useful drives the search engine Google, for example (Brin & Page, 1998). Bibliometric interest has recently concentrated upon counts of links to entire web sites or national domains. Ingwersen (1998) introduced the key link count metric, the Web Impact Factor (WIF), which has since been used in various guises to drive research into this phenomenon. Academic web sites have been the most intensively studied and it is now known that aggregates of links to an institution correlate with measures of its research output, once size is taken into account (Thelwall, 2001d, Thelwall, 2002b). This paper analyses links at a finer-grained level of detail, counting not all links to an institution but instead all links between *pairs* of institutions. This allows an exploration into patterns of inter-university linking to investigate in particular whether source and target institutional size and research are useful indicators and, if so, the most effective model that can be derived from using these as predictor variables.

The majority of previous research into web link metrics has focussed on sites with academic content: university web sites (Smith, 1999; Darmoni *et al.*, 2000; Thelwall, 2000; Thomas & Willet, 2000; Vreeland, 2000; Thelwall, 2001d; Thelwall, 2002b); e-journals (Harter & Ford, 2000; Kim, 2000); journal web sites (Soualmia *et al.*, 2002); and subject-specific web sites (Larson, 1996; Hernández-Borges *et al.*, 1999), perhaps motivated by the analogy with citations (Davenport & Cronin, 2000). Entire countries have also been subjected to web link analysis, however (Ingwersen, 1998). In order to develop meaningful metrics for the academic web, a theory of

¹ Journal of Documentation, 2002: 58(6), 683-694.

hyperlink creation must first be constructed. This will need to be derived from the results of several different types of activities and recent research into these areas is described in brief below. This section also serves to introduce concepts that will be alluded to later.

Reliably counting link pages Most research into web links has used search engines to collect the raw data but technical concerns have been raised with the reliability of the figures produced (Rousseau, 1999; Snyder & Rosenbaum, 1999; Thelwall, 2000; Bar-Ilan, 2001; Björneborn & Ingwersen, 2001; Thelwall, 2001a). Recent results indicate, however, that AltaVista has become more reliable as a data source, at least for the academic web (Thelwall, 2001b; Thelwall, 2001d; Thelwall, 2002b). It must also be noted that any automatic crawler will only cover a subset of the web and the results should be seen as applying only to the crawlable pages.

Developing and testing metrics Ingwersen's (1998) early work on the WIF described and assessed several different versions. One was the external relative WIF, which is a simple count of all pages outside a site that contain a link to it (inlink pages), divided by the number of pages in this target site. A version of the external relative WIF was created specially for academic web sites (Thelwall, 2001a). This uses the number of full-time equivalent academic members of staff at the institution instead of a site page count as the denominator of the calculation. Early attempts to correlate versions of the WIF with research ratings for universities did not get positive results (Smith, 1999; Thelwall, 2000; Thelwall, 2001a) but later studies have found a statistically significant correlation with average research ratings for UK university institutions and for Australian universities (Smith & Thelwall, 2002; Thelwall 2001d; Thelwall, 2002b).

Identifying reasons for link creation Academic web sites contain material created for many different purposes (Middleton *et al.*, 1999), and studies of the targets of academic links have found that these pages were of many different types (Cronin *et al.*, 1998), including recreational, with only a tiny minority containing academic content equivalent to a journal article (Thelwall, 2001d). For the similar issue of URL citation in traditional articles, Kim (2000) found that reasons for use extended those for citations to print sources to include medium-specific ones. Zang (2001) found several factors that inhibit URL citing, including self-perceived lack of ability to use the Internet.

Dealing with anomalies The lack of quality control on the web means that some individuals publish large quantities of web pages, even with outlinks to other sites, greatly altering the results of link and page counts. One provisional attempt to deal with this issue has been produced, (Thelwall, 2001c) which is a stopgap solution to identify entire sites that have inlink or outlink profiles that are unusual in some way.

This paper addresses the second area above. Firstly an investigation into whether the size and research quality of the *source* institution also affects the number of outlinks to other academic web sites will be reported. This is in contrast to the focus of Web Impact Factor studies, which deal only with the *target* institution. Size is an obvious likely indicator of the number of outlinks created, but research quality is perhaps more contentious. Through an analogy with citation practices, it seems logical to propose that institutions conducting more research will tend to create more outlinks (links to external sites), but the fact that the number directed at online articles is negligible (Thelwall, 2001d) undermines this argument. It may be the case, however, that higher research institutions simply produce more web output or have more reason to link to other universities because of the collaboration that is a common part of scholarly activity. In this paper, nine different models that attempt to

predict inter-university link counts will be assessed for goodness of fit against a large dataset. Each model will be based upon some or all of the source and target institutional size and research quality variables. The most successful model will then be further investigated to discuss the theoretical and practical reasons for the extent of its effectiveness in estimating link counts.

Method

The study will be based upon link counts between 86 pairs of UK university institutions. Academic staff numbers were obtained from Noble (1999). Average institutional research quality ratings were obtained from the Times Higher Education Supplement league tables (Mayfield University Consultants, 2000). Both of these sources are probably only estimates of the quantity being measured. AltaVista was used to obtain link counts between all pairs of UK universities except St Andrews, for which it did not return correct results. This is the same data as was used in Thelwall (2001b) to investigate patterns of university interlinking, and the universities identified in that study as being outliers were excluded from the data. It was decided to use the same data rather than collect a fresh set (which could be easily prepared automatically using software already written) because of ethical and practical considerations: it involves over 7,000 queries to AltaVista, which is a heavy exercising of its services when excessive use could possibly lead to their withdrawal. The outliers had either unusual web sites or were the source or target of a large number of links from a single causative factor. Excluding link counts between a university and itself this gave a data set of 7,310 points. It should be emphasised that the accuracy of the data is dependent on the correctness of AltaVista's algorithms and refers not to entire web sites, but only to that proportion that is indexed by the search engine. It is, unfortunately, not of a set of independent observations because the mechanism for obtaining link counts overlaps for each one. Specifically, there will be many single pages on a web site that contain more than one outlink, and if AltaVista's crawler finds such a page, then the count of inlinks for each of its target universities would increase by one. More generally, there may be sites that AltaVista indexes particularly poorly, perhaps for design reasons, and the link counts for all targets from these sites may be low as a result. The implication of the non-independence of the data is that statistically valid conclusions cannot be inferred from it and the results of calculations undertaken will, therefore, serve to inform a discussion of the underlying structure rather than to directly accept or reject test hypotheses.

Link counts were entered into SPSS as the dependant variable, with up to six independent variables: source and target research rating; source and target web site size; source and target academic staff numbers. A series of single and multiple linear regressions were then performed in order to produce evidence of causative relationships. In the nature of the technique, linear relationships are being sought, but significant results can also be obtained from other non-linear, non-random phenomenon, such as logarithmic or exponential trends.

Results

Model number	Variable(s) and/or products of variables	Model fitted	R
1	R_F	$-1.9 + 10.5 R_F$	0.20
2	R_T	$-5.0 + 11.6 R_T$	0.22
3	$R_F S_F$	$11.1 + 0.00046 R_F S_F$	0.23
4	S_F	$-4.3 + 0.032 S_F$	0.24
5	S_T	$-8.6 + 0.036 S_T$	0.28
6	$R_T S_T$	$7.9 + 0.00045 R_T S_T$	0.28
7	$R_F S_F, R_T S_T$	$-10.2 + 0.00056 R_T S_T + 0.00047 R_F S_F$	0.36
8	R_F, S_F, R_T, S_T	$-69.2 + 7.3 R_T + 0.029 S_T + 6.8 R_F + 0.026 S_F$	0.41
9	$R_F S_F R_T S_T$	$9.8 + 0.000,000,013 R_F S_F R_T S_T$	0.45

Table 1. A list of models fitted to the UK university interlinking data. R_F, R_T are the staff-weighted RAE averages for the source and target institution respectively. S_F, S_T are the full-time equivalent academic staff numbers at the source and target institution respectively. $N=7,310$. R values are given as descriptive statistics only.

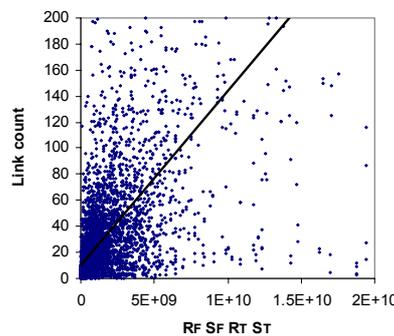
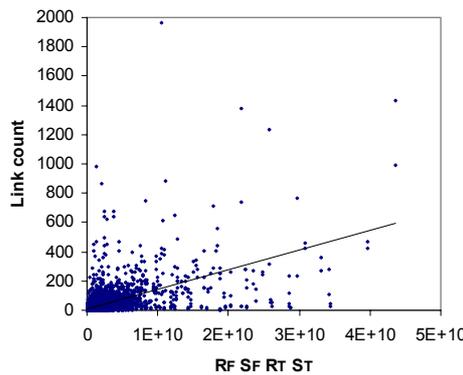
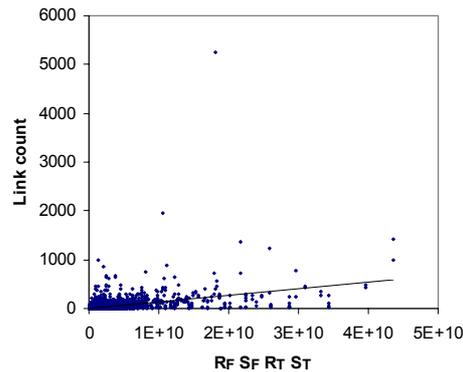
Table 1 shows a list of the regression equations fitted and their residuals. For all models except 7 and 8 (with multiple independent variables) the R column value is also the standard Pearson correlation coefficient. Because the individual link counts are not independent and all the variables are significantly non-normal, the data cannot be subjected to rigorous tests (which all would pass with significance $p < 0.001$, for example slope t values vary from 17.7 for equation 1 to 43.3 for equation 9) it is believed that all the values are large enough to satisfy doubts arising from the non-independence and non-normality of the data. The existence of a relationship is further justified below for the critical ninth model through both theoretical discussion and an examination of the relevant graph.

Several different statistical techniques were applied to the data to corroborate the conclusions.

- All seven variables used in the different models in Table 1 were fitted to a single regression equation for link counts. This gave only a slightly improved R of 0.51 and the fitted equation included counterintuitive negative coefficients for some variables.
- The testing was repeated after (a) splitting the data set into half and analysing both halves and (b) eliminating the main outlier (the highest link count value, see Figure 1). The results did not change significantly.
- A regression was run with the natural logs of R_F, S_F, R_T and S_T as the four independent variables and the logs of the link counts as the dependent variable. The link counts had 0.01 added to them to avoid log of zero errors. The logarithm-transformed data was still significantly non-normal. This regression estimates *powers* for R_F, S_F, R_T and S_T in a purely multiplicative model but does not take into account a constant term. The powers produced were reasonably close to 1 in all cases (0.742, 1.056, 0.903, 0.920 respectively, giving a model where link counts are proportional to $R_F^{0.742} S_F^{1.056} R_T^{0.903} S_T^{0.920}$). This provides some corroborative evidence for model 9 from a different perspective but, more importantly, indicates that all the four variables make quite similar contributions to the quadruple product.
- The main two models were compared using the non-parametric Spearman correlation test. In this test, link counts were correlated with the simplest additive model $R_F S_F + R_T S_T$ and the multiplicative model $R_F S_F R_T S_T$, producing very high

results for each one (0.673 and 0.752 respectively, $p < 0.001$ for both). This is a non-parametric test and so is not influenced by the distribution of the data, but is not robust because of the non-independence of the observations. Again it supports the multiplicative model over the additive one and is useful because it is less effected than regression techniques by the extreme outliers in the data set.

Figures 1 to 3 show link counts plotted against $R_F S_F R_T S_T$ at several levels of detail. The multiple similar graphs are useful because the 7,310 points overlap on the larger scale graphs to the extent that some of the information on them is lost.



Figures 1 to 3 show link counts between 86 university web sites plotted against $R_F S_F R_T S_T$ at full scale and zoomed in to two different levels of detail ($1E+10 = 10,000,000,000$).

The graphs do not show an immediately obvious linear trend, but merit further inspection. The high correlation coefficient comes partly from the clustering around

the origin. The disproportionate number of low $R_F S_F R_T S_T$ values is a mathematical effect of multiplication: for a high value, both source and target institutional research and size must be large, but for a low value, only one needs to be small. Essentially, the regression equation is good at predicting low link counts when $R_F S_F R_T S_T$ is small, say in the bottom 20% of its range, but for higher values link counts are more widely scattered. This is consistent with an underlying linear relationship where bigger $R_F S_F R_T S_T$ values result in higher average link counts, although all values are clearly subject to variation. The non-zero constant term in the regression equation is potentially an artefact of the underlying assumptions of linear regression: that deviations from the predicted equation should average zero. This is because negative link counts are impossible and, therefore, particularly for the low values of $R_F S_F R_T S_T$ that dominate the graph the extent of the deviation below the predicted value is limited.

Discussion

The findings of the statistical experiments suggest that, after taking into account a constant of approximately 9, the propensity of one university's pages to link to those of another is very approximately proportional to the product of their (academic staff) sizes and average research ratings. This model gives a higher R-value than the others in Table 1 and has the advantage over its nearest competitor of having only one effective explanatory variable from a regression point of view. The most complex regression equation that was fitted, having nine independent variables, did give a better fit to the data, but the small increase in R was not considered sufficient to justify the extra model complexity. Unfortunately, the non-independence of the data points and non-normality of the variables precludes giving a formal statistical statement of confidence levels for these conclusions. The square of the R-value gives the proportion of the variation in link counts that is predictable from the regression equation and so the model explains 20% of the variance in link counts. This is quite a low figure, despite its argued significance, but high values are unlikely in the context of the relatively low predicted numbers for many of the between-site counts and the probable existence of other trends, for example geographic (Thelwall, 2002a) and site management based.

The model is consistent with both the average quantity of outlink pages produced (per academic member of staff) in an institution and inlink pages (per academic member of staff) being proportional to the average research quality of its academic staff. It has been suggested that the reason for research quality of the target institution correlating with inlink counts to it may be to do with its reputation, since links tended to be of a general nature or to departments, researchers or research groups rather than to academic work (Thelwall, 2001d). If this is the case, then a different explanation must be found for the relationship between outlink counts and quality of the source institution. Two alternatives are that higher research quality academics publish more on the web or that their pages contain more outlinks. Since academics do not create all of the web pages on a site, other explanations are possible concerning the other creators of web content. Perhaps administrative staff / webmasters / students at higher research quality institutions create more pages or link more to external sites, or maybe there are more of the first two categories at higher research institutions. Further research is needed to address this question.

At an abstract level, a simple theory for the behaviour of scholars can be proposed that would provide some justification for the link count model. The theoretical assumption is that every academic member of staff (or support staff and students working explicitly or implicitly with them) creates a variable quantity of web

content, each component part of which has an equal probability of linking to each part of the web content of all other academics. Under the assumption that better researchers were likely to produce more web content, this would fully explain the link count model (apart from the constant term already discussed). This content size-based explanation is deliberately vague about the units of measurement, and is only a preliminary attempt at a theory. It will need modification, for example, to deal with the proposal that link targets reflect more the reputation of the creators than the actual content of the pages.

The model proposed for the underlying tendency to link between web sites has been justified through its ability to explain the correlation between $R_F S_F R_T S_T$ and link counts. Inspection of the graph, however, does cause some problems. A graph with a predominantly linear trend, but with points high above the trend line would be easy to explain, since anomalies through the creation of large numbers of pages with outlinks to the same university are easy to create. The occurrence of points far below the trend line are much more difficult to explain. Why do some pairs of large, high quality institutions have very low levels of interlinking? Table 2 lists some of the counts furthest below the predicted value.

From	To	Actual links	Predicted links
University of Wales Bangor	University of Cambridge	30	469
University of Cambridge	University of Wales Bangor	45	469
Oxford Brooks University	University of Cambridge	16	394
University of Cambridge	Oxford Brooks University	21	394
University of Cambridge	University of Lancaster	104	469
University of Oxford	University of Wales Bangor	29	394
University of Wales Bangor	University of Oxford	42	394
University of Southampton	University of Wales Bangor	31	359
University of Southampton	University of Lancaster	53	358
University of Wales Bangor	University of Southampton	55	359

Table 2. The ten link counts that were furthest below the predicted value

Bangor appears in six of the rows and so it may well be the case that this university's use of the web is unusual. The results were rechecked on 24 September 2001, when only 8 links were found to Cambridge out of a reported 428,522 pages. Cambridge and Oxford Brooks create fewer than expected links to each other, and Lancaster attracts few links from Cambridge and Southampton. Of course, in a data set as large as 7,310 the lack of any anomalies would be more surprising than the appearance of some, but the pattern of repeated institutions suggest that this is not a sufficient explanation. Possible reasons include a lack of overlap of research specialisms between pairs of universities or problems with AltaVista's indexing software. A further potential explanation is that these sites are missing a highly linking and linked to type of page, such as student web pages, society pages, or informal academic pages. An assessment of the size of the websites indexed by AltaVista was conducted to investigate if the issue seemed to be related to the amount of material published. If web publication is proportional to both staff numbers and RAE rating, then the calculation

$$\frac{\text{Pages Indexed by AltaVista}}{\text{RAE} \times \text{Academic Staff Numbers}}$$

should be approximately constant across the universities, with those having a larger web presence than expected recording higher than average values. As explained above, counts of web pages are very unreliable as an assessment of the size of the content of a web site, but this is the only measure currently available. The formula was evaluated for all universities in the study, with the unexpected discovery that all the universities mentioned in Table 2 were above average, with the exception of the University of Oxford. In fact Oxford Brookes' result was the largest of all UK university institutions covered, and Bangor was third largest. If the explanation for the below average link counts in Table 2 are in fact due to missing content on the web site, then more powerful tools are needed to assess this, or a detailed study of the sites must be undertaken to identify the omissions.

The proposed general model for university interlinking

The following is proposed as a general model for university interlinking, on the basis of the findings and discussion above.

$$L_{F,T} = \alpha + \beta R_F S_F R_T S_T$$

Where $L_{F,T}$ is the number of web hyperlinks from institution F to institution T , R is a measure of research quality and S is a measure of institutional size, such as academic staff numbers. The constants α and β are expected to vary from nation to nation. The inclusion of the constant α is justified theoretically on the grounds that some links are not quality or size dependent, for example those from lists that intend to be exhaustive lists of universities or specific types of university department. The following conditions are set down as requirements for this model to be applicable to a country.

1. Outliers are removed from the data. In other words web sites with a disproportionately high source or target of links created or copied by an individual or small group should be excluded, as should sites that are controlled to the extent that informal linking is not allowed.
2. The model should be applied to multidisciplinary universities in countries with a significant number, say 20 as an arbitrary minimum. The reason for this is that universities which are not multidisciplinary are expected to link less to other institutions that do not have overlapping specialisms.
3. Web site sizes should be sufficiently large to allow a natural averaging process to work. The figure 5,000 pages as an arbitrary minimum median web site size is suggested.
4. The country's universities are sufficiently geographically scattered that geographic factors do not obscure any research-related trend.
5. The country's universities vary significantly in quality.

Important sources of variation outside the scope of the model

It has been shown that geographic proximity affects research collaboration (Katz, 1994) and also link count profiles (Thelwall, 2001c). This has not been taken account of in the model and so it is to be expected that some of the variation seen in Figures 1 to 3 is due to this. It is speculated that three further important sources of discrepancies would be: the degree of overlap of specialism; the size of departments making use of the web, particularly computing; and site management issues such as whether student pages are permitted. It may also be the case that access to appropriate computing

resources is a factor, making subjects with lower quality equipment and support being less able to make a mark on the web.

Summary and further work

Exploration of the data for university interlinking has led to the suggestion of a specific model for national university interlinking. Although the constants in the formula are expected to vary across nations and over time, the basic premise is that link counts will be approximately proportional to the quadruple product of academic staff numbers and research quality for both source and target universities. The model is expected to apply internally to the multidisciplinary universities of countries where these are sufficiently evenly geographically scattered and the websites are sufficiently developed, after removing outliers. The model indicates that size and research quality are both indicators of the propensity to create outlinks and to be the target of inlinks. The purpose of investigating the data and proposing the model is to move forwards understanding of the phenomenon of academic web links by providing a practical and plausible explanation for a significant part of the variance in the data. This also opens up more questions for investigation, including the following that have arisen from the discussion of the results.

1. Do universities in other countries adhere to the model proposed for interlinking?
2. Why is research quality an indicator of higher outlinking, and who creates the outlinking pages?
3. Why do some universities and university pairs have much lower link counts than would be expected by the model? Is there a type of web page that is missing from such a site, is there a lack of disciplinary overlap, or is this a purely statistical phenomenon?
4. In the case of asymmetric link counts between universities is there an identifiable reason for this, and are there institutions that tend to supply information, (being predominantly the target of links) and others that tend to use information, (being predominantly the source of links)?

It must be stressed that these questions are in addition to the ongoing problem of developing a methodology to clean the link data (by removing problematic data in a theoretically and methodologically sound way) so that meaningful conclusions can be drawn from it.

Acknowledgements

Many thanks to the referees and Liwen Vaughan for helpful comments on an earlier draft of the paper.

References

- Björneborn, L. & Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics*, 50(1), 65-82.
- 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.
- Darmoni, S. J., Thirion, B., Douyere, M., Challoub, C., Leroy, J. P. (2000). Mesure de l'impact des sites Web: le Web Impact Factor, l'exemple des CHU français. *La Revue du Praticien - Médecine Générale* 14 (516): 2079-2080.
- Soualmia, L.F., Darmoni, S.J. Le Duff, F., Douyère, M., & Thelwall, M. (2002, to appear). Web Impact Factor: a bibliometric criterion applied to medical

- informatics societies' Web sites, Medical Informatics in Europe MIE2002 congress (to be held in Budapest, Hungary, August 25-29).
- Davenport, E. & Cronin, B. (2000). The citation network as a prototype for representing trust in virtual environments. In: Cronin, B. & 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.
- Harter, S. P. & Ford, C. E. (2000). Web-based analysis of e-journal impact: approaches, problems and issues. *Journal of the American Society for Information Science and Technology*, 51(13), 1159-1176.
- Hernández-Borges, A. A., Macías-Cervi, P., Gaspar-Guardado, M. A., Torres-Álvarez de Arcaya, M. L., Ruiz-Rabaza, A. & Jiménez-Sosa, A. (1999). Can Examination of WWW Usage Statistics and other Indirect Quality Indicators Distinguish the Relative Quality of Medical Web Sites? *Journal of Medical Internet Research*, 1(1). Available: <http://www.jmir.org/1999/1/e1/index.htm>
- Ingwersen, P. (1998). The calculation of Web Impact Factors. *Journal of Documentation*, 54(2), 236-243.
- Katz, J. S. (1994). Geographical proximity in Scientific collaboration, *Scientometrics*, 31(1), 31-43.
- Kim, H. J. (2000). Motivations for hyperlinking in Scholarly electronic articles: a qualitative study. *Journal of the American Society for Information Science*, 51(10) 887-899.
- Larson, R. R. (1996). Bibliometrics of the World Wide Web: An Exploratory Analysis of the Intellectual Structure of Cyberspace. ASIS 96. <http://sherlock.berkeley.edu/asis96/asis96.html> (visited 4 August 2001).
- Mayfield University Consultants (2000). League Tables 2000. *The Times Higher Education Supplement*, April 14, II-III.
- 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.
- The Noble Publishing Company (1999). *Noble's Higher Education Financial Yearbook 1999*, Edinburgh: Noble.
- Rousseau, R., (1999). Daily time series of common single word searches in AltaVista and NorthernLight, *Cybermetrics*, 2/3. Available: <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html>
- Smith, A. G. (1999). A tale of two web spaces: comparing sites using Web Impact Factors. *Journal of Documentation*, 55(5), 577-592.
- Smith, A. G. & Thelwall, M. (2002). Web Impact Factors for Australasian Universities, *Scientometrics*, 54(1-2), 363-380.
- Snyder, H. & Rosenbaum, H. (1999). Can search engines be used for web-link analysis? A critical review. *Journal of Documentation*, 55(4), 375-384.
- Thelwall, M. (2000). Web Impact Factors and search engine coverage. *Journal of Documentation*, 56(2), 185-189.
- Thelwall, M. (2001a). Results from a Web Impact Factor crawler. *Journal of Documentation*, 57(2), 177-191.
- Thelwall, M. (2001b). The responsiveness of search engine indexes. *Cybermetrics*, 5(1). <http://www.cindoc.csic.es/cybermetrics/articles/v5i1p1.html> (visited 28 March 2001)
- Thelwall, M. (2001c). An initial exploration of the link relationship between UK university web sites, University of Wolverhampton.

- Thelwall, M. (2001d). Extracting macroscopic information from web links. *Journal of the American Society of Information Science and Technology*, 52 (13), 1157-1168.
- Thelwall, M. (2002a, to appear). Evidence for the existence of geographic trends in university web site interlinking, *Journal of Documentation*, 58(5).
- Thelwall, M. (2002b). A comparison of sources of links for academic Web Impact Factor calculations. *Journal of Documentation*, 58(1), 60-72.
- Thomas, O. & Willet, P. (2000). Webometric analysis of departments of Librarianship and information science. *Journal of Information Science*, 26(6), 421-428.
- Vreeland, R. C. (2000). Law libraries in hyperspace: a citation analysis of world wide web sites. *Law Library Journal*, 92(1), 9-25.
- Zang, Y. (2001). Scholarly use of internet-based electronic resources, *Journal of the American Society for Information Science and Technology*, 52(8), 628-654.