Co-inlinking to a Municipal Web space: A Webometric and Content Analysis

Kim Holmberg

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

Abstract

It is known that there are significant correlations between linking and geographical patterns. Although interlinking patterns have been studied in various contexts, co-inlinking patterns on the Web have only been studied as indicator of business competitive positions. This research studies the use of co-inlinks to local government Web sites, assesses whether co-inlinking follows geographic patterns and investigates reasons for creating the co-inlinks. Strong evidence was found that co-inlinking is more frequent to municipalities in the same functional region than to municipalities in different functional regions, indicating that this geographic aspect influences co-inlinking, even though geographic co-inlinking was not a strong trend overall. Because the functional regions are created based on cooperation between the municipalities, we have indirectly been able to map cooperation from co-inlinking patterns on the Web. The main reason to create co-inlinking links to municipalities was that the source of the links wanted to show a connection to its region.

Keywords

Webometrics; hyperlinks; co-inlinking; colinks; link creation, local government, geography

Introduction

Hyperlinks on the Web can be created for probably as many reasons as there are Web sites. It is reasonable to assume that for instance hyperlinks to a municipal Web site in Finland are not created for the same reasons as hyperlinks to the British model Jordan's Web site. There is no control or quality assurance for what can be published and what kind of hyperlinks can be created on the Web. Yet there seems to be some underlying information in the hyperlinks that can be used in webometric studies to map trends and relationships.

In webometric studies, hyperlinks are assumed to indicate some offline relationships between the organizations represented by the Web sites. Link counts have been shown to correlate with measures describing business performance (Vaughan, 2004) and with research ratings (Smith & Thelwall, 2002; Li, Thelwall, Musgrove & Wilkinson, 2003). Different types of linking have also been studied. Interlinking has been found to correlate with geographic patterns (Thelwall, 2002; Holmberg & Thelwall, 2007) and co-inlinking has been found to be a possible tool for mapping business competitive patterns (Vaughan & You, 2005; 2008). Links have been shown to have sometimes an offline creation motivation and hence links are a valuable source of information for webometric research.

This research studies whether co-inlinking to Web sites follows geographic patterns, using a case study of Finland Proper (a region in the South-Western part of Finland). It investigates how neighboring municipalities and municipalities in the same functional regions are co-inlinked together, and why the co-inlinks have been created. The results increase knowledge about co-inlinking and about how and why local administration Web sites are linked to.

Literature review

An overview of linking terminology for webometric purposes has been given by Björneborn and Ingwersen (2004). In their proposed terminology, two sites receiving links from a third

Proceedings of ISSI 2009, pages 254–263.
Edited by B. Larsen and J. Leta.

site are said to be colinked or to have co-inlinks. In similar manner, two sites linking to a third are said to be colinking or to have co-outlinks. We propose an extension of this terminology: the analysis of sites that are colinked is *co-inlink analysis* and the analysis of sites that are colinking is *co-outlink analysis*.

Patterns in linking practices Some earlier studies have investigated whether linking follows geographic patterns. An investigation of international interlinking between Asia-Pacific University Web sites showed the central role of Australia and Japan in the region (Thelwall & Smith, 2002). Within a single country, Thelwall (2002) found evidence of geographic trends in interlinking between universities in the UK. He found that the level of interlinking decreases with distance. Tang and Thelwall (2003) explored disciplinary differences in academic Web site interlinking. They found large disciplinary differences in the use of the Web. They also discovered strong correlations between inlink counts and research impact, but little evidence of geographic trends in interlinking. Holmberg and Thelwall (2007) discovered strong evidence of the influence of geographic adjacency in interlinking between local government Web sites in Finland Proper and that the links were created because of official reasons. Interlinking between local government Web sites was found to give a geopolitical view of the relationships between the municipalities. In summary, several earlier studies have found strong evidence for geographic patterns to have an impact on (direct) links.

In term of colinks, in 1996 Larson ran a cocitation analysis of earth-science-related Web sites. This was actually a co-inlink analysis, but at the time Larson called it cocitation analysis. Multidimensional scaling techniques were used and pages of similar topics were found to group together. Thelwall and Wilkinson (2004) studied whether colinks (both co-inlinks and co-outlinks) could be used to find academic Web sites of similar topic. Only a small improvement was discovered when using a combination of different types of links in comparison to using interlinking alone. Using colinks was found to improve the recall rather than precision of information retrieval. Vaughan and You (2005, 2008) used co-inlink data to collect link counts to pairs of business Web sites and mapped the relationships using multidimensional scaling techniques. Clear clustering of close competitors could be viewed from the maps, suggesting that co-inlinking could be used to map offline relationships such as business competitive positions.

Motivations for hyperlink creation Link counts and maps do not reveal much about the motivations for linking because links could indicate positive, supporting actions or negative opinions. Only by determining the reasons why the links have been created can conclusions be drawn about patterns discovered from links and link networks. Because of the multitude of content on the web and the many possible reasons for hyperlinking, there is no uniform classification scheme that could be used for all webometric research. So far, researchers have created their own classification schemes that have best suited the goals of their research.

In the earliest study, Kim (2000) used author interviews and found 19 different motivations for link creation in scholarly electronic articles. These were grouped into scholarly, social and technological classes. Park (2002) surveyed Korean webmasters for reasons for creating hyperlinks and found that credibility and usefulness were important factors. Wilkinson et al. (2003) found that links between academic Web sites were created for scholarly reasons and could be used as evidence of informal scholarly communication. Bar-Ilan (2005) developed a classification scheme that incorporated several different aspects of both source and target pages and the context in which the links were created. Chu (2005) studied how and why links to academic Web sites were created and categorized link creation motivations into four

groups: teaching/learning, research, service and homepage. Vaughan et al. (2007) classified co-inlinks to university Web sites in Canada according to contents of the page and the context in which the links were created. They found that in 94% of the cases the universities connected by co-inlinks were also related academically, giving even more evidence about the possibilities of using co-inlinks to analyze existing offline relationships.

In a commercial context, Vaughan et al. (2006) studied motivations for creating links to business Web sites and found that most of the links were created for business purposes. The three researchers all classified links according to the content of the source page, the context in which the link had been created and the country of origin. Competing businesses were discovered to be often co-inlinked, suggesting the value of co-inlink analysis to map competitive positions of the businesses.

Background

In 2007 the region of Finland Proper had 54 municipalities. The municipalities are on NUTS-5 level on the *Nomenclature of Territorial Units for Statistics* (Nomenclature des Unités Territoriales Statistiques - NUTS) scale (Statistical Regions of Europe, 2006). Turku is the largest municipality in the region and has a population of about 170,000 people, which is about 37% of the whole population in the region. Some of the municipalities in the region are very small: 45 have less than 10,000 residents and 22 have less than 2,000 residents. The two smallest municipalities in the region, Iniö and Velkua, both have about 250 residents. Finland is at the moment in the middle of a municipal reform where municipalities are merging into new, larger municipalities.

The region of Finland Proper has five functional regions that the municipalities belong to. The functional regions are on NUTS-4 level and the whole region of Finland Proper is on NUTS-3 level. Existing cooperation between the municipalities and commuting for work have been used as criteria when forming the functional regions. Because of this, municipalities in the same functional region have to be close to each other. The functional regions in the region of Finland Proper are: Loimaa (with 10 municipalities), Salo (with 11 municipalities), Turunmaa (8 municipalities), Turku (18 municipalities) and Vakka-Suomi (7 municipalities).

Research questions

This research has two goals: to study co-inlinking patterns to local government Web sites in the region of Finland Proper and to investigate whether co-inlinking follows geographic patterns; and to study motivations for creating co-inlinking links. The first goal is investigated from two different angles. First co-inlinking to local government Web sites of neighboring municipalities is assessed to see whether it is more frequent than to municipalities further apart. Second, the frequency of municipalities in the same functional region being co-inlinked is compared to co-inlinking with municipalities from different functional regions. The second part of this study is a content analysis of link creation motivations with a suitable classification for co-inlinking to local government Web sites. The goals of this research are in the following three research questions:

1. Does co-inlinking to local government Web sites follow geographic patterns?
2. Are municipal Web sites in the same functional regions co-inlinked together more frequently than with municipalities in other functional regions?
3. What are the motivations to create co-inlinks to municipal Web sites?

Methods

Data collection and analysis In 2007, when the link data was collected, there were 54 municipalities in the region of Finland Proper. A total of 1,431 queries (=((54*54)-54)/2) were submitted to MSN/Live Search to collect the link data. Advanced features, which were still available at that time, made it possible to retrieve all pages that had links to two Web sites and that were not in the researched Web sites. This excluded all navigational links inside the researched Web sites. These 1,431 queries retrieved a total of almost 78,000 pages as indexed by MSN that had outlinks to the Web sites of at least two of the municipalities in Finland Proper. The average number of co-inlinks that a pair of municipalities received was 54, the maximum number of co-inlinks was 1,005 and the minimum was 7 co-inlinks.

Instead of using multidimensional scaling techniques and exact co-inlink counts to the municipalities, the data was converted into a binary data matrix. To convert the data into a binary matrix a level of co-inlinks was chosen that could be considered as significant enough to indicate a strong connection between the co-inlinked municipalities and that would at the same time exclude a certain number of noise or unwanted pages (such as link lists to all municipalities in Finland) from the analysis. The binary matrix and the graph drawn from the matrix would then show the strongest co-inlinking to the municipalities, while the weakest connections would be disconnected nodes in the graph. The average number of co-inlinks was chosen as a level of significance and used in the conversion. All pairs of municipalities that had 54 co-inlinks or more were marked with a 1 in the matrix and pairs of municipalities that had less than 54 co-inlinks were marked with a 0. The co-inlinking was visualized on top of a geographic map, so that links between the municipalities and the neighborhood of the municipalities are simultaneously visible (Figure 1). A matrix of the neighboring relationships between the municipalities was done by indicating neighborhood, or shared borders, with 1 in the matrix and non-neighbors with 0. This gave two binary squared matrices of the same size, one for co-inlinking and one for neighboring relationships. QAP (Krackhardt, 1992) was used to test similarity between the matrices and to calculate the probability of getting as good match by accident. To study how well co-inlinking matched neighborhood the two matrices were combined. Only the overlapping 1's were included in the new matrix, clearly showing the co-inlinking that matches with the shared borders of the municipalities (Figure 2).

In order to answer the second research question the co-inlinking matrix was permuted so that the municipalities were grouped according to the functional regions. Co-inlinking to the municipalities inside the functional regions could then be studied separately from the co-inlinking to municipalities in different functional regions. The proportion of all possible links that were present in the groups was used to create a block density matrix (Table 1) that can be used in an additional way to analyze co-inlinking (Hanneman, 2005). Diagonal values representing self-links were ignored in the calculations.

Classification of link creation motivations A random sample of 184 source pages was used in the classification. The link creation motivations were classified according to the content on the source page, content on target page and link type. The content of the source pages was classified according to the topic and purpose of the pages. If the target page was the homepage of the municipality then this was used as a category. Other target pages within the municipal Web sites were classified according to topic. The link type could be a link list of various kinds and lengths or the link could be in text. The relationship of the two co-inlinked municipalities was also studied from the source pages. This criterion is important when studying co-inlinking, because it is from this relationship on the source page that the two links form a relationship between the two target Web sites and municipalities they belong to.

Results

Co-inlinking to the region of Finland Proper In the binary co-inlinking matrix there are 670 ties or links present, which is 23.4% of all possible links. The density of the matrix is therefore 23.4%. Turku is the most connected municipality with its 36 links. Other municipalities have between 25 and 0 links. This co-inlinking matrix is visualized in Figure 1 below. No visible patterns or clusters can be seen in this network. Some of the links connect municipalities close to each other while other links connect municipalities from completely opposite sides of the region. The density of the matrix based on neighborhood, or shared borders, is 11% and there are 316 links present. QAP shows a simple match of 80.5% between the matrices and that the probability of getting the same match or better by accident is p=0.000, giving strong evidence that co-inlinking follows geographic patterns.

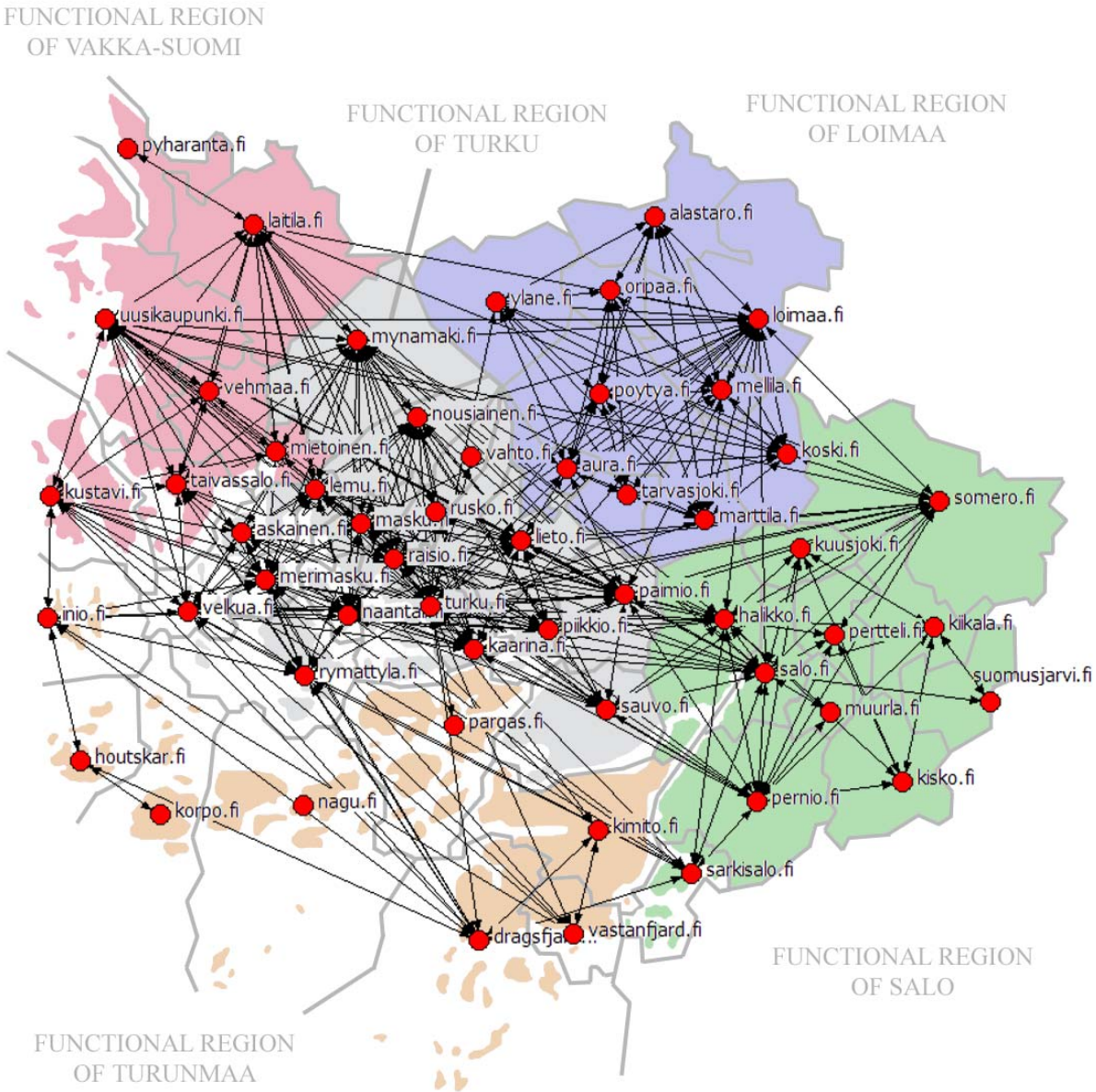

Figure 1. Co-inlinking in the region of Finland Proper

The two matrices were then combined to get a matrix containing the co-inlinking that matched neighborhood in the region. There are 214 links present in the matrix created from combining neighboring links and co-inlinks. The density of this new matrix is 7.5%. In other words, of the 316 municipalities with shared borders between neighbors, 214 were also combined by co-inlinking. A total of 67.7% of the possible connections matching neighborhood in the region were actually present. A total of 456 co-inlinks did not match with neighborhood. Only about 32% of the co-inlinks match therefore with geography. The combined matrix is visualized in Figure 2 below. Although co-inlinking matches quite well with the neighboring relationships in the region, more than two thirds of the co-inlinking links did not match. Using information retrieval terminology we could say that the recall was quite good, but the precision was not.

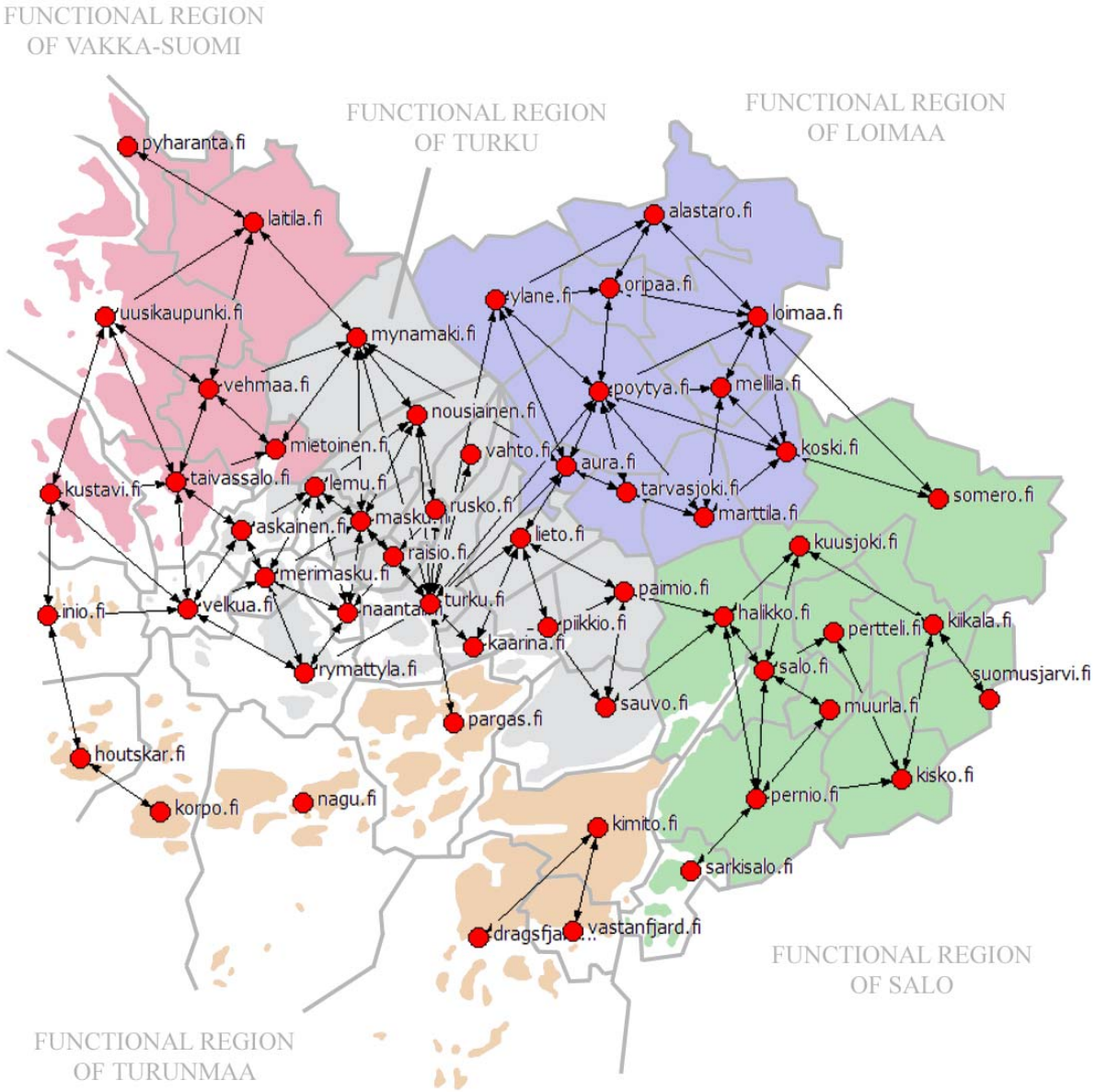

Figure 2. Co-inlinking that matches neighborhood

With this we can answer the first research question by stating that there are some indications for co-inlinking to follow geographic patterns, but that a high match could be due to the fact that there are so many connections present in the co-inlinking matrix.

Co-inlinking to the functional regions in the region of Finland Proper From grouping municipalities belonging to a functional region together, the percentage of co-inlinked sites found is presented in the block density matrix in Table 1 below. A total of 422 co-inlinking links were present between municipalities belonging to same functional regions and only 248 co-inlinking links create a connection between municipalities in different functional regions. A total of 63% of the co-inlinking links are creating a connection between municipalities in the same functional region, giving strong evidence for the role of the functional regions in co-inlinking to local government Web sites.

In the functional region of Loimaa 97.8% of the possible connections were present (Table 1). Only the connection between Alastaro and Tarvasjoki was missing. A total of 76.2% of the connections in the functional region of Vakka-Suomi were present. In Turku region 73.9% of the possible connections were present. In Salo region over half of the possible connections were present. In the functional region of Turunmaa less than one third of the possible connections were present and in the case of Turunmaa, the municipalities in the region were not that well connected with other municipalities in other functional regions either. Only 10.7% of possible connections between the municipalities in Turunmaa were connected to municipalities in Vakka-Suomi, which was the region that Turunmaa had most connections with. The municipalities in Vakka-Suomi were much better connected to municipalities in Turku region: 28.6% of possible connections were present. The municipalities in Turku region were also quite well connected with municipalities in Salo region: 18.2% of possible ties were present. The results clearly show that the municipalities in the region of Finland Proper are best co-inlinked together with other municipalities in the same functional regions, answering the second research question. Because municipalities in functional regions are close to each other, this also gives more evidence for the first research question.

Table 1. The percentage of co-inlinking out of all possible co-inlinking within and between the functional regions

	Loimaa	Salo	Turku	Turunmaa	Vakka-Suomi
Loimaa	**97.8%**	3.6%	9.4%	0.0%	4.3%
Salo	3.6%	**54.5%**	18.2%	1.1%	10.4%
Turku	9.4%	18.2%	**73.9%**	9.0%	28.6%
Turunmaa	0.0%	1.1%	9.0%	**28.6%**	10.7%
Vakka-Suomi	4.3%	10.4%	28.6%	10.7%	**76.2%**

Co-inlinking motivations A clear majority of the links came from pages with link lists of various lengths and various topics. These lists included colinks that combined neighboring municipalities but also municipalities that were further apart. When classifying the link types it was found that a total of almost 87.4% (159 pages) of the total amount of pages with co-inlinks had the co-inlinks in link lists. Two of the lists (1.1%) were about events, three (1.6%) were about information sources and seven lists (3.8%) were about contact information. 68 of these lists (40%) included four or less municipalities. 18 municipalities (9.9%) were linked to in text. Six of these pages (3.3%) were biographic pages where the municipalities that the writer had lived in were mentioned.

The relationship of the colinks on the source page was also investigated. On almost 60% (69 pages) of the source pages the links to the municipalities were included because the municipalities were in the same geographic region or because they were close to each other. In some cases the geographic relationship was even clearer and well defined. In 6 cases (5.2%) it was obvious that the municipalities had been linked to because they were in the region of Finland Proper. In five cases (4.3%) the municipalities were in Finland (link lists to every municipality in Finland) and in four cases (3.4%) they were on the same island. In 13 cases (11.2%) the municipalities were members of the same association or campaign. In 19 cases (16.4%) it was not possible to determine any relationship between the municipalities linked to.

Most of the target pages were homepages of the municipalities. A total of 122 target pages (68.9%) had links to municipal homepages and 55 (31%) were content pages of various topics inside the municipal web sites. Ten links (5.4%) targeted library sites or library pages within the municipal Web site. Eight links (4.3%) were to schools or information about schools in the municipality.

As already mentioned most of the retrieved source pages containing the colinks contained different types of lists of links. The source pages, although including link lists of various kinds and lengths, were of various topics: tourist information (21 pages, 11.4%), list of members in various associations (17 pages, 9.2%), links to schools and various educational resources (14 pages, 7.6%), business information and resources (14 pages, 7.6%),), sports and sports associations (12 pages, 6.5%), information about museums (9 pages, 4.9%), members of library networks (9 pages, 4.9%) and information about health care (8 pages, 4.3%). 22 pages (12%) contained lists of links with no common factor. Ten pages (5.4%) contained a list of municipalities in the region or close by.

Discussion and conclusions

One goal of this research was to study whether co-inlinking to local government Web sites follows geographic patterns. Co-inlinking to municipal Web sites was shown to follow geographic patterns, and co-inlinking was strongest within the (geographically-organised) functional regions, suggesting that the main trend was for geo-political linking. The results suggest that neighborhood has an impact on linking, but only in about one third of the cases. This research gave clear evidence that the majority of co-inlinking to municipalities was to municipalities in the same functional region and that the municipalities in the functional regions are very well connected to each other. Only about 27% of the co-inlinks were between municipalities in different functional regions. The only functional region that was not as well connected as the others was the functional region of Turunmaa. This lack of connections may be because of the fact that cooperation and commuting between the municipalities in this region is difficult because of the long distances caused by the archipelago. This also supports our conclusion that because municipalities in functional regions are joined by cooperation and closeness, geography in fact is an important factor in co-inlinking.

We also found some more evidence of the impact of geography on co-inlinking when the motivations to create links were classified. 60% of the links were created because the target municipalities belonged to a certain geographic region or were close by. The amount of link lists in the source pages was a surprising discovery, as was the amount of homepages in the target pages. It is not usual to target a specific content page in the municipal Web sites as most links go directly to the homepage. It is likely that the main reason to link to local

government Web sites is to recognize a connection with the geographic region and to indicate that the source has a connection with that region.

Although the present study gave some clear evidence about geographic patterns in co-inlinking, it has some limitations. Some of the earlier studies that have classified linking motivations have used more than one classifier. Using only one classifier means that the interpretations of the motivations are subjective to the classifier's own opinions. Nevertheless, in this research the majority of links did not leave much choice for interpretation as they were in link lists. The second limitations and suggestion for future research concerns the conversion of the link data to binary form. Using distances between the cities/municipalities or the distances when driving from one city to another and exact link counts would be a logical continuation of this research and it would increase knowledge about the effects of geography and distance on co-inlinking.

The results suggest that co-inlinking can be used to map some offline relationships and that they do reflect some connections between the target organizations. In fact, the present research have indirectly been able to map cooperation between the municipalities, as we have showed that co-inlinking is strongest within the functional regions, which are formed to reflect existing cooperation.

References

Bar-Ilan, J. (2005). What do we know about links and linking? A framework for studying links in academic environments. *Information Processing & Management*, vol. 41, no. 4, pp. 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, pp. 77-103.

Björneborn, L. & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, vol. 55, no. 14, pp. 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*, vol. 27, no. 1, pp. 8-27.

Hanneman, Robert A. and Mark Riddle. 2005. *Introduction to social network methods*. Riverside, CA: University of California, Riverside. Available at http://faculty.ucr.edu/~hanneman/, retrieved on January 18, 2009.

Holmberg, K. & Thelwall, M. (to appear, 2009). Local government web sites in Finland: A geographic and webometric analysis, *Scientometrics*.

Kamada, T. & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, vol. 31, no. 1, pp. 7-15.

Kim, H. J. (2000). Motivations for hyperlinking in scholarly electronic articles: a quantitative study. *Journal of American Society for Information Science*, vol. 51, no. 10, pp. 887-899.

Krackhardt, D. (1992). A caveat on the use of the Quadratic Assignment Procedure. Journal of Quantitative Anthropology, 3, 279-296.

Larson, R. (1996). Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of cyberspace. In *Proceedings of the 59th Annual Meeting of the American Society for Information Science*, pp. 71-78.

Li, X., Thelwall, M., Musgrove, P. & Wilkinson, D. (2003). The relationship between the WIFs or inlinks of Computer Science Departments in UK and their RAE ratings or research productivities in 2001. *Scientometrics*, vol. 57, no. 2, pp. 239-255.

Park, H.W. (2003). Examining the determinants of who is hyperlinked to whom: a survey of webmasters in Korea. *First Monday*, vol. 7, no. 11.

Smith, A. & Thelwall, M. (2002). Web impact factors for Australasian Universities. *Scientometrics*, vol. 54, no. 1-2, pp. 363-380.

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

Tang, R. & Thelwall, M. (2003). U.S. academic departmental Web-site interlinking in the United States Disciplinary differences. *Library & Information Science Research*, vol. 25, pp. 437-458.

Thelwall, M. (2002). Evidence for the existence of geographic trends in university Web site interlinking. *Journal of Documentation*, vol. 58, pp. 563-574.

Thelwall, M. & Smith, A. (2002). Interlinking between Asia-Pacific University Web sites. *Scientometrics*, vol. 55, no. 3, pp. 363-376.

Thelwall, M. & Wilkinson, D. (2004). Finding similar academic Web sites with links, bibliometric couplings and colinks. *Information Processing & Management*, vol. 40, no. 3, pp. 515-526.

Vaughan, L. (2004). Exploring website features for business information. *Scientometrics*, vol. 61, no. 3, pp. 467-477.

Vaughan, L. & You, J. (2005), Mapping business competitive positions using Web co-link analysis. In: Ingwersen, P. & Larsen B. (Eds), *Proceedings of ISSI 2005 – the 10th International Conference of the International Society for Scientometrics and Informetrics,* pp. 534–543, Stockholm, Sweden, July 24–28, 2005.

Vaughan, L. & You, J. (2008). Content assisted web co-link analysis for competitive intelligence. *Scientometrics*, vol. 77, no. 3, pp. 433-444.

Vaughan, L., Gao, Y. & Kipp, M. (2006). Why are hyperlinks to business Websites created? A content analysis. *Scientometrics*, vol. 67, no. 2, pp. 291-300.

Vaughan, L., Kipp, M. & Gao, Y. (2007). Why are Websites co-linked? The case of Canadian universities. *Scientometrics*, vol. 72, no. 1, pp. 81-92.

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*, vol. 29, no. 1, pp. 49-56.