Germany in the European academic web space

José Luis Ortega[1] Isidro F. Aguillo[1]

22 May 2008

Abstract

This communication tries to analyse and visualize the German academic web space. It intends to show the structural characteristic of the network, showing groups and detecting the highest degree universities. Its relationships with the remaining European countries are also described. A network graph was built from the link relationship among 110 German universities. This data were extracted from Yahoo! Search and modelled through Pajek 1.02. Results show that the German network rests on a dense core of 46 universities, where the Berliner universities have a central position. Several geographical and thematic groups have been also detected. We conclude that the use of social network techniques allow us to uncover the structural dimension of an academic web space.

1 Introduction

Science is developing new ways of communication their results. World Wide Web has been the most extensive and recent tool to improve this process. E-journals, scientific repositories, web databases are examples of a new scientific production and a new way to strike up relationships with other scientific agents. Web indicators have arisen with the purpose of measuring this scientific activity on the Web and uncovering new models which allow us to assess and valuate scientific production through web based indicators. (Scharnhorst and Wouters, 2006)

Recently, structural indicators based in network analysis are being proposed in order to extract the position and environment of a journal (Leydesdorff, 2007) or an institution in a defined scientific network (Scimago, 2007). In the Web, these indicators have showed the structural properties such as Small World phenomenon (Björneborn, 2001; 2003) or the shape of the European Academic web space (Ortega, Aguillo, Cothey and Scharnhorst, 2008).

Several studies have been developed in order to describe the local academic web space of Spain (Thelwall and Aguillo, 2004; Ortega and Aguillo, 2007), Canada (Vaughan, 2006), United Kingdom (Thelwall, 2002; Thelwall and Price, 2003) or the Nordic countries (Almind and Ingwersen, 1997; Ortega and Aguillo, 2008). However, Germany is one of the most important countries in the European Higher Education web space. It has a central position in the European network university system, as well as it contributes with the largest number of contents and links (Ortega, Aguillo, Cothey & Scharnhorst, 2008). So, the main position of the German network encourages us to study in depth the structure and relationships of this one.

2 Objectives

We intend to visualize the local environment of the German academic web space and its relationship with the European countries which belong to the EU 15. Through social network analysis techniques we try to explain how is structured the German network, which are the

[1]Cybermetrics Lab. (CCHS-CSIC), Joaquín Costa, 22. 28002 Madrid {jortega, isidro}@cindoc.csic.es

H. Kretschmer & F. Havemann (Eds.): Proceedings of WIS 2008, Berlin
Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting
Humboldt-Universität zu Berlin, Institute for Library and Information Science (IBI)

main universities and if we be able to identify cluster by regions or by matters.

3 Methods

The analysis of the German academic web space is within the framework of the analysis of the European one (Ortega, Aguillo, Cothey and Scharnhorst, 2008). 535 universities of the 14 European countries (EU except Luxembourg) in 2004 were selected from *Webometrics Ranking of World Universities*[2]. This site ranks 3,000 universities according two main criteria: size (number of pages and rich files) and visibility (number of incoming links). This set of European universities was mapped according to the link relationships among them. From the first 1,000 universities ranked in this web site, we extracted the European ones, obtaining 535 universities web domains. *Yahoo! Search*[3] was used to obtain the outcoming links data of each university domain (Aguillo, Granadino, Ortega, & Prieto, 2006). It was used because allows to combine several search operators. The following query was used to extract the links in August of 2005:

+site:{university domainA}
+linkdomain:{university domainB}

From this set of data we extracted the nodes and the links between German universities (110) and the remaining European university domains.

A graph was built through the link matrix retrieved from the search engine to illustrate the topology of the network and its connectivity degree. We have used Pajek 1.02 to visualise the network. We have selected a cut-off of more than 50 links to improve the network visualization. We have also used the Fruchterman-Reingold algorithm to lay out the network because it is the fastest algorithm energizing large networks (Nooy, Mrvar and Batagelj, 2005).

Several variables were added to the graph in order to improve the visualization and to visually show the relationships between different variables. The arc size shows the number of links that point from/to a university web domain. The node size represents the amount of web pages that each university makes available on the web and are retrieved through search engines. The node colour displays the federal state (*Bundesländ*) of each university. This geographical classification was developed through the NUTS (1 level) codes (Eurostat, 2005). And finally the shape of each node informs about the thematic area in which each university is specialized. Hence, circles are general universities; boxes are universities of applied sciences and technical universities; diamonds are biomedicine and agricultural universities; and the triangles are social sciences and humanities universities, including business schools and arts & media schools.

During the data extraction process, we have noticed that there are universities that have some problems with their web domains. For example, the Technical University of Munich owns three different domains (mytum.de, tu-muenchen.de, tum.de). Other similar example is the University of Osnabrück which uses two domains (uos.de and uni-osnabrueck.de). This causes data scattering and a low representation of these universities.

Several Social Network Analysis (SNA) measures were used to analyze the resulting graphs.

Degree: measure the number of lines incident with a node. This can be normalized (nDegree) by the total number of nodes in the network. In a directed network such as the Web we can count only the incoming links (InDegree) or the outcoming links (OutDegree). In Webometrics, this measure allows us to detect the visibililty of a web domain (Cothey, 2005; Kretschmer & Kretschmer, 2006).

Betweenness: measures the intermediation degree of a node to keep the network connected, that is to say, the capacity of one node to connect only those nodes that are not directly connected to each other. Its normalization is the percentage over the total number of nodes in the network. From a webometric point of view, this measure allows us to detect hubs or gateways that connected different web networks (Faba-Pérez, Zapico-Alonso, Guerrero-Bote & Moya-Anegón, 2005).

K-Cores: is a sub-network in which each node has *k* degree in that sub-network. K-Cores allow us to detect groups with a strong link density. In free-scale networks, i.e. the Web, the core with the highest degree is the central core of the net-

[2] http://www.webometrics.info
[3] http://search.yahoo.com

work, detecting the set of nodes where the network rests on (Seidman, 1983).

4 Results

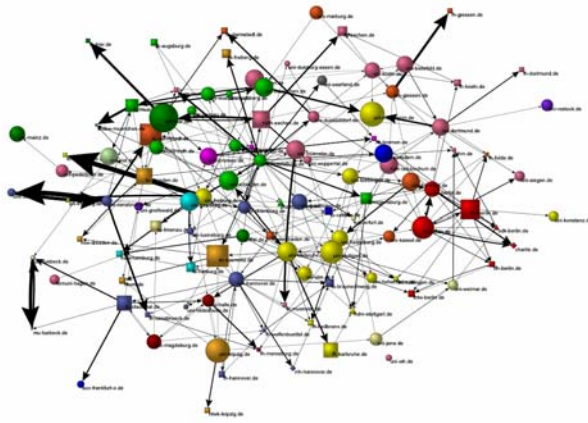

Figure 1. German network of university web domains (110 nodes; arcs weight≥ 50).

The German academic network (Figure 1) shows two significant patterns:
1. There is a geographical relationship: picture shows several geographic groups set by colours. Thus, we can detect a Berliner group in red mainly shaped by Humboldt (hu-berlin.de), Applied Sciences (tu-berlin.de) and Free (fu-berlin.de) universities, being 8 universities. We also can appreciate a Bavarian group in green light centred by Regensburg (uni-regensburg.de) and Erlangen (uni-erlangen.de) universities. The Rhenanian cluster in red light shows a less cohesive set, connecting other regional groups. In this group we can highlight the RWTH Aachen (rwth-aachen.de) and Münster (uni-muenster.de) universities. Finally, a Badenian clique in yellow grouped around the Stuttgart (uni-stuttgart.de) and Karlsruhe (uni-karlsruhe.de) universities.
2. There are thematic relationships: graph shows groups of universities of applied sciences (*Fachhochschulen*) and technical universities (*Technische Universitäten*). The axis Chemnitz (tu-chemnitz.de) and Clausthal (tu-clausthal.de) technical universities, shows these relationships across different regions.
It is surprising that the University of Trier (uni-trier.de) is the largest German university in number of web pages. This is due to that it houses the Digital Bibliography & Library Project (dblp.uni-trier.de) one of the most important international database on Computing and Information Sciences. Contrarily, we observe the under presentation of the Technical University of Munich, which have a remote position and their structural indicators are not comparable with other principal German universities. Maybe, this is caused by the use of different web domains, as we comment before.
According to the centrality measures (indegree, outdegree and betweenness), we can appreciate (Tables 1, 2 and 3) that the universities (*Universitäten*) and technical universities (*Technische Universitäten*) are the main educational institutions in the German academic web space. Also, the Berliner universities are located in the first positions in the centrality measures rank.

Table 1. First five normalized indegree rank of the German universities web domains.

web domain	*InDegree*	*NInDegree*
uni-karlsruhe.de	65	59.633
hu-berlin.de	61	55.963
fu-berlin.de	61	55.963
uni-regensburg.de	60	55.046
tu-berlin.de	60	55.046

Table 1 shows the five universities with largest centralization InDegree. The Karlsruhe University is the university with highest indegree (59.63), followed by Humboldt (55.96) and Free (55.96) universities of Berlin. Notice that the Berliner universities are the most linked in the German network.

Table 2. First five outdegree rank of the German universities web domains.

web domain	*OutDegree*	*NOutDegree*
tu-berlin.de	63	57.798
hu-berlin.de	60	55.046
fu-berlin.de	60	55.046
uni-hamburg.de	59	54.128
uni-karlsruhe.de	59	54.128

According to the centralization OutDegree, the Berliner ones are also the universities with highest scores. We can indeed stand out the Technical University of Berlin (57.80) and the Humboldt (55.05) and the Free (55.05) universities of

Berlin as well. As the InDegree values, the Berliner universities have again the highest scores.

Table 3. First five betweenness rank of the German universities web domains.

web domain	*Betweenness*	*nBetweenness*
uni-hannover.de	1033.765	8.782
uni-karlsruhe.de	690.299	5.864
tu-berlin.de	516.952	4.391
fu-berlin.de	508.26	4.318
hu-berlin.de	491.707	4.177

Table 3 shows the five universities with largest Betweenness and normalized Betweenness degree. This indicator allows us to detect gateway web universities that mediate between separated web clusters (Ortega and Aguillo, 2008). Although the Berliner universities are still in the first positions, the Hannover University emerges with the largest nBetweenness degree (8.78). This can be observed in the Figure 1, where the Hannover University mediates between several universities from different regions.

We have used K-cores technique to detect the centre of the German network, due to show the group of nodes with a largest degree between themselves. K-cores show that there is a set of 46 universities with a degree of 36 between them. This allows us to identify the core of the German network, which is dominated exclusively by the most highlight universities (*Universitäten)* such as the Berliner universities, Hamburg University or Karlsruhe University.

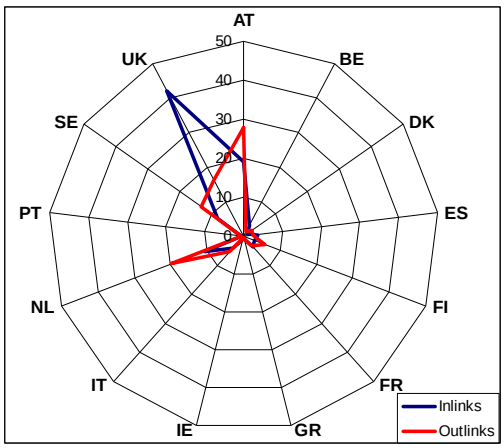

Figure 2. Percentage of inlinks and outlinks per each European country.

Figure 2 shows a spider graph where we can see the distribution of inlinks and outlink percentage from/to the remaining EU countries. Firstly, we appreciate that the outlinks and inlinks follow different pattern. Outlinks are addressed mainly to Austria (27.92%) and The Netherlands (19.7%), while the Inlinks come from United Kingdom (42.01%) and Austria (19.05%). Hence, Austria is the principal country related with the German web network.. This allows us to confirm the cultural and geographical relationships that there are in the European academic web (Ortega, Aguillo, Cothey and Scharnhorst, 2008).

5 Conclusions

The use of social network analysis techniques allows us to analyse and describe the structural features of complex web networks. These have showed that German network is quite dense and compact, because it has a dense core set up by 46 universities with a 36 degree. It rests on a large group of highlighted universities, where stand out the Berliner ones such as Humboldt and Technical universities. The network visualization has made possible to see geographical and thematic relationships between them, detecting the regional clusters (Berliner and Badenian sub-networks). The visualization of these type of networks allow us appreciated multiple characteristic of the academic web space of a country, related to other indicators such as web pages indexed.

References

Aguillo, I. F., Granadino, B., Ortega, J. L., & J. A. Prieto (2006). Scientific research activity and communication measured with cybermetrics indicators. *Journal of the American Society for Information Science and Technology* 57 (10), 1296–1302.

Almind, T. C. and P. Ingwersen (1997). Informetric Analyses on the World Wide Web: Methodological Approaches to 'webometrics'. *Journal of Documentation* 53 (4), 404–426.

Björneborn, L. (2001). Small-world linkage and co-linkage. In *Proceedings of the 12th ACM Conference on Hypertext and Hy-*

permedia, New York, USA, pp. 133–137. ACM Press.

Björneborn, L. (2003). Small-World Link Structures across an Academic Web Space: A Library and Information Science Approach, Copenhagen: Royal School of Library and Information Science.

Cothey, V. (2005), Some preliminary results from a link-crawl of the European Union Research Area Web. In *Proceeding of the 10th International Conference of the International Society for Scientometrics and Informetrics*, Karolinska University Press, Stockholm.

Eurostat (2005) Nomenclature of territorial units for statistics - NUTS Statistical Regions of Europe. Luxembourg: European Union http://ec.europa.eu/comm/eurostat/ramon/nuts/home_regions_en.html

Faba-Pérez, C., Zapico-Alonso, F., Guerrero-Bote, V. P. & F. de Moya-Anegón (2005). Comparative analysis of webometric measurements in thematic environments. *Journal of the American Society for Information Science and Technology* 56 (8), 779–785

Kretschmer, H. and T. Kretschmer (2006). A New Centrality Measure for Social Network Analysis Applicable to Bibliometric and Webometric Data. In *International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting*

Leydesdorff, L. (2007). "Betweenness Centrality" as an Indicator of the "Interdisciplinarity" of Scientific Journals. *Journal of the American Society for Information Science and Technology* 58 (9), 1303–1319.

Nooy, W. d., Mrvar, A., & V. Batagelj (2005). Exploratory social network analysis with pajek. Cambridge, UK: Cambridge University Press.

Ortega, J. L. and I. F. Aguillo (2007). La Web académica española en el contexto del Espacio Europeo de Educación Superior: Estudio exploratorio. *El profesional de la información* 16 (5), 417–425.

Ortega, J. L. and I. F. Aguillo (2008). Visualization of the Nordic academic web: Link analysis using social network tools. *Information Processing & Management* (in press)

Ortega, J. L., Aguillo, I. F., Cothey, V., & A. Scharnhorst (2008). Maps of the academic web in the European higher education area – an exploration of visual web indicators. *Scientometrics* 74 (2), 295–308

Scharnhorst, A. and P. Wouters (2006). Web indicators – a new generation of S&T indicators? *Cybermetrics* 10 (1), paper 6. http://www.cindoc.csic.es/cybermetrics/articles/v10i1p6.html

Scimago group (2007). Scimago Journal & Country Rank. http://www.scimagojr.com/

Seidman, S. B. (1983). Network structure and minimum degree. *Social Networks* 5, 269–287.

Thelwall, M. (2002). The Top 100 Linked-to Pages on Uk University Web Sites: High Inlink Counts Are Not Usually Associated With Quality Scholarly Content. *Journal of Information Science* 28 (6), 483–491.

Thelwall, M. and I. F. Aguillo (2003). La salud de las Web universitarias españolas. *Revista Española de Documentación Científica* 26 (3), 291–305.

Thelwall, M. and L. Price (2003). Disciplinary Differences in Academic Web Presence - a Statistical Study of the UK. *Libri* 53 (4), 242–253.

Vaughan, L. (2006). Visualizing linguistic and cultural differences using Web co-link data. *Journal of the American Society for Information Science and Technology* 57 (9), 1178–1193.