

Liwen Vaughan, Margaret E.I. Kipp, Yijun Gao
Faculty of Information and Media Studies
University of Western Ontario
London, Ontario, N6A 5B7, Canada

Why are Websites Co-linked?

The Case of Canadian Universities

Abstract: A random sample Web pages that linked to a pair of Canadian universities was retrieved. The content of the page as well as the context of the link were manually examined to record the following variables: language, country, type of Website, and the reason for co-linking.

Résumé : Un échantillon aléatoire de pages web reliées à une paire d'universités canadiennes a été repéré. Le contenu de la page, aussi bien que le contexte du lien ont été manuellement examinés pour enregistrer les variables suivantes : langue, pays, type de site web et raison du co-lien.

1. Background of the Study

Many Webometrics studies have been conducted to examine the nature of Web hyperlinks. Among the various types of Websites, academic Websites are most frequently studied. Several studies have found that the number of hyperlinks (inlinks) to university Websites correlates with the quality of the university using varying definitions of quality. For example, in UK studies, quality is measured by the university's research performance as shown in UK government's official research assessment exercise (RAE) ratings (Thelwall, 2001). In the Canadian context, quality is measured by the Maclean's ranking of Canadian universities (Vaughan & Thelwall, 2005). The reliability of the quality measurement notwithstanding; the studies have shown that, when links are aggregated on a macro level, we can find patterns of linking and uncover useful information.

Vaughan (2006) further studied patterns of co-links (If page X and page Y are both linked to by page Z, then X and Y have a co-link) to Canadian university Websites. Multidimensional scaling (MDS) was used to analyze and visualize co-link data. MDS mapping results of the data sets accurately reflected the ways Canadians see the universities and clearly showed the linguistic and cultural differences within Canadian society. This shows that co-links are not created at random and that Web co-links can be a measure of the similarity or relatedness of sites being co-linked.

While these quantitative studies showed us the patterns of the Web links, qualitative studies are needed to gain a better understanding of the nature of the Web link. Parallel to the quantitative studies, various qualitative studies have been conducted to find out why hyperlinks are created, i.e. the reasons or motivations for creation of a Web link. Most of

these studies used a content analysis approach (e.g. Thelwall, 2003; Vaughan, Gao, Kipp, 2006; Wilkinson et al 2003) but other approaches, such as personal interviews (Kim, 2000), have also been used. All these studies examined the motivations for the creation of inlinks but none analyzed reasons for co-linking. The current study fills this gap. It builds on the Vaughan (2006) study by taking a content analysis approach to find out why co-links were created. Specifically, are the co-linked sites related and how are they related? It is hoped that the findings from this study will complement or triangulate findings from the quantitative study and provide us with a deeper understanding of the Web co-link phenomenon.

2. Methodology

2.1 Websites in the Study

This study used the same set of universities as that in the Vaughan (2006) study. Thus all universities ranked in the Maclean’s Guide to Canadian Universities (Johnston, 2004) were candidates for the study. URLs of the university Websites were obtained from the Maclean’s Guide and the Association of Universities and Colleges of Canada (2005). Among the 47 universities in Maclean’s Guide, five have multiple URLs in the form of redirect or alias. These URLs cause problems in collecting co-link data and they were thus excluded from the previous study (Vaughan, 2006) and consequently the current study. Among the 42 universities studied, 36 are English, four are French, and two are bilingual. See Appendix 1 for the list of the universities and related information.

2.2 Search Engines and Query Syntax

Yahoo! was used for collecting co-link data. Other major search engines that were considered, but eliminated as a choice for data collection, include Google and MSN. Google’s “link” query term cannot be combined with the “site” query term to search for external inlinks as is stated in Google’s API documentation “no other query terms can be specified when using this special query term” (Google, 2006). MSN can perform external inlink search but it can only display the first 250 retrieved items at the time of data collection (Aug. 2005) while Yahoo! can display the first 1000 items.

Yahoo! has two link search commands: link and linkdomain. The “link” command will only search for links pointing to the homepage while the “linkdomain” command will find all links pointing to the site in question including the homepage and all other pages of the site. The study intends to find co-links both to the homepage and to non-homepages and to make a comparison of the two types of co-links. Using two imaginary URLs www.abc.ca and www.xyz.ca as an example, the queries used to search for co-links are show in Table 1.

To search for	Query Used
Co-links to the homepages	(link:http://www.abc.ca –site:abc.ca) AND (link:http://www.xyz.ca –site:xyz.ca)
Co-links to non-homepages	(linkdomain:abc.ca –link:http://www.abc.ca) AND (linkdomain:xyz.ca –link:http://www.xyz.ca)

Table 1 – Search Queries

It should be noted that the query used to search for co-links to non-homepages, as shown in Table 1, does not guarantee that the co-links found are external links, i.e. links coming from sites other than the two sites in question. For example, a page in site www.abc.ca could have links to another page in site www.abc.ca and a page in site www.xyz.ca. These co-links were eliminated when they were encountered and replaced with another randomly selected co-link page.

2.3 Sampling and Classification Process

We searched for co-links between all pairs of the universities in the study alternating between searching for co-links to homepages and to non-homepages. One co-link was randomly selected from each set of search results to be manually classified. Each of these linking pages (page that initiated the co-links to the pair of universities) was classified in the following aspects: language, country, type of site, and reasons or motivations for creating the co-link.

Determining the language of the page is straightforward. Country is determined by the country of the page owner rather than the country designation in the URL. Consistent with a previous study (Vaughan, Gao, & Kipp, 2006), the type of the Website was manually assigned based on the type of organization responsible for the content of the page and not based on the domain name (e.g. .com or .org). The following five types of Websites were identified: commercial, organizational (e.g. non-profit organizations, associations, conferences), governmental, educational (including universities, colleges, schools, and professors' homepages), and personal. The classification of language, country, and type of Websites was done by only one of us as the coding of these variables was fairly objective and straightforward.

Both the content of the Webpage as well as the context of the links were examined to determine the reason or motivation for creating the co-link. As there was no prior study that classified motivations of co-link creation, we needed to develop the classification scheme from the ground up. Based on the experience of our previous study that developed a classification scheme of motivations for linking to commercial Websites (Vaughan, Gao, & Kipp, 2006), we used the following induction process to develop the classification scheme based on grounded theory (Berg, 1995, p.179-181). The first author examined a group of co-linking pages (pages that initiated co-links) and identified six preliminary categories with brief notes defining each category. Then the second and the third authors classified a group of co-link pages using the preliminary categories and revised the categories as needed. They did this exercise together to achieve a consistent understanding of the categories. They also recorded examples for each category. All three authors then discussed and finalized a classification scheme of the following five categories: general, teaching, research, association, and not related. See Appendix 2 for the definition of categories and the examples that illustrate each category. Only 3.8% of pages classified fall into the category "association" (the two universities are co-linked because they have related academic, professional or student associations). In the data analysis stage, we decided to merge this very small category with the category of "general" (the two universities are co-linked in a list of universities, university libraries, university ranking, or education and research facilities) because being related by "associations" is similar to other types of "general" relationship in that it is not specifically related to teaching or research. Thus the final categories in the "results"

section of this paper are “general”, “teaching”, “research”, and “not related”, which provides a clearer picture of if and how co-linked universities are related.

A total of 859 co-linking pages were classified. The second and the third authors classified the first 41 pages together in the process of developing the classification scheme. Then they classified the remaining 818 pages independently. The inter-coder consistency rate in classifying these 818 pages was 98%; this is a very high consistency rate that assures the validity and reliability of the data. This consistency rate is higher than that achieved in previous studies of Web page and Web link classification (Bar-Ilan, 2003; Crowston & Williams, 2000; Thelwall et al, 2003; Wilkinson et al, 2003).

3. Results

3.1 Reasons for Co-linking

The reasons or motivations for co-linking to a pair of universities are summarized in Table 2. In only 5.9% of the cases (the “not related” category), the reason that the two universities were co-linked was accidental (i.e. they happened to be mentioned on the same Web page). For the remaining 96.1% cases, the reasons that the two universities were co-linked were either for their relationship in teaching/research or in other general ways such as student organizations or libraries. It is worth noting that teaching is a much more common reason than research for co-linking. This reflects the general observation that the use of Websites for teaching (e.g. course Websites) is much more common than use of Websites for research (e.g. report research findings).

General	Teaching	Research	Not related	Total
484 (56.3%)	271 (31.5%)	53 (6.2%)	51 (5.9%)	859 (100%)

Table 2 - Reasons for Co-linking

3.2 Types of Sites

Canadian university Websites were co-linked by various types of sites as shown in Table 3. Not surprisingly, the most common type of site was educational sites but this type of site does not constitute a majority (43.3%). What is somewhat surprising is that commercial sites outnumbered government sites by about 10 times. Even personal sites (7.7%) are much more common than government sites (2.7%).

Educational	Organizational	Governmental	Commercial	Personal	Total
372 (43.3%)	158 (18.4%)	23 (2.7%)	240 (27.9%)	66 (7.7%)	859 (100%)

Table 3 - Types of Sites

3.3 Country

For the purpose of succinct presentation, European and Asian countries were merged by continent while countries in South America, Australia etc. were merged into a category called “others” (see Table 4 for the country distribution of co-linking pages). In 8.8% of the cases we were unable to determine the country from the Web page (“unknown” category in Table 4).

It is worth noting that the majority of links were from outside Canada; a sign that Canadian universities were very visible internationally. Links from the U.S.A. outnumbered those from Europe and Asia combined. Geographical vicinity seems to be a factor in co-linking as is in the case of simple inlinking (Thelwall, 2002; Vaughan & Thelwall, 2005)

Canada	U.S.A.	Europe	Asia	Others	Unknown	Total
370 (43.1%)	230 (26.8%)	100 (11.6%)	51 (5.9%)	32 (3.7%)	76 (8.8%)	859 (100%)

Table 4 - Country of Co-linking page

3.4 Language

The language of the Web pages that initiated co-links to Canadian universities is shown in Table 5 where “mixed languages” means that two or more languages were present on the page. The most common type of mix is English and another language (e.g. English and Chinese) but the mix of English and French did not occur. Among the 42 universities in the study, 36 (85.7%) are English, 4 (9.5%) are French and 2 (4.8%) are bilingual in English and French. Given this situation, the dominance of English in Table 5 is not surprising but the French language seems to be under-represented among the co-linking pages.

English	French	Other	Mixed Languages	Total
749(87.2%)	48(5.6%)	36 (4.2%)	26(3.0%)	859(100%)

Table 5 - Language of Co-linking page

3.5 Relationship between Reasons for Co-linking and Types of Sites

Table 6 shows the cross tabulation of reasons for co-linking and the types of sites that initiated the co-links. A chi-square test shows that there is a significant relationship between the two variables ($p < 0.01$). Comparing the observed frequency vs. the expected frequency (shown in brackets in each cell of Table 6), we can see co-links from educational sites are more likely to be created for teaching and research purposes and less likely to be in the “not related” category. This means that if we can limit data collection to educational sites, we can get data that better show teaching and research relationships of co-linked universities. However, implementing this strategy of data collection will be very difficult without the very time-consuming process of manual filtering as educational

sites in some countries such as Canada do not have clear designation of .edu or .ac in their URLs.

Reasons for Co-linking	Types of Sites					Total
	Educational	Organizational	Governmental	Commercial	Personal	
General	170(209.6)	93(89.0)	19(13.0)	170(135.2)	32(37.2)	484
Teaching	147(117.4)	49(49.8)	3(7.3)	54(75.7)	18(20.8)	271
Research	41(23.0)	4(9.7)	0(1.4)	3(14.8)	5(4.1)	53
Not Related	14(22.1)	12(9.4)	1(1.4)	13(14.2)	11(3.9)	51
Total	372(372.0)	158(158)	23(23.0)	240(240.0)	66(66.0)	859

Table 6 - Relationship between Reasons for Co-linking and the Types of Sites (Number shown in brackets in each cell is the expected frequency)

4. Conclusions and Discussion

The study found that in over 94% of cases, the two co-linked universities were related academically. Here academic relationship can be either in general ways (e.g. university libraries or student organizations) or specifically in teaching or research (38%). This confirms results from quantitative studies that Web co-links can be a measure of the similarity or relatedness of sites being co-linked and that Web co-link analysis can thus be used to study relationships among universities.

The majority of co-links came from sites outside Canada, a sign that Canadian universities are very visible internationally. Educational sites are the most common type of site that generated co-links to Canadian universities. Furthermore, co-links from this type of site have a higher chance (51%) of showing teaching or research relationships. If co-link data are filtered to be from educational sites only, the data will contain better information on teaching and research relationships among universities and thus will be better objects for Web data mining.

Acknowledgement: This study is part of a larger project funded by the Initiative on the New Economy (INE) Research Grants program of the Social Sciences and Humanities Research Council of Canada (SSHRC).

References

- Association of Universities and Colleges of Canada. 2005. Canadian Universities. http://www.aucc.ca/can_uni/our_universities/index_e.html. (accessed January 17, 2005)
- Bar-Ilan, J. 2004. A microscopic link analysis of academic institutions within a country- the case of Israel. *Scientometrics* 59, no. 3: 391-403.
- Berg, B.L. 1995. *Qualitative research methods for social science*, 2nd ed. Boston: Allyn and Bacon.
- Crowston, K. & Williams, M. 2000. Reproduced and emergent genres of communication on the world wide web. *The Information Society* 16: 201-215.
- Google. 2006. *Google Web APIs Reference*, http://www.google.com/apis/reference.html#2_2 (accessed April 11, 2006).
- Johnston, A.D. (editor) 2004. *Maclean's guide to Canadian universities 2004*, Toronto, Canada: Rogers Publishing.
- Kim, H. J. 2000. Motivations for hyperlinking in scholarly electronic articles: a qualitative study. *Journal of the American Society for Information Science* 51, no.10: 887-899.
- Thelwall, M. 2003. What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation. *Information research* 8, no.3, <http://informationr.net/ir/8-3/paper151.html> (accessed April 11, 2006).
- Thelwall, M. 2002. Evidence for the existence of geographic trends in university web site interlinking. *Journal of Documentation* 58, no.5: 563-574.
- Thelwall, M. 2001. Extracting macroscopic information from Web links. *Journal of the American Society for Information Science and Technology* 52, no.13: 1157-1168.
- Thelwall, M., Vaughan, L., Cothey, V., Li, X. and Smith, A. G. 2003. Which academic subjects have most online impact? A pilot study and a new classification process. *Online Information Review* 27, no.5: 333-343.
- Vaughan, L. 2006. Visualizing linguistic and cultural differences using Web co-link data. To appear in *Journal of the American Society for Information Science and Technology*.
- Vaughan, L., Gao, Y. & Kipp, M. 2006. Why are hyperlinks to business Websites created? A content analysis. *Scientometrics* 67, no.2: 291-300.

Vaughan, L. & Thelwall, M. 2005. A modeling approach to uncover hyperlink patterns: The case of Canadian universities. *Information Processing & Management* 41, no.2: 347-359.

Wilkinson, D., Harries, G., Thelwall, M. & Price, E. 2003. Motivations for academic Web site interlinking: Evidence for the Web as a novel source of information on informal scholarly communication. *Journal of Information Science* 29, no.1: 59-66.

Appendix 1 – Universities in the Study

Name	URL	Language	Province
Acadia University	http://www.acadiau.ca	English	NS
University of Alberta	http://www.ualberta.ca	English	AB
Bishop's University	http://www.ubishops.ca	English	QC
Brandon University	http://www.brandonu.ca	English	MB
The University of British Columbia	http://www.ubc.ca	English	BC
Brock University	http://www.brocku.ca	English	ON
University of Calgary	http://www.ucalgary.ca	English	AB
Carleton University	http://www.carleton.ca	English	ON
Concordia University	http://www.concordia.ca	English	QC
Dalhousie University	http://www.dal.ca	English	NS
University of Guelph	http://www.uoguelph.ca	English	ON
Lakehead University	http://www.lakeheadu.ca	English	ON
Laurentian University of Sudbury	http://www.laurentian.ca	bilingual	ON
Université Laval	http://www.ulaval.ca	French	QC
McGill University	http://www.mcgill.ca	English	QC
McMaster University	http://www.mcmaster.ca	English	ON
Memorial University of Newfoundland	http://www.mun.ca	English	NF
Université de Moncton	http://www.umoncton.ca	French	NB
Université de Montréal	http://www.umontreal.ca	French	QC
Mount Allison University	http://www.mta.ca	English	NB
Mount Saint Vincent University	http://www.msvu.ca	English	NS
University of New Brunswick	http://www.unb.ca	English	NB
University of Northern British Columbia	http://www.unbc.ca	English	BC
University of Ottawa / Université d'Ottawa	http://www.uottawa.ca	bilingual	ON
University of Prince Edward Island	http://www.upei.ca	English	PEI
Queen's University	http://www.queensu.ca	English	ON
The University of Regina	http://www.uregina.ca	English	SK
Ryerson University	http://www.ryerson.ca	English	ON
St. Francis Xavier University	http://www.stfx.ca	English	NS
St. Thomas University	http://www.stthomasu.ca	English	NB
University of Saskatchewan	http://www.usask.ca	English	SK
Université de Sherbrooke	http://www.usherb.ca	French	QC
Simon Fraser University	http://www.sfu.ca	English	BC
University of Toronto	http://www.utoronto.ca	English	ON
Trent University	http://www.trentu.ca	English	ON
University of Victoria	http://www.uvic.ca	English	BC

University of Waterloo	http://www.uwaterloo.ca	English	ON
The University of Western Ontario	http://www.uwo.ca	English	ON
Wilfrid Laurier University	http://www.wlu.ca	English	ON
University of Windsor	http://www.uwindsor.ca	English	ON
The University of Winnipeg	http://www.uwinnipeg.ca	English	MB
York University	http://www.yorku.ca	English	ON

Appendix 2 - Classification Scheme of Reasons for Co-linking

Definition of the category and an example that illustrates the category is provided

General (the two universities are co-linked in a list of universities, university libraries, university ranking, or education and research facilities)

Example: Web page

<http://www.webpan.com/msauers/libdir/countries/canada.html> is a list of libraries that includes libraries of Dalhousie and St. Francis Xavier. Thus these two university websites are co-linked.

Teaching (the two universities re co-linked in a list of academic departments, course offerings, course material, or ranking of departments)

Example: Web page http://web.uvic.ca/econ/depts_Uvic.html lists economics departments including that of Carleton and Wilfrid Laurier University. These two universities are co-linked by this Web page.

Research (the two universities are co-linked because of research projects, research programs, or professor's research websites)

Example: the Website of a biology lab at Berkley

<http://plantbio.berkeley.edu/~bruns/> links to biology labs at Guelph and Regina.

Associations (the two universities are co-linked because they have related academic, professional or student associations)

Example: Acadia and Laval websites are co-linked as they both appear on the Web page <http://www.cpa.ca/Info.htm> which is a list of student associations.

Not Related (the two universities are co-linked because they are mentioned together by chance, or for reasons that have nothing to do with the university as an institution)

Example: Guelph and Laval are co-linked as they both appeared on Web page <http://www.websurveyor.com/company/websurveyor-customers-edu.asp> which lists customers of the company.