Patterns of Web Linking to Heterogeneous Groups of Companies:

The Case of Stock Exchange Indexes

Abstract

Purpose – To extend co-link analysis to Websites of heterogeneous companies belonging to different industries and countries. To cluster companies by industries and compare results from different countries.

Design/methodology/approach – Websites of 255 companies that belong to five stock exchange indexes were included in the study. Data on co-links pointing to these Websites were gathered using Yahoo!. Co-link data were analyzed using multidimensional scaling (MDS) to generate MDS maps that would position companies based on their co-link counts.

Findings –Comparisons of results across different countries and economies showed the following overall pattern: companies whose businesses are information based tend for form well defined clusters while companies operating on a more traditional business model tend not to form clear groups. A comparison between EU zone and the U.S. suggests that the EU economy is not well integrated yet.

Practical implications – Findings from the study suggest the possibility of using co-link analysis to distinguish the information based industries from traditional industries.

Originality/value –Extended co-link analysis from a single industry to heterogeneous industries with global and complex business phenomena.

Keywords Competitive intelligence, Web data mining, co-link analysis, Webometrics

Paper type Research paper

1. Introduction

Thirty years ago Michael Porter (1980, 1985) developed the idea that the creation and maintenance of competitive advantages are key factors of business success. Despite of the evolution of business theories over the last three decades, competitive strategy remains as a fundamental element in the management of a company. Competitive Intelligence (CI) is the result of the general acceptance of this important managerial thinking. Today CI plays a key role in a company's strategic effort of achieving and sustaining its competitive advantages (Heppes and du Toit, 2009). According to Kahaner (1996), CI consists of a systematic plan to obtain and analyze information about competitors and general trends in the industry. The abundance of information on developed economies, especially with the development of digital technologies and the Internet, has created new opportunities and challenges for companies so that they need to monitor changes around them in order to compete in better conditions. In the last decade, Web hyperlink analysis has been used for CI purposes with promising results (Reid, 2003; Tan *et al*., 2002).

Inlink and co-inlink are basic Webometric concepts to understand the nature of Web hyperlinks. According to Björneborn and Ingwersen (2004), an inlink, also called back link, is a link pointing to a Webpage, e.g. page X has an inlink coming from page Z. If page X and page Y both have inlinks from page Z, then page X and page Y are co-inlinked. Co-inlink analysis is referred simply as co-link analysis later in this paper.

Previous research has established the relationship between the number of links pointing to a business website and the company's financial performance, e.g. revenue and profit (Vaughan and Wu, 2004). Moreover, direct links pointing to websites have proved to be a relevant measure of similarity (Thelwall and Wilkinson, 2004). However, business competitors link each other very rarely in order to avoid diverting Web traffic to rival companies (Shaw, 2001; Vaughan, Gao and Kipp, 2006). This fact leads to the use of co-links rather than inlinks to study competitive positions. As Vaughan (2006) points out, co-link data are more robust than inlink data as the former are less easily manipulated. Content analysis has been carried out to study the motivations for co-link creation (Vaughan, Kipp and Gao, 2007). Co-link analysis has also been demonstrated to be a useful tool to reveal the cognitive or intellectual structure of a particular field of study (Zuccala, 2006), analogously to co-citation analysis in bibliometrics (Small, 1973).

Web co-link analysis for business information started with a single industry (Vaughan and You, 2006) or on a detailed picture of the competitive landscape of a sector within an industry (Vaughan and You, 2008; Vaughan and You, 2009). The methodology developed in these papers has also been tested and verified in other countries and other industries, e.g. China's chemical industry and electronics industry (Vaughan, Tang and Du, 2009). Romero-Frías and Vaughan (2009) extended the use of co-link analysis into the banking industry in the US in order to test the feasibility of combining page content with co-link data to monitor financial crisis. Parallel to these studies on commerical Websites, research on heterogenous Websites has been carried out to test the triple helix theory on the Web (Stuart and Thelwall, 2006; García-Santiago and de Moya-Anegón, 2009).

This paper reports a study that extended co-link analysis to Websites of heterogeneous companies belonging to five stock exchange indexes. These companies are the biggest in their respective economies and therefore have a significant presence on the Web and are likely to receive attention from users, companies, government and other economic agents. When studying a single industry, co-link analysis allowed us to visualize business competition within the industry. When we applied co-link analysis to companies belonging to different industries in this study, we may see not only competition within an industry but also alliances or other relationships between different industries. The scope of this study is much broader than those carried out before, which gives us an opportunity to explore the issues that were not examined in previous co-link studies. We compared results from the five different stock exchange indexes that represent different economic, geographical and cultural backgrounds. We analyzed the Eurostoxx 50 from two different perspectives of country and industry. This allowed us to observe the degree of economic integration in the Euro-market. We also tried to observe if there is any clear pattern, common to all the stock exchanges, about the type of business activities that attract more Web links and other activities that attract fewer Web hyperlinks due to their particular business features. We observed that there is one main variable that could determine a generalised pattern between economic activities,

this is, the degree in which the activity is information centered. This applies mainly to IT, media, and financial companies, among others. These industries are in an ongoing process of business model transformation, e.g. the deep crisis in the media sector due to digitalisation of information. The paper is exploratory in nature with an intent to open new lines of research on the application of Webometric methodology to business studies.

2. Methodology

2.1. Selection of companies to study

All companies in the following stock exchange indexes are candidates of the study: Dow Jones Industrial (30 US companies), the Dow Jones Euro Stoxx 50 (Euro zone companies), CAC 40 (Paris), FTSE 100 (London), and Ibex 35 (Madrid). The complete list of all these 255 companies is omitted due to a space consideration; however, a selection of companies is included in Appendixes 1 to 5 to facilitate the understanding of the findings (Figures 1 to 7). Table 1 provides an overview of the indexes in the study: country, number of companies, composition retrieval date, URL and the appendix in which the companies are listed. Readers can refer to the URLs in the Table for the companies in each of the indexes, however changes in the composition are expected over the time. In order to facilitate the identification of companies that are omitted in the appendixes, the company labels in the MDS maps are the tickers used in the stock exchanges for each business.

Table 1. Information about the stock exchange indexes in the study

Index	Country / Region	Number of companies	Composition retrieval date	URL	Appendix
Dow Jones Industrial	U.S.A.	30	17/02/2009	http://www.djaverages.com/	1
FTSE 100	U.K.	100	24/02/2009	http://uk.finance.yahoo.com/q/cp?s=^FTSE	2
Eurostoxx 50	12 Eurozone countries	50	17/02/2009	http://www.stoxx.com/indices/components.html?symbol=SX5E	3
CAC 40	France	40	19/02/2009	http://www.euronext.com/trader/indicescomposition/composition-4411-EN-FR0003500008.html?selectedMep=1	4
IBEX 35	Spain	35	19/02/2009	http://www.bolsamadrid.es/ing/mercados/acciones/accind1_1.htm	5

The Website address of each of these companies was collected from the stock exchange Website and then manually checked to ensure its correctness. The vast majority of companies in the study have only one URL for their Websites. For the few that have alternative URLs in the form of alias or redirect, we checked each URL to find out which one has more inlinks and used that one for collecting inlink data. We considered including both URLs in data collection. However, Yahoo!, the search engine used for data collection, cannot handle the complex query syntax for collecting co-link data using two URLs.

Industry classification in each stock exchange differs. For example, the Industry Classification Benchmark (ICB) is used globally (though not universally) to divide the market into increasingly specific categories. Currently it is used by Dow Jones, FTSE and several other markets around the globe. However, other markets, e.g. the Spanish one represented by IBEX 35, use its own industrial classification. This heterogeneity, together with the scarce number of companies for some industries in the indexes, obliges to merge companies into general groups such as IT, Media or Financial. We understand that this classification used for practical purposes in the study is subject to discussion. Therefore, brief notes about the native industrial classification for each company is provided in Appendixes 1 to 5.

2.2. Collection of Web link data

Of the three major search engines, Google, Yahoo! and MSN Live Search, only Yahoo! could be used for data collection for the study. Google's inlink search only returns a sample of all inlinks that the Google database records (Google, 2009). Another problem is that Google cannot filter out internal inlinks (inlinks originated from within the Website itself such as "back to home" type of links) as the query term 'link" cannot be combined with any other query terms (Google, 2006). In other words, it cannot report the external inlink counts that the study needed. MSN Live Search used to have inlink search functions but the service was turned off around March 2007 (Live Search, 2007). At the time of data collection, winter 2009, Yahoo! is the only option for collecting inlink data as required by the study.

Because search engines of different countries may have databases that favour Websites of the host countries (Vaughan and Thelwall, 2004), we considered using the Spanish version of Yahoo! to retrieve inlinks to the companies in the Spanish stock exchange and the French version of Yahoo! for French companies. However, tests of these versions of search engine showed that they returned the same inlink search results as that from the global version of Yahoo!. Unlike the Chinese version of Yahoo! which has a database that is different from the global version of Yahoo!, the Spanish and French version of Yahoo! just had a different interface but the same underlying database. So the global version of Yahoo! (www.yahoo.com) was used for all data collection.

Yahoo! has two inlink search query terms, link and linkdomain. The "link" query term finds links to a particular page (e.g. link:http://www.abc.com finds links to the homepage of www.abc.com) while the linkdomain query term retrieves all links that point to all pages of a particular Website or domain including the homepage. We used the linkdomain query term for data collection because all links, not just links to homepage are of relevance to the study. The query syntax for the data collection is illustrated in Table 2 using the example URLs of www.abc.com and www.xyz.com. We truncated the www portion of the URLs in the queries to capture all links to all subdomains such as mail.abc.com. The "-

site:abc.com" part of the query is to filter out internal links coming from within the domain of abc.com itself.

Table 2. Illustration of Yahoo! Queries for Data Collection

Types of links searched for	Query
Inlinks to www.abc.com	linkdomain:abc.com –site:abc.com
Co-links between www.abc.com and www.xyz.com	(linkdomain:abc.com –site:abac.com) (linkdomain:xyz.com –site:xyz.com)

Since co-links involve a pair of Websites, the co-link data are collected in the form of a matrix with row x and column y of the matrix representing the number of co-links between URL x and URL y. We collected co-link data by stock exchanges, i.e. companies in the same stock exchange are in the same co-link matrix. Thus there are five co-link matrices for the five stock exchanges.

2.3. Methods of data analysis

Each co-link matrix was analyzed using multidimensional scaling (MDS) to generate a MDS map. MDS uses a heuristic method to place companies with higher co-link counts closer in the resulting MDS map. Since similar or related companies are more likely to receive co-links (two unrelated companies such as a food company and an IT company has a very small chance of being co-linked), the number of co-links between a pair of companies is a measure of their relatedness. Therefore, similar or related companies will be placed closer in the MDS map. The positions of the companies in the map only reveal their relative relationship to each other, e.g. if we modify the list of companies in the analysis, the positions would also change. Therefore, it is important to highlight that there is no absolute meaning in the position of a business in the map, but only if we take into consideration its position in relation to other companies. We hoped to use MDS the map to see clusters of companies and to reveal their market positions. We also hoped to examine of the effectiveness of our methods by comparing the MDS maps from the five stock exchanges.

The raw co-link count collected from Yahoo! needed to be normalized to obtain a relative measure of the relatedness of companies. This is necessary because a co-link count of ten is large if the two companies in question each had very few inlinks, e.g. 15. On the other hand, if each company received thousands of ininks, then the 10 co-links that they shared represented a very small portion. The normalization was done by applying the Jaccard index in the following way.

$$\text{Normalized co-link count} = n_{A \cap B} / n_{A \cup B}$$

Where A is the set of the Web pages that linked to Website X
      B is the set of the Web pages that linked to Website Y

$n_{A \cap B}$ is the number of pages that linked to both Website X and Website Y, i.e. the raw co-link count

$n_{A \cup B}$ is the number of pages that linked to either Website X or Website Y.

The normalized co-link matrices were feed into SPSS for MDS analysis. The MDS map for the FTSE 100 was too crowded to see clusters of companies clearly for a meaningful interpretation. So we reduced the number of companies from 100 to 80 and then to 35. We ranked the companies by the number of inlinks and selected the top 80 and 35 companies for co-link MDS analysis. Both the top 80 and top 35 maps are included (Figure 3 and 4) as they both are useful to understand the overall patterns discovered. One company in Euro 50, Generali, had no co-links with most other companies so it was omitted from the co-link analysis. The stress values of MDS analysis are 0.03 for Dow Jones, 0.07 for Euro 50 (49 companies), 0.06 for CAC 40, and 0.05 for Ibex 35, 0.05 for FTSE top 80 and 0.06 for top 35. These stress values are all fairly low, which suggests a good fit between the data and the MDS map positions.

3. Results

The five stock exchange indexes are representative of different economies and differ in industry composition and in social, cultural and political backgrounds. In this section we present a sketch of each stock and then compare them to obtain general findings and patterns. Limitations and future research prospects are discussed in the next section.

As pointed out above, co-link count is a similarity measure which has been used in previous papers to evaluate competitive positions of companies within a single industry. Extending those earlier studies, the current study analyzed heterogeneous companies in terms of activity and, therefore, our first expectation is to find clusters of similar companies, that is, companies belonging to the same industry are expected to be grouped.

3.1. Dow Jones Industrial

Dow Jones Industrial is made of the top 30 companies in the USA. The composition at the time of the study and the industry affiliation of companies are in Appendix 1. Figure 1 is the MDS map for this stock. Only significant clusters are drawn to make the interpretation clearer.

Three major groups of companies are discernable:

1. IT group includes the following companies: fixed line telecommunications (Verizon, AT&T), computer services and hardware (IBM, HP), semiconductors (Intel) and software (Microsoft). Positions of companies within this cluster are also meaningful. The three hardware companies are closer to each other and the same is true for the two telecommunications companies. Microsoft is the only software company here and it occupies a distinctive and domination market position, so it is distant from other companies.

2. Financial group is made of banks and other financial services. There are three banks in the index Bank of America (BAC), Citigroup (C), JP Morgan (JMP) that are located within a triangle in the upper part of the map. Clustering is not perfect as two other companies appear in the same space.

A fourth company to take into consideration is American Express that is placed down in the map closer to IT group. A common feature of financial companies in the study is that they seem to be located in a similar position as shown in Fig. 5 and 7 later.

3. Other three clusters are well identified (pharmaceutical, oil & gas, transport). Scattered around these three clusters are other companies belonging to various other industries. A common feature of these companies and the companies in the three clusters is that they belong to traditional industries whose activities are no so focused on information.

The positions of two other companies are worth noting. As the only media conglomerate in Dow Jones, Disney does not compete directly with any other companies in the map. The nature of its activity, with most of its products and services likely to be digitalised, implies a strong presence on the Web which is also used as one of its main distribution channels. It received 6,070,000 external inlinks, second only to Microsoft (46,600,000). The fact that the company received millions of inlinks but had a very small percentage of co-links with other companies, together with the lack of competition with other companies in the index, explains Disney's outskirt position very well. Although there is no other media company in the map to form a cluster with Disney, it is circled in Fig. 1 for comparisons with media companies in other maps later. Like Disney, Wal-Mart is positioned as an outlier and attracted 1,210,000 external inlinks. It is worth to note how a retailer company as Wal-Mart received so many inlinks. This could be explain by the evolution in the business model that turn to be more focused on Internet selling and distribution of products. It is the only company other than the IT and Media ones that received more than 1,000,000 inlinks.

Regarding inlink counts, it is worth to note that companies on the right side of the map are the ones with more inlinks. All companies (except Verizon) received more than 1,000,000 inlinks. Companies with fewer inlinks (less than 150,000 inlinks) are located at the left side of the map. Companies in the middle of the map have mostly between 200,000 and 300,000 inlinks. It is perhaps not surprising that IT companies and consumer oriented companies (Disney and Wal-Mart) are more visible on the Web and therefore they attract more inlinks. A study of Chinese Web inlink patterns also found that Websites of consumer oriented companies received more inlinks (Vaughan, Tang & Du, 2009).

Figure 1. MDS map for Dow Jones Industrial
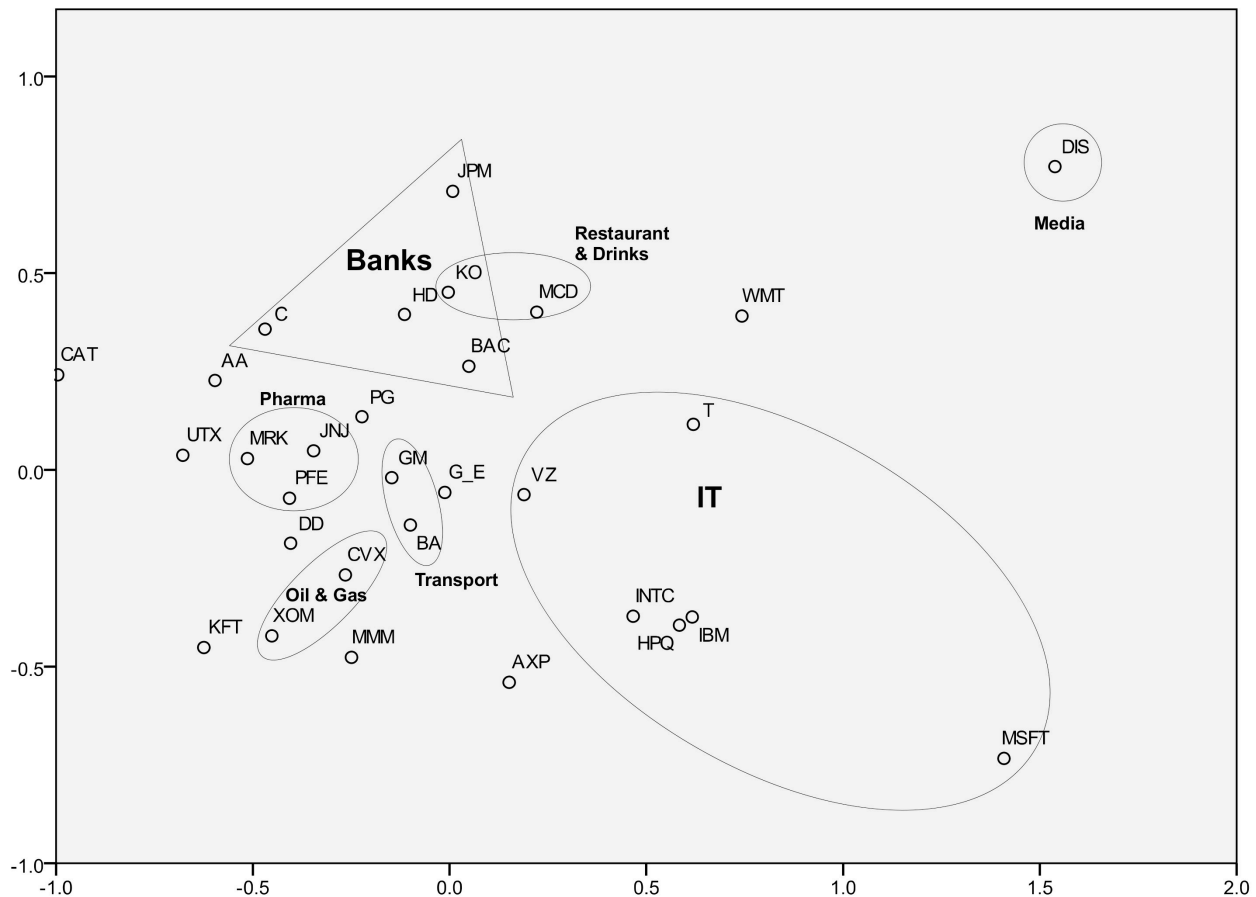
3.2. FTSE 100

FTSE 100 companies represent more than 80% of the market capitalization of the whole London Stock Exchange and constitutes the most widely used UK stock market index. Figure 2 shows the MDS map for the top 80 companies of the index in terms of inlinks. Although the high concentration of points in the center of the map makes interpretation unfeasible, this picture is especially useful as it shows clearly how some well defined groups are placed at the outskirts of the map. These clusters are IT, Media and Leisure companies. Because these IT companies are not as homogenous as the IT companies in the Dow Jones, they are not grouped into a single space; however, they are all located in a peripheral position of the map, as observed for Dow Jones. In this case, three out of the four most important media companies in the index are located in a common area. Only Pearson remains very close but out of this group; however its affiliation is much clearer in Fig. 3. Leisure group of companies is made of user oriented companies such as: Carnival, Thomas Cook and Tui Travel (travel agents), and Intercontinental Hotels (hotels). Travel business is one of the activities that have implemented e-commerce intensively. Although not showed in the map, financial companies are relatively well clustered in an intermediate position in the right part of the cloud of points.

Figure 2. MDS map for FTSE 100 (Top 80 companies in terms of inlinks)
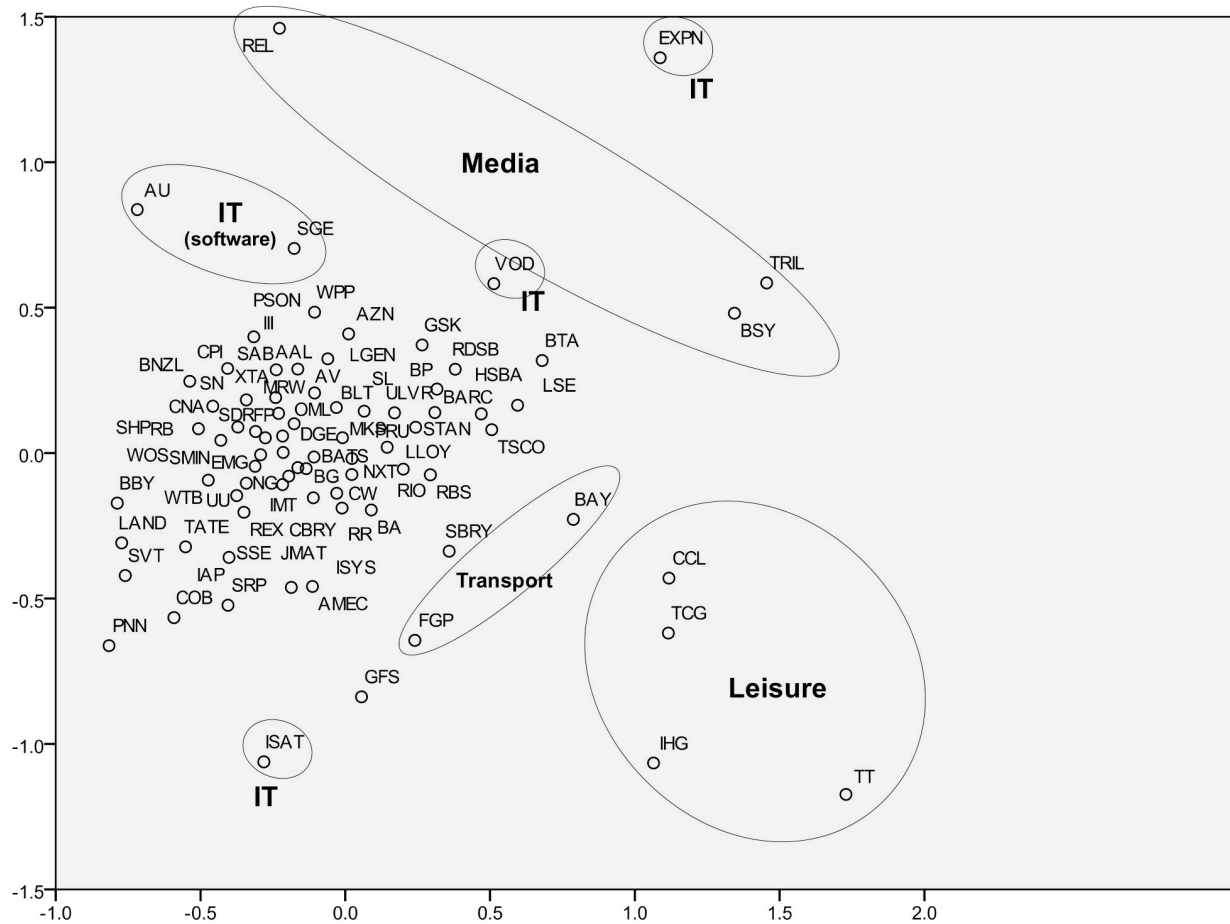

Figure 3 is the result for the data set of the top 35 FTSE companies in terms of inlink count. This selection probably overweighs some industries that by nature are likely to receive more inlinks (such as Media or IT). This approach provides an opportunity to test if more meaningful clusters will emerge from this microscopy approach. The resulting MDS map is a diverse representation of companies showing consistent clustering patterns that we have also found in other indexes. The results suggest that IT, Media and Leisure companies are placed at the outskirts of the map, with banks again placed adjacent to them. Clusters here are not clearly distinguished, probably due to the heterogeneity of activities and the high number of companies included in the IT and Media groups. For instance, media companies include such a distinct businesses as TV, books, press or a marketing services company (not strictly a media but strongly connected).

Figure 3. MDS map for FTSE 100 (Top 35 companies in terms of inlinks)
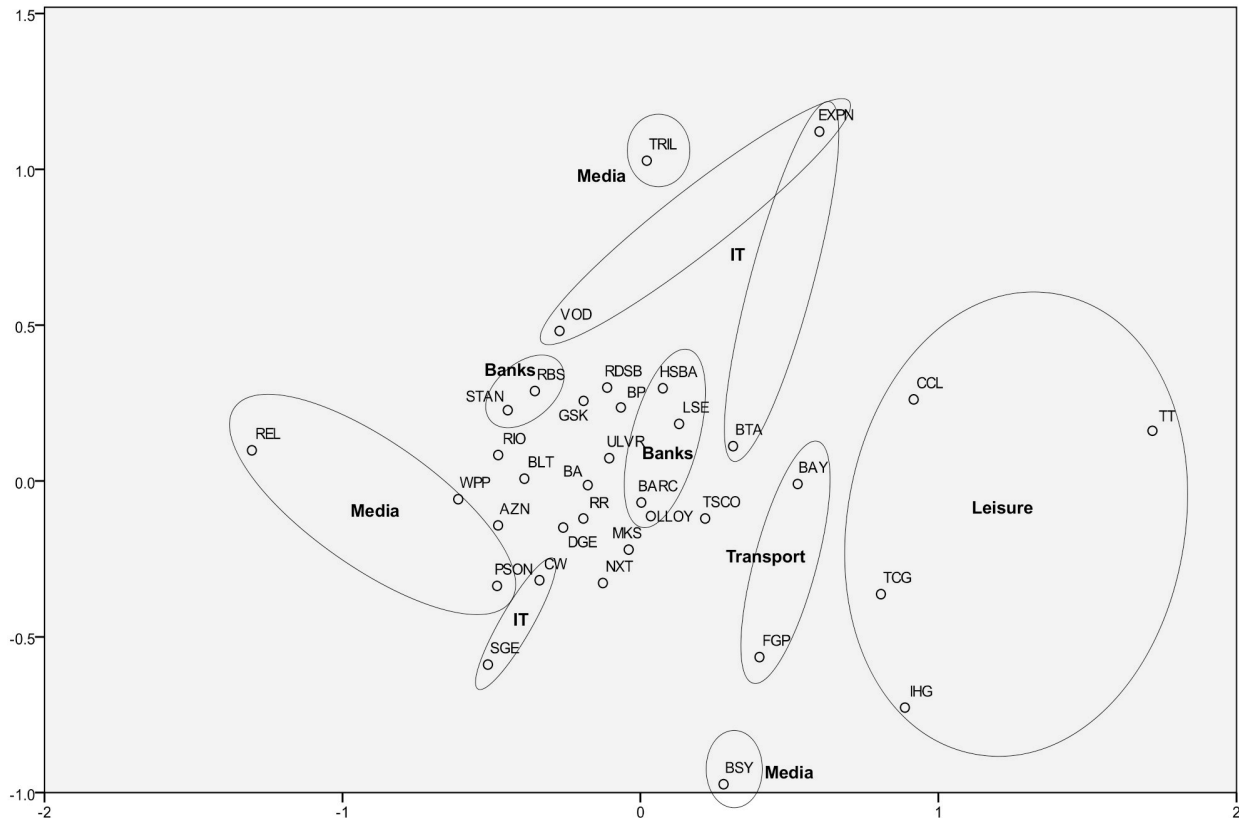
3.3. Eurostoxx 50

Dow Jones Eurostoxx 50 is a stock index that intends to provide a blue-chip representation of Supersector leaders in the Eurozone. It is composed of 50 companies from seven countries: France (19), Germany (13), Netherlands (5), Italy (6), Spain (5), Finland (1) and Luxemburg (1).

This index is heterogeneous in terms of countries and industries. Therefore two different interpretations based on these criteria are developed. Firstly, Figure 4 shows the MDS map for Eurostoxx 50 according to the country location of the companies. Companies belonging to the same country appear to be together. This is especially clear for French companies. In the German space we can find a number of companies belonging to other countries (dashed line), but we will see later that this could be explain in some cases by the strength of industrial pattern. Most Spanish companies are grouped very close. Dutch companies occupy a central position in the map between German and French companies. This position reflects the real geographical and economic situation in the Eurozone, where Dutch companies, due to the small size of its domestic market, tend to be more international. The only Luxembourgish company in this map, ArcelorMittal (ISPA), is embedded into the French space. This could be explained by the fact that Arcelor is the result of a merge in 2002 of three companies: Aceralia (Spain), Usinor (France) and Arbed (Luxembourg). This merger combined with the geographical proximity justifies its position.

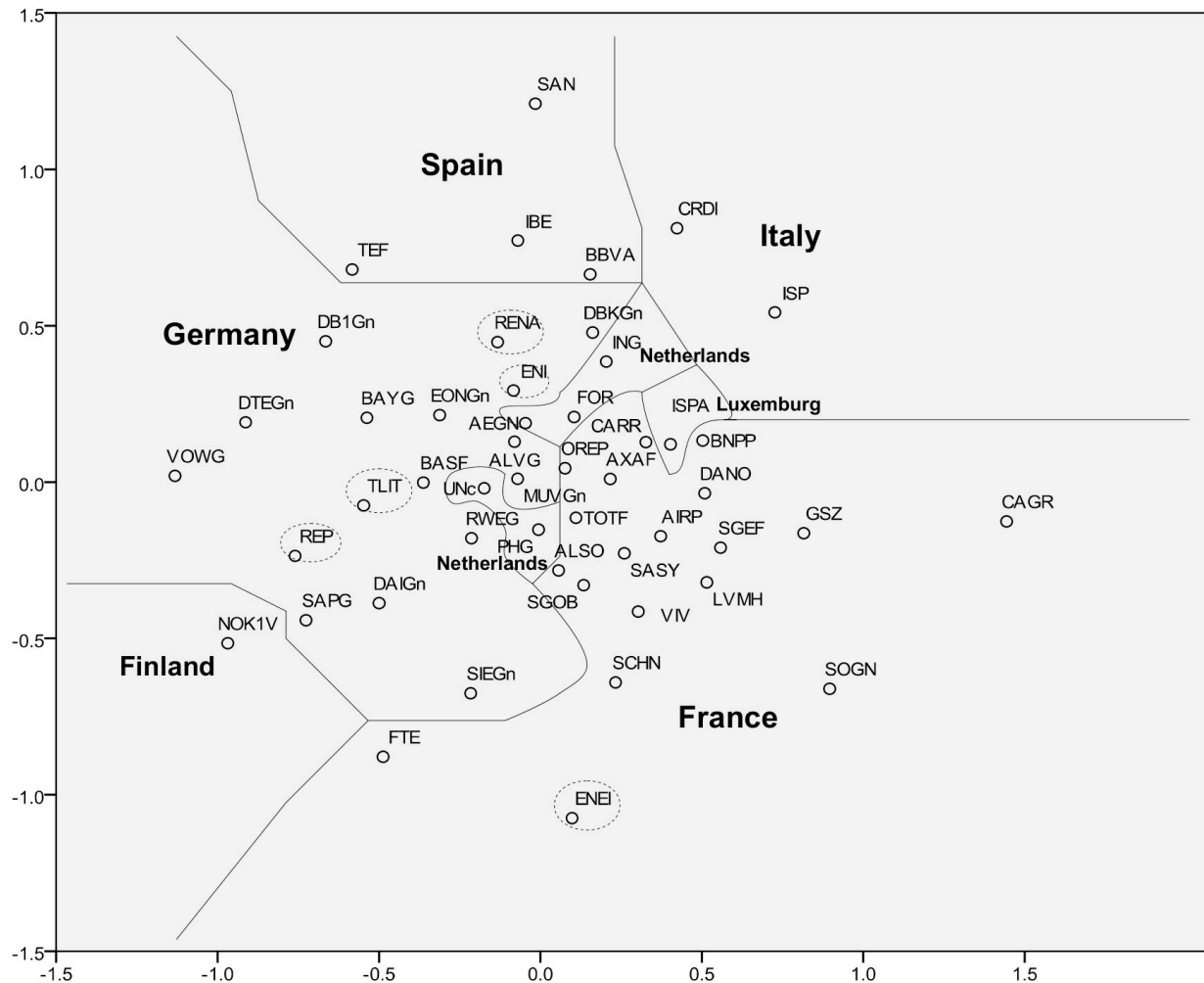Figure 4. MDS map for Eurostoxx 50, interpreted by countries.

Figure 5 contains the same MDS map interpreted according to industries. Only two main groups are clearly identified: IT and Financial groups. No media or leisure companies are included in this index. Financial group is mainly made of banks and insurance companies. All of them, with the exception of Deutsche Boerse (DB1Gn) that is close to IT companies, are together in a vast area regardless of country origin. There are also three companies (dashed line) from other industries that are inside of this cluster. IT companies are placed in the left part of the map including six companies from five countries (two of them are German). The existence of this cluster of IT companies explains why the German area of the map (Figure 4) presents some companies from other countries. IT has consistently immerged as a clearly defined cluster of companies in all stock exchanges in the study. This could be explained by the nature of its activity that is based intensively on information content and infrastructure. The financial companies in this map also form a clear cluster, probably due to its information based activity. The rest of the companies do not show a clear pattern in terms of industries, consistent with the results in other maps.

Market integration in the Eurozone is likely to be deeper in these information sectors that are not restricted by geographical location for the provision of services and goods.

Figure 5. MDS map for Eurostoxx 50, interpreted by industries
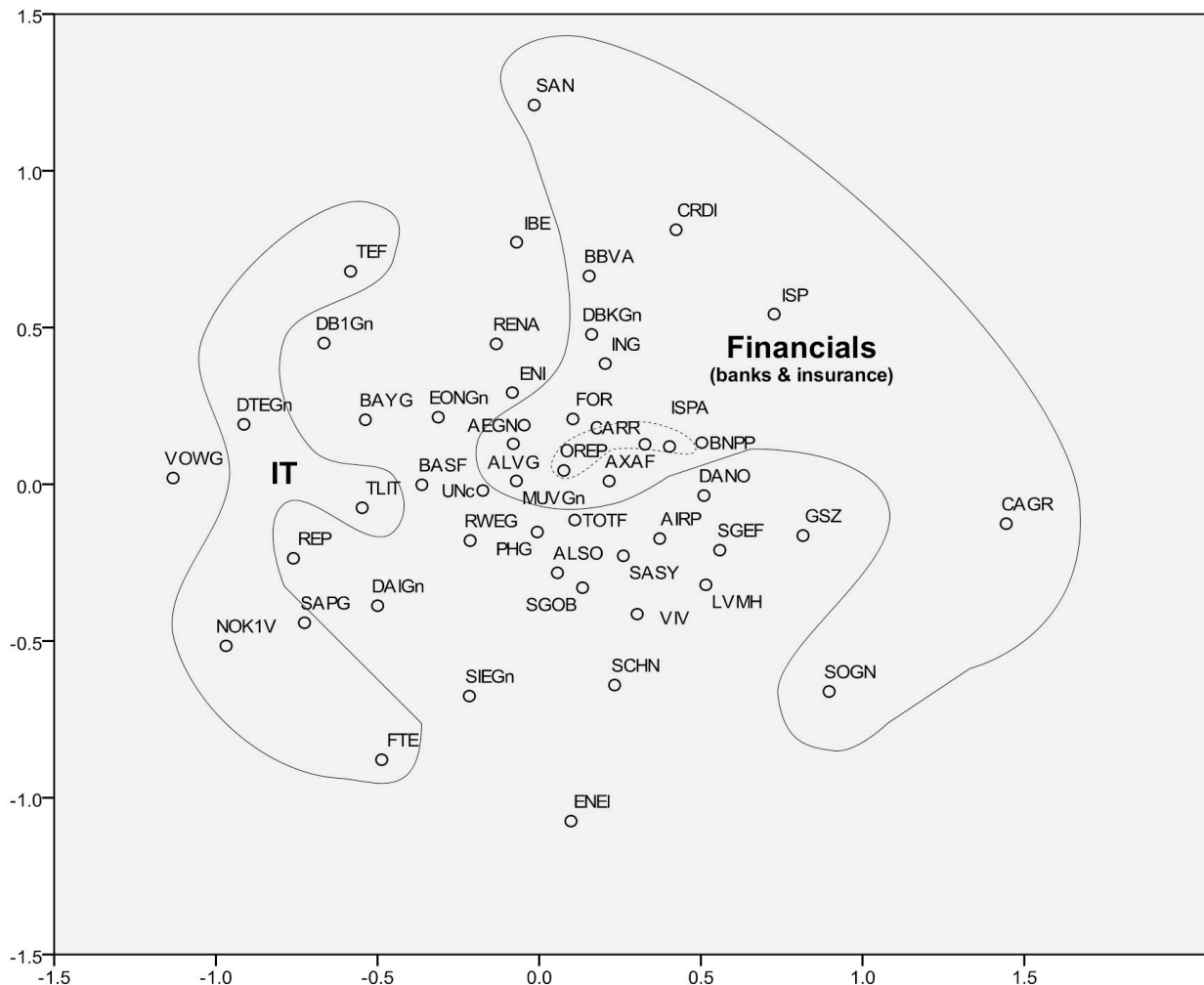

3.4. CAC 40

The CAC 40 is a benchmark French stock market index that represents a capitalization-weighted measure of the 40 most significant values among the highest market caps on the Euronext Paris Bourse. Figure 6 is the MDS map for this data set. It shows that IT companies are in a cluster that is far removed from other companies. The position and the fact that these companies are top inlinked are coherent with findings in other indexes. Accor (AC), a hotel company (Leisure industry) receives a high number of inlinks and is placed very close to IT companies. Banks are more spread in the map but two of them are located near the IT companies: BNP Paribas (BNP) and Société Générale (GLE). Another bank, Crédit Agricole (ACA) is at the border of the map. However, a few companies in other industries present positions not expected at the outskirt of the map: Peugeot (UG, automobiles), Suez (SEV, waste & disposal services), Vallourec

(VK, industrial machinery). However a result inconsistent with previous findings is that Vivendi (VIV), a media company, is placed in the middle of the map. Also Lagardere (MMB), a group with very diverse activities but with a significant publishing segment is also placed in the center very close to Vivendi. Moreover, inlink counts for these companies are significantly low in relation to other companies in different industries and especially with other media companies in other indexes. These unexpected results could partially be explained by the language problem. Further research on this particular market is necessary to get a better understanding of some particular companies.

Figure 6. MDS map for CAC 40
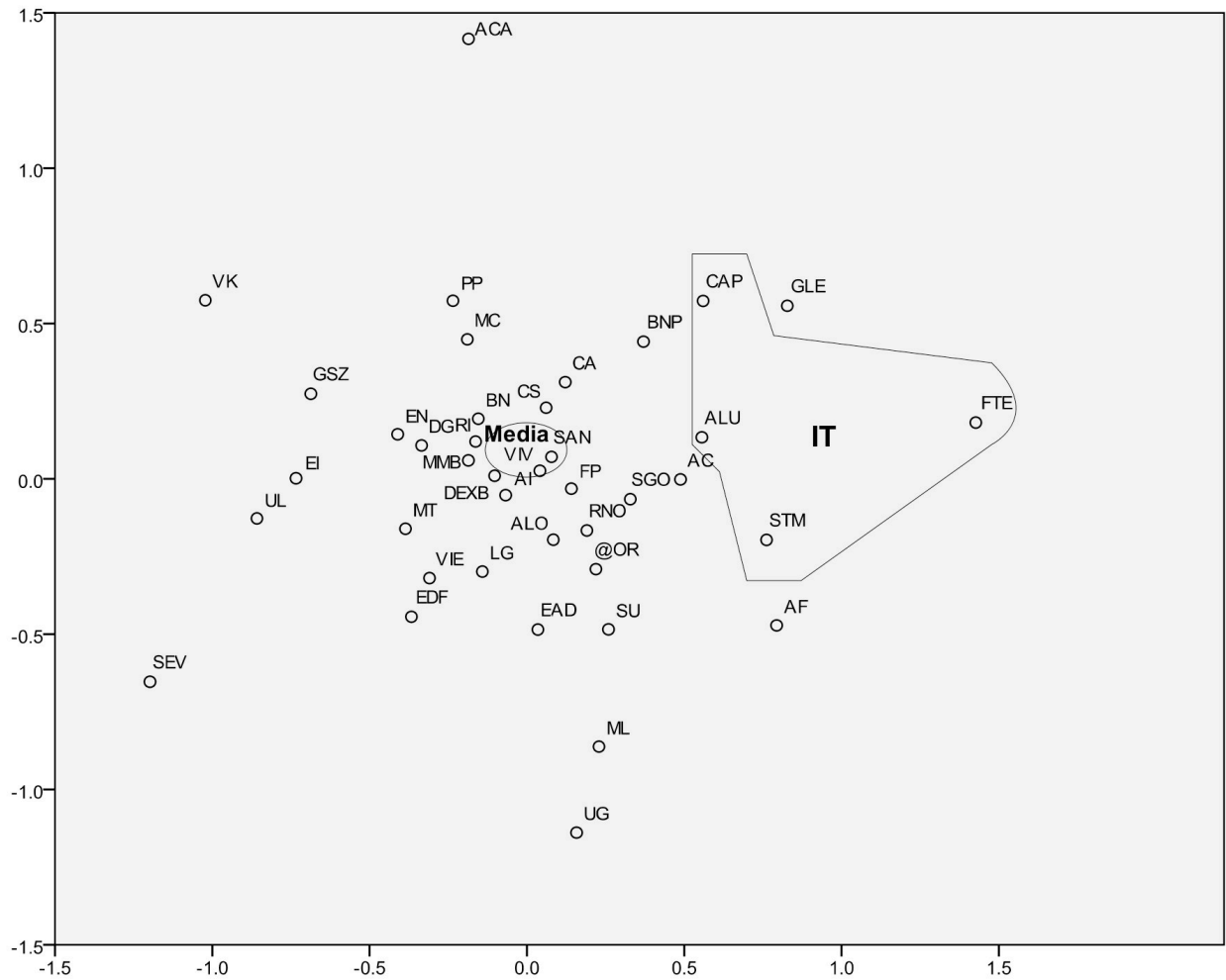
3.5. Ibex 35

The IBEX 35 is the benchmark stock market index of the Madrid Stock Exchange, comprising of the 35 most liquid Spanish stocks. Figure 7 confirms some of the patterns discovered in other stocks in this study.

The only media company TL5 is placed at the outskirt of the map. The cluster with the two IT companies is placed near it. Financial companies, including banks, insurance and traders, form a clear cluster. The rest of the companies in the center of the map belong to different industries with no clear pattern or not enough number of companies to form clusters. However, two of the key industries in Spain are clearly defined in the map: Construction and Gas & Electricity. Construction has been one of the leading industries in the country during the last decades and today it is the one that is suffering economic crisis more deeply.

Figure 7. MDS map for IBEX 35

4. Conclusions and future research

The originality of this study is based on the application of co-link analysis to heterogeneous companies listed in five major stock exchanges indexes. Previous co-link studies examined companies within a single industry with successful results in mapping competitive positions. However, due to the diverse industrial affiliation of the companies in this paper, the expectation is companies in the same industry to appear together in clearly identified clusters, according to the similarity principle used in previous research. Competitive analysis could also be developed but only within a previously identified cluster of

homogeneous companies. However, this is not possible in this study due to the scarce number of companies belonging to the same industry. So far, studies based on heterogeneous types of organisations have only been developed in order to test the triple helix theory in the Web, but their perspective is not purely commercial or economic. Previous section examined the results individually, but there are some overall exploratory conclusions that are promising in order to perform future research and to advance in the application of Webometric techniques to business studies.

Individual MDS maps show that the expectation of finding same industry companies grouped together is only partially fullfilled. Industries whose business model is more information centered form distinct clusters located at a position that is distant from other companies located in more central position. Information centered industries, according to our interpretation, comprise mainly four groups: companies based in the production of information contents (so called Media in the study), companies providing information infrastructures and services (IT), companies very intensive in applying e-commerce techniques (Leisure companies in the tourism industry) and Financial companies. Financial companies, including banks, insurance and other financial services, are information intensive in the sense that they do not usually provide physically tangible services or that they only operate in virtual venues (e.g. stock exchange companies, banks that are mainly present on the Internet, etc.).

In addition to their location in the MDS maps, media and IT companies receive more inlinks than any other companies that belong to more traditional industries. This particular inlink pattern reflects a business model that is highly exposed to Internet. This is true of financial companies to some extent. Generally speaking, they are second only to IT companies in the number of inlinks received. This explain why, in most countries, financial companies are located in a position between the center of the map, where more traditional companies are placed, and the IT and media companies. Therefore, generally speaking, companies that are located away from the centre of the map are the ones that depart from a traditional business model that is slower, by its own industrial nature or by a management decision, to move to an online context. This explanation is supported by the use of the Jaccard index in our data analysis. Because information centered companies receive many inlinks but relatively fewer of them are co-links with other companies, the Jaccard index scores tend to be lower. That is why they are positioned away from the center, at the outskirt of the map.

But why do different patterns exist based on information? The so called information centered companies have undergone a process of digitalization with the result of an increasing presence on the Web. The more dramatic changes are affecting the information content production industry (Media industry, e.g. press, music, cinema, and so on). These companies need to reinvent their business models in order to face the significant digitalization process of the last two decades. In most of the cases intellectual property issues are brought up. Companies that in the past seemed to be stable and to enjoy dominant positions in the market have found themselves competing in a totally new and challenging economic and social environment.

Eurostoxx 50 is the only index in the study that is made of companies belonging to different countries. This fact allowed us to analyze economic integration in the Eurozone. Results show that interpretation by countries of Eurostoxx 50 map fits well the geographical dimension on the economies involved. Except

for the information based industries (IT and Financial) no industrial clusters are clearly identified in the map. As a contrast, Dow Jones map is more clearly defined in terms of industries, underlining the deeper economic integration of the US market. This suggests that the Eurozone economy is not as tightly integrated yet. Perhaps future studies will reveal less country division as the trend of integration continues. An effective economic integration is likely to be easier in industries that are based on information technology. This integration is fostered by the rapid process of digitalization of EU countries. However, there could be some limitations in the Eurostoxx 50 co-link analysis due to language differences.

This paper reveals that Web co-link data constitute a readily available source of information to obtain new insights into the business environment. The study of heterogeneous companies belonging to different industries allows managers, financial analysts and other stakeholders to locate their company's activity in relation to others and to identify differing business modes and to follow their evolution over time. If a company is trying to increase their presence online, co-link analysis could be used by managers to monitor the progresses of their efforts by observing how the co-link MDS maps change. Therefore, co-link analysis could become a managerial tool useful for companies changing their IT strategies. Finally, this paper used online data to provide confirmatory evidence of the changes in business models of information centered industries due to the digitalization process of the last decades.

Conclusions are mainly exploratory but relevant enough to develop a whole set of future confirmatory studies and to open a new direction in the webometric research of commercial and economic issues. Additional evidence about the suggested explanation of different business models in terms of information is needed. Another important research direction to pursue is to monitor the changes of the MDS maps by collecting and analyzing data over time. Given the rapid development of the Internet in the last decade, it is likely that years ago IT and Media companies were not located in such an external position in the maps. Evolution in the coming years could provide confirmation about the transformation in the business models. Same longitudinal perspective could be applied to monitor changes in the Eurozone in order to find out whether country lines become less visible and industry clusters become clearer. Finally, a content analysis study to examine the motivations of Webpage inlinking could be useful to determine if there are significant differences between information centered and traditional business model companies.

Although the paper explored the use of co-link analysis for heterogeneous groups of companies, providing new ways of studying commercial and economic issues, the study has some limitations, most of which are beyond our control. The interpretation of the results could differ due to factors such as different official industrial classifications, alternative criteria to include companies within a cluster, or specific knowledge of the researcher about a particular economic context. The industrial composition of the indexes differs due to the economic characteristics of each country. Also the number of companies in each index is not homogeneous. For example, FTSE 100 includes several leisure companies although Dow Jones Industrial does not. Sometimes only a single company is representative of an industry and no cluster can be identified. However the position of the particular company can be interpreted in comparison to similar companies in other indexes. Finally, language differences could affect the results

due to a potential language bias of the search engine used for data collection, especially for Eurostoxx 50 that includes companies from different countries.

References

Björneborn, L. and 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.

García-Santiago, L. and de Moya-Anegón, F. (2009), "Using co-outlinks to mine heterogeneous networks", *Scientometrics*, Vol. 79 No. 3, pp. 681-702.

Google (2006), "Google SOAP Search API Reference", available at: http://www.google.com/apis/reference.html#2_2 (accessed 3 June 2009).

Google (2009), "Links to your site", available at: http://www.google.com/support/webmasters/bin/answer.py?hl=en&answer=55281 (accessed 3 June 2009).

Heppes, D. and du Toit, A. (2009), "Level of maturity of the competitive intelligence function. Case study of a retail bank in South Africa", *Aslib Proceedings: New Information Perspectives*, Vol. 61 No. 1, pp. 48-66.

Kahaner, L. (1996), *Competitive Intelligence – How to Gather, Analyze, and Use Information to Move Your Business to the Top*, 7th ed., Touchstone, New York, NY.

Live Search (2007), "We are flattered, but...", available at: http://www.bing.com/community/blogs/search/archive/2007/03/28/we-are-flattered-but.aspx (accessed 3 June 2009).

Porter, M. E. (1980), *Competitive Strategy: Techniques for Analyzing Industries and Competitors*, Free Press, New York, NY.

Porter, M. E. (1985), *The Competitive Advantage: Creating and Sustaining Superior Performance*, Free Press, New York, NY.

Reid, E. (2003), "Using Web link analysis to detect and analyze hidden Web communities", in Vriens, D. (Ed.), *Information and communications technology for competitive intelligence*, Ideal Group, Hilliard, OH, pp. 57-84.

Romero-Frías, E. and Vaughan, L. (2009), "Financial Distress of U.S. Banking Industry Viewed through Web Data", *12th International Conference on Scientometrics and Informetrics ISSI* 2009, Rio de Janeiro, Brazil, July 14-17, 2009.

Shaw, D. (2001), "Playing the links: interactivity and stickiness in .com and "not.com" websites", *First Monday*, Vol. 6 No, 3, available at: http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/837/746 (accessed 10 May 2009).

Small, H. (1973), "Co-citation in the scientific literature: a new measure of the relationship between two documents", *Journal of the American Society for Information Science*, Vol. 24 No. 4, pp. 265-269.

Stuart, D. and Thelwall, M. (2006), "Investigating triple helix relationships using URL citations: a case study of the UK West Midlands automobile industry", *Research Evaluation*, Vol. 15 No. 2, pp. 97–106.

Tan, B., Foo, S. and Hui, S. C. (2002), "Web information monitoring for competitive intelligence", *Cybernetics and Systems*, Vol. 33 No. 3, pp. 225-251.

Thelwall, M. and Wilkinson, D. (2004), "Finding similar academic Web sites with links, bibliometric couplings and colinks", *Information Processing and Management*, Vol. 40, pp. 94–101.

Vaughan, L. (2006), "Visualizing Linguistic and Cultural Differences Using Web Co-Link Data", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 9, pp. 1178-1193.

Vaughan, L. and Thelwall, M. (2004), "Search engine coverage bias: Evidence and possible causes", *Information Processing & Management*, Vol. 40 No. 4, pp. 693-707.

Vaughan, L., and You, J. (2006), "Comparing business competition positions based on Web co-link data: The global market vs. the Chinese market", *Scientometrics*, Vol. 68 No. 3, pp. 611–628.

Vaughan, L., and You, J. (2008), "Content assisted web co-link analysis for competitive intelligence", *Scientometrics*, Vol. 77 No. 3, pp. 433-444.

Vaughan, L., and You, J. (2009), "Keyword enhanced Web structure mining for business intelligence", *Lecture Notes in Computer Science*, Vol. 4879, pp. 161–168.

Vaughan, L., and Wu, G. Z. (2004), "Links to commercial websites as a source of business information", *Scientometrics*, Vol 60 No 3, pp. 487–496.

Vaughan, L., Gao, Y. and 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. and Gao, Y. (2007), "Are co-linked business web sites really related? A link classification study", *Online Information Review*, Vol. 31 No. 4, pp. 440–450.

Vaughan, L., Tang, J. and Du, J. (2009), "Examining the robustness of Web co-link analysis", To appear in *Online Information Review*, 33, (5).

Zuccala, A. (2006), "Author Cocitation Analysis Is to Intellectual Structure A Web Colink Analysis is to...?", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 11, pp. 1487-1502.

Appendix 1. Dow Jones Industrial companies

Labels	Company	Main URL	Industry	Linkdomain count
MMM	3M Co.	http://www.3m.com	Diversified Industrials	251,000
AA	Alcoa Inc.	http://www.alcoa.com	Aluminum	42,900
AXP	American Express Co.	http://www.americanexpress.com	Consumer Finance	1,120,000
T	AT&T Inc.	http://www.att.com	Fixed Line Telecommunications	2,700,000
BAC	Bank of America Corp.	http://www.bankofamerica.com	Banks	364,000
BA	Boeing Co.	http://www.boeing.com	Aerospace	247,000
CAT	Caterpillar Inc.	http://www.cat.com	Commercial Vehicles & Trucks	227,000
CVX	Chevron Corp.	http://www.chevron.com	Integrated Oil & Gas	233,000
C	Citigroup Inc.	http://www.citigroup.com	Banks	128,000
KO	Coca-Cola Co.	http://www.coca-cola.com	Soft Drinks	214,000
DD	E.I. DuPont de Nemours & Co.	http://www.dupont.com	Commodity Chemicals	141,000
XOM	Exxon Mobil Corp.	http://www.exxonmobil.com	Integrated Oil & Gas	264,000
GE	General Electric Co.	http://www.ge.com	Diversified Industrials	262,000
GM	General Motors Corp.	http://www.gm.com	Automobiles	242,000
HPQ	Hewlett-Packard Co.	http://www.hp.com	Computer Hardware	2,900,000
HD	Home Depot Inc.	http://www.homedepot.com	Home Improvement Retailers	236,000
INTC	Intel Corp.	http://www.intel.com	Semiconductors	2,070,000
IBM	International Business Machines Corp.	http://www.ibm.com	Computer Services	3,830,000
JNJ	Johnson & Johnson	http://www.jnj.com	Pharmaceuticals	91,300
JPM	JPMorgan Chase & Co.	http://www.chase.com	Banks	232,000
KFT	Kraft Foods Inc. Cl A	http://www.kraft.com	Food Products	142,000
MCD	McDonald's Corp.	http://www.mcdonalds.com	Restaurants & Bars	636,000
MRK	Merck & Co. Inc.	http://www.merck.com	Pharmaceuticals	253,000
MSFT	Microsoft Corp.	http://www.microsoft.com	Software	45,500,000
PFE	Pfizer Inc.	http://www.pfizer.com	Pharmaceuticals	130,000
PG	Procter & Gamble Co.	http://www.pg.com	Nondurable Household Products	272,000
UTX	United Technologies Corp.	http://www.utc.com	Aerospace	42,800
VZ	Verizon Communications	http://www22.verizon.com	Fixed Line	332,000

	Inc.		Telecommunications	
WMT	Wal-Mart Stores Inc.	http://www.walmart.com	Broadline Retailers	1,310,000
DIS	Walt Disney Co.	http://www.disney.go.com	Broadcasting & Entertainment	5,500,000

Appendix 2. FTSE 100 (top 35 inlinked companies)

Labels	Company	Main URL	Industry	Linkdomain count
BSY	B Sky B Group	http://www.sky.com	Media; Television	2,040,000
BARC	Barclays	http://www.barclays.co.uk	Financials; Bank	62,500
BAY	British Airways	http://www.britishairways.com	Transports; Airlines	241,000
BT-A	BT Group	http://www.bt.com	Telecommunications	451,000
CW	Cable & Wireless	http://www.cw.com	Telecommunications	20,300
CCL	Carnival	http://www.carnival.com	Leisure; Travel Agents	69,800
EXPN	Experian	http://www.experian.com	Economic intelligence services	423,000
FGP	Firstgroup	http://www.firstgroup.com	Transports	30,600
HSBA	HSBC Hldg	http://www.hsbc.com	Financials; Bank	188,000
IHG	Intercont Hotels	http://www.intercontinental.com	Leisure; Accommodation	142,000
LLOY	Lloyds Banking Grp	http://www.lloydstsb.com	Financials; Bank	31,300
LSE	LSE Group	http://www.londonstockexchange.com	Financials; Stock Exchange	149,000
PSON	Pearson	http://www.pearson.com	Media; Books	22,100
REL	Reed Elsevier	http://www.reed-elsevier.com	Media; Press	828,000
RBS	Royal Bank Scotland Group	http://www.rbs.com	Financials; Bank	26,600
SGE	Sage Group	http://www.sage.com	IT; Software	73,000
STAN	Standard Chartered	http://www.standardchartered.com	Financials; Bank	37,400
TCG	Thomas Cook Group	http://www.thomascook.com	Leisure; Travel Agents	52,000
TRIL	Thomson Reuters	http://www.reuters.com	Media; Specialist Retailing	10,400,000
TT	Tui Travel	http://www.tuitravelplc.com	Leisure; Travel Agents	92,600
VOD	Vodafone Group	http://www.vodafone.com	Telecommunications; Mobile	148,000
WPP	WPP	http://www.wpp.com	Services; Advertising	34,700

Appendix 3. Eurostoxx 50

Labels	Company	Main URL	Country	Industry	Linkdomain count
AEGN	Aegon	http://www.aegon.com	NL	Financials, Life Insurance	11,600
ALVG	Allianz	http://www.allianz.com	DE	Financials, Insurance	38,600
AXAF	Axa	http://www.axa.com	FR	Financials, Insurance	32,300
BBVA	Banco Bilbao Vizcaya Argentaria	http://www.bbva.com	ES	Financials, Bank	33,800
SAN	Banco Santander	http://www.santander.com	ES	Financials, Bank	366,000
BNPP	BNP Paribas	http://www.bnpparibas.com	FR	Financials, Bank	112,000
CAGR	Crédit Agricole	http://www.credit-agricole.com	FR	Financials, Bank	26,200
DBKGn	Deutsche Bank	http://www.db.com	DE	Financials, Bank	76,900
DB1Gn	Deutsche Boerse	http://www.deutsche-boerse.com	DE	Financials, Investment Services	88,800
DTEGn	Deutsche Telekom	http://www.telekom.de	DE	Telecommunications, Mobile	222,000
FOR	Fortis	http://www.fortis.com	NL	Financials, Bank	21,800
FTE	France Telecom	http://www.francetelecom.com	FR	Telecommunications, Fixed Line	885,000
SOGN	Société Générale	http://www.societegenerale.fr	FR	Financials, Bank	52,800
ING	ING grp	http://www.ing.com	NL	Financials, Insurance	57,400
ISP	Intesa Sanpaolo	http://www.intesasanpaolo.com	IT	Financials, Bank	40,900
MUVGn	Muenchener Rueck	http://www.munichre.com	DE	Financials, Insurance	8,870
NOK1V	Nokia	http://www.nokia.com	FI	Technology, Technology Hardware & Equipment	1,340,000
TLIT	Telecom Italia	http://www.telecomitalia.com	IT	Telecommunications, Fixed Line	5,140
TEF	Telefonica	http://www.telefonica.es	ES	Telecommunications, Fixed Line	279,000
CRDI	Unicredit	http://www.unicreditgroup.eu	IT	Financials, Bank	58,000

Appendix 4. CAC 40

Labels	Company	Main URL	Industry	Linkdomain count
AC	Accor	http://www.accor.com	Leisure; Travel; Hotels	75,900
ALU	Alcatel-Lucent	http://www.alcatel-lucent.com	Technology Hardware & Equipment	69,100
BNP	BNP Paribas	http://www.bnpparibas.com	Financials, Bank	111,000
CAP	Cap Gemini	http://www.capgemini.com	Software & Computer Services	77,900
ACA	Crédit Agricole	http://www.credit-agricole.com	Financials, Bank	26,300
FTE	France Teleco	http://www.francetelecom.com	Fixed Line Telecommunications	881,000
MMB	Lagardere	http://www.lagardere.com	Media; others	5,890
UG	Peugeot	http://www.peugeot.com	Automobiles	47,200
GLE	Société Générale	http://www.societegenerale.fr	Financials, Bank	51,500
STM	STMicroelectronics	http://www.st.com	Technology Hardware & Equipment	75,000
SEV	Suez Environnement	http://www.suez-environnement.com	Industrial Goods & Services	21,800
VK	Vallourec	http://www.vallourec.com	Industrial Goods & Services	2,600
VIV	Vivendi	http://www.vivendi.com	Media	23,500

Appendix 5. IBEX 35

Labels	Company	Main URL	Industry	Linkdomain count
ANA	Acciona	http://www.acciona.es	Construction	23,700
ACS	Actividades Construcciones y Servicios	http://www.grupoacs.com	Construction	3,710
BBVA	Banco Bilbao Vizcaya Argentaria	http://www.bbva.es	Bank	214,000
SAB	Banco de Sabadell	https://www.bancsabadell.com	Bank	5,860
BTO	Banco Español de Credito	http://www.banesto.es	Bank	33,200
POP	Banco Popular Español	http://www.bancopopular.es	Bank	18,200
SAN	Banco Santander Central Hispano	http://www.santander.com	Bank	367,000
BKT	Bankinter	https://www.bankinter.com	Bank	8,490
BME	Bolsas y Mercados Españoles	http://www.bolsasymercados.es	Investment services	15,800
MAP	Corporacion Mapfre	http://www.mapfre.com	Insurance	24,400
ELE	Endesa	http://www.endesa.es	Utilities	61,500
FER	Ferrovial	http://www.ferrovial.es	Construction	13,100
FCC	Fomento de Con. y Contratas	http://www.fcc.es	Construction	20,600
GAS	Gas Natural sdg	http://www.gasnatural.com	Utilities	28,800
TL5	Gestevision Telecinco	http://www.telecinco.es	Media & Advertisement	378,000
IBE	Iberdrola	http://www.iberdrola.es	Utilities	50,600
IDR	Indra, serie A	http://www.indra.es	Electronics & Software	37,700
OHL	Obrascon Huarte Lain	http://www.ohl.es	Construction	1,520
REE	Red Electrica de España	http://www.ree.es	Utilities	19,900
SYV	Sacyr Vallehermoso	http://www.gruposyv.com	Construction	6,330
TEF	Telefonica	http://www.telefonica.es	Telecommunications	279,000
UNF	Union Fenosa	http://www.unionfenosa.es	Utilities	18,400

Romero-Frías, E. & Vaughan, L. (2010). "Patterns of Web Linking to Heterogeneous Groups of Companies: The Case of Stock Exchange Indexes". *Aslib Proceedings*, 62(2): 144-164. Postprint for research purposes.