

The Journal

Cybermetrics News

Editorial Board

Guide for Authors

Issues Contents **The Seminars** **The Source**Scientometrics Tools R&D Policy & Resources **VOLUME 10 (2006): ISSUE 1. PAPER 1****Analyzing hyperlinks networks: The meaning
of hyperlink based indicators of knowledge**

production

**Gaston Heimeriks* & Peter Van den Besselaar****

* NIWI, Royal Netherlands Academy of Arts and Sciences

** Science System Assessment Unit, Rathenau Instituut, the Netherlands

** Amsterdam School of Communications Research, University of Amsterdam

E-mail: Gheimeriks@gmail.com ; P.vandenbesselaar@rathenau.nl**Abstract**

Hyperlinks are the most commonly used attributes to study web sites and structures on the web. In this paper, we analyze and compare hyperlink networks using a variety of linking units on different levels of aggregation and specificity. Focusing on the scientific web, we selected the following linking units: countries, universities, departments and individuals. This paper discusses whether and how the heterogeneous linking patterns might provide information about knowledge production and its context.

Keywords

Hyperlink analysis, academic web, multi-level, mapping

1 Introduction

On a very large scale, the Web is a potential source of data that can be used for developing indicators and statistics about processes of knowledge production and dissemination. In this paper we explore what is considered to be one of the most important sources for webometric indicators: hyperlinks. Hyperlinks, or links, have been used to indicate recognition, trust, importance, impact and internationalization of a web site (Thelwall 2002), and there has been a strong growth of hyperlink analysis studies in the Internet community (Park and Thelwall 2003). Our substantial interest here is in using hyperlink networks to study the development of research fields and the relationship between research organizations and the relevant institutions in their environment (users of knowledge, funding institutions, regulatory bodies, and media). The underlying idea is that collaboration and information exchange between these various organizations are also reflected in the hyperlink networks.

However, the interpretation of hyperlinks and hyperlink networks remains problematic. What are the nodes that carry the hyperlink networks? In other words, what is an appropriate level of analysis for hyperlink studies? What do hyperlink networks teach us? This study aims to explore how hyperlinks can be used to study social and cognitive processes in knowledge production. One of the main problems inherent in performing web-based studies on knowledge production is the absence of clear sociological or scientometric units of analysis on the web. The offline world provides relatively unambiguous units of analysis in the form of individuals and institutions as well as publications and journals. Not only do these phenomena not always have clear equivalents on the web, but also the way actors relate to each other through hyperlinks may be different from their relationships in the offline world. In this paper we

provide a systematic exploration of the kind of information that can be obtained in studying hyperlinks on various levels. We will study hyperlinks and hyperlink networks on various levels of institutional aggregation and on various levels of disciplinary specialization. By analyzing and comparing the resulting networks, we hope to gain insight into the socio-cognitive meaning of hyperlinks and link-structures, and into the appropriate scope and level for hyperlink studies. In completing such an analysis, we can ask: What do we learn from links? What associations are created on the web, and which patterns are emerging? In this paper we focus on the scientific web, and more specifically on the institutions of higher education in the EU; this class of organizations was chosen as it has a relatively developed presence on the Internet. Even so, the results may also be relevant for other parts of the WWW.

2 Webometrics

The emergence of ICTs gave rise to a new kind of quantitative analysis called "webometrics." Most of the attention in webometrics has been focused on hyperlinks. As Tang and Thelwall (2003) point out, scholars have studied hyperlinks from several disciplines using very different perspectives. In theoretical physics, for example, the emphasis has been on developing mathematical models for the underlying properties of the network created by pages and links, in addition to other topologically similar types of networks (Adamic and Huberman 2000). Web links have also been the subject of information retrieval research by information scientists. Such studies have used Web-link structures essentially on their own (Brin and Page 1998), or in combination with textual analysis (Kleinberg 1999). Tang and Thelwall (2003) stress that even though empirical evidence is lacking for the efficacy of using links in this context, the widespread belief in their usefulness is fuelled by qualitative analyses and case studies (Brin and Page 1998), as well as the effectiveness and popularity of Google (Thelwall 2002), a search engine with a link-based algorithm claimed as the reason for its success.

Academic Web sites contain material created for many different purposes (Middleton 1999), and studies of the targets of academic links have found that these pages were of many different types (Cronin, Snyder et al. 1998) including recreational, with only a tiny minority containing academic content equivalent to a journal article (Thelwall 2001). For the similar issue of URL citation in traditional articles, Kim (2000) shows that reasons for use extended beyond those used in print citations to include medium-specific ones. Zang (2001) finds several factors that inhibit URL citing, one of which includes the self-perceived lack of ability to use the Internet.

Research on scholarly communication from the information science perspective has focused on the use of techniques derived from bibliometrics to study conceptually coherent networks of pages or sites; such networks include groups of countries (Ingwersen 1998) and groups of universities (Thelwall 2001). Web links are a phenomenon of interest to bibliometricians through their analogy with citations, and to others because of their use in Web navigation and search engines. Hyperlink research is an increasingly important area in information sciences based on the idea that hyperlinks provide information that is not accessible to traditional bibliometrics (Cronin 2001). However, scientific communications in electronic media are less codified and more heterogeneous than their print equivalent because the operations that define the boundaries of the scientific system (like peer review procedures and acceptance or rejection of papers by editorial boards) do not always apply to electronic communications. It is known that very few links on university Web sites are targeted at scholarly expositions and yet, at least in the UK and Australia, a correlation has been established between link count metrics for universities and measures of institutional research (Thelwall 2002).

The most developed area of study in academic hyperlinking is that of an entire group of universities within a single country (Thelwall 2002 and Thelwall and Tang, 2003). Because science is probably one of the areas where national differences in communication habits may be less pronounced, this might provide an interesting starting point for developing hyperlink-based indicators. In many social situations, communication is subject to more national preferences than in science. Increasingly, science is very much a global undertaking, and research teams and science communities are often multinational. Furthermore, English is predominantly the language of choice,

due to the dominance of scientists from the USA as well as English language journals in many fields. However, the systems of higher education and research (and their funding) are also organized at the national level, as are many aspects of career development (Barjak 2004). A survey (Thelwall, Tang et al. 2003) shows that English is the dominant language both for linking pages and all other pages: in a typical country, approximately half the pages were in English. Additionally, normalized interlinking shows that international interlinking throughout Europe transpires in English, and additionally in Swedish in Scandinavia. Furthermore, linking occurs between countries sharing the same language. A study on the structure of the international Internet flows (Barnett 2004) concludes that language and culture are significant determinants of hyperlinking relationships.

Departments have been the targets of Webometric analysis before, but early research did not produce statistically significant findings. One recent study demonstrates that the counts of links to LIS departments in the United States are correlated with their U.S. News and World Report ranking (Chu and Thelwall 2002). A later study finds a similar connection between research and in-link counts for U.K. computer science departments (Li 2003). Nonetheless, no causative connection is claimed between research and link targeting. Most academic Web-link studies have investigated the reason for the statistical association between links and research to some extent, and one recent article focuses on this single issue (Wilkinson 2003). Although the vast majority of links between U.K. universities were connected in some way with scholarly activities (including teaching), less than 1% were formally referencing journal-quality publications. Counts of links, therefore, represent predominantly an amalgamation of many different causes loosely related to research and various forms of informal scholarly communication. Nevertheless, web data can be meaningful in mapping the aspects of knowledge production.

In a recent case study (Heimeriks and Van den Besselaar submitted) we aimed at mapping relevant scholarly communications of a research group operating in an application-oriented techno-scientific environment. This study explored the opportunity of using inlinks, outlinks, incoming emails, outgoing emails, project co-operations and the co-authored publications of a research group as indicators for the context of knowledge production and dissemination. By focusing on the shape of the networks and the intensity of the communication over these networks, we were able to identify the role of computer-mediated communications in relation to print and other traditional media. Non-electronic media mainly function within the disciplinary network, but electronic media enlarge the network to the users of the produced knowledge. The outgoing communications also showed the application context of the research of this group. Electronic media, therefore, create new means of access, traceable by inlinks and log-files, to the knowledge produced by research groups for a variety of users. Inlinks provide us with interesting information about the academic environment of the knowledge production as well as the (academic and non-academic) context of users, and visits to the web site present the network of the (non-academic) users of the output.

Taking the web site of a research group as point of departure, the results of our research suggest that the individual Web page is not necessarily the correct or only useful unitary entity for the purpose of analyzing the Web. For example, inter-site links have been singled out as more important than intra-site links (Kleinberg 1999), showing the need for alternative perspectives. Moreover, by returning results organized by site, search engines also implicitly recognize that the Web is not a collection of unrelated pages (Thelwall 2002). From a Webometrics perspective, aggregating pages into clusters using alternative document models based upon directories, domains and multi-domain sites has previously been found to be a fruitful technique (e.g., Thelwall 2001, 2002).

When acknowledging these indicators, it seems appropriate to explore the potential for clustering information on the Web based on different levels of aggregation rather than that of the page. (Thelwall 2003) coined this as a "layered approach" to investigate the community structure of a section of the Web in a systematic way, in order to ascertain, in principle, whether this approach is viable.

Secondly, as Thelwall points out, the structure of the Web is studied, modeled and visualized within a variety of disciplines, such as communication studies, social network analysis, computer science, geography, information science,

physics and sociology (Tang and Thelwall 2003). In this paper, we start from a more sociological perspective to explore the kind of information that can be derived from academic hyperlink patterns on the web, on different levels of aggregation and specialization. Hyperlinks hold meaning in two different ways: First of all, hyperlinks are attributes of a web page or web site. This enables us to map the websphere that emerges from the set of outlinks and inlinks of a site. What types of actors are involved in the environment of a site? What role does geographical proximity play? Additionally, to what extent are the number and nature of inlinks and outlinks related to other characteristics of the site? We will investigate some of the relationships one may expect to follow from the linking behavior of individuals and organizations.

We approach hyperlinks as associations between web pages, web sites or webspheres. Communications between actors (e.g., between co-producers of knowledge, between collaborators in research projects, or between users and producers of knowledge) are not randomly distributed but take place in clusters or communities. Here, we will use various methods to delineate the hyperlink networks, using a variety of definitions of the nodes. The question to be answered is: to what extent do the hyperlinks constitute networks that hold a meaningful interpretation? We are interested in discovering whether these hyperlink communities are motivated by cultural, geographical or cognitive reasons.

3 Data and methodology

Foremost, the WWW is a large collection of individual web pages. In studying the web it is often useful to take collections of web pages as a unit. However, the question quickly arises as to how aggregation of individual pages into meaningful units can take place. An obvious starting point is to take the web sites of individuals or institutions as the unit of analysis. Apart from the technical problem of identifying the boundaries of a web site, we are confronted with the conceptual issue of defining a web site: is it a personal site, or an institutional one? Furthermore, we can distinguish between sites of research groups, of departments, of faculties and of universities. Another possible way to define the unit of analysis is geographical, e.g. by taking the national or the regional scientific web as a unit of analysis. Finally, one may argue that the boundaries should be determined topologically, that is by using the density of the network of pages as criterion (Thelwall 2003).

In this paper, we distinguish between the various types of nodes that are hyperlinked to each other. In the literature, many different types of nodes are used. They differ with regards to their level of aggregation: country, university, department, or individual researcher, and in terms of generality: all disciplines, one discipline, or a research topic. Finally, the delineation of the nodes differ in a geographical sense: one region, one country, a few countries, the EU, or the whole world. In terms of the geographical boundaries, we take the EU as the domain. In this paper we use the following types of nodes:

1. The national academic web is defined as the hyperlink network that emerges from all of the web pages of all academic institutions in one European country.
2. The national disciplinary web is defined as the hyperlink network that emerges from all of the web pages of all academic departments in one country devoted to one discipline. This is one level of generality lower than the first network.
3. A university web site is defined as the hyperlink network that emerges from the set of all pages of one university. This is one level of aggregation and one level of generality lower than the first network. We will perform this analysis for all European universities and for all universities in one country (here: the Netherlands).
4. A departmental (disciplinary) web site is defined as the hyperlink network that emerges from all pages of one department. This is one level of aggregation and one level of generality lower than the first network.
5. The individual researchers' web site is the hyperlink network between the different individual web sites. This is one level of aggregation lower than the previous network.

In the first part of this study, we started with data collected by automatic intelligent agents scanning various search engines. This data set contained 1064 academic Web sites from 22 European countries, including all countries of the European Union. With an autonomous intelligent agent operating on the Alta Vista search engine and for every academic site, we measured the number of links to every other site of the data set and the number of internal links.

At a later stage, a much larger dataset became available. Most of the data in this study is based on web data collected in the EICSTES project¹. The data consist of information about the fifteen² EU member states' universities, departments of the universities and their outgoing links, in addition to descriptions of the site characteristics (Arroyo, Pareja et al. 2003). Once web sites were identified and selected, some basic information was collected, using software tools called 'mappers': we recorded the name, the institution they belong to and the URL that identifies them. Mappers simply crawl the web starting from a certain site, follow the trace of its embedded links and register the objects found in this process³. The software program used to construct the database of European universities is Microsoft Site Analyst⁴. All URLs were classified in three ways: an Institutional coding that provides a classification of the type of entity based on a survey of the higher education systems in the European Union; a geographical coding using the NUTS classification (Nomenclature of Territorial Units for Statistics) of EUROSTAT; and a thematic coding, according to UNESCO codes in science and technology domains. These codes have a 3-level structure. The first two digits refer to the general fields, and the third and fourth digit refer to scientific fields. The last two digits refer to subfields. In this study, the first 4 digits are used for the delineation of the fields. Here, we focus on the field of Artificial Intelligence for mapping the disciplinary networks on the Web. The field of AI is defined rather broadly in this UNESCO classification, and therefore includes a large number of departments from computer science. For our purposes we organized the data in an ORACLE database. This enabled us to include all data (links and site description) in a single database. The data consist of actor descriptions and hyperlinks that enable us to construct networks. In order to study the various dimensions of hyperlink properties, an ORACLE database enables us to select all outlinks from a certain country, a set of departments, etc.

The network analyses in this study focus on two dimensions of the hyperlink networks. First, we performed relational analysis that concentrate on the emerging clusters between organizations that maintain hyperlink relations. We use the so-called cosine algorithm for determining the association (proximity) of two nodes. As the number of nodes is often very large, we used network visualizations by means of bibliometric software BibTechMon^{®5} (BTM). BTM is a flexible tool for analyzing and visualizing large networks in various dimensions, and it is based on a "mechanical spring model" (Kopcsa 2000). This enables a transformation into 2-dimensional map. These relational maps provide information about the cliques and cohesive subgroups into which a network can be divided.

Second, we carried out a positional analysis of the hyperlink network of European academic organizations. This analysis focuses on the similarity of linking patterns of groups of nodes, rather than the existence of direct associations between the nodes. Positional analysis of the networks is based on factor analysis; multi dimensional scaling plots can visualize the results⁶. For these analyses we use SPSS PC 11.5. The positional analyses focus on more qualitative features of the networks (Burt 1982). Such an analysis enables us to identify structural similarity such as roles in the networks: For example, nodes may occupy similar positions in a network without maintaining a relationship between them.

4 Networks on different levels of aggregation and specialization

4.1 The hyperlink network of national academic webs

To analyze the hyperlink network of national academic webs, automatic intelligent agents carried out data collection by scanning various search engines. These data were used to measure the number of outlinks among the university sites in the 15 EU countries.

Aggregating these outlinks to the country level gives the hyperlink network of national academic webs. This network is complete, in that all countries link to all others, as expected. Therefore, the relationships do not show any pattern. Visualization of the relational hyperlink network (figure 1) shows this, and additionally the strength of the relationships between the EU countries in academic hyperlinking. Not surprisingly, a strong correlation exists between the size of the country and the number of relationships it maintains with other countries; the size of the nodes represents this phenomenon.

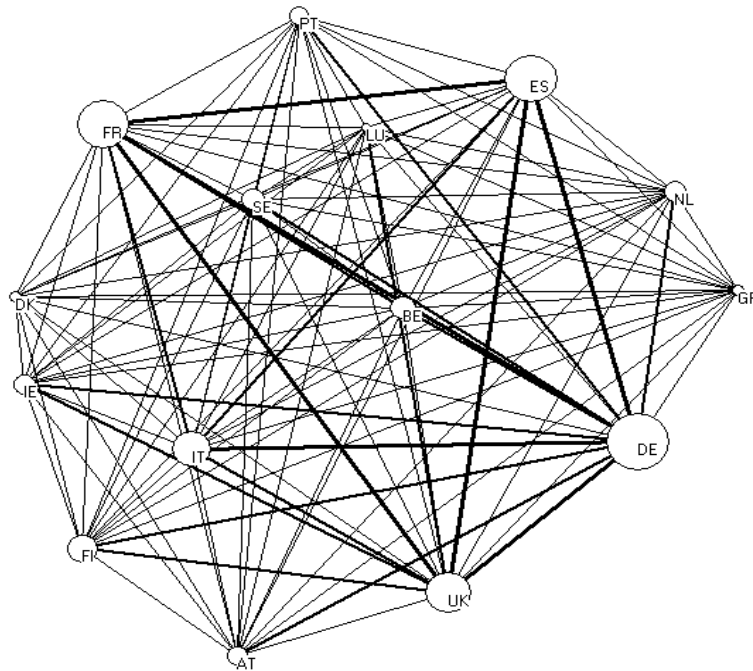


Figure 1. The European network of academic outlinks

The relational network has no structure or cohesive subgroups within the network. This may be different in the positional network, which informs us about countries that do have similar linking patterns. Factor analysis of the matrix of links between all European universities on the country level (thus excluding the domestic links) reveals two very pronounced clusters that can be identified as a Germanic and an Anglo-Saxon cluster, in addition to several other clusters.

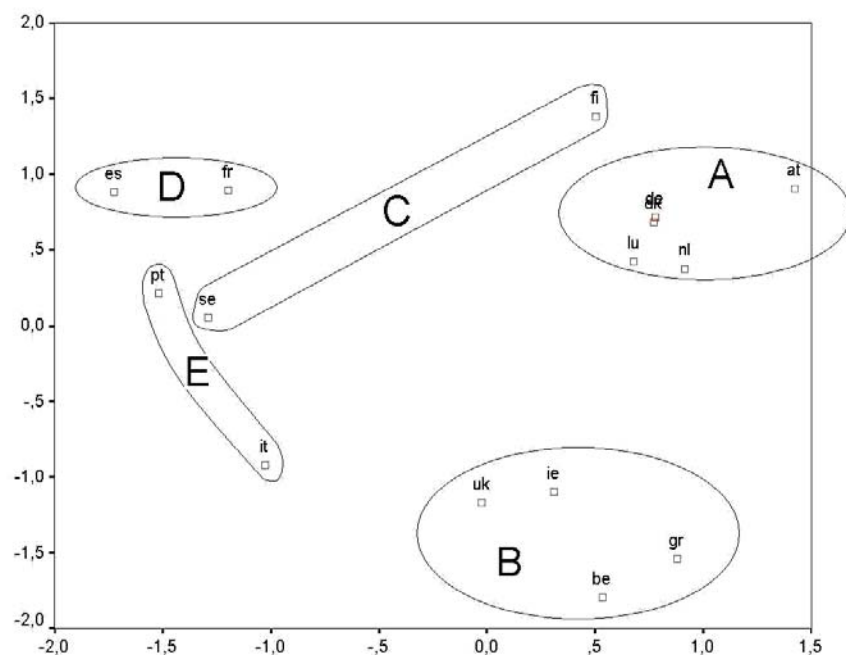


Figure 2. Structural equivalent positions –in the inlink network (EU)

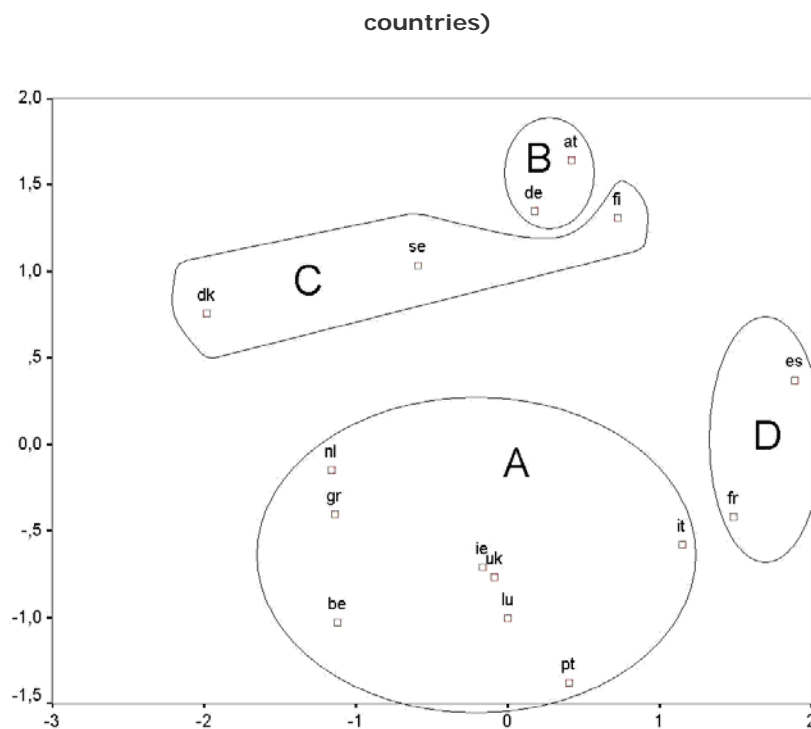


Figure 3. Structural equivalent positions –in the outlink network (EU countries)

A factor analysis of the country-by-country “situation” matrix was carried out for the incoming as well as for the outgoing direction. The factor analysis of the inlinks to European universities suggests a strong geographical and/or language bias. The most pronounced cluster (with the highest explained variance) is made up of Germany, Denmark, Austria, Luxembourg and the Netherlands. The second most important cluster contains the UK, Ireland, Belgium and Greece. This cluster is clearly organized around hyperlink patterns to English-language sites. This is obvious for the UK and Ireland, but also Greece, which is well known to communicate (only) in English in international contexts. Other factors, such as Internet penetration (among the lowest in Greece and among the highest in the UK) fail to explain the composition of this cluster. Belgium shows interfactorial complexity: It contributes to all clusters probably because of its bilingual and bicultural nature. Belgium has its highest loading (0.56) on factor B and its second loading (0.42) on factor D, which also contains France. The smaller clusters are Scandinavian (Sweden and Finland in cluster C) and Southern European (Spain and France in cluster D, and Portugal and Italy in cluster E).

The two multidimensional scaling plots (figures 2 and 3) show the positioning of the EU countries according to the incoming links of their university sites.

The positional clusters emerging from the outlink communication patterns show some resemblance with the inlink structure. However, the most significant cluster contains the UK, Ireland, Luxembourg, Greece, Portugal, Belgium, the Netherlands and Italy. It seems that the language, more than geographical distance, is the decisive operation underlying this pattern. This is confirmed by the composition of the other clusters: Austria and Germany (cluster B), Denmark, Sweden and Finland (cluster C) and France and Spain (cluster D).

Note that the clustering here is a positional one, indicating that countries in one cluster are similar in terms of their international out/in-link pattern.

4.2 The hyperlink network of national disciplinary webs

We repeated the positional and relational analyses on the level of some individual disciplines. Again we found relations between all countries, and again this does not provide us with a structure. The map (figure 4) shows the strength of European countries in the network of knowledge production in Artificial Intelligence (AI). The largest nodes in the aggregated outlink network in Artificial Intelligence are France, Germany, the UK and Spain. The

size of the nodes in the visualization is proportional to the number of links in the node. The figure indicates that there is hardly any relational structure in the network, as all countries have outlink-relations with each other.

Also here, the positions in the network shed a different light on the situation. Factor analyses of the AI (out)linking matrix between countries again reveal a clustering that seems to be based on cultural and geographical proximity⁷ (fig. 5).

On the highest level of countries, the emerging structures of outlinks between departments in the same discipline also provide a measure of the European network in the science system. As on the level of universities, the positions are based on geographical proximity and language. The outlink patterns between countries within the same discipline suggest a tri polar system with the UK, France and Germany as the central nodes. A separate Scandinavian cluster is positioned between these three clusters.

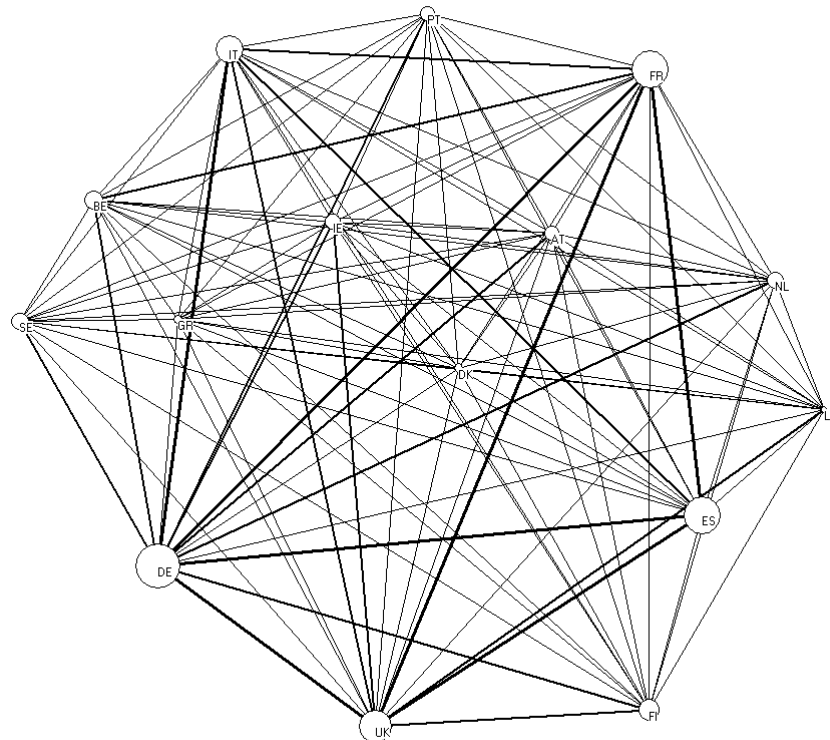


Figure 4. The European outlink network in AI - aggregated on country level

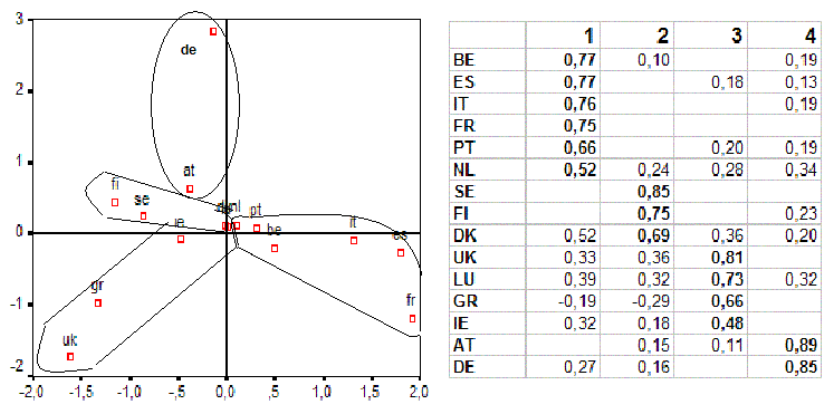


Figure 5. Structural equivalent positions in the European outlink network in AI

4.3 The hyperlink network of universities web sites

On the next, more detailed, level we focused on the hyperlink network between EU universities. The relational analysis of the outlink network suggests that many clusters of universities can be identified, but the pattern does not seem to be based on “cognitive” structures. Although many international hyperlink relationships exist, the national preference in linking relationships seems dominant. The visualization (figure 6) shows a strong core of linked universities that emerges as centered around three dominant countries: the UK (shown in grey in the upper right part of the graph), Germany (on the left side) and France (on the lower right side). The Scandinavian universities are mostly positioned between the German and British universities. Most of the network is language-based, as Austrian universities are adjacent to the German ones, the Irish are close to the UK, and Italian, Spanish, Portuguese and Belgium universities are alongside the French cluster.

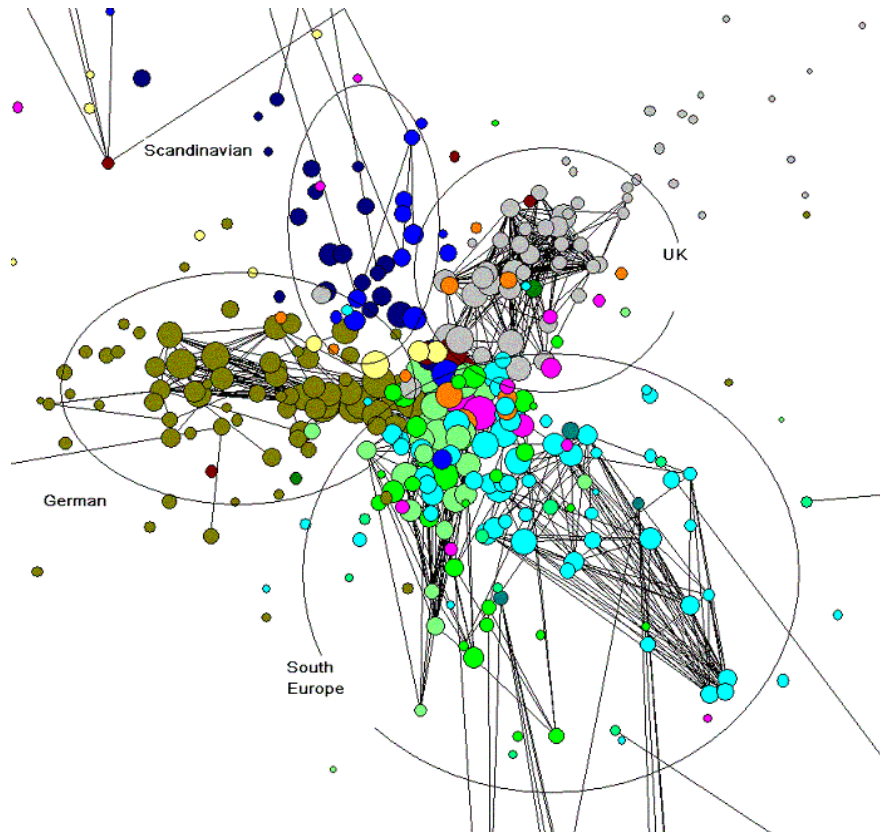


Figure 6. The core of the outlink network between EU universities

The network of European universities shows a strong level of national preference: universities from the same country are likely to maintain hyperlink relationships. Furthermore, many of the international linking patterns that do exist reflect a local (neighboring) preference. The figure shows distinct clusters of southern European countries (France, Italy and Spain), Germanic countries (Germany, Austria and the Netherlands), Scandinavian countries (Sweden, Denmark and Finland) and a British cluster (UK, Ireland). However, in the center of the graph, a set of universities that maintain a wider link behavior seems to emerge.

Positional analysis was carried out on this core set of 220 universities with all of their outlinks to other EU-based universities. Factor analysis was used to cluster the universities based on similarity in out-link patterns. The scree plot suggests the forcing of a 12-factor solution. The resulting clusters of universities are very much country-based factors, with the first factors: UK (F1), Fr (F2), DE (F3), SP (F4), IT (F5), NL/BE (F6) and SE/FI (F7). The remaining factors are mixed. Allowing more factors does not influence this outcome very much, although the smaller countries became more scattered over the factors. On the other hand, allowing fewer factors creates a geographically and language-based pattern, as the three factor solution resulted in a Francophone/Latin factor, a UK/Scandinavian factor and a Germanic factor. In other words, the positions of universities in the outlink network are very much nationally based. Elsewhere (Polanco 2004), similar results were obtained with network analysis of European universities as nodes. The association analysis method applied here makes use of web site links in order to produce a representation of the structure of the associations

of sites. The resulting clusters of universities mostly hold a national orientation; clusters typically contain universities from one country.

4.4 The hyperlink network of university web sites in one country

Next, we focus on the European orientation of Dutch universities and colleges in terms of their linking structure. We constructed a matrix of all outlinks from the Dutch higher educational institutes to all other European universities, and another matrix containing all inlinks from European universities to Dutch sites. Factor analysis of these matrices reveals clusters of Dutch universities with a similar European orientation as indicated by links.

Factor analyses in both the inlink and outlink dimensions resulted in a cluster of larger universities that hold very similar linking patterns, only differing in their links to small, local neighboring organizations. Smaller institutes with a more homogenous disciplinary focus, however, group together in thematic clusters; e.g., all teacher schools can be found in one cluster.

At a lower level of analysis, we concentrated on all Dutch universities and their national out-linking patterns. The network visualization of the outlink relations between Dutch universities clearly shows that large organizations with broad orientations occupy central positions in the linking network.

Figure 7 shows a core of large universities that are densely connected by outlinks. The smaller colleges and schools for professional education are located in the periphery of the map; they maintain hyperlink relationships mostly with large universities in their geographical vicinity. When smaller organizations establish hyperlinks with other smaller organizations, geographical distance and thematic focus are the determining factors.

Positional analysis confirms these results. Again, we used factor analysis in two dimensions to distinguish between the structure emerging from the inlinks and the outlinks. The results indicate two relevant dimensions in the clusters of universities: geographical proximity and intellectual focus. Especially grouped together are those organizations with a focused orientation such as teacher schools, conservatories, technical schools and art schools. Disciplinary or thematic similarity results in similar linking patterns, as with language on a higher level of aggregation.

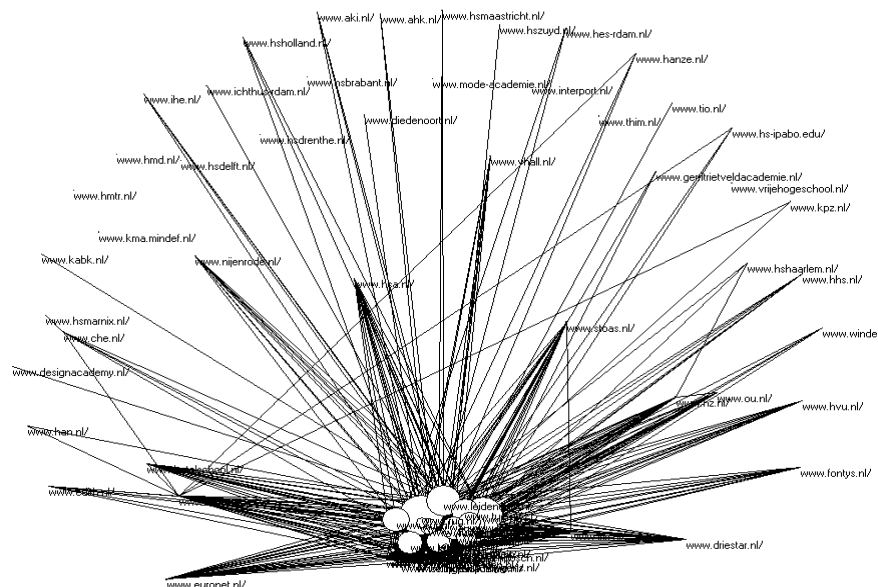


Figure 7. The outlink network between Dutch universities

In general, the analyses indicated that a number of factors seemed to determine the linking structure of national universities: the size of the

organization (the number of links), geographical distance and thematic focus. Large institutes are generally too heterogeneous in their intellectual focus to be distinguished in terms of their linking structure.

4.5 The hyperlink network of departmental disciplinary web sites

We used the outlinks of the AI departments to construct the hyperlink network between departments. The relational analyses of the hyperlink network aims to provide information about the level of “internationalization” of communications in the disciplines.

We still see some national and language orientation in the network; however, it is much less pronounced than in the universities network. The outlink network of European departments in AI shows a core subgroup that consists of departments from different countries. We therefore may conclude that the content of the research field - the intellectual structure - influences the linking behavior much more on the lower level of departments. Nevertheless, the language orientation is still visible, and therefore the network simultaneously reflects geographical and language proximity and the cognitive structure of the research field.

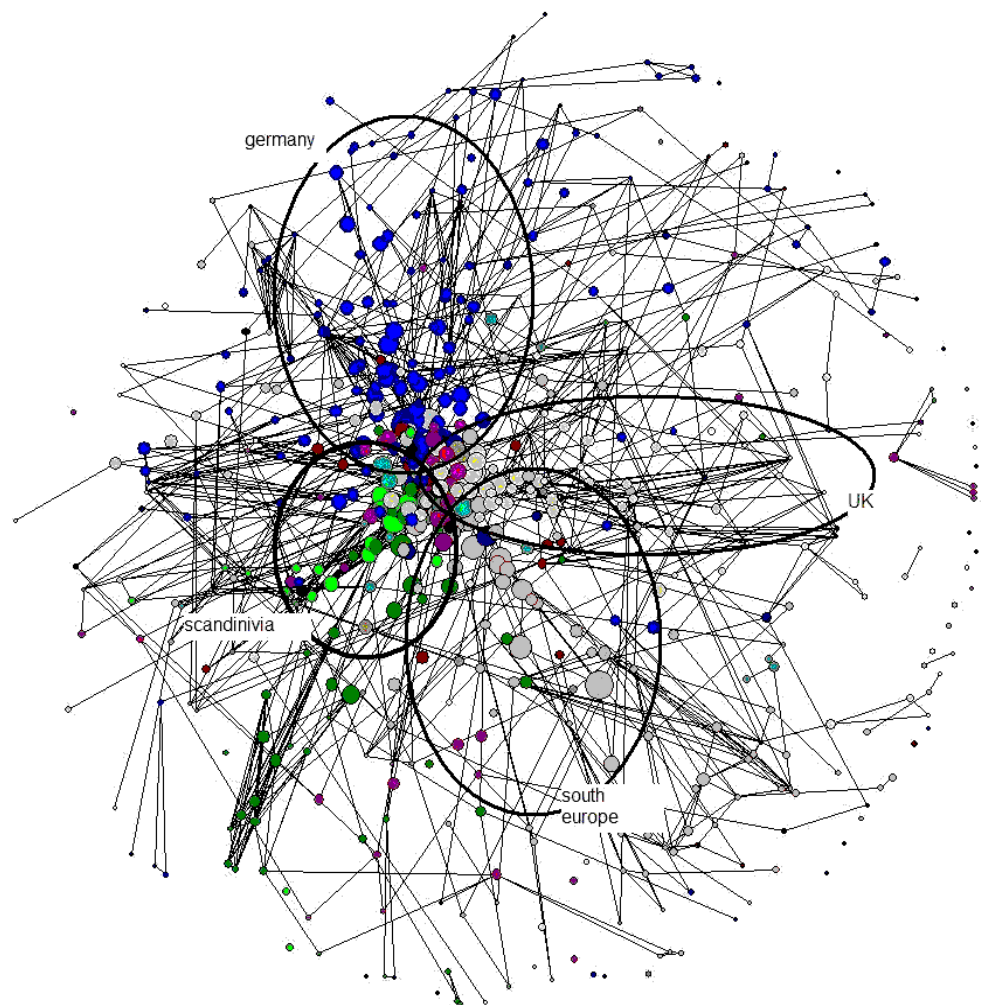


Figure 8. Network of outlinks between European departments in AI

Positional analysis brings to the fore the extent to which departments share a common set of hyperlinked “references”. This may indicate a more fine grained division in sub-fields. Factor analysis of the core set of 250 departments (the densely connected set in the center of figure 8) shows a strong common orientation in linking patterns: the first factor is dominated by UK departments, but it consists of 114 departments from all European countries. This also suggests that the departmental hyperlink network reflects more international (and therefore cognitive) structure than the hyperlink network of European universities.

4.6 The hyperlink network of departmental disciplinary web sites in one country

On the European level, in addition to cognitive orientations, the national orientation still seems to play a significant role. In this section, we used the outlinks of the departments in artificial intelligence in the Netherlands to construct the hyperlink network between departments. The relational analysis of the hyperlink network provides information about the communicational preferences within the same country.

The relational hyperlink networks show a more pronounced clustering into subgroups in which a network can be divided. Comparing the hyperlink network of Dutch departments in AI with the previous networks shows that two clusters determine the linking structure. The core of the network (circled in figure 9) contains departments from all universities. The cluster at the top of the graph consists of departments primarily from Twente University.

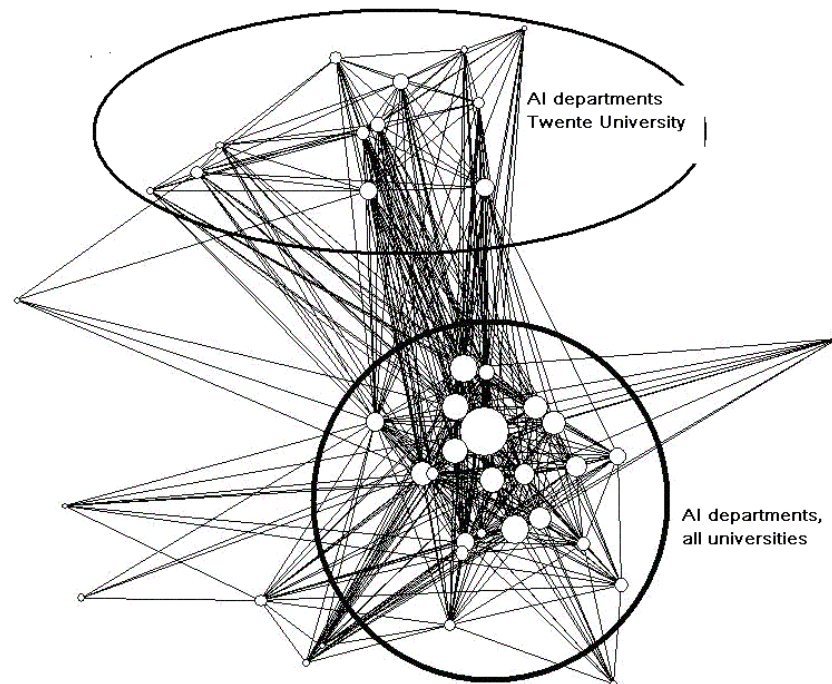


Figure 9. Network of outlinks between Dutch departments in AI

Positional analysis brings to the fore the extent to which departments share a common set of hyperlinks within the set of Dutch departments in AI. Factor analysis shows a very strong common orientation in linking patterns; the most important factor consists of 13 departments from all universities.

4.7 Hyperlink networks of individual people's homepages

In social network analysis, it has been shown that personal homepages provide a glimpse into the social structure of university communities. Not only do they reveal who knows whom, but they also provide a context, be it a shared professor, hobby, or research lab.

However, studying the "scientific dimension" of hyperlink networks in the context of the WISER project⁸, a social network analysis of German immunology researchers in one discipline, did not result in any significant linkages. The study attempted to compare collaboration between what was shown in the traditional literature and was indicated by reciprocal linking in the web. The finding was that, contrary to expectations that linkage would have been greater than in the paper world, inter linkage on the web was almost completely absent. A similar methodology was applied to the individual linking structure of an in-residence researcher in a computer science department. Factor analysis of the matrix of links between all sites linked by the researcher's site, as well as all sites that link to the individual researcher's site, failed to show any structure. The rotated factor solution did not cluster any web sites into a joint factor. In other words, it indicated that at this level of analysis, links represent the heterogeneous profile of individual interests, and cannot be used as a cognitive indicator.

5 Conclusions and discussion

Links and link patterns on the web can be used for generating indicators of web presence, web content and web impact. There are various questions to be answered in order to derive useful indicators: Which level of web links is meaningful? What do they mean? What are the meaningful indicators? In this paper we analyzed the hyperlink networks on different levels of aggregation and in different domains. The results of this analysis are summarized in table 1.

Nodes	Network	
	Structural equivalent positions:	Relations:
National academic web spheres (EU)	Language / region	- no structure -
National disciplinary networks (EU)	Language / region	- no structure -
University web sites in 15 countries (EU)	Country	Country
Department web sites in one field (EU)	Core of an international set of departments	Core set
	Some national orientation remains	Country
University web sites in one country	Region	Research field
	Research field	Size
Department web sites, one field, one country	Core group of departments from all universities	University
	Local group of departments at one technical university	Core set
Individuals	- no structure -	- no structure -

First, on the higher (country) level of aggregation, no relational structure exists: all countries link with all others. However, the various countries have different positions in the European hyperlink networks, based on geographical and language proximity. Countries in the same region and with related languages do link and are linked in similar ways.

One level lower, the hyperlink network of European universities is also highly geographical and language specific: universities link much stronger with universities in the same country (and with a similar language) than with universities in other countries (and languages). Additionally, their link patterns are similar to other universities in their home country. We find a tri-polar system with the UK, France and Germany as main nodes on the three axes. A separate Scandinavian cluster is positioned in between the German axis and the English axis. Hyperlinking has local preferences. Not only do hyperlinks show a preference for the same language and geographical proximity, but they also reflect a preference for the same country. This local preference can be found at all levels of aggregation and specificity.

In other words, at these levels the data do not show an emerging single European communication system. Obviously, the traditional forces of integration - proximity and language - are still dominant. This finding is in line with a study on the structure of the international Internet flows, which concluded that language and culture are significant determinants of hyperlinking relationships (Barnett 2004). However, a longitudinal study is needed to find out whether Europeanization increases at the level of the scientific web. Also, including non-EU universities in the analysis may enable us to find out whether European sub-networks have stronger associations with each other than, for example, with the US.

Secondly, things change at the next level of specialization: the hyperlink networks between departments in a specific research field. Analyzing the case of Artificial Intelligence, it becomes clear that language and country remain

important, but much less than on the university and country level. Although the analysis again shows the three-polar system, we now also find a large core which consists of departments from all countries. Furthermore, the positional analysis does indicate that at this level, the hyperlinking patterns are no longer characterized by geographical and language proximity. This indicates that if the research domain is restricted, the characteristics of the domain, and the position of the departments in this research domain are reflected in the hyperlink network. A common outlink orientation of a set of departments may therefore indicate a shared intellectual focus.

However, as Kling and McKim suggest, linking characteristics may differ widely across disciplines; the Internet introduced a wide variety of new means of communications that have been adopted in different ways (reflecting disciplinary specific needs) across disciplines (Kling and McKim 2000). This is also our finding (Heimeriks 2005) as we show that the use and orientation of web-based communications differs strongly between scientific disciplines. The explanation of these differences lies in the wide variety of data analysis tools, data storage, processing capacity, software tools, information delivery technologies and electronic networks. Together, these create a communicative plurality and communicative heterogeneity that increasingly reflect discipline-specific patterns and needs. In other words, ICTs introduced a wide variety of new means of knowledge production and communication; consequently, the use of electronic communications increasingly varies from discipline to discipline.

Thirdly, analyzing hyperlink networks of universities in one country supports these findings. We do find hyperlink networks between institutions in the same region, indicating the local dimension of hyperlinking. We also find similar linking patterns for types of institutions. The smaller institutions within a similar field (e.g., art schools, or teacher schools) do have comparable hyperlink networks. The more specific the focus of the institution (so the more an institution is like a department), the more pronounced the clustering. In other words, the positions of similar institutes are also similar in the web link structure. The large universities also have similar positions, but with unspecific networks.

Analyzing the departments in one field (AI) in a single country, we find a general cluster of AI groups in all universities, and one local network of AI groups in one technical university. Also here, the main sub-network represents the research field, but a smaller geographically oriented sub-network also exists.

Finally, the hyperlink network of an individual researcher shows personal interests at a micro level. However, research into the "scientific dimension" of personal hyperlink networks does not result in any significant patterns.

In conclusion, the difference between the university hyperlink networks and the department hyperlink networks is interesting. It suggests that the university hyperlink network is the sum of lower level networks of the departments within the university. As different fields exhibit very different hyperlink patterns, the university as an aggregate of heterogeneous research fields is not a useful unit of analysis for mapping cognitive structures. Our comparison shows that departmental web sites are most suited for mapping processes of knowledge production. At the level of the research group, analyzing the linking architectures may provide us with detailed information about the context of knowledge production, the geographical distribution of the nodes, the patterns of collaboration and the types of actors involved (including funding organizations, publishers, users of knowledge, etc).

In order to gain more insight in the role of web-based communications, it would be useful to specify the position of the hyperlink network on the web with respect to other communication networks in the "real world". This type of analysis endeavors to provide insight into the relationship of web indicators with traditional indicators for networks of scientific communications (Heimeriks & van den Besselaar, submitted). Does the presence of the Internet and the World Wide Web inform us about inter-organizational networks "in real life", or are the real and virtual communication networks relatively loosely coupled? From the analyses discussed in this paper, it becomes clear that a proper delineation is of utmost importance for conducting hyperlink based analysis. The research presented here indicates that linking structures have a predominantly local orientation. Even the

patterns that emerge on the country level are mostly based on linguistic and geographical vicinity. Although locally oriented, the research department is the most appropriate unit of analysis to map the cognitive dimensions of scientific hyperlink communications. Not only does the research department represent the smallest organizational unit that participates in institutional linkages, but it also provides the most unambiguous unit of analysis in different (online and offline) networks of knowledge production. This comparison of networks in various media is needed to further improve our understanding of the meaning of hyperlink based indicators.

Acknowledgements

The research underlying this paper has been partially funded by the European Commission, grant IST-1999-20350, the EICSTES project. The project aims at developing of Web-based indicators for knowledge production, communication, and use. The work has partly been funded by the EC, grant IST-1999-20350. Partners are CINDOC (Spain), ARC Seibersdorf (Austria), DTI (Denmark), INIST-CNRS (France), NIWI-KNAW (Netherlands), and University of Surrey (UK). The authors would like to thank Viv Cothey, Arie Rip, Eleftheria Vasileiadou and Paul Wouters for useful comments and suggestions on earlier drafts of this paper.

Notes

1. See www.eicstes.org
2. Situation before the entrance of the 10 new member states in 2004.
3. For more information on the process of data gathering see Arroyo et al 2003, available at:
http://www.eicstes.org/EICSTES_PDF/Deliverables/Web%20data%20description.PDF
4. This is a shareware version that is included by the manufacturer in the Back Office pack and was developed from a former software called Webmapper.
5. Developed at the Department of Technology Management of the Austrian Research Center Seibersdorf.
6. The small sets of factors will be shown; however in some cases the visualizations are not shown when the number of nodes and factors become too large.
7. In this factor analysis the number of links between different departments is also included.
8. Personal communication with Viv Cothey.

References

- Adamic, L. A. and B. A. Huberman (2000). "Power-law distribution of the World-Wide Web." **Science** 287: 2115.
- Arroyo, N., V. M. Pareja, et al. (2003). D3.2: Description of Web data in D3.1. Madrid, CINDOC.
- Barjak, F. (2004). On the integration of the Internet into informal science communication. Solothurn, University of Applied Sciences Northwestern Switzerland.
- Barnett, G. A. (2004). "The Structure of International Internet Flows." unpublished.
- Brin, S. and L. Page (1998). "The anatomy of a large scale hypertextual web search engine." **Computer Networks and ISDN Systems** 30(1-7): 107-17.

Burt, R. S. (1982). **Towards a structural theory of action**. New York, Academic Press.

Chu, H. S. and M. Thelwall (2002). "Library and information science schools in Canada and USA: A Webometric perspective." **Journal of Education for Library and Information Science** 43: 110-125.

Cronin, B. (2001). "Bibliometrics and beyond: some thoughts on Web-based citation analysis." **Journal of Information Science** 27(1): 1-7.

Cronin, B., H. W. Snyder, et al. (1998). "Invoked on the Web." **Journal of the American Society for Information Science** 49(14): 1319-28.

Schneider, S., & Foot, K. (2002). Online structure for political action: Exploring Presidential campaign Web sites from the 2000 American election. **Javnost - The Public**, 9(2): 43-60.

Heimeriks, G., (2005) Knowledge Production and Communication in the Information Society. Mapping communications in heterogeneous research networks. Unpublished PhD thesis University of Amsterdam

Heimeriks, G. J. and P. Van den Besselaar (submitted). New media and communication networks in knowledge production - a case study.

Ingwersen, P. (1998). "The calculation of Web impact factors." **Journal of Documentation** 54: 236-243.

Kim, H. J. (2000). "Motivations for hyperlinking in scholarly electronic articles: a qualitative study." **Journal of the American Society for Information Science** 51(10): 887-99.

Kleinberg, J. (1999). "Authoritative sources in a hyperlinked environment." **Journal of the Association for Computing Machinery** 46: 604-632.

Kling, R. and G. McKim (2000). "Not Just a Matter of Time: Field Differences and the Shaping of Electronic Media in Supporting Scientific Communication." **Journal of the American Society for Information Science** 51(14): 1306-1320.

Kopcsa, A., Schiebel E. (2000). "Science and technology mapping: A new iteration model for representing multidimensional relationships." **Journal of the American Society for Information Science** 49: 7-17.

Li, X., M Thelwall, P Musgrove and D Wilkinson (2003). "The relationship between the links/Web impact factors of computer science departments in UK and their RAE (Research Assessment Exercise) ranking in 2001." **Scientometrics** 57(2): 239-255.

Middleton, I., McConnell, M., Davidson, G. (1999). "Presenting a model for the structure and content of a university World Wide Web site." **Journal of Information Science** 25(3): 219-27.

Park, H. W. and M. Thelwall (2003). "Hyperlink Analyses of the World Wide Web: A Review." **Journal of Computer-Mediated Communication** 8(4). <http://jcmc.indiana.edu/vol8/issue4/park.html>

Polanco, X. M. A. B., Dominique Besagni and Ivana Roche. Clustering and Mapping European University Web Sites Sample for Displaying Associations and Visualizing Networks. 2004.

Tang, R. and M. Thelwall (2003). "U.S. academic departmental Web-site interlinking in the United States Disciplinary differences." **Library & Information Science Research** 25(4): 437-458.

Thelwall, M. (2001). "Extracting macroscopic information from Web links." **Journal of the American Society for Information Science and Technology** 52: 1157-1168.

Thelwall, M. (2002). "In praise of Google: Finding law journal Web sites." **Online Information Review** 26: 271-272.

Thelwall, M. (2002). "A research and institutional size-based model for national university Web site interlinking." **Journal of Documentation** 58(6): 683-694.

Thelwall, M. (2003). "A layered approach for investigating the topological structure of communities in the web." **Journal of Documentation** 59(4): 410-429.

Thelwall, M., R. Tang, et al. (2003). "Linguistic patterns of academic Web use in Western Europe." **Scientometrics**, 56(3): 417-432.

Wilkinson, D., G. Harries, M. Thelwall and E. Price (2003). "Motivations for academic Web site interlinking: Evidence for the Web as a novel source of information on informal scholarly communication." **Journal of Information Science**, 29: 59-66.

Zang, Y. (2001). "Scholarly use of Internet-based electronic resources." **Journal of the American Society for Information Science and Technology**, 52(8): 628-54.

Received 15/Dec/2005

Accepted 31/Jan/2006

DISCUSSION

What do hyperlinks mean: the value of hyperlink-networks as indicators of knowledge production

Viv Cothey