

# Web Links as an Indicator of Research Output: a Comparison of NZ Tertiary Institution Links with the Performance Based Research Funding Assessment.

Alastair G. Smith\* and Mike Thelwall\*\*

\**Alastair.Smith@vuw.ac.nz*

School of Information Management, Victoria University of Wellington  
PO Box 600, Wellington, New Zealand

\*\**M.Thelwall@wlv.ac.uk*

School of Computing & Information Technology, University of Wolverhampton, Wulfruna Street,  
Wolverhampton WV1 1SB, UK

## Abstract

Data from a crawl of New Zealand Universities is used to study links between the institutions, and the extent to which web based measures correlate with a measure of research performance. The research measure used was the Performance Based Research Fund (PBRF) Quality Score, arrived at in the 2003 research assessment exercises by the NZ Government's Tertiary Education Commission (TEC). There is a moderate correlation between the Quality Score and the FTE-corrected link counts (link counts divided by numbers of full time equivalent academic staff), though it is not clear whether counting on the basis of pages, directories, or domains gives the best measure. Some exceptions to the correlation could be explained in terms of the special characteristics of the institutions. The pattern of linking was not uniform, and could indicate special relationships between institutions. Suggestions for further research are made.

## Introduction

This paper reports research in progress to examine web linking as an estimator of the research performance of New Zealand universities. Measuring the influence of research web sites is important, given the increasing centrality of the web for scholarly communication. Web links may also potentially be used to estimate the research quality or productivity for units for which no direct data is accessible to the investigator.

There has been growing interest in studies of web linking as an analogy of citations in traditional bibliometrics, for example Ingwersen (1998) proposed the Web Impact Factor (defined below) as a measure of the influence of a web site or domain, analogous to the Journal Impact Factor. Subsequent studies have indicated that web links are made for rather different reasons than conventional citations, and research continues into the significance of web link based measures and their applications (Björneborn, 2003; Faba-Perez, Guerrero-Bote & De Moya-Anegón, 2003; Ortega Priego, 2003; Prime, Bassecouard, & Zitt, 2002; Rousseau, 1997; Vaughan & Wu, 2003), recently summarised by Björneborn (Björneborn & Ingwersen, 2004; Thelwall, Vaughan & Björneborn, 2005).

A number of studies have investigated web links to and between institutions as a measure of research quality, and made comparisons between link measures and research output (e.g. An & Qiu, 2003; Heimeriks, Hörlesberger & van den Besselaar, 2003; Smith & Thelwall, 2002; Thelwall & Harries, 2004). Wilkinson *et al.* (Wilkinson, Harries, Thelwall, & Price, 2003) found that 90% of a sample of university links were research or education related but that only about 1% were equivalent to citations, showing that link counts might relate to research but do not directly measure research productivity or quality. Various measures of linking as a basis for research comparisons of universities have been compared, including the following (Thelwall, 2004):

- Individual web pages
- Directories
- Domains
- Universities

It has been found that in some countries counting links by directories or domains is significantly better than counting links by pages (Thelwall, 2004), whereas in others there is little difference (Thelwall & Tang, 2003). The difference occurs in countries where there are several pairs of sites with a large number of links between them caused by interconnecting databases, or large high-linking multi-institution collaborative project sites.

A problem with evaluating web link counts as research estimators is finding an objective measure of research performance. Comparisons across institutions and discipline areas are difficult; while publication and citation rates yield quantitative data, arguments arise over their significance. Thelwall (2004) used the UK's Research Assessment Exercise (RAE), which combines data relating to publications, research funding, PhD awards, etc. to rank universities. In 2003, New Zealand tertiary institutions underwent a similar research assessment exercise to determine part of the extent of government funding, the Performance Based Research Fund (PBRF) assessment. This produced a ranking of institutions, staff, and subject specialities (Tertiary Education Commission, 2004). This provides a new opportunity to evaluate the connection between research and links outside the UK.

In the current study, NZ university website interlinking is compared with the PBRF assessment. The following research questions are addressed:

- Are web links between universities related to research performance in New Zealand?
- Which link counting measures relate best to research productivity in New Zealand?

### **Methodology**

Gathering data about web links has been carried out in the past using commercial search engines such as AltaVista. This approach is convenient and has produced some interesting results. However, there is a danger in relying upon on data collected for non-webometric purposes using opaque algorithms (e.g., Bar-Ilan, 1999, 2001; Björneborn & Ingwersen, 2001). AltaVista has useful search operators and the ability to combine sets using Boolean logic (although the detailed processing of these requests was erratic). However, given the choice, a specialist crawler is preferable.

The current study used data produced by the Wolverhampton web crawler specifically designed for bibliometric studies (Thelwall, 2003). It crawls web sites from their home page by repeatedly following links within the same site. Each new page is compared to all other downloaded pages and deleted if found to duplicate any page previously fetched. The data was gathered during a crawl of all NZ University domains in December 2003.

- University of Auckland, Auckland, auckland.ac.nz
- Auckland University of Technology, Auckland, aut.ac.nz
- University of Canterbury, Christchurch, canterbury.ac.nz
- Lincoln University, Christchurch, lincoln.ac.nz
- Massey University, Palmerston North, Auckland and Wellington, massey.ac.nz
- University of Otago, Dunedin, otago.ac.nz
- Victoria University of Wellington, Wellington, vuw.ac.nz
- Waikato University, Hamilton, waikato.ac.nz

An ongoing debate in webometric studies of this type is the unit of analysis. Traditional bibliometrics uses the document, or the journal as a unit. In webometrics, at least four possibilities are available (Thelwall, 2004):

- Individual webfile/page, e.g. <http://www.sim.vuw.ac.nz/staff/pak-yoong.aspx>, a specific page describing a researcher at Victoria University.
- Directory, e.g. <http://www.sim.vuw.ac.nz/research/> a directory containing information about research in the School of Information Management at Victoria University of Wellington.
- Domain, e.g. <http://www.sim.vuw.ac.nz> the domain name for the School of Information Management.
- University, e.g. <http://www.vuw.ac.nz> the domain for Victoria University of Wellington.

Web pages are an obvious basis for comparison, but have some faults. The number of pages in a site may be determined by web design policies rather than the amount of information, since a document may either exist as one file, or be broken into screen sized units. Directories and domains may reflect quite specific documents or projects, or may contain a wide range of information. As mentioned above, previous studies (Thelwall, 2004) indicate that the domain level analysis may be the "best"; however this may depend on the country context, objectives of the comparison, etc.

An early measure proposed (Ingwersen, 1998) was the Web Impact Factor (WIF). In his study a WIF was calculated by dividing the number of page inlinks by the number of pages crawled. Note that

the WIFs calculated by us are very low, since they only use links from other NZ universities, not from the global Internet.

In the current study, data was gathered relating to the different units of analysis, and their relationship to the PBRF rankings compared.

The purpose of the PBRF exercise was to allocate government funding to tertiary institutions, based on their research output. As part of this exercise, a quality evaluation of the tertiary institutions was conducted. In determining the overall PBRF research funding, completion rates of research degrees, and external funding were also taken into account. The PBRF quality evaluation has been published (Tertiary Education Commission, 2004) by the Tertiary Education Commission (TEC) a NZ government agency set up to oversee tertiary education. In 2003/4 every academic staff member at a NZ Tertiary institution submitted an evidence portfolio consisting of three parts:

- Research output. Up to 50 outputs from the previous six years were listed, and up to four nominated as significant outputs. Outputs could include publications, conference papers, exhibitions, performances, software, etc, and could be digital.
- Indications of peer esteem, such as awards, prizes and citation counts.
- Contributions to the research environment, such as research student supervision, reviewing of publications, and conference organisation.

Portfolios were assessed by subject based peer review panels to arrive at a numeric score, using a weighting of publications 70%, peer esteem 15%, and contributions to the research environment 15%. The numeric scores were translated to a grade: R (research inactive), C (good quality research), B (very good quality research), A (world class research). In the present context it is worth noting that in some cases web based measures were incorporated into the PBRF rankings. For example numbers of link counts to online publications were included in the “indications of peer esteem” sections of PBRF portfolios.

The ratings have been aggregated to provide overall scores for institutions, using a weighting:

- R=0
- C=2
- B=6
- A=10

TEC admit that this weighting is arbitrary and a different regime could produce different results. We follow their expert judgement as the best available choice for weightings.

In the current study, link measures were compared with both the overall research output (the sum of weighted portfolio rankings for the institution) and the quality score (the average portfolio ranking per full time staff member).

## **Results**

Because of the small size of the sample (8 universities) the correlations calculated can only be descriptive, and conclusions can only be tentative. Appendix 1 summarises the data. Link counts are of the number of connections made between instances of the particular unit. For example, the 259 domain inlinks to auckland.ac.nz means that there were 259 links from domains external to auckland.ac.nz that connected to domains within auckland.ac.nz, after excluding duplicates (i.e. links sharing a common source domain and sharing a common target domain). If there were 15 links made from sim.vuw.ac.nz pages to lbr.auckland.ac.nz pages, this would count as one inlink for the link count.

To what extent do the various measures of linking correlate to the overall PBRF research output?

Table 1 shows the Pearson correlation for various measures and the PBRF Quality Score. In order to compensate for institution size, the number of inlinks or outlinks were divided by the number of full time equivalent (FTE) academic staff, to make a valid comparison with the PBRF Quality Score.

Table 1: Correlations between web-based measures and the PBRF Quality Score

Page Inlinks/FTE	0.54
Directory Inlinks/FTE	0.53
Domain Inlinks/FTE	0.59
Page Outlinks/FTE	0.54
Directory Outlinks/FTE	0.58
Domain Outlinks/FTE	0.71
WIF	-0.09

There is a moderate degree of correlation between the PBRF Quality Score and the FTE corrected inlinks and outlinks, but it is not possible to conclude which of the bases (page, directory, domain) gives the best correlation. The strongest correlation is between the FTE corrected domain outlinks, and the PBRF Quality score. Possibly this indicates that a strong research institution makes links to the outside world.

There is a poor correlation between the WIF and the PBRF Quality Score. This is consistent with the recent UK finding that better researchers receive more links because they produce more pages, rather than because the number of inlinks per page (i.e. the WIF) is larger; according to this logic, the WIF should be approximately constant (Thelwall & Harries, 2004).

The relationship between a web based measure and the PBRF Quality Score was explored for specific institutions. Figure 1 compares the domain inlinks per FTE with the quality scores published by TEC.

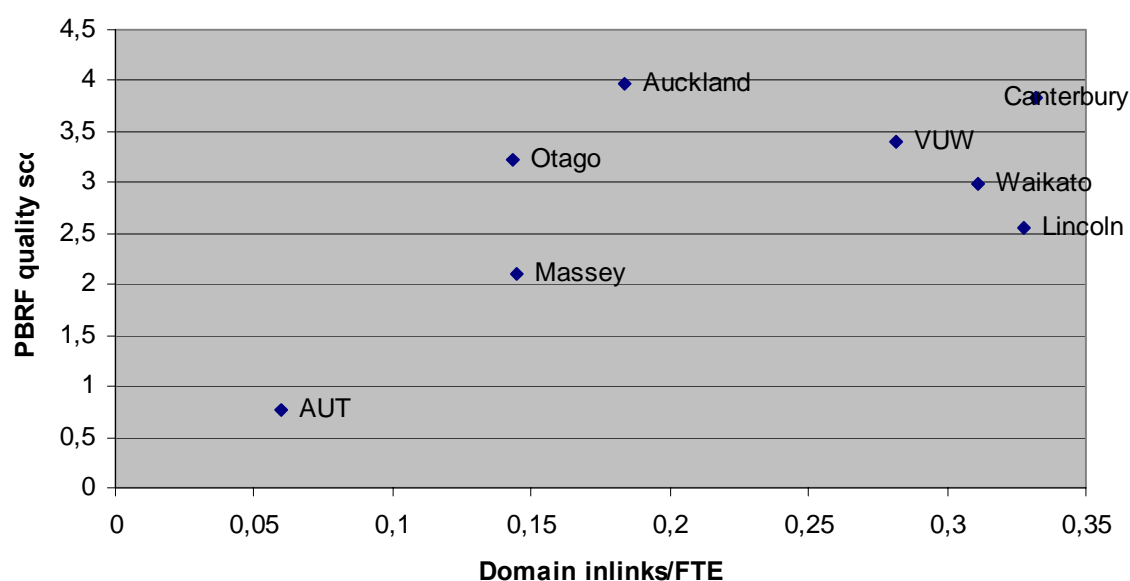


Figure 1: Domain inlinks per FTE compared with PBRF quality scores

Although there is a broad correlation between the PBRF scores and the domain links/FTE there are interesting exceptions:

- Auckland and Otago have high PBRF quality scores but relatively low inlink scores. This could reflect their academic recognition lying in conventional research publication, rather than internet presence. This makes sense since Auckland is the largest university, with a strong conventional research record in many areas; and Otago, while a smaller institution, is the home of the best established medical school and has a strong medical research record.
- Waikato and Lincoln have relatively high inlink scores in relation to their PBRF quality scores, indicating that their Internet presence is stronger than their conventional research publication. In the case of Waikato this could reflect the fact that it is a relatively new university (founded in the 1960s) but has a strong computer science department, for example housing the New Zealand Digital Library project. Lincoln may be a statistical anomaly due to its small number of staff.

This discussion indicates that although the overall correlations between web based measures and PBRF Quality Scores is not high, the deviations from the trend can be explained in terms of the characteristics of the individual institutions.

An examination of numbers of page-based links between institutions (Appendix 2) reveals that there is not a uniform pattern of linking, and that in most cases there one or two institutions make a significant proportion of the links to another. For example:

- Half of the links made to Auckland (ranked first on the PBRF Quality Score) are from Victoria (ranked third); and vice versa. This could reflect a high degree of common interest.
- Almost half of the links made to Canterbury (ranked second) and to Massey (ranked seventh) are from Auckland, although Canterbury and Massey each make up only about 10% of the links to Auckland.
- Auckland and Victoria each make about one third of the links to Otago (ranked fourth), and to Waikato (ranked fifth), although Otago and Waikato do not originate a high proportion of the links to Auckland and Victoria.

## **Conclusions**

The 2003 PBRF research evaluation exercise provided a valuable opportunity to test the usefulness of web based research measures for university domains. Although the sample is small, it is interesting that link counts broadly correlate with the PBRF measures and the results are consistent with prior findings for the UK. This is valuable because it gives some confidence that the previous UK findings may have a wider international applicability.

The results also suggest that the number of domain inlinks per staff member is a useful measure of Internet impact of the institution; while this correlates broadly with the PBRF quality score, some exceptions are explainable in terms of the known characteristics of the institutions.

The pattern of linking between institutions is not uniform, and strong linking in one direction is not necessarily reflected in other directions, although there is an indication of mutual linking between Auckland and Victoria, ranked first and third respectively in the PBRF Quality Scores.

While it is not surprising that larger institutions have both good research rankings and significant web linkages, such a pattern still needs empirical verification and this study provides supporting non-UK evidence. The correlations found indicate the value of investigating the significance of web link based measures, but also that there are other factors that affect an institution's web presence.

This is a preliminary examination of the relationships between crawl data of the .ac.nz webspace and the PBRF results. Future work that can be done includes:

- Compare PBRF rankings and link counts for subject areas. Previously, a team has examined differences in linking patterns for different disciplines, using universities in Australia and Taiwan as examples (Thelwall, Vaughan, Cothey, Li, & Smith, 2003). The published PBRF rankings for subject areas could be compared to link counts for these subject areas. There are methodological issues here in identifying the URLs associated with subject areas in an institution. For example while Professor Ian Witten's PBRF research output is associated with the Department of Computer Science at Waikato University, much of his research output lies in the New Zealand Digital Library (<http://www.nzdl.org>) rather than in the departmental domain (<http://www.cs.waikato.ac.nz>).
- Compare PBRF rankings with links from other communities (e.g. universities in the UK or US). This study was restricted to links between NZ universities, and thus does not reflect their recognition (or otherwise) from the wider academic world. Crawl data from Australian and UK universities could be mined for links to .ac.nz institutions.
- Extend the crawl to other tertiary institutions, e.g. polytechnics. The PBRF exercise included a total of 22 institutions. However the significant research outputs in the PBRF exercise came from the eight universities, rather than from the other institutions.

## References

- An, L., & Qiu, J. (2003). Research on the relationships between Chinese journal impact factors and web impact factors and external web link counts. *Journal of the China Society for Scientific and Technical Information*, 22(4), 398-402.
- Bar-Ilan, J. (1999). Search engine results over time - a case study on search engine stability. *Cybermetrics*, 2/3(1). <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p1.html>
- Bar-Ilan, J. (2001). Data collection methods on the Web for informetric purposes: a review and analysis. *Scientometrics*, 50(1), 7-32.
- Björneborn, L. (2003). *Small-world link structures across an academic Web space: a library and information science approach*. PhD Thesis. Copenhagen, Denmark: Royal School of Library and Information Science.
- Björneborn, L., & Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics*, 50(1), 65-82.
- Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, 55(14), 1216-1227.
- Faba-Perez, C., Guerrero-Bote, V. P., & De Moya-Anegón, F. (2003). Data mining in a closed Web environment. *Scientometrics*, 58(3), 623-640.
- Heimeriks, G., Hörlesberger, M., & van den Besselaar, P. (2003). Mapping communication and collaboration in heterogeneous research networks. *Scientometrics*, 58(2), 391-413.
- Ingwersen, P. (1998). Web Impact Factors. *Journal of Documentation*, 54(2), 236-243.
- Ortega Priego, J. L. (2003). A Vector Space Model as a methodological approach to the Triple Helix dimensionality: A comparative study of Biology and Biomedicine centres of two European national research councils from a Webometric view. *Scientometrics*, 58(2), 429-443.
- Prime, C., Bassecoulard, E., & Zitt, M. (2002). Co-citations and co-sitations: A cautionary view on an analogy. *Scientometrics*, 54(2), 291-308.
- Rousseau, R. (1997). Sitations: An exploratory study. *Cybermetrics*, 1(1), Retrieved November 20, 2003, from <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html>
- Smith, A. G., & Thelwall, M. (2002). Web Impact Factors for Australasian universities. *Scientometrics*, 54(3), 363-380.
- Tertiary Education Commission. (2004). *Overview and Key Findings: Performance-Based Research Fund: Evaluating Research Excellence: the 2003 assessment*. [http://www.tec.govt.nz/downloads/a2z\\_publications/pbrf\\_report\\_overview.html](http://www.tec.govt.nz/downloads/a2z_publications/pbrf_report_overview.html)
- Thelwall, M. (2001). Extracting macroscopic information from web links. *Journal of the American Society for Information Science & Technology*, 52(13), 157-1168.
- Thelwall, M. (2003). A free database of university web links: Data collection issues. *Cybermetrics*, 6/7(1). <http://www.cindoc.csic.es/cybermetrics/articles/v6i1p2.html>
- Thelwall, M. (2004). Methods for reporting on the targets of links from national systems of university Web sites. *Information Processing & Management*, 40(1), 125-144.
- Thelwall, M., & Harries, G. (2004). Do the web sites of higher rated scholars have significantly more online impact? *Journal of the American Society for Information Science and Technology*, 55(2), 149-159.
- Thelwall, M., & 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.
- Thelwall, M., Vaughan, L. & Björneborn, L. (2005). Webometrics. In: *Annual Review of Information Science and Technology* 39, 81-135.
- Thelwall, M., Vaughan, L., Cothey, V., Li, X., & Smith, A. G. (2003). Which academic subjects have most online impact? A pilot study and a new classification process. *Online Information Review*, 27(5), 333-343.
- Vaughan, L., & Wu, G. (2003). Link counts to commercial Web sites as a source of company information. *Proceedings the 9th International Conference of Scientometrics and Informetrics*, 321-329.
- Wilkinson, D., Harries, G., Thelwall, M., & Price, L. (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(1), 49-56.

## Appendix 1: summary of data

Name	Pages	page inlinks directory inlinks	domain inlinks	site inlinks	page outlinks directory outlinks	domain outlinks	site outlinks	PBRF Quality Score	FTE staff	PBRF overall output	WIF	Domain inlinks/FTE	
auckland.ac.nz	69720	1011	636	259	7	991	712	270	7 3.96	1412	5591	0.015	0.183
aut.ac.nz	8575	87	74	34	6	84	81	53	7 0.77	568	437	0.010	0.060
canterbury.ac.nz	29157	368	320	196	7	417	344	197	7 3.83	590	2260	0.013	0.332
lincoln.ac.nz	3590	191	141	64	7	44	32	23	6 2.56	195	500	0.053	0.328
massey.ac.nz	25734	404	334	177	7	248	196	143	7 2.11	1226	2586	0.016	0.144
otago.ac.nz	43255	429	367	168	7	405	345	210	7 3.23	1175	3795	0.010	0.143
vuw.ac.nz	79241	639	440	163	7	1048	675	197	7 3.39	579	1964	0.008	0.281
waikato.ac.nz	35594	557	428	167	7	449	355	135	7 2.98	536	1598	0.016	0.311

## Appendix 2: Link counts between sites

To From	auckland.a c.nz	aut.ac.nz	canterbury. ac.nz	lincoln.ac.n z	massey.ac .nz	otago.ac.n z	vuw.ac.nz	waikato.ac. nz	Total
auckland.ac.nz	0	19	121	40	163	134	354	160	991
aut.ac.nz	22	0	7	5	10	11	14	15	84
canterbury.ac.nz	112	5	0	68	41	54	73	64	417
lincoln.ac.nz	15	0	11	0	3	3	3	9	44
massey.ac.nz	82	5	22	18	0	45	39	37	248
otago.ac.nz	121	9	67	21	49	0	64	74	405
vuw.ac.nz	512	37	71	23	80	124	0	198	1045