

# Local government web sites in Finland: A geographic and webometric analysis

Kim Holmberg\* and Mike Thelwall\*\*

\* *kim.holmberg@abo.fi*  
Information Studies, Åbo Akademi University,  
Tavastgatan 13, 20500 Åbo, Finland

\*\* *m.thelwall@wlv.ac.uk*  
School of Computing and Information Technology, University of Wolverhampton,  
Wulfruna Street, Wolverhampton WV1 1SB, UK

## Abstract

It has been shown that information collected from and about links between web pages and web sites can reflect real world phenomena and relationships between the organizations they represent. Yet, government linking has not been extensively studied from a webometric point of view. The aim of this study was to increase the knowledge of governmental interlinking and to shed some light on the possible real world phenomena it may indicate. We show that interlinking between local government bodies in Finland follows a strong geographic, or rather a geopolitical pattern and that governmental interlinking is mostly motivated by official cooperation that geographic adjacency has made possible.

## Keywords

Webometrics; hyperlinks; link creation motivations; local government; geography

## Introduction

The web is an enormous source of information, both of a visible kind in the form of the content of web pages, and also of a more hidden kind, for example through the connections that hyperlinks create between different web sites and the organizations they represent. The research field of webometrics tries, among other things, to create new knowledge from this hidden information and to understand what kinds of real world phenomena it may represent.

Webometrics is still a young field of research and that is why a majority of webometric studies, after Almind and Ingwersen (1997) set the foundations for webometrics (and also gave the research field its name), have focused on developing and improving methods for data collection, processing and interpreting the results. These include studies about the use and coverage of commercial search engines for informetric purposes (Bar-Ilan, 1999; Ingwersen, 1998; Lawrence & Giles, 1999; Rousseau, 1997; Snyder & Rosenbaum, 1999; Thelwall, 2001a), interlinking between academic web sites on both national and international level (Thelwall, 2002a; Thelwall & Smith, 2002), networks and small worlds on the web (Björneborn, 2004; Björneborn & Ingwersen, 2004) and the use of links as indicators of business performance and business competitive positions (Vaughan, 2004; Vaughan & Wu, 2004; Vaughan & You, 2005). The results have collectively increased knowledge of how search engines function and how data gathered from the web should be processed and interpreted, how links and networks they create between web sites can be used to improve web information retrieval and how links can be used to indicate real world phenomena both in academia and business.

Although academic and, to a lesser extent, commercial web sites have been the subject of many webometric studies, much less has been written about government web sites. In this paper we seek to redress this imbalance through a case study of local government web sites in the region of Finland Proper (Varsinais-Suomi in Finnish) in the southwestern part of Finland. From the perspective of local government, we believe that geography is particularly important and hence we focus on this issue.

## Literature review

*Geography* Some previous research has investigated geographic factors affecting linking. In a study of academic interlinking in the UK Thelwall (2002a) showed that universities close to each other tend to interlink more than distant ones. The exact coordinates of the locations of the universities were used to

calculate the distances between them for this research. Later, Tang and Thelwall (2003) found only limited evidence of geographic trends in interlinking within the web sites of three different U.S. academic disciplines.

International interlinking has also been occasionally studied, giving an alternative perspective on geographic factors. A study of the links between universities in UK, Australian and New Zealand gave an interesting finding: that New Zealand was rather isolated on the web (Smith & Thelwall, 2002). A larger-scale study of international interlinking covered the whole Asia-Pacific region through its universities and mapped the connections in network diagrams (Thelwall & Smith, 2002). This showed the central roles of both Australia and Japan, indicating that geographic proximity was less significant than geopolitical role in this context.

*Motivations* To be able to interpret link counts or network diagrams created based on links, it is essential to understand the underlying link creation motivations. Kim (2000) investigated authors' motivations for creating links in e-journals and found that some motivations were similar to those previously discovered for citing in printed scientific journals. Those that were new were related to accessibility of electronic sources. In a classification of academic hyperlinking, five types of creation motivations were identified: ownership, social, general, navigational and gratuitous (Thelwall, 2003). In a more extensive study Wilkinson, Harries, Thelwall and Price (2003) classified a random collection of 414 links between UK universities. Although 90% of the links were related to scholarly activities, only 1% of the links were equivalent to citations in scientific journals. This suggests that interlinking between academic web sites is evidence of informal rather than formal scientific communication. In a case study of interlinking between Israeli academic institutions Bar-Ilan (2004; 2005) proposed a multi-faceted classification of link types, in which she included many different aspects of the links and the source and target pages, i.e. source page types, target page types, creators, intention of linking and the relationship between the link area and the target. Bar-Ilan's categorizations give a very complex picture and cover many different aspects of academic link creation. Chu (2005) crystallized her findings about reasons for hyperlinking into one sentence: "One links to a site for what it is about", meaning that links are made to point to other sites that are relevant for the outlinking site. From the above results we can conclude that academic interlinking is based on informal communication with parties tending to be close to each other. This suggests that geographical distance is a component in informal scientific communication.

*Government links* E-Government, in terms of information technology use, has been studied from several aspects. Moon (2002) studied how e-government has evolved among municipalities and discovered that the size of the municipalities and cities (by population) and age of the web sites were positively associated with the adoption of e-government. Petricek, Escher, Cox and Margetts (2006) developed a methodology to quantitatively evaluate structural characteristics of e-government as indicators for web site navigability and 'nodality'. Their methodology gave meaningful evaluations of government web sites in a cheap and efficient way using freely available web link data. Government links alone have not been studied extensively before with webometric methods, but there is some research about university-government links and university-industry-government linking. Stuart and Thelwall (2005) investigated university-to-government links as indicators of collaboration between UK universities and government. Although a correlation between university's research productivity and number of outlinks was found, there was no evidence of a causative connection between university's research productivity and the number of university-to-government links. In a later study Stuart and Thelwall (2006) examined whether web URL citations could be used as 'weak benchmarking indicators' to investigate collaboration between universities, industry and government. As a part of this study directional graphs based on the URL citations showed clustering between the district councils (LAU 1<sup>1</sup>) and their respective county councils (NUTS-3<sup>1</sup>) and also between

---

<sup>1</sup> The Nomenclature of Territorial Units for Statistics (NUTS) is a uniform scale that makes it possible to compare regional statistics in European Union (Statistical Regions of Europe, 2006).

neighboring local government bodies. The study showed that county councils tend to link more often to other county councils and equally district councils tend to link more often to other district councils under the same county council. There were also geographic patterns in government to university URL citations. Although URL citations correlate statistically with some real world phenomena, a closer analysis of the URL citations indicated that these were more often created to point to information sources than to reflect cooperation or other relationships between the studied organizations, which was also the case in Stuart and Thelwall (2005).

### **Background**

The regional and local administration levels in Finland are divided into six provinces, 20 regions and 431 municipalities. The provinces are on level 2 on the *Nomenclature of Territorial Units for Statistics* (Nomenclature des Unités Territoriales Statistiques - NUTS) scale (Statistical Regions of Europe, 2006). The regions are on NUTS-3 level and municipalities are on NUTS-5 level. The more unofficial regions called functional regions are placed on NUTS-4 level. The NUTS-4 level is equal to the newer LAU 1 level (*Local Administrative Units*) and NUTS-5 equals LAU 2.

The region of Finland Proper is located in the southwestern part of Finland and is slightly smaller than Wales, Israel or New Jersey. About half of this area is land and the rest is sea. In January 2006 the population was 456,000 people. The city of Turku is the capital of the region and has over 170,000 residents. The region of Finland Proper has five functional regions (Figure 1).

- Loimaa is known for agricultural and provisions production.
- Salo is a leader in high technology industry and contains some of mobile phone manufacturer Nokia's main plants.
- Turunmaa (Åboland in Swedish) attracts tourism for its unique archipelago.
- Vakka-Suomi hosts Finland's car factory and main metal industry.
- Turku, the capital of Finland Proper, is an old university city with high quality research and manufacturing, e.g. in biotechnology.

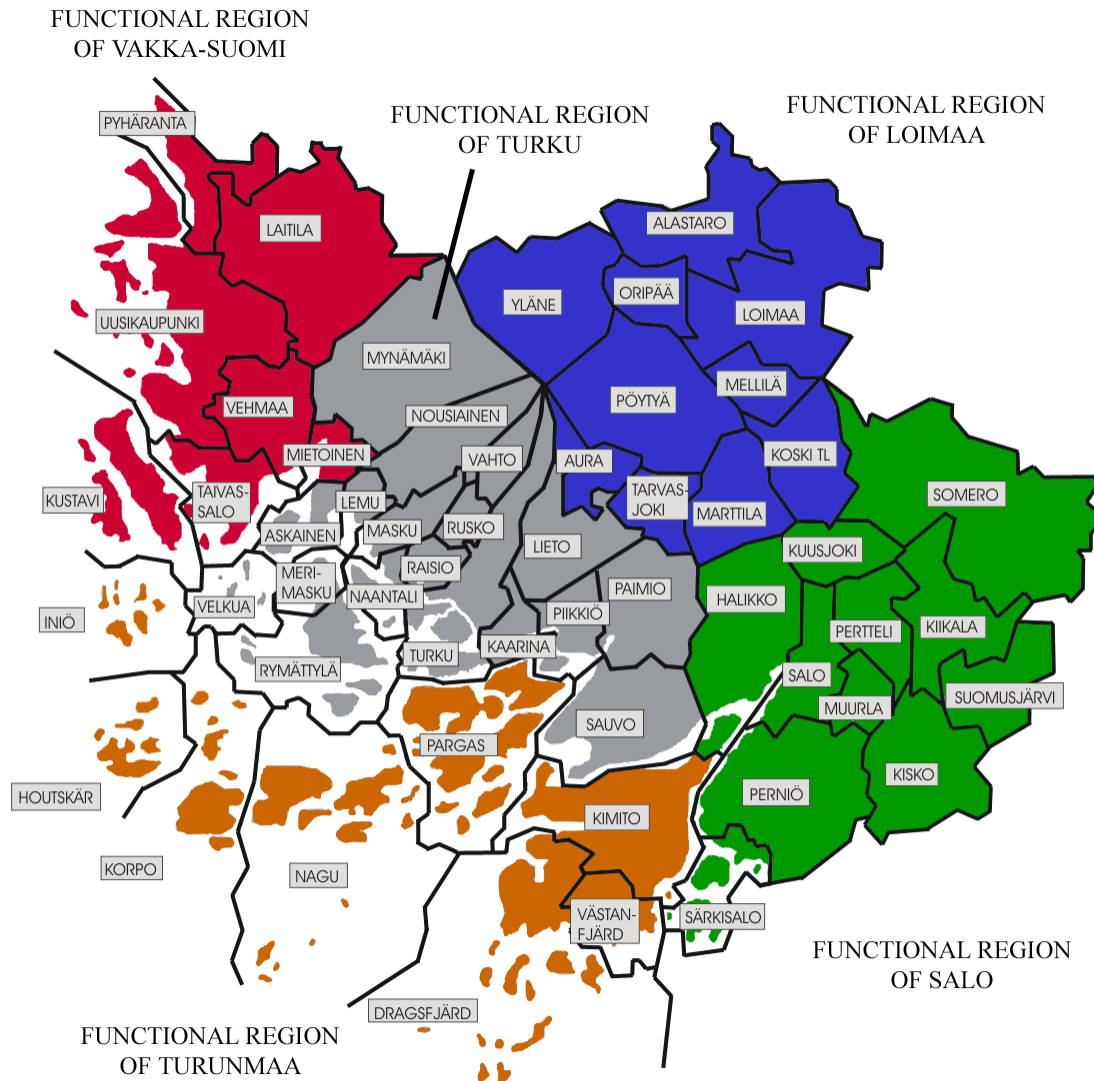


Figure 1. Functional regions in the region of Finland Proper (Varsinais-Suomi)

### Research Questions

From a broad perspective, the goal of this research is to contribute to understanding government-related web linking, both to help develop effective webometric techniques and to build understanding of the phenomenon on an international scale. The aim of this case study is to map possible geographic trends in interlinking between municipalities in Finland Proper and to study what motivates the interlinking between municipalities. Can interlinking between municipalities be used to trace some other existing trends or phenomena? The following research questions drive this investigation.

1. Does local government web site interlinking in Finland Proper follow geographic lines?
2. Why do local government web sites in Finland Proper interlink?

### Methods

*Data collection* There are three possible methods for collecting link data. The first is to visit every web site and manually collect the information needed about the content and links. This can be very strenuous and only suitable to study a very small set of web sites. It is not practical to collect the link data manually from the 54 web sites in this study. The second method is to use commercial search engines to count the occurrences of links and pages. It has been shown that no search engine covers

the whole web, and that coverage of major search engines may be as low as 16% (Lawrence & Giles, 1999). Even if later studies have found much higher coverage rates (Gulli and Signorini, 2005) it is still unclear exactly how well commercial search engines cover the web. It has also been shown that search engines have coverage biased towards the U.S. (Vaughan & Thelwall, 2004). This is a problem because if we used a search engine then we couldn't be sure if all the pages and links would be covered. The third and final method is to use a web crawler that will automatically and independently follow every link within a given web site and collect information about the links and the content on the pages. A crawler also has its disadvantages. The crawler may not be able to index as many types of non-html pages and dynamically created pages as a commercial search engine might be able to. We still used this approach because it gave us greater control of the results and hence was a better solution than using commercial search engines (Thelwall, Vaughan & Björneborn, 2005).

*Data modeling* A problem with using link data is that a single person can create several links in a directory or even on a single web page, which would skew the data collected about the links. To improve the quality of link data, the 'Alternative Document Model' (ADM) concept has been developed (Thelwall, 2002b; Thelwall & Wilkinson, 2003). When using ADMs, duplicate links are combined at different hierarchical levels of a web site's file structure. This minimizes the effect of links that are created repeatedly because of a web design decision. For instance, in this study the municipality of Lieto had a link on almost every page of the site to the web based map service hosted on the web site of Turku. Using the site level of ADM removes such duplicate occurrences of links, leaving only one for the whole site. Counting links at the site ADM level in the above example would give only one outlink from Lieto to Turku. In this study we are only interested in the existence or the lack of a link between sites and therefore use the site level ADM.

*Visualization* The data about interlinking links were collected with a special information science web crawler in July 2006 and the site ADM level of counting links was applied on the data with another software distributed freely in the same package as the crawler (Thelwall, 2001b, 2002c). The link data was converted to a squared binary 54x54 site-by-site matrix with BibExcel (Persson, 2006). In this matrix, 1 indicates the existence of a link and 0 indicates the lack of a link. The outlinks can be read from the rows and the inlinks can be read from the columns in the matrix, one row and column per municipality. The matrix was imported to Ucinet (Borgatti, Everett & Freeman, 2002) for analysis and visualized as a two-dimensional network map (Figure 2) with Pajek (Batagelj & Mrvar, 2003). In the map nodes represent municipal websites and the lines represent hyperlinks between them. Arrows indicate the direction of the hyperlinks. The locations of the nodes and the distances between the nodes were calculated with the commonly used Kamada-Kawai spring-based algorithm, (Kamada & Kawai, 1989), via Pajek (Figure 3). Kamada-Kawai positions nodes close to other nodes that they are linked to by pulling them closer as if the links were springs. A network map was chosen rather than a proximity-based mapping technique, such as Multi-Dimensional Scaling, because there are few enough links for the network diagram to be clear and hence the network diagram is easier to interpret because it directly reflects the raw data (i.e. 1s and 0s of the matrix).

*Geographic testing* To study possible geographic trends in interlinking between municipalities we created a simple binary matrix based on shared borders of the municipalities. Another possibility would have been to use distance between cities and municipalities. This approach would have been problematic because cities seldom lie in the geographic center of the municipality. Also the very large archipelago has definitively a big influence on the geographic distances between municipalities. To create a binary matrix based on shared borders gives a simple way to compare interlinking with geography. Strictly speaking our matrix is geopolitical rather than geographic. In the matrix a neighbor or a shared border is indicated with a 1 and the lack of a shared border is indicated with a 0. In a geographic matrix based on neighbors, the relationship, or the neighborhood, is always reciprocal. So if Nagu is a neighbor of Pargas, then Pargas has to be a neighbor of Nagu. The matrix is therefore symmetric, in contrast to the interlinking matrix, which does not have to be symmetric. We used QAP (described below) to test for similarity between the geographic and link matrices.

*Classification* To answer the second research question we took a random sample of links between the municipalities and visited both the source pages and the target pages to determine the underlying motivations to create links. We took a random sample of up to 20 interlinking outlinks from each of the 54 municipalities. This would have given a total of 1080 links, but because most of the municipalities did not have 20 outlinks to other municipalities, only 496 links were found. Some of the links were located on a single source page, giving 337 source pages to start from. We used the combined content and intention of the sources, the links and the targets as a basis for our categorization. As previous studies have shown, there is no uniform model for categorizing links, so we developed our own categorization designed for links between local government bodies and reflecting the duties and the responsibilities of municipalities. Once the categories were completed we grouped the categories based on a mutual intention or purpose of the content of the categories.

## Results

*Geography* When counting links with the site level ADM, there were 315 links between municipal web sites. When converted to a site-by-site matrix with BibExcel, there were 315 links or ties present in the matrix, which is 11.0 percent of all possible ties. This is also the density of the matrix. For a binary squared matrix this also measures the probability that any given tie between two random municipalities is present. Figure 2 reveals the central role that Turku plays in the area and that Turku is the “capital” of the region, even on the web. Turku has the most outlinks to (53) and receives the most inlinks from (47) other municipalities.

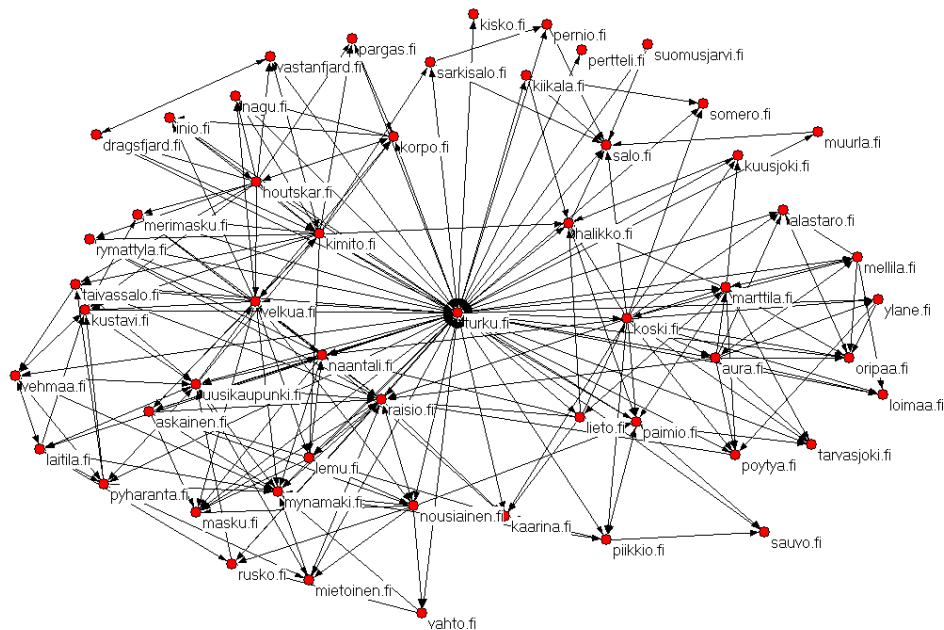


Figure 2. Interlinking municipalities in the region of Finland Proper (Varsinais-Suomi)

The municipalities are located quite clearly in clusters according to their functional regions, with the exception of the functional region of Turku and the functional region of Vakka-Suomi. The border between these two functional regions is like an invisible barrier, which is also the case in the real world. The municipalities in the functional regions of Turunmaa Loimaa both create quite strongly connected clusters. The municipalities in the functional region of Salo also create a clear cluster, but this cluster is not as well connected. It is almost only the links from and to Salo (which is the biggest city in the functional region) that keep the functional region connected to the rest of the municipal web space in the region of Finland Proper. Figure 3 below shows the interlinking between municipalities drawn with Pajek and using the Kamada-Kawai algorithm. The functional regions are drawn by hand, showing clearly that the regions do not overlap.

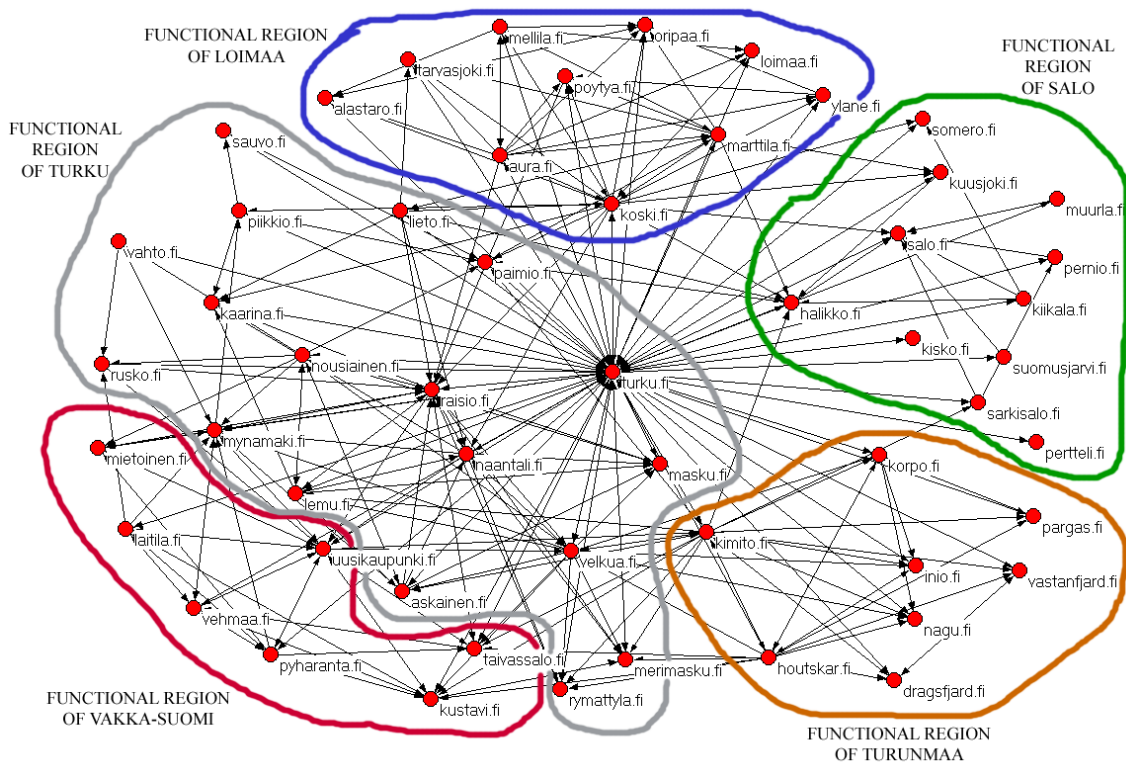


Figure 3. Interlinking and the functional regions in the region of Finland Proper (Varsinais-Suomi)

When comparing the matrix of neighboring municipalities with the matrix of interlinking between municipalities in the same region, we found that the matrices matched each other to 87.2% accuracy. In other words, there is a probability of 87.2% that for any value in any given cell in one of the matrices, the same value would be found in the corresponding cell from the second matrix. This match was tested with a dyadic matrix correlation test called the Quadratic Assignment Procedure or QAP (Krackhardt, 1992). QAP matrix correlation calculates the probability of getting the same match by accident, compared to random permutations of the rows and columns in the matrices. QAP indicated that a match of 80.4% between the matrices would occur if there was no underlying similarity. This high figure is due to both matrices having mainly zeros. The difference between 87.2% and 80.4% is big enough that the chance of getting a match of 87.2% or over between the matrices is approximately  $p=0.000$ . This is strong statistical evidence that geographic proximity plays an important role in creating links between the studied municipalities, answering the first research question.

*Link creation motivations* A total of 337 links between municipalities were visited and categorized, via their source and target pages, with the results shown in Table 1. In fact, both the source and target pages had the same topic in a majority of cases, something that has not been true for academic source and target pages (Bar-Ilan, 2004, 2005).

Table 1. Categorization of the motivations to link

Number of links	Type of motivation for linking	Percent of total	Official Cooperation	Geographic motivation
59	Map service, hosted by Turku municipality	17.51%		17.51%
43	Rescue services in the region, hosted on Turku municipality's website	12.76%	12.76%	12.76%
27	Link list, to closest municipalities (about 2-8 municipalities)	8.01%		8.01%
22	Libraries, agreement of shared resources	6.53%	6.53%	6.53%
22	Tourist information, attractions, link list to the closest municipalities	6.53%	6.53%	6.53%
21	Education, agreement about placing pupils	6.23%	6.23%	6.23%
18	Link list	5.34%		
17	Health care in the region	5.04%	5.04%	5.04%
12	Education, some cooperation and link list to close municipalities	3.56%	3.56%	3.56%
11	Consumer information, external information source	3.26%	3.26%	3.26%
11	Database of Turku Library	3.26%	3.26%	3.26%
9	Culture, links to closest municipalities	2.67%		2.67%
9	Business information services, area development, external information	2.67%	2.67%	2.67%
9	Nature and environmental protection	2.67%	2.67%	2.67%
9	Recreation and hobbies in close municipalities	2.67%		2.67%
8	Businesses, organizations and associations in close municipalities	2.37%		2.37%
8	Social service information	2.37%	2.37%	2.37%
6	Agriculture, cooperation, external information	1.78%	1.78%	1.78%
5	Waste management and recycling	1.48%	1.48%	1.48%
4	Sports in close municipalities	1.19%	1.19%	1.19%
3	Living, links to close municipalities	0.89%	0.89%	0.89%
2	Local newspapers	0.59%		0.59%
1	Traffic, local buses in Turku	0.30%	0.30%	0.30%
1	Unemployment, links to close municipalities	0.30%	0.30%	0.30%
<b>337</b>	<b>Total</b>	<b>100.00%</b>	<b>60.84%</b>	<b>94.67%</b>

Two trends were discovered: links were often motivated by official cooperation between municipalities and links were often motivated by the geographic closeness of the municipalities. These two groups overlapped, as can be seen from Table 1 above. The group of links motivated by official cooperation includes links to services that the municipalities had bought from other municipalities or that they had agreed to join together for. For instance, some municipalities do not have Swedish-speaking schools even if they have Swedish-speaking residents. These have an agreement with another adjacent municipality to buy a Swedish-speaking education for their residents. A total of almost 61% of the interlinking is motivated by similar official cooperation between the municipalities. Examples of links motivated by geographic proximity are local rescue services and tourist attractions. A total of almost 95% of the links are motivated for various reasons by the fact that the municipalities are close to each other. In answer to the second research question, we can conclude that the results strongly indicate that links are primarily created to reflect official cooperation and that geographic closeness is a factor in the vast majority of cases.

### Discussion and conclusions

The aim of this study was to investigate if interlinking between municipalities in the region of Finland Proper would correlate with the geography of the region and also why these links were created. We found that interlinking correlates quite strongly with geography. The most interesting finding, however, is that the interlinking tends to be motivated by official cooperation that geographic closeness has made possible, in contrast to Stuart and Thelwall's (2006) UK study. Although Stuart and Thelwall discovered linking to reflect some real world phenomena, motivations for creating links and URL citations did not support a cause-and-effect relationship. A possible reason for this difference may be the fact that Stuart and Thelwall studied larger local authorities (NUTS-3 and LAU 1) while in our study we concentrated on municipalities (LAU 2). Some of the municipalities in our study were very small (only 245 residents in one municipality) and hence more dependent on cooperation with other municipalities. This cooperation also seems to be reflected on the web. As a result, we can conclude that it is possible that interlinking can be used to map official cooperation and networks based on cooperation between municipalities and cities in Finland.

Our case study has produced unusually clear-cut results for a webometric analysis, both in terms of identified geographic factors and the link creation motivations. Although two weaknesses of the classification study are the use of a single classifier and the need to infer link creation motivations



rather than directly asking the link creators why the links were created, its geographic results are at least corroborated by the QAP test. Nevertheless, there is a general limitation: it is not clear to what extent the results can be generalized. Presumably, similar results would be found for any other area of Finland but perhaps there are few other countries that have as extensive a system of local government (Finland has the highest UN democracy index) and as extensive use of the web (Finland is one of the most online countries in the world). This hints that Finnish online local government may even be a best case for the world, rather than a typical example. If this were to be true, as future research will hopefully reveal, then the results would be useful as an extreme case of what is possible.

## References

- Almind, T. C. & Ingwersen, P. (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. Retrieved February 28, 2007 from <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p1.pdf>.
- Bar-Ilan, J. (2004). A microscopic link analysis of academic institutions within a country - the case of Israel. *Scientometrics*, 59(3), 391-403.
- Bar-Ilan, J. (2005). What do we know about links and linking? A framework for studying links in academic environments. *Information Processing & Management*, 41(4), 973-986.
- Batagelj, V. & Mrvar, A. (2003). *Pajek - Analysis and visualization of large networks*. In: *Graph Drawing Software*, M. M. Jünger, P., Editor. Berlin: Springer, 77-103.
- Björneborn, L., *Small-world link structures across an academic Web space: A library and information science approach*. PhD Thesis. Royal School of Library and Information Science, Copenhagen, Denmark, 2004. Retrieved February 28, 2007 from <http://vip.db.dk/lb/phd/phd-thesis.pdf>.
- Björneborn, L. & Ingwersen, P. (2004). Towards a basic framework for webometrics. *Journal of American Society for Information Science and Technology*, 55(14), 1216-1227.
- Borgatti, S. P., Everett, M. G. & Freeman, L. C. (2002). *Ucinet for Windows: Software for Social Network Analysis*. Harvard, MA: Analytic Technologies.
- Chu, H. (2005). Taxonomy of inlinked Web entities: what does it imply for Webometric research? *Library & Information Science Research*, 27(1), 8-27.
- Gulli, A. & Signorini, A. (2005). *The indexable web is more than 11.5 billion pages*. *Proceedings of the 14<sup>th</sup> international World Wide Web conference (WWW2005)*. [CD Version] Retrieved February 28, 2007 from [http://www.di.unipi.it/~gulli/papers/f692\\_gulli\\_signorini.pdf](http://www.di.unipi.it/~gulli/papers/f692_gulli_signorini.pdf).
- Ingwersen, P. (1998). The calculation of Web impact factors. *Journal of Documentation*, 54(2), 236-243.
- Kamada, T. & Kawai, S. (1989). An Algorithm for Drawing General undirected Graphs. *Information Processing Letters*, 31(1), 7-15.
- Kim, H. J. (2000). Motivations for hyperlinking in scholarly electronic articles: a quantitative study. *Journal of American Society for Information Science*, 51(10), 887-899.
- Krackhardt, D. (1992). A caveat on the use of the Quadratic Assignment Procedure. *Journal of Quantitative Anthropology*, 3, 279-296.
- Lawrence, S. & Giles, C. L. (1999). Accessibility of information on the web. *Nature*, 400, 107-109.
- Moon, M. Jae (2002). The Evolution of E-Government among Municipalities: Rhetoric or Reality? *Public Administration Review*, 62(4), 424-433.
- Persson, O. (2006) Bibexcel - a toolbox for bibliometricians (Version 2006-05-11). [Computer software]. Retrieved May 11, 2006 from <http://www.umu.se/inforsk/Bibexcel/>.
- Petricek, V., Escher, T., Cox, I. J. & Margetts, H. (2006). *The Web structure of e-government - developing a methodology for quantitative evaluation*. *Proceedings of the 15th international conference of World Wide Web*. Edinburgh, Scotland: ACM Press. Retrieved February 28, 2007 from <http://www2006.org/programme/files/pdf/1041.pdf>.
- Rousseau, R. (1997). Sitations, an exploratory study. *Cybermetrics*, 1(1), paper 1.
- Smith, A. & Thelwall, M. (2002). Web impact factors for Australasian universities. *Scientometrics*, 54(3), 363-380.
- Snyder, H. & Rosenbaum, H. (1999). Can search engines be used for web-link analysis? A critical review. *Journal of Documentation*, 55(5), 577-592.
- Statistical Regions of Europe, 2006. *Nomenclature of territorial units of statistics*. Retrieved November 22, 2006 from [http://ec.europa.eu/comm/eurostat/ramon/nuts/home\\_regions\\_en.html](http://ec.europa.eu/comm/eurostat/ramon/nuts/home_regions_en.html)

- Stuart, D. & Thelwall, M. (2005). *What can university-to-government web links reveal about university-government collaborations? Proceedings of the 10th International Conference of the International Society of Scientometrics and Informetrics*. Stockholm, Sweden: Karolinska University press. Vol. 1, 188-192.
- Stuart, D. & Thelwall, M. (2006). Investigating triple helix relationships using URL citations: a case study of the UK West Midlands automobile industry. *Research Evaluation*, 15(2), 97-106.
- Tang, R. & Thelwall, M. (2003). U.S. academic departmental Web-site interlinking in the United States: Disciplinary differences. *Library & Information Science Research*, 25, 437-458.
- Thelwall, M. (2001a). The responsiveness of search engine indexes. *Cybermetrics*, 5(1), paper 1.
- Thelwall, M. (2001b). A web crawler design for data mining. *Journal of Information Science*, 27(5), 347-359.
- Thelwall, M. (2002a). Evidence for the existence of geographic trends in university Web site interlinking. *Journal of Documentation*, 58(5), 563-574.
- Thelwall, M. (2002b). Conceptualizing documentation on the Web: an evaluation of different heuristic-based models for counting links between university web sites. *Journal of American Society for Information Science and Technology*, 53(12), 995-1005.
- Thelwall, M. (2002c). Methodologies for crawler based web surveys. *Internet Research: Electronic networking and applications*, 12(2), 124-138.
- Thelwall, M. & Smith, A. (2002). A study of the interlinking between Asia-Pacific University Web sites. *Scientometrics*, 55(3), 335-348.
- Thelwall, M. (2003). What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation. *Information Research*, 8(3). Retrieved February 28, 2007 from <http://informationr.net/ir/8-3/paper151.html>.
- Thelwall, M. & Wilkinson, D. (2003). Three target document range metrics for University Web sites. *Journal of American Society for Information Science and Technology*, 54(6), 490-497.
- Vaughan, L. & Thelwall, M. (2004). Search engine coverage bias: evidence and possible causes. *Information Processing & Management*, 40(4), 693-707.
- Thelwall, M., Vaughan, L. & Björneborn, L. (2005). *Webometrics*. In: *Annual Review of Information Science and Technology*, B. Cronin, Editor. Medford, NJ: Information Today Inc. Vol. 39, 81-135.
- Vaughan, L. (2004). Exploring website features for business information. *Scientometrics*, 61(3), 467-477.
- Vaughan, L. & Wu, G. (2004). Links to commercial web sites as a source of business information. *Scientometrics*, 60(3), 487-496.
- Vaughan, L. & You, J. (2005). *Mapping business competitive positions using web co-link analysis. Proceedings of the 10th International Conference of the International Society of Scientometrics and Informetrics*. Stockholm, Sweden: Karolinska University press. Vol. 2, 534-543.
- 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 scholarly communication. *Journal of Information Science*, 29(1), 49-56.