Research Evaluation, volume 15, number 2, August 2006, pages 97–106, Beech Tree Publishing, 10 Watford Close, Guildford, Surrey GU1 2EP, England

Researching co-operation

Investigating triple helix relationships using URL citations: a case study of the UK West Midlands automobile industry

David Stuart and Mike Thelwall

The growth in the importance of co-operation between universities, industry and government to contemporary research has created a need to understand how cross-sector research collaborations occur in practice. Hence quantitative methods are needed to identify and assess the strength of connections within particular university–industry–government groups. The Web is a free source of timely information about all three sectors and hence should be tested for whether it can deliver relevant information. In this project, the potential use of web URL citations, collected through Google's API, as weak benchmarking indicators to estimate the levels of collaboration between different organisations, is explored through a case study of the automobile industry in the UK West Midlands region. URL citation practices are found to differ, depending on whether the website belongs to an academic, commercial, or governmental organisation. While URL citations may sometimes reflect real-world relationships, most real-world relationships probably do not result in an URL citation, particularly in the commercial sector.

David Stuart (Email: dp_stuart@hotmail.com) and Mike Thelwall (Email: m.thelwall@wlv.ac.uk) are at the School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton, WV1 1SB, UK.

THERE HAS BEEN A SHIFT in contemporary research practice in the academic community, with universities working more in collaboration with other types of organisation. This shift is described by Gibbons *et al*'s (1994) 'Mode 2 Science', as well as Etzkowitz and Leydesdorff's (1997) 'triple helix' of university–industry–government relations. The move towards greater collaboration is regarded by policy-makers as a positive change, as can be seen by its importance within the European Framework Programme (Widhalm *et al*, 2001). Hence it is useful to be able to measure how these collaborations are appearing (Nowotny *et al*, 2001), so that organisations can compare themselves with other organisations to determine their own relative strengths and weaknesses, and to help policy-makers and researchers to understand and monitor the process.

The Web is a free source of timely information about all three sectors of the triple helix and hence should be tested for whether it can deliver relevant information regarding inter-sector collaborations. Until now most studies have focused on intra-sector collaboration. This can be attributed to the size of the Web, necessitating a narrow focus, and the difficulties in drawing conclusions for heterogeneous groups of organisations. Within the business sector, relationships have been found between online and offline phenomena, for example between an organisation's inlinks and profitability for homogenous businesses (Vaughan, 2004; Vaughan and Wu, 2004) but not for heterogeneous businesses (Vaughan and Wu, 2004); differences in the

linking behaviours of organisations from different sectors as well as different fields is likely to be even greater.

While there have been a number of investigations focusing on intra-sector interlinking within the academic community (eg Thelwall and Smith, 2002; Thelwall, 2003) it is not yet clear how this interlinking relates to collaboration. Kretschmer and Aguillo (2004) have shown that academic collaboration visible on the Web follows the same rules as collaboration through bibliometric data to some extent; using the web visibility of multi-authored publications. Co-authorship is a limited form of collaboration however, and does not take into account the wider, more subtle, forms of collaboration that do not lead to shared formal scholarly publications.

In an inter-sector investigation, Heimeriks *et al* (2003) investigated the web links between organisations involved in European Union-funded research using two levels of each selected organisation's website. They found that the network of web links is different to the networks expressed through co-authorship and collaboration, attributing this to the Web being the social and public interface for research organisations. This important study needs to be corroborated with further research with different geographic areas and sizes, and with different methods.

An alternative in determining what is represented by web links is through a classification of the web links. One study showed that evidence for the transfer of academic knowledge to the commercial sector was sparse (Thelwall, 2004c), although another showed that links could reveal examples of academic knowledge transfer to the general public (Thelwall and Harries, 2004), which is the fourth component of a revised triple helix model (Leydesdorff and Etzkowitz, 2003).

Leydesdorff and Curran's (2000) investigation into triple helix relations on the Web searched for the broad terms "university", "industry" and "government" within various sets of web pages, using hit counts as potential indicators of triple helix behaviour. An investigation into the web pages is needed to determine the extent to which the co-occurrences of the text reflect genuine triple helix relationships.

This exploratory study investigates the potential use of URL citations (defined below) as "weak benchmarking indicators" to estimate levels of collaboration between organisations within the different sectors of the triple helix model: university, industry and government. Weak benchmarking indicators are indicators that may provide useful information, but have not met the required reliability and validity criteria to be used in an evaluative role (Thelwall, 2004a). In other words their values are indicative or suggestive rather than definitive.

One useful application is in identifying outliers: websites that do not fit the expected pattern for their type.

Data collection methods in webometrics

Data collection is a key issue in web-based research. Search engines have been used extensively in webometric research since the term was first coined (Almind and Ingwersen, 1997), and before (Larson, 1996). The reason for their extensive use is that their search interfaces provide access to large quantities of web pages, far more than individual academics and research groups could harvest, even if using a web crawler. This is facilitated by website managers who often want their sites to be crawled and indexed by search engines, while they may see academic crawlers as a hindrance and want to keep them out. In recent years search engine crawlers have become more sophisticated and are now capable of crawling pages in a variety of formats such as Word documents and PDF files. Unfortunately the details of the crawling and ranking algorithms of search engines tend to be hidden, both to keep a competitive edge over rivals and to prevent spamming. This has resulted in much research on search engines and the results they provide (Bar-Ilan, 2004).

As search engines play an important role in Internet research, bias is an obvious cause for concern. While it has been acknowledged that a level of bias does exist within search engines, the attributed reasons differ. A recent study concluded that biased coverage in favour of US websites, compared with those of China, Taiwan and Singapore, appeared to be caused by technical and historical reasons, for example, early adoption of the Web in the USA (Vaughan and Thelwall, 2004), while other studies have suggested that political and economic motivations may also be involved (Introna and Nissenbaum, 2000; Van Couvering, 2004). The result of this bias in search engines is that the greater the diversity between the organisations and the countries that are being examined, the less reliable are any conclusions that are drawn from the results.

In addition to questions regarding a search engine's crawling and indexing methods, there have been a number of questions regarding the consistency with which search engines provide results. A number of early webometric studies found search engines to be inconsistent in terms of differences between search engines and differences for the same search engine at different points in time (Rousseau, 1999; Bar-Ilan, 1999; Lawrence and Giles, 1999). Although subsequent research suggested that there had been some improvement with the stability of AltaVista's results (Thelwall, 2001), these hopes were ill-founded (Wouters and Gerbec, 2003). Consistency of results is important if results are to be validated, but limitations in the measurement of the Web are unavoidable. There will always be pages that have not been indexed, or have changed since they were last indexed, whatever methods are used to collect source data. What is important is that the uncertainty is recognised, is minimised as far as possible, and the remaining

level of uncertainty is not disproportionate to the data that is being collected.

Google's web API (Applications Programming Interface) service allows for the automatic retrieval of information from the Google database by up to 1,000 queries per day <www.google.com/apis/>. This is a valuable tool for the gathering of information from a commercial search engine without having to laboriously type in each of the queries one at a time, since 'scraping' search engine results is not normally allowed, that is, retrieving web pages from a search engine for particular queries and extracting the information needed automatically. An initial study by Mayr and Tosques (2005) has shown that Google API has similar reliability problems to those that have been found with other search engines; as well as utilising a smaller and less regularly updated database than the standard Google interface. Nevertheless, it is a logical tool for large-scale webometric research.

Research methodology

This initial exploratory study into university–industry–government relations focuses on one specific industry in a small area of the UK, the automobile industry in the West Midlands, to address the question: Can URL citations be used as weak benchmarking indicators to determine levels of collaboration between organisations within the different sectors of the triple helix model? This is a case study style of research, which is appropriate for a new research topic.

Car production is an established industry in the West Midlands and, by focusing on the more significant companies in the automobile industry, the region and industry provides a number of websites that is not too large for the capabilities of the Google API. The West Midlands region incorporates 13 higher education institutions (West Midlands Higher Education Association, 2005), 39 principal local authorities (DirectGov, 2005), and the automotive industry unit of the Department of Trade and Industry highlights on their website 14 organisations that play a significant role in the automobile industry in the region (DTI, 2005). For convenience, all higher education institutions in the study will be described as universities, even though most do not hold this legal status (eg not having the power to award PhDs).

Google allows the retrieval of pages that link to a specified page, but not to a specified website, so it would not be of use to this study. Note also that links that do not echo the link target URL in the text of the web page will be missed by this method. This matching process has been described as *URL citation* (Kousha and Thelwall, 2005) to distinguish it from linking, although the two may often be the same and have similar characteristics.

Organisational home pages can have multiple URLs. Some organisations use more than one domain name. For example the University of Wolverhampton utilises two domains:

- <.wlv.ac.uk>
- <.wolverhampton.ac.uk>.

It is also the case that some small organisations use space on another organisation's server, and do not have a separate domain name. Information on the different domain names for universities is readily available (University of Wolverhampton, 2005) and, where applicable, multiple domain names were used as search terms. Each of the industry and local government websites were also visited to determine whether the identified domain was still the principal focus. When a page was redirected to another URL both the old and the new URLs were used as search terms.

The Google web Application Programming Interface allows queries to be sent to the Google web API database in a number of different programming languages, and allows the retrieval of a range of information for each of the queries, including the estimated total number of hits for a particular query, and the retrieval of the URLs for a limited number of the hits (the first 1,000).

Number of pages indexed

As Google does not crawl and index every web page of every website it is necessary to determine that each of the websites involved in this investigation have had sufficient pages crawled and indexed to allow for conclusions to be drawn about URL citations placed on the sites. A program was written to send queries to Google API, utilising the Google's "allinurl:" facility, for example: <allinurl:.wlv.ac.uk>. This query would list indexed pages with <.wlv.ac.uk> in their URL. The "allinurl:" facility was used as opposed to "site:" as the functionality of the "site:" facility was found to often be inaccurate unless used in conjunction with some other keyword; it often states that a query produced no hits despite the addition of a more restrictive keyword producing hits.

Number of URL citations between websites

Another program was written to determine the number of URL citations between the different organisations. This was accomplished by utilising the Google Web Application Programming Interface phrase search and site restricted search capabilities, for example: <".wolverhampton.gov.uk" site:.wlv.ac.uk>. This query would retrieve information from the University of Wolverhampton domain that included <.wolverhampton.gov.uk>, the domain name of the Wolverhampton local government website. The appearance of an organisation's URL in a web page is likely to often be a link to that organisation's website.

In the mean number of URL citations from one organisational sector to another, there are wide variations, with the strongest connections being among the universities

Confirmatory URL citation analysis

In order to interpret URL citation counts the link analysis methodology uses a classification of the purposes of web links (Thelwall, 2004b). The Google API limits the URLs that can be received to any 10 consecutive hits in the first 1,000 returned. While it would be possible to retrieve a random sample of non-consecutive URLs from the first 1,000 hits, the retrieval would mean repetition of similar queries and, as the number of automatic queries that can be sent in one day are restricted, it would mean a lengthy data collection exercise. Even selecting 10 consecutive URLs from a random position in the first 1,000 hits would double the time required; an initial query would be needed to establish how many URL citations there were before a random group of 10 could be extracted on a second query.

We chose not to include an extensive content analysis of the web pages matched by the Google API. Instead, where patterns emerged further investigations were made through browsing the relevant websites. This approach focuses on individual cases of significance, rather than gaining a global picture, and has been described as significant anomaly identification (Thelwall and Price, 2005).

Results

Size of websites

Table 1 gives the estimated number of pages that have been indexed by the Google API database. While there is a wide variety in the size of the websites, there is some overlap. Although the average size of an industry website is smaller than both the average government and university website, the largest industry websites are bigger than the smallest government and university websites.

URL citation practices

Using the estimated number of hits from the Google API, Table 2 shows the mean number of URL citations from one organisational sector to another. There are wide variations, with the strongest connections being among the universities, and the weakest connections all involving the industry websites. To show that the differences are not caused by large URL citation counts between a single pair of institutions, Table 3 shows the percentage of possible relationships that are reflected by at least one URL citation. Comparing the interlinking of websites at different levels — that is, at the page level and the site level — has been used in other studies to find which provides the best indicator of research productivity (Thelwall, 2002b). Table 3 reinforces the pattern exhibited in Table 2. The variation in the average number of URL citations between different organisations cannot be attributed solely to excessively large URL citation practices between any particular pair of institutions.

Government-to-government URL citations

There was a relatively high mean number of URL citations between the 39 principal local authorities, although these figures have been skewed by certain councils that were found to URL-cite more highly than others. Of the 1,482 relationships that could be expressed via an URL citation from one institution to another (each of the 39 local authorities could have a relationship with each of the other 38 local authorities), 1,267 did not have any URL citations.

Figure 1 shows the interconnections between the local authorities' websites, with the size of the arrows being used to indicate the number of URL citations represented. The organisations' positions on the diagram have been determined by the Kamada and Kawai (1989) algorithm, which tries to position connected sites close together although ignoring the strength of interconnection between a pair of

Table 1. Estimated web site sizes as reported by the Google API on 30 June 2005

	Minimum	Maximum	Mean	Median
Government (39)	105	57,500	4,707	1,530
Industry (14)	1	3,756	870	489
University (13)	699	186,000	36,197	11,780

Table 2. The mean number of URL citations between organisations, by sector

		To		
		Government	Industry	University
From	Government	1.371	0.005	0.450
	Industry	0	0	0
	University	0.673	0.126	10.090

Table 3. The percentage of possible relationships utilised

		To		
		Government	**Industry**	**University**
from	Government	215/1482 = 14.5%	3/546 = 0.5%	52/507 = 10.3%
	Industry	0	0	0
	University	79/507 = 15.6%	11/182 = 6.0%	102/156 = 65.4%

organisations (Leydesdorff and Vaughan, 2006). While <birmingham.gov.uk> can be seen to be the centre of the network, with URL citations to 29 of the 38 other government websites, 1,693 of the 2,032 inter-government URL citations are to or from <warwickshire.gov.uk>. The large number of web URL citations from <warwickshire.gov.uk> to the district councils who cover the same area is due in part to an extensive database of links detailing all council services in the district, including those provided by the district councils. These links are often for information purposes, rather than an expression of the two organisations working together.

The clustering of the websites in Figure 1 corresponds with the two-tier structure of some local authorities (see Table 4). While some local authorities are unitary authorities, others work on a two-tier system with the county council providing certain services, and district councils providing other services. The websites of the county councils can be seen as being more connected with the other county councils and unitary authorities than the district councils that cite predominantly the URLs of other district councils under the same county council.

Among the websites researched there were only three URL citations found from a government website to an industry website: one from <warwickshire.gov.uk> to <jaguar.com>; one from <birmingham.gov.uk> to <mira.co.uk>; and one from <malvernhills.gov.uk> to <morgan-motor.co.uk>.

Inspection of the relevant web pages found that the hit for <warwickshire.gov.uk> was an email address in a database of local businesses rather than a link; bizarrely this was the email address of a chiropodist, and it was unclear why the email address was in the <jaguar.com> domain. The link to <mira.co.uk> was from a library information services database of local businesses, and did not express any particular relationship. The <malvernhills.gov.uk> link to <morgan-motor.co.uk> did reflect the type of relationship that could be expected between local government and business; it was on a section of the site with the aim to promote the region, and promote partnerships between business and the community.

Government-to-university URL citations

While there are a number of URL citations from local authority websites to university websites, there were fewer than for government-to-government URL citations. Nineteen local authorities didn't cite a single university, and no local authority cited the

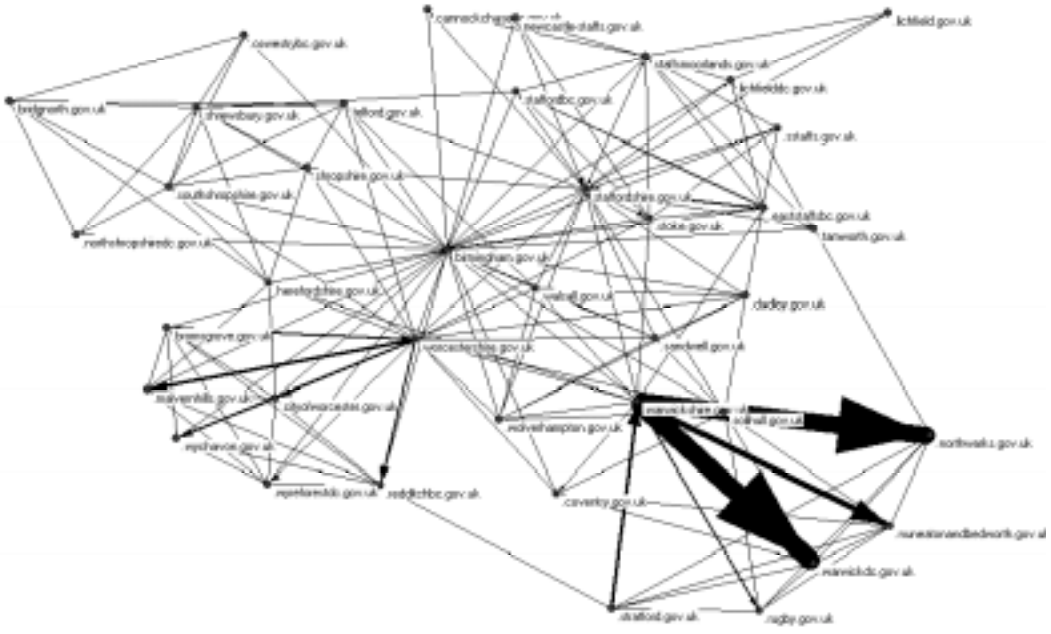

Figure 1. URL citations between local authorities' websites in the West Midlands

Table 4. County councils with their respective district councils

County council	District councils
<.worcestershire.gov.uk>	<.bromsgrove.gov.uk>; <.malvernhills.gov.uk>; <.redditchbc.gov.uk>; <.cityofworcester.gov.uk>; <.wychavon.gov.uk>; <.wyreforestdc.gov.uk>
<.warwickshire.gov.uk>	<.warwickdc.gov.uk>; <.northwarks.gov.uk>; <.nuneatonandbedworth.gov.uk>; <.rugby.gov.uk>; <.stratford.gov.uk>
<.staffordshire.gov.uk>	<.cannockchasedc.gov.uk>; <.eaststaffsbc.gov.uk>; <.lichfield.gov.uk>; <.lichfielddc.gov.uk>; <.newcastle-staffs.gov.uk>; <.sstaffs.gov.uk>; <.staffordbc.gov.uk>; <.staffsmoorlands.gov.uk>; <.tamworth.gov.uk>
<.shropshire.gov.uk>	<.southshropshire.gov.uk>; <.bridgnorth.gov.uk>; <.northshropshiredc.gov.uk>; <.oswestrybc.gov.uk>; <.shrewsbury.gov.uk>

Newman Higher Education college website, <newman.ac.uk>. While it is not surprising that it is not cited as much as the large, prestigious universities, it *is* surprising that no URL citations were found from the local government websites, especially as it is on the outskirts of Birmingham, and the institutions in the metropolitan borough of the West Midlands form the centre of the graph. Of the 507 relationships that could be expressed with web URL citations, only 52 were. As can be seen in Figure 2, once again the URL citations follow geographic trends.

Industry URL citations

No URL citations were found from any of the organisations in the industry sector to any of the other organisations in the study.

University-to-government URL citations

The results (Figure 3) show higher education institutions to be the most outward-looking of the three types of organisation. Only the Birmingham College of Food, Tourism and Creative Studies, <bcftcs.ac.uk>, didn't cite any local authorities. Twelve local authorities' websites weren't cited by any West Midlands university: <northshropshiredc.gov.uk>, <bridgnorth.gov.uk>, <oswestrybc.gov.uk>, <shropshire.gov.uk>, <lichfielddc.gov.uk>, <rugby.gov.uk>, <southshropshire.gov.uk>, <wychavon.gov.uk>, <northwarks.gov.uk>, <malvernhills.gov.uk>, <nuneatonandbedworth.gov.uk>, <redditchbc.gov.uk>.

University-to-industry URL citations

There was a small amount of URL citing from university websites to industry websites (Figure 4).

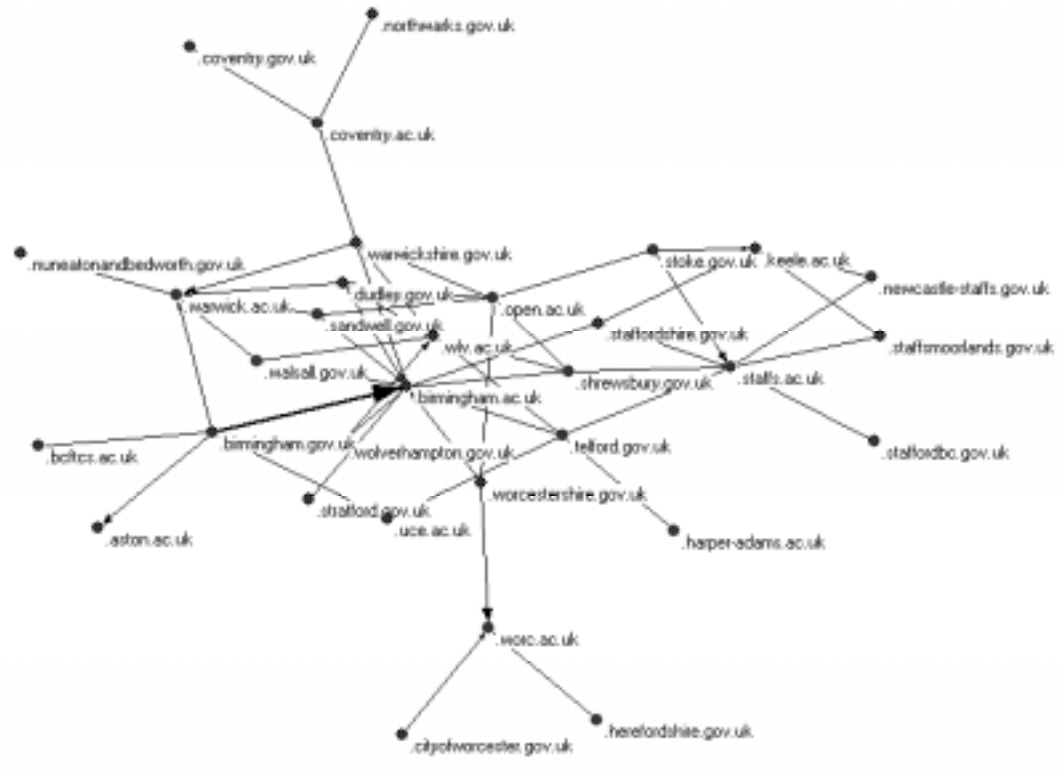

Figure 2. Government-to-university URL citations

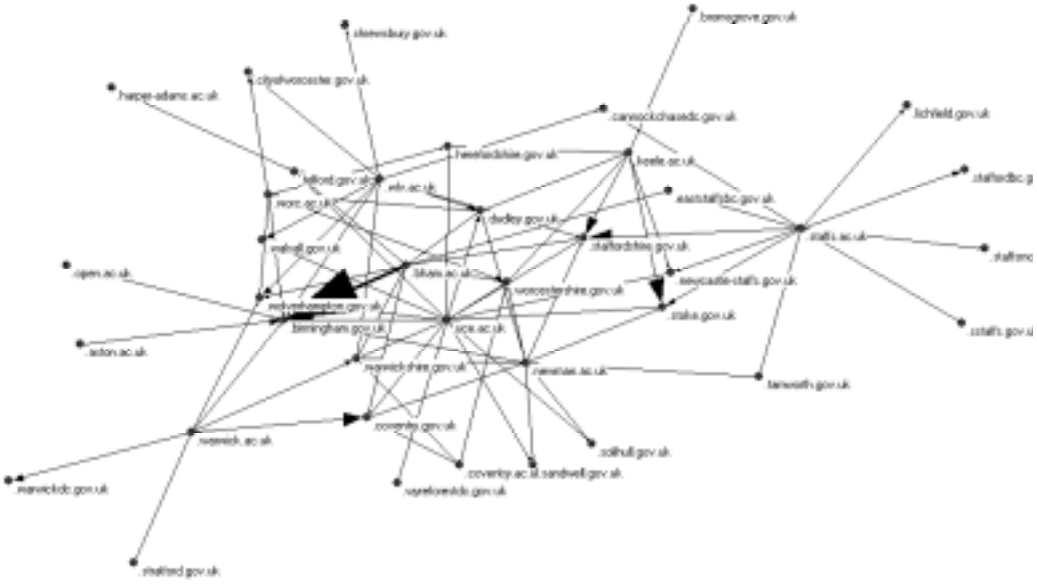

Figure 3. University-to-government URL citations

Seven of the universities didn't cite to any of the industry sites and eight of the industry sites weren't cited to. While a variety in citing practices could be expected due to different universities having differing academic focuses; there are also universities with specific automotive courses, but no URL citations. For instance, Aston University has a high research rated engineering department, with a course in automotive product design, but the Google API found no URL citations from <aston.ac.uk> to any of the industry websites in this study. This affirms that URL citations do not fully reflect real-world connections.

University-to-university URL citations

Unsurprisingly the university sector is highly interconnected, although two universities don't cite any of the other universities: Birmingham College of Food, Tourism and Creative Studies, and Harper Adams University College (Figure 5). The mean number of institutions that the other 11 universities cite is 9.27. All universities are cited by at least three other universities.

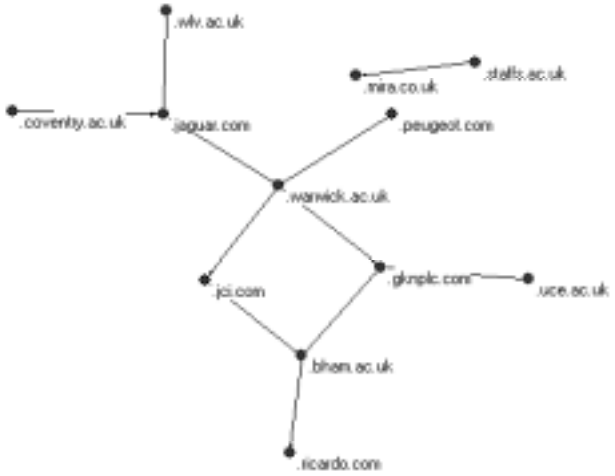

Figure 4. University-to-industry URL citations

Discussion

The results show that there are major differences in the URL citation practices of websites in the university, industry and government sectors, as would be expected between heterogeneous organisations (Vaughan and Wu, 2004). Most notable is the lack of URL citations both to and from the industry websites that have been found using Google API. It is not surprising that there are fewer URL citations from the industry websites, especially to competitors, as they have much to gain by increasing the stickiness of their website rather than encouraging surfers to move on to other websites (Shaw, 2001). That no URL citations were found from any industry website to any of the local government or university websites shows that URL citations are not a suitable way of indicating collaboration, especially in the context of an EU-funded, university-based project that gave websites to automotive supply chain companies in the area (Costello *et al*, 1999).

The weak network reinforces the findings of Heimeriks *et al* (2003). Their study did not address the differences in linking behaviour of the different sectors, although it was noted that the linked network was an unrelated set of organisations. This is attributable to the proportion of the websites that were crawled in their study producing even fewer results, which may have been exacerbated by the differences that may arise due to URL citations being searched for rather than web links.

While the lack of URL citations could be due to an active attempt to keep web users on the website,

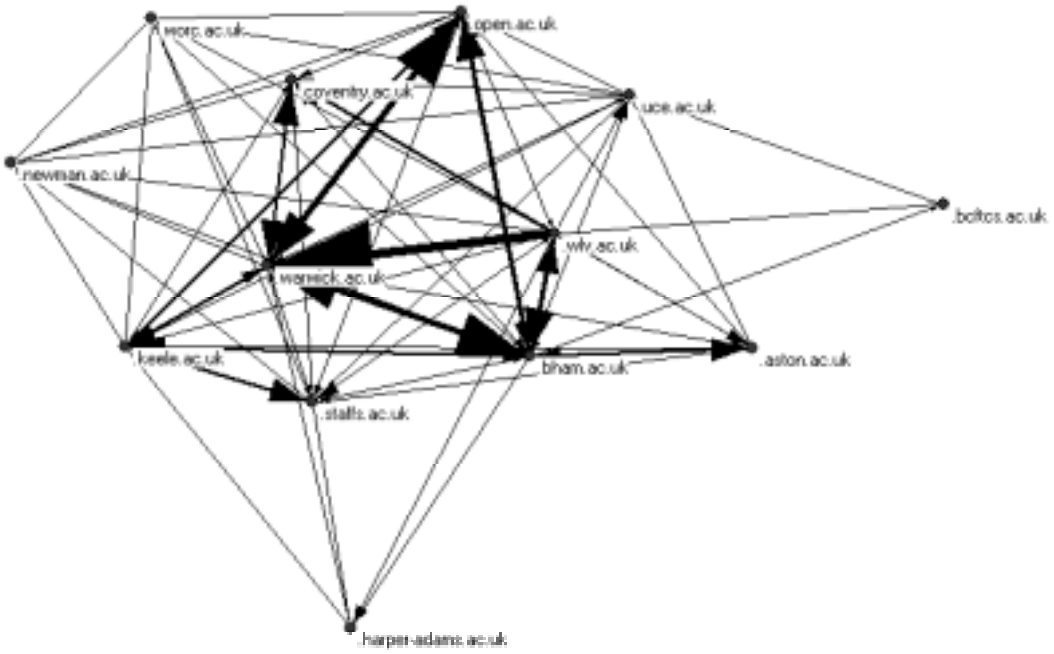

Figure 5. University-to-university URL citations

it could also be caused indirectly due to the type of information placed on the Web, and who places the information. It could be that the collaboration that occurs between an organisation and local government and universities are not seen as worthy as highlighting on a website that will be seen internationally, or that the collaborations are mentioned, but do not take the form of web URL citations.

It is also possible that the URL citations exist, but there are restrictions in the indexes of the pages by: the organisations not wanting the pages indexed, and utilising the robots.txt protocol; Google choosing not to index the pages; or the pages not being in a format that can be read and indexed by the Google robot. While it is the case that the industry websites have fewer pages indexed by Google API, there are still a significant number of pages indexed. It seems unlikely that technical reasons such as these could account for the large differences in URL citation practices found. It is more likely that websites for the three sectors have different priorities. Table 5 gives a possible simple type variation.

It is also possible that there were a number of URL citations that were not found as they were to domains that were not incorporated within the study. Organisations in the industry sector may have a diverse web presence, with many different websites for different subsidiaries or for different purposes. This raises the question of what comprises an organisation and which websites should be investigated. Affiliations between websites are often established to emphasise the credibility of the websites, with financial websites occupying a central position in the network (Park *et al*, 2002).

While more URL citations were found from local government and universities to the industry websites, these were still relatively few. This is worrying and highlights quite a disconnected web. That seven of the

It could be that collaboration between an organisation and local government and universities is not seen as worthy as highlighting on a website that will be seen internationally

Table 5. Suggested website types for university, industry and government sectors

Sector	Website main functions	Possible common uses for links and URL citations
University	Communication between individuals and groups; providing access to useful online information; promoting the organisation (Middleton *et al*, 1999)	Pointing to other information sources; acknowledging collaboration
Industry	A marketing tool (Shaw, 2001) providing commercial and non-commercial information, entertainment; aiding transactions (Huizingh, 2000)	Accessing services such as online payment forms; reflecting organisational hierarchies
Government	Promotional; content provider; aiding transactions (Musgrave, 2004)	Pointing to information of use to the public; reflecting organisational hierarchies

14 industry sites weren't URL-cited by any of the local authorities or universities is not particularly healthy from the organisations' perspectives. Even if a lack of web URL citations does not reflect a lack of collaboration, it does reflect a significant lack of web presence, something that is important to the promotion of their own organisations, and the region.

Local governments' concern with local issues is highlighted by their URL citation patterns. The directional graph showing the citing between the different local authorities shows clusters of the district councils, with their respective county councils, as well as citing between those authorities that are next to each other. URL citation patterns from government to universities also follow a geographic pattern. Analysis of the URL citations finds that they do not necessarily reflect relationships between the organisations, but are often used as sources of information and the numbers are very much reliant on the URL citation or linking policy of the individual site. It is likely that the URL citation patterns would differ for government institutions at a regional level.

The university sector was by far the most heavily connected; more outward-looking to the other sectors, and heavily connected to the other institutions within the university sector. This is not unexpected, both because of the early adoption of the Internet by the academic sector, and their being involved in the dissemination of information. They are still relatively disconnected, however. Note that the link model developed for university website interlinking (Thelwall, 2002a) is symmetrical and so the asymmetric nature of links between sectors is a significant indicator that their websites are used for different purposes.

As the numbers of web URL citations between the organisations is small, there is an increased need for transparency in the way search engines work, and how much or how often each organisation's website is indexed; just a handful of URL citations could change an organisation's position relative to others in the field. The limited number of URL citations between sites does mean that a random sample of pages could be selected that cite from one organisation to another. In the vast majority of cases there are fewer than 10 URL citations between any two websites, and where there are more than 10, it is fewer than the 1,000 the Google API allows to be retrieved.

When looking at a small selection of websites, even within a limited geographical area, there are fewer connections than may be expected in the real world. Therefore it is necessary to look at larger numbers of websites for a reliable picture to emerge, and with the number of queries necessary growing exponentially, it is not practical to use a similar methodology on a larger scale.

Previous studies have found that the majority of university outlinks do not show a working relationship between the organisations concerned, but rather are placed for other reasons such as pointing to information resources (Stuart and Thelwall, 2005). Analysis of a small selection of the web URL citations from local authority websites would seem to confirm the same type of relationships. That the URL citations still produce clusters of known real-world relationships suggests that these relationships are reflected as URL citations, but not necessarily ones that explicitly state the relationship.

Conclusions

This exploratory study has investigated the potential use of web URL citations as weak benchmarking indicators to determine levels of collaboration between organisations within the different sectors of the triple helix model. The results show that URL citation only partially reflects real-world relationships because of different purposes for websites between sectors, and also because there is no imperative to create links or URL citations to advertise a collaboration. It seems likely that the use of web links and URL citations would be as one variable used in conjunction with one or more other manifestations of relationships between organisations, and that web-based indicators will always give only a partial picture of collaboration, particularly collaboration concerning industry.

There is a need for more research into the web use of the non-academic community: how they utilise their websites and their reasons for link placement. The differences between the university, industry and government sectors have been emphasised, and a far more extensive classification of websites is likely to be needed if meaningful assertions are to be made about what sort of weighting can be ascribed to different web links or URL citations. The weighting of web links is clearly needed, as a link from some sites is clearly of more value than a link from another.

Even if enough information were known for appropriate weightings to be ascribed to the different meanings of web links or URL citations, it would be necessary for other manifestations to be taken into account. While URL citations provide useful information, they are not enough on their own. The small number of connections between the websites means that it is necessary for subtler connections to be analysed, and future research into the triple helix manifestations on the Web may find text analysis rather than link analysis or URL citation analysis a more productive avenue to follow: searching for occurrences of one organisation's name on another organisation's websites; or co-occurrences of organisations' names on other websites.

References

Almind, T C and P Ingwersen 1997. Informetric analyses on the world wide web: methodological approaches to 'webometrics'. *Journal of Documentation*, **53**(4), 404–426.

Bar-Ilan, J 1999. Search engine results over time: a case study on search engine stability. *Cybermetrics*, **2/3**(1), paper 1.

Available at <http://cybermetrics.cindoc.csic.es/pruebas/v2i1p1.htm>, last accessed 4 October 2005.

Bar-Ilan, J 2004. The use of Web search engines in information science research. In *Annual Review of Information Science and Technology*, ed. Blaise Cronin, Vol. 38, pp. 231–288. Medford, NJ: Information Today.

Costello, P, S Garner, G Homer and D Thompson 1999. Can the Internet provide the West Midlands automotive industry with the ultimate lean production tool? A case study of the Auto Lean Project. *Second International SMESME Conference: Manufacturing and Business Systems Group*. University of Plymouth. 29–31 March 1999.

DirectGov 2005. *Local Councils in the West Midlands*. Available at <http://www.direct.gov.uk/QuickFind/LocalCouncils/LocalCouncilArticle/fs/en?CONTENT_ID=4003650&chk=BVWOSS>, last accessed 12 June 2005.

DTI< Department of Trade and Industry 2005 *Auto Industry: West Midlands.* Available at <http://www.autoindustry.co.uk/regions/westmidlands/2>, last accessed 12 June 2005.

Etzkowitz, H and L Leydesdorff 1997. *Universities and the Global Knowledge Economy: A Triple Helix of University-Industry-Government Relations*. London: Pinter.

Gibbons, M, C Limoges, H Nowotny, S Schwartzman, P Scott and M Trow 1994. *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*. London: Sage.

Heimeriks, G, M Horlesberger and P van den Besselaar 2003. Mapping communication and collaboration in heterogeneous research networks. *Scientometrics*, **58**(2), 391–413.

Huleingh, E K R E 2000. The content and design of websites: an empirical study. *Information and Management*, **37**(3), 123–134.

Introna, L D and H Nissenbaum 2000. Shaping the Web: why the politics of search engines matters. *Information Society*, **16**(3), 129–185.

Kamada, T and S Kawai 1989. An algorithm for drawing general undirected graph. *Information Processing Letters*, **31**(1), 7–15.

Kousha, K and M Thelwall 2005. Motivations for URL citations to open access library and information science articles. In *Proceedings of ISSI 2005: The 10th International Conference of the International Society for Scientometrics and Informetrics. Vol. 1*, eds. P Ingwersen and B Larsen, pp. 67–77. Stockholm: Karolinska University Press.

Kretschmer, H and I Aguillo 2004. Visibility of collaboration on the Web. *Scientometrics*, **61**(3), 405–426.

Larson, R 1996. Bibliometrics of the World Wide Web: an exploratory analysis of the intellectual structure of cyberspace. *Proceedings of AISIS 96*, 71–78.

Lawrence, S and C L Giles 1999. Accessibility of information on the Web. *Nature*, 400, 8 July, 107–109.

Leydesdorff, L and M Curran 2000. Mapping university-industry-government relations on the Internet: the construction of indicators for a knowledge-based economy. *CYBERmetrics*, **4**(1), paper 2. Available at <http://cybermetrics.cindoc.csic.es/pruebas/v4i1o2.htm>, last accessed 12 June 2005.

Leydesdorff, L and H Etzkowitz 2003. Can "the public" be considered as a fourth helix in university-industry-government relations? Report of the fourth triple helix conference. *Science and Public Policy*, **30**(1), 55–61.

Leydesdorff, L and L Vaughan 2006, to appear. Co-occurrence matrices and their applications in information science: extending ACA to the Web environment. *Journal of the American Society for Information Science and Technology.* Available at <http://www.leydesdorff.net/aca/aca.pdf>, last accessed 12 January 2006.

Mayr, P and F Tosques 2005. Google Web APIs: an instrument for webometric analyses? [poster] In *Proceedings of ISSI 2005: The 10th International Conference of the International Society for Scientometrics and Informetrics. Vol. 1*., eds. P Ingwesen and B Larsen, pp. 677–687. Stockholm: Karolinska University Press.

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

Musgrave, S 2004. The community portal challenge: is there a technology barrier for local authorities? *Telematics and Informatics*, **21**(2004), 261–272.

Nowotny, H, P Scott and M Gibbons 2001. *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge: Polity.

Park, H W, G A Barnett and I Y Nam 2002. Hyperlink-affiliation network structure of top websites: examining affiliates with hyperlink in Korea. *Journal of the American Society of Information Science and Technology*, **53**(7), 602–611.

Rousseau, R 1999. Daily time series of common single word searches in AltaVista and NorthernLight. *CyberMetrics*, 2/3(1). Available at <http://cybermetrics.cindoc.csic.es/pruebas/v2i1p2.htm>, last accessed 12 June 2005.

Shaw, D 2001. Playing the links: interactivity and stickiness in .com and "not.com" websites. *First Monday*, 6(3). Available at <http://firstmonday.org/issues/issue6_3/shaw/index.html>, last accessed 12 June 2005.

Stuart, D and M Thelwall 2005. What can university-to-government web links reveal about university-government collaborations? In *Proceedings of ISSI 2005: The 10th International Conference of the International Society for Scientometrics and Informetrics. Vol. 1*., eds. P Ingwesen and B Larsen, pp.188–192. Stockholm: Karolinska University Press.

Thelwall, M 2001. The responsiveness of search engine indexes. *CyberMetrics*, **5**(1). Available at <http://cybermetrics.cindoc.csic.es/pruebas/v5i1p1.htm>, last accessed 12 June 2005.

Thelwall, M 2002a. A research and institutional size based model for national university website interlinking. *Journal of Documentation*, **58**(6), 683–694.

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

Thelwall, M 2003. Web use and peer interconnectivity metrics for academic websites. *Journal of Information Science*, **29**(1), 1–10.

Thelwall, M 2004a. Weak benchmarking indicators for formative and semi-evaluative assessment of research. *Research Evaluation*, **13**(1), April, pp. 63–68.

Thelwall, M 2004b. *Web Link Analysis: An Information Science Approach*. New York: Academic Press.

Thelwall, M 2004c. Can the Web give useful information about commercial uses of scientific research? *Online Information Review*, **28**(2), 120–130.

Thelwall, M and G Harries 2004. Can personal web pages that link to universities yield information about the wider dissemination of research? *Journal of Information Science*, **30**(3), 243–256.

Thelwall, M and E Price 2005, submitted. A generic lexical framework for link list and URL list analysis.

Thelwall, M and A Smith 2002. Interlinking between Asia-Pacific university websites. *Scientometrics*, **55**(3), 363–376.

University of Wolverhampton 2005. *UK University Names and Domain Names*. Available at <http://cybermetrics.wlv.ac.uk/database/university_lists.htm>, last accessed 12 June 2005.

Van Couvering, E 2004. New media? The political economy of internet search engines. [Presented at the 2003 Conference of the International Association of Media and Communications Researcher]. Available at <http://personal.lse.ac.uk/vancouve/IAMCR-CTP_SearchEnginePoliticalEconomy_EVC_2004-07-14.pdf>, last accessed 12 June 2005.

Vaughan, L 2004. Exploring website features for business information. *Scientometrics*, **61**(3), 467–477.

Vaughan, L and M Thelwall 2004. Search engine coverage bias: evidence and possible causes. *Information Processing and Management*, **40**(4), 693–707.

Vaughan, L and G Wu 2004. Links to commercial websites as a source of business information. *Scientometrics*, **60**(3), 487–496.

West Midlands Higher Education Association 2005. Available at <http://www.wmhea.ac.uk>, last accessed 12 June 2005.

Widhalm, C, M Topolnik, A Kopcsa, E Schiebel and M Weber 2001. Evaluating patterns of co-operation: application of a bibliometric visualisation tool to the Fourth Framework Programme and the Transport Research Programme. *Research Evaluation*, **10**(2), August, 129–140.

Wouters, P and D Gerbec 2003. Interactive Internet? Studying mediated interaction with publicly available search engines. *Journal of Computer Mediated Communication*, **8**(4). Available at <http://www.ascusc.org/jcmc/vol8/issue4/wouters.html>, last accessed 20 June 2005.