International Journal of Scientometrics, Informetrics and Bibliometrics
ISSN 1137-5019

VOLUME 11 (2007): ISSUE 1. PAPER 4

Interdisciplinary relationships in the Spanish academic web space: A Webometric study through networks visualization

José Luis Ortega, Isidro F. Aguillo

Cybermetrics Lab
CCSH - CSIC
Madrid, Spain
email: **jortega@cindoc.csic.es**
email: **isidro@cindoc.csic.es**

Abstract

The aim of this work is to describe the interdisplinary research relationships among several Spanish university departments and research groups located in the Spanish web space. 699 web sites from 2390 were selected according to whether a web site receives or links to at least one of the rest. The links between them were extracted with a commercial crawler in 2004 and then analysed to built a complex directed network of in-links and out-links. The results show that the Spanish academic web space is weakly interconnected both at the level of groups and departments, and that the relationships between disciplines can be appreciated through network graphs. The use of network graphs is a suitable technique to show the transversal relationships among disciplines and to detect incipient research fronts in the web space. The web presence of Experimental and Technological Sciences is higher than Social Sciences and Humanities as well.

Keywords

Webometrics, World Wide Web, Network Analyses, interdisciplinary research, Spanish universities

1. Introduction

Scientific interdisciplinarity may be understood as the converge of two o more disciplines in a same field to investigate the same phenomenon sharing their methodologies and theoretical background (OECD 1972). This fact implies scientific dynamics that connect different and distant disciplines. The relationships between different research disciplines allow the knowledge's growth and the emergence of new disciplines (Dogan & Pahre 1990). For this reason the scientific interdisciplinarity is considered as a necessary process in the development of a research system (EC 2000).

Several studies have focused on identifying interdisciplinary relationships between scientific stakeholders. From a scientometric point of view, Leydesdorff (2007) used the betweenness centrality measure as an indicator to study the interdisplinarity in 7,379 scientific journals, finding that this is a suitable measure in local citation environments. In a similar way, Rafols and Meyers (2007) used Social Networks Analysis (SNA) centrality measures to analyse the interdisciplinarity in bionanoscience articles, and showed that these measures are more suitable to detect cross-disciplinary relationships.

The interdisciplinary relationships can be modelled as a network in which different disciplines are connected themselves to face new research lines. The World Wide Web appear as a good medium to detect these relationships because it is a complex network of links which connect web sites from separate disciplines. Björneborn (2004; 2006) introduced the small world theory to the Web research to describe the cross-topic connectors across the UK academic Web space. However, the size in number of pages and the numerous relationships generated by the links make difficult to analyse this environment through classical statistical tools. The Visualization of Information (Chen, 2003) is an emergent discipline whose purpose is to present conceptual entities and their relationships in a visual way to obtain an in-depth knowledge of these entities.

As the Web is a logical network of nodes and links, SNA has been a technique used to study its topological characteristics and to represent it visually. Several studies have focused on these characteristics. Albert, Jeong and Barabasi (1999) estimated the diameter of the Web, i.e. number of links to cover whole web. These same authors (Barabasi, Albert, & Jeong, 2000) discovered that the Web shown scale-free network properties. Thus, from this characteristic, Broder et al. (2000) defined the bow-tie model to describe the linkage features of web regions. This model is mainly set up by four components: the SCC or Strong Connected Component, this is the zone where all the nodes are connected among themselves; IN component, the zone where the nodes link to SCC but they are not reached from SCC; OUT component, the nodes are linked from SCC but they do not link to SCC; and TENDRILS, nodes that link to other nodes outside the sample. Several authors have studied the national

networks from this point of view, such as Thelwall and Wilkinson (2003) with United Kingdom, Australia and New Zealand; Baeza-Yates et al. (Baeza-Yates, Castillo, & López, 2005; Baeza-Yates & Castillo, 2007) with Chile and Spain. Bjorneborn (2004) focused in the UK academic web sub-sites to develop a variation of the bow-tie model, the *corona* model, where he detected nodes that connect directly from IN zone to OUT zone.

2. Objectives

This study intends to detect and describe interdisciplinary relationships on the Web. The hyperlink relationships among 699 web sites of departments and research groups from Spanish universities are used to show the cross-boundary connections between disciplines on the Web. How these relationships are presented and what research fronts arise from different disciplines are the aims of this paper. We also plan to see the suitability of networks visualization to represent the distribution of subjects in academic web sites and to analyse the interdisciplinary relationships among web sites. Finally, we want to describe the global subject distribution in the Spanish academic web space.

3. Methods

For A set of 2390 academic websites, which represent the institutional sites of departments and research groups of 52 Spanish universities were studied (EICSTES, 2006). These websites were described and indexed institutionally, geographically and thematically. Unesco Nomenclature (Unesco, 1988) was used to assign topics to each department and research group website. Next, the links between these websites and their web pages were extracted with a commercial crawler, Microsoft Site Explorer (Arroyo, 2004), during 2004 and then prepared to build a complex directed network of in-links and out-links. From this set, 699 sites were selected if they link to or receive at least one link from the rest of them. This network was analysed according to the web graph theory. Pajek 0.97 networks software was used to analyse the network and NetDraw 2.28 was used to show the results according to several characteristics and variables. The colour of the nodes express the subject according to the Unesco Nomenclature, the arcs width shows the frequency of links and the size the number of pages.

4. Results

We wanted to know how the disciplines are distributed along the network of departments and research group websites. We use the bow-tie model to see the weight of each discipline in each zone of the model. We have grouped these disciplines in five thematic areas to detect clearly the different distributions. Table 1 shows these five new categories and their Unesco equivalents. Notice that Computing is not a Unesco discipline. It has to be created because this discipline contains a high number of pages that are spread out as a category of Computer Science inside Mathematics, and Computer Technology inside Technological Sciences.

Thematic Areas	*Disciplines*
Experimental Sciences	Mathematics, Astronomy, Physics, Chemistry, Earth and Space Science
Technological Sciences	Technological Sciences, Computing
Social Sciences	Demography, Economic Sciences, Geography, Juridical Sciences and Law, Linguistics, Pedagogy, Political Science, Sociology
Life Sciences	Life Sciences, Medical Sciences, Agricultural Sciences
Humanities	Anthropology, History, Arts and Letters, Philosophy

Table 1. Thematic areas and their respective disciplines

4. 1. Bow-tie model

Table 2 contains the number and percentage of nodes in each zone group by thematic areas. In general, Table 2 shows that the Spanish academic web space, at the level of departments and research group websites, is not strongly interconnected. According to the Bow-tie model, 699 (29,24%) web sites are connected among themselves, so the size of the ISLAND component (70,75%) is close to the Spanish web (Baeza, Castillo & López, 2005). The percentage of nodes in the SCC (Strong Connected Component) is quite small (18.74%) and the sum of the percentages of the IN (22.03%) and OUT (25.46%) components is quite large, just like Tendrils (25.89%). The size of OUT and Tendrils lets us to appreciate that there are a lot of links which point out of the sample. So this shows that the sample is barely connected. It is possible that a lot of websites are enclosed within other larger web spaces which covers websites from similar disciplines or "communities" (Flake, Lawrence & Giles, 2000; Flake, Lawrence, Giles & Coetzee 2002), or they are more connected with regional or local institutions websites through gateways (Ortega, Aguillo, Cothey & Scharnhorst, 2008; Thelwall, 2002). Disconnected components (7.87%) are websites linked with other ones but disconnected with the SCC component.

	SCC	*IN*	*OUT*	*Tendrils+Tubes*	*Disconn.*	**Total**
Experimental Sc.	56	54	65	91	26	292
%	*8,01*	*7,73*	*9,30*	*13,02*	*3,72*	*38,05*
Technological Sc.	35	40	18	59	9	161
%	*5,01*	*5,72*	*2,58*	*8,44*	*1,29*	*21,75*
Social Sc.	24	35	39	30	6	134
%	*3,43*	*5,01*	*5,58*	*4,29*	*0,86*	*18,31*
Life Sc.	11	16	36	1	7	71
%	*1,57*	*2,29*	*5,15*	*0,14*	*1,00*	*9,16*

Humanities	5	9	20	0	7	41
%	*0,72*	*1,29*	*2,86*	*0,00*	*1,00*	*4,86*
Total	131	154	178	181	55	699
%	*18,74*	*22,03*	*25,46*	*25,89*	*7,87*	*100*

Table 2. Distribution by thematic areas of websites along the bow-tie zones

At the level of thematic areas, Table 2 shows that Experimental Sciences (38.05%), Technological Sciences (21.75%) and Social Sciences (18.31%) are the main areas in the sample of networked websites. Although the distribution of thematic areas along the bow-tie zones is quite homogenous, we can highlight that Experimental Sciences (8.01%) and Technological Sciences (5.01%) are the predominant in SCC zone, while Social Sciences, Life Sciences and Humanities are mainly spread over OUT and IN zones. Hence, we can notice that the Experimental and Technological Sciences are the main core of the sample of the Spanish department and research group websites, while the other areas are more peripheral.

4.2. Interdisciplinary Networks

Three sample network maps (Economics, Biology and Chemistry) have been chosen to show the interdisciplinary relationships and the main research fronts in the Web for each discipline. These have been selected because they show a representative sample of these relationships in the Web.

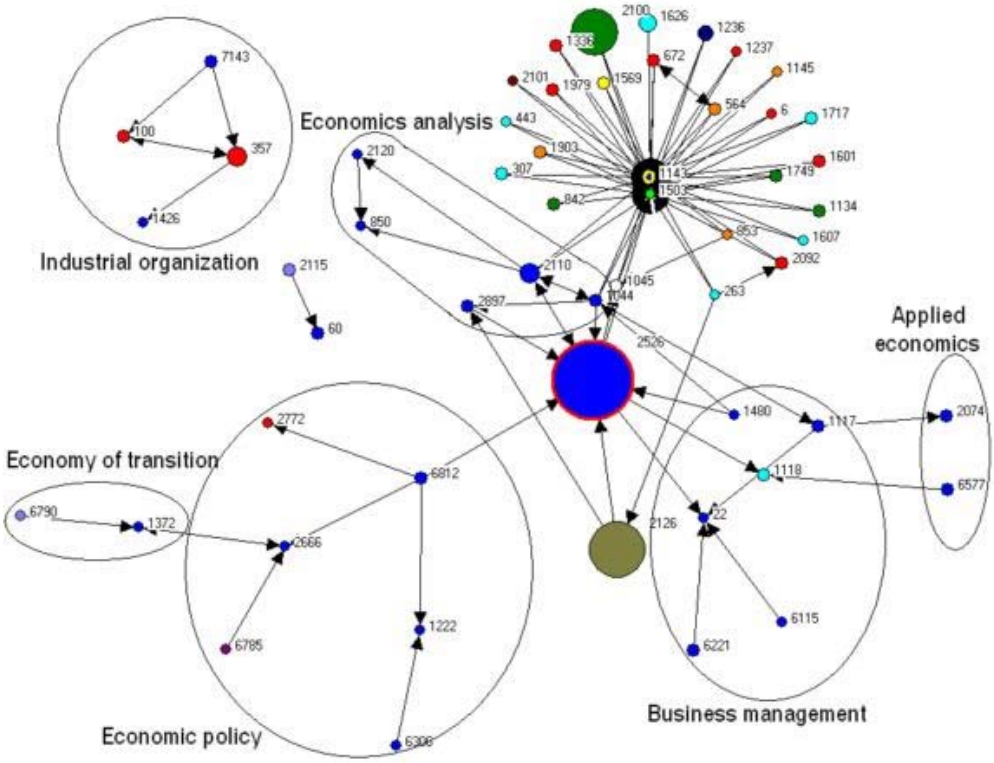

Figure 1. Interdisciplinary network graph of Economics

Figure 1 shows the network graph of Economics (in blue). We can see different research or academic lines such as Industrial Organization, Business Management or Economics Analysis. We can also see the presence of nodes of Mathematics or Physics (red, olive green) that express statistical or econometric relationships. There are relationships between similar fronts such as Economic Policy with Economy of Transition or Business Management and Applied Economics. We can highlight the centrality degree of the Department of Economics and Business of the Universidad Pompeu Fabra (2526) (red crown) which organises and connects the different lines of the Economic web network in Spain. We can also note the node 1143, Universidad de Murcia's Department of Economics (yellow crown), which is strongly linked from several nodes and different disciplines, mainly from mathematic units. This suggests that this department is oriented to the econometric activities.

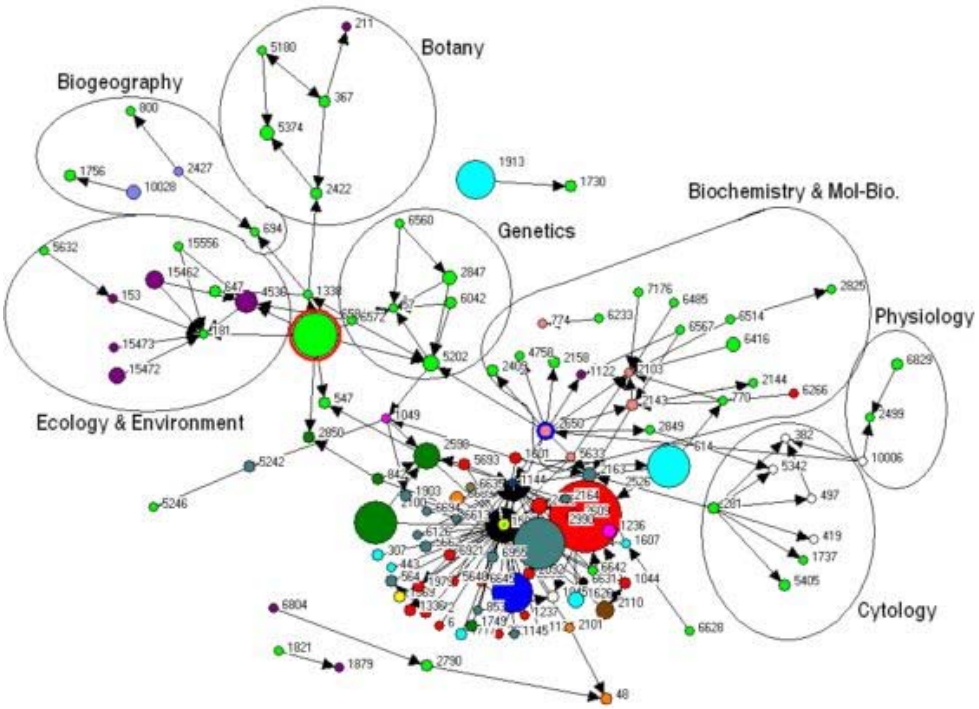

Figure 2. Interdisciplinary network of Biology

The next Figure 2 shows the interdisciplinary network of Biology (in green) departments and research groups. Several units are identified and their relationships with other groups of others disciplines. For instance, the Ecology and Environment cluster is linked with the Earth Science websites (purple), the Cytology and Physiology groups are linked with the Medicine research groups (white) and the Biochemistry and Molecular Biology set is connected with the Chemistry sites (light red). However, the Genetics group does not make relationship with other disciplines. This graph also shows the main nodes in the network according to the centrality degree. The Department of Biology of the Universitat de les Illes Balears (6584) (red crown) is the central node in the sub-network of the Natural Biology, acting as mediator between Botany, Biogeography, Ecology and Genetics. However, the Department of Biochemistry and Molecular Biology of the Universitat de Barcelona (2650) (blue crown) is the website that organise the Microbiology sub-network, with the sets of Biochemistry, Cytology and Physology. As in the Figure 3, we can see the Unspecific Immunitary System of Teleostei Research Group in Universidad de Murcia (1503) (yellow crown) strongly connected to different websites because it works as an information reference resource in the biology field.

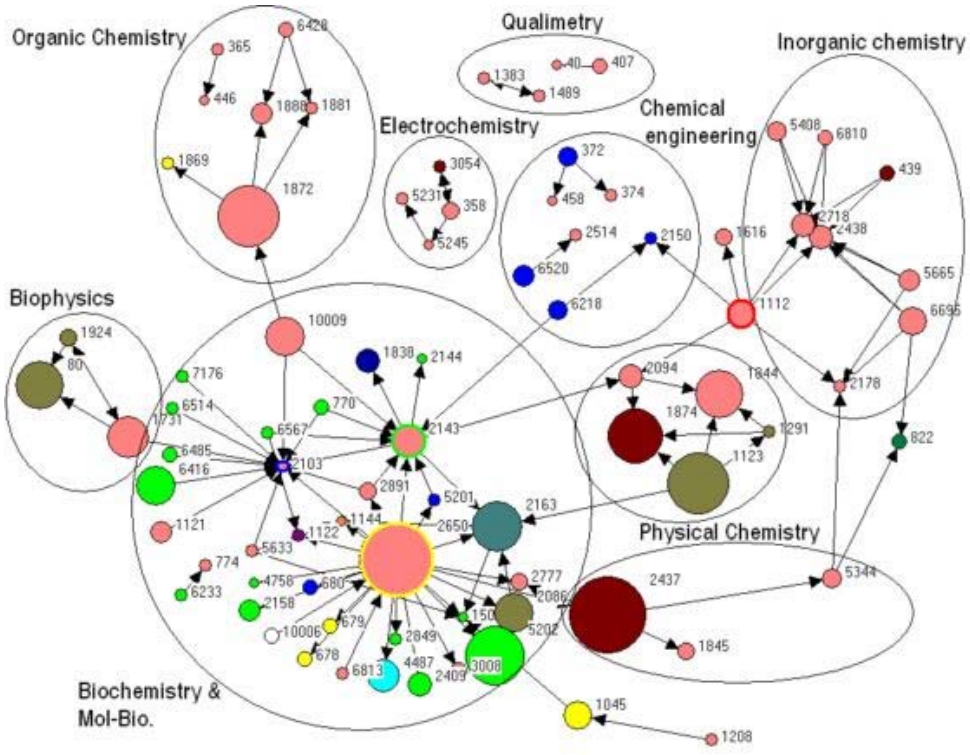

Figure 3. Interdisciplinary network of Chemistry

Finally, Figure 3 shows the relationships of the chemical websites (in salmon). Several groups are identified in this graph such as Organic Chemistry, Inorganic Chemistry or Qualimetry. We can see the relationships of different research fields with other disciplines such as Chemical Engineering with Engineering or Biophysics with Physics. The main group is Biochemistry which has a strong relationship with biological websites and it is also the most linked set. According to the centrality degree we can highlight the Department of Biochemistry and Molecular Biology of the Universitat de Barcelona (2650) (yellow crown), the Department of Biochemistry of the Universidad Autónoma de Madrid (2103) (blue crown) and the Department of Biochemistry and Molecular Biology of the Universidad Complutense de Madrid (2143) (green crown) which are the three edges of the Biochemistry sub-network. In the same way, the Chemistry Department of the Universitat de les Illes Balears (1112) (red crown) is the central node that connects the Inorganic chemistry, Physical Chemistry and Chemical Engineering clusters.

5. Discussion

One of the principal problems when analysing interdisciplinary relationships is the thematic classification of the units because this may affect to the detection of cross boundary relationships. Leydesdorff (2007) found this problem with the ISI (Institute for Scientific Information) categories when he studied interdisciplinarity in scientific journals. In our case, we have found similar problems with the Unesco Nomenclature. For instance, Technological Sciences discipline cover a wide range of topics such as Physics Engineering, Chemistry Engineering, etc., which can belong to other disciplines. Computing discipline was created to merge scattered topics related with Computer Science. We think that this approach based on links relationships on the Web, just like Leydesdorff (2007) with citations between journals, could allow us to solve this problem and to represent the relationships between disciplines. Unlike to Leydesdorff (2007), we have used a visual inspection method to present these relationships because the sample is quite small and the network centrality measures may produce some artifacts.

The bow-tie model shows a network less interconnected than previous characterizations (Broder *et al.* 2000; Thelwall & Wilkinson 2003) because the size of SCC component is quite smaller than the other components. However, the SCC component in our study is closer to the same component in the Baeza-Yates, *et al.* (2005) study about the Spanish web. We consider that this may be due to the own characteristics of the Spanish web. Even so, the sample studied covers a small part of the Spanish web, even of the Spanish academic web, and the large size of OUT and Tendril informs us these websites may be connected with other wider web spaces such international research web space or Spanish web space.
Our results also show that there is a great proportion of Experimental Sciences and Technological Sciences websites. This result correlate with those obtained in previous studies about the academic web in Australia and Taiwan (Thelwall, Vaughan, Cothey, Li and Smith, 2003) and United Kingdom (Thelwall, Harries and Wilkinson, 2003). And they are also similar to the print scientific production in Spain (Gomez, *et al.*, 2007) and Europe (EC, 2003). We think this imbalance between Technological and Experimental Sciences, and Humanities and Social Sciences may be caused by different habits when it comes to publish web contents and the use of different format and media to communicate their results. Nevertheless, this great presence of Experimental and Technological Sciences may determine our results because they may show more visible interdisciplinary relationships related to Experimental and Technological Sciences than other sciences. Thus, the Economics map show less interdisciplinary relationships than the Biology and Chemistry ones.

6. Conclusions

The network graph allows us to detect research fronts and their relationships in the Web. Through these three graphs we have seen the interdisciplinary relationships, the presence of these disciplines in the Web and the detection of new research fronts. Thus, we consider that the use of network graph and Visualization of Information techniques allows to map the scientific presence and their relationships on the Web.

The characterization of the academic web space in Spain through the Bow-tie model has led us to describe the thematic distribution of the Spanish academic web space and detect the significant presence of Experimental and Technological Sciences. It has also shown that the main or SCC zone is smaller than in previous studies, probably due to the small size of the sample and the wide range of disciplines covered by the sample. On the contrary, the size of the Tendrils component confirms that the network of research groups and departments is weakly connected.

10. References

Albert, R., Jeong, H., & Barabasi, A.L. (1999). Internet - Diameter of the World-Wide Web. **Nature**, 401(6749): 130-131.

Arroyo, N. (2004). Métodos y herramientas para la extracción de datos en Cibermetría. El software académico y comercial. [MS thesis]. University of Salamanca.. **http://eprints.rclis.org/archive/00007207/**

Baeza-Yates, R., & Castillo, C. (2007). Características de la web chilena 2006. [Technical report] Santiago: Center for Web Research, University of Chile. **http://www.ciw.cl/material/web_chilena_2006/Web_Chilena2006.pdf**

Baeza-Yates, R., Castillo, C., & López, V. (2005). Characteristics of the Web of Spain. **Cybermetrics**, 9(1): Paper 3
http://www.cindoc.csic.es/cybermetrics/articles/v9i1p3.html

Barabasi, A.L., Albert, R., & Jeong, H. (2000). Scale-Free Characteristics of Random Networks: the Topology of the World-Wide Web. **Physica a**, 281(1-4): 69-77.

Björneborn, L. (2004). Small-World Link Structures across an Academic Web Space: A Library and Information Science Approach. [PhD thesis] Copenhagen: Royal School of Library and Information Science.

http://www.db.dk/lb/home_uk.htm

Björneborn, L. (2006). Mini small worlds of shortest link paths crossing domain boundaries in an academic Web space. **Scientometrics**, 68(3): 395-414

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., & Wiener, J. (2000). Graph Structure in the Web. **Computer Networks-the International Journal of Computer and Telecommunications Networking**, 33(1-6): 309-320.

Chen, C. (2003). Mapping Scientific Frontiers: The Quest for Knowledge Visualization. London: Springer-Verlag.

Gomez, I., Fernandez, M. T., Bordons, M. & Morillo, F. (2007). La actividad científica del CSIC a través del Web of Science: Estudio bibliométrico del periodo 2001-2005. Madrid: CINDOC. **http://www.cindoc.csic.es/investigacion/xx/csic1.pdf**

Dogan, M. & Pahre, R. (1990). Creative marginality: innovation at the intersections of social sciences. Bouler, CO: Westview Press.

European Commission (2000). Towards a European research area. Luxembourg: EC.

European Comission (2003) Third European Report on Science & Technology Indicators 2003. Luxembourg: EU.

EICSTES (2006) European Indicators, Cyberspace and the Science-Technology-Economy System. Retrieved 21 March, 2007 from **http://www.eicstes.org/**

Flake, G.W., Lawrence, S., Giles, C.L., & Coetzee, F.M. (2002). Self-Organization and Identification of Web Communities. **Computer**, 35(3): 66-71

Flake, G.W., Lawrence, S., & Giles, L. (2000). Efficient Identification of Web Communities. Proceeding of the Sixth International Conference on Knowledge Discovery and Data Mining (ACM SIGKDD-2000): 150-160.

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): 1303-1322

OECD (1972). Interdisciplinarity: Problems of Teaching and Research in Universities. Washington: OECD Publication Center.

Ortega, J.L., Aguillo, I.F., Cothey, V., & Scharnhorst, A. (2008). Maps of the academic web in the European Higher Education Area - an exploration of visual web indicators. **Scientometrics**, 74(2) **http://internetlab.cindoc.csic.es/cv/11/EU_Web_maps/EU_Web_maps.htm**

Rafols, I., & Meyer, M. (2007). Diversity measures and network centralities as indicators of interdisciplinarity: case studies in bionanoscience. Proceedings of the 11th International Conference of the International Society for Scientometrics and Informetrics

Thelwall, M. (2002). Evidence for the existence of geographic trends in university web site interlinking. **Journal of Documentation**, 58(5): 563-574.

Thelwall, M., Harries, G., & Wilkinson, D. (2003). Why Do Web Sites From Different Academic Subjects Interlink?. **Journal of Information Science**, 29(6): 453-471.

Thelwall, M., Vaughan, L., Cothey, V., Li, X.M., & Smith, A.G. (2003). Which Academic Subjects Have Most Online Impact? A Pilot Study and a New Classification Process. **Online Information Review**, 27(5): 333-343.

Thelwall, M., & Wilkinson, D. (2003). Graph structure in three national academic webs: power laws with anomalies. **Journal of the American Society for Information Science and Technology**, 54(8): 706-712.

Unesco (1988). Proposed International Standard Nomenclature for Fields of Science and Technology. Retrieved 21 March, 2007 from **http://unesdoc.unesco.org/images/0008/000829/082946eb.pdf**

Received June 1st 2007
Accepted November 6th 2007
