The websites of the biomedical research centres of the Autonomous Community of Castile and León. A cybermetric analysis

Fernando Sánchez-Pita [CV] [ ] [ ] Instituto de Estudios de la Ciencia y la Tecnología. Universidad de Salamanca / fernandosp@usal.es

José Luis Alonso-Berrocal [CV] [ ] [ ] Departamento de Informática y Automática. Universidad de Salamanca / berrocal@usal.es

Abstract: Institutional websites are sources of content and a space that brings together users and institutions with common interests and objectives. This article offers a cybermetric analysis of the design of the websites of the biomedical research centres of Castile and León, which contributes to the study of one of the main foundations that underpin the information society, i.e. the access to information through such websites. This quantitative study reveals some of the communication tools implemented in this type of websites. This study is based on the evaluation of these websites' topological measures, properties and positioning values. In addition, this article offers a graphical representation of these various aspects under analysis in order to facilitate the understanding of the results.

Keywords: Cybermetrics; Internet; Web analysis; Institutional communication.

How to cite this article in bibliographies / References
F Sánchez-Pita, JL Alonso-Berrocal (2013): "The websites of the biomedical research centres of the Autonomous Community of Castile and León. A cybermetric analysis", at *Revista Latina de Comunicación Social*, 68. La Laguna (Tenerife): Universidad de La Laguna, pages 383 to 419, retrieved on ___ de ___th of ____ of 2_______, from http://www.revistalatinacs.org/068/paper/982_Salamanca/16_Sanchezen.html
DOI: 10.4185/RLCS-2013-982en/CrossRef link

Contents: 1. introduction. 2. Methods. 3. Results. 4. Conclusions. 5. List of references.

Translation by **CA Martínez Arcos**, Ph.D. (Universidad Autónoma de Tamaulipas)

1. Introduction

In the last decades thee has been a fundamental change in the Spanish science and technology system due, largely, to strategic decisions for the introduction and implementation of information and communication technologies (Zamora; Aguillo; Ortega; Granadino, 2007).

In this new scenario, in which scientific contents are more accessible than ever, and are no longer restricted by schedules or lack of building facilities, the web resources have become an essential part of what the Spanish Foundation for Science and Technology (FECYT) has called e-science, i.e.,

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

internet-mediated scientific activities. The online distribution of content is part of a new scenario which combines traditional media and users (Izquierdo Castillo, 2012) and a new category of users, the so-called digital natives.

This new reality revives the idea of the info-sphere, proposed by the futuristic author from the 1980s, Alvin Toffler, in which information is the main living creature in the cyberspace.

Thanks to the Internet, communication and dissemination have become invaluable elements, due to the great possibilities that are provided by the new working environment to the information professionals. The source of such possibilities is websites (Fernández, Aguirregoitia, Boix, 2011) and this study focuses on the examination of the institutional websites of the public biomedical research centres from Castile and León.

For research centres, Internet has become the force driving their evolution. Internet has enabled the development of new teaching and communication models in centres involved in educational programmes. This allows the development of distance learning programmes and allows the centre to develop beyond their geographical scope of action. In this way, research centres can disseminate news and advertise their services beyond their geographic limitations, which increases their field of action and, therefore, their capabilities.

Taking into account the research dimension, which is an essential part of research centres, Internet has facilitated the dissemination of results, and the showcase of the research capacity of institutions, through such tools as institutional websites, repositories and personal webpages, among others. These digital tools have enabled messages to be directional and have changed the roles played by the different actors involved in the communication processes (Martínez Sanz, 2012). In addition, these internet tools add prestige to research institutions, which allows them to obtain new human and economic resources.

Internet’s content management and dissemination capacities have increased the competitiveness between different types of institutions, including research centres, which attempt to attract new resources. This activity has become the subject matter of many studies and disciplines, particularly cybermetrics, which has been used to explain and evaluate the presence of research centres on the web and the scientific and educational activities that are carried out on the Internet (Thelwall, 2004).

The presence of the research activity on the Internet has been studied by cybermetrics Lab of the Spanish National Scientific Research Council (*CSIC* according to its initials in Spanish), which produces four global rankings on the web presence of research centres, hospitals, universities, repositories and business schools, which are very relevant for this study. The rankings produced by cybermetrics lab are updated twice a year, normally at the end of January and July, and the information presented in this study is the one published in July 2012.

It should be noted that these rankings are determined according to cybermetric parameters, related to web presence, and not according to issues related to the centres’ service or work quality. However, in cases where the web performance of an institution does not meet its expectations, based on the performance of other areas, the leaders of such centres should reconsider their web policies, which can lead to an increase in the volume and quality of content published on the internet.

According to the results published by cybermetrics Lab on July 2012 about research centres, the top positions are occupied by US institutions. The first Spanish centre to appear in the top 100 research centres is the CSIC itself, in the 11th position, followed by the Miguel de Cervantes Virtual Library,

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

in the 35th place, the Carlos III Health Institute (in the 45th position), Catalonia's Educational Telematics Website (51st), and the Institute for Catalan Studies (94th).

The published data show the low web presence of Spanish research centres. Moreover, of the 7532 research centres examined by the Lab, only three qualify as objects of study for our study, biomedical research centres from Castile and León, and all of them occupy very low positions: the Institute for Applied Ophthalmology, which is the 4003rd place, the Institute for Biology and Molecular Genetics, in the 5459th place, and the Institute for Biotechnology de León, in the 7153rd place.

Thus, this study aims to provide new data on the actual web presence of public biomedical research centres from Castile and León. This study is based on a cybermetric approach, which applies Social Network Analysis, which in turns provides revealing data, supported with graphic representations, about the quantitative aspects of the construction and construction trends of the websites of these research centres.

1.1. Justification

The objective of this study is to identify the quantitative aspects of the websites operated by the public biomedical research centres from Castile and León.

The centres responsible for the websites under analysis, due to their public nature, have the implicit duty of communicate their activities and scientist results. The communication function is not always performed with the necessary dedication and seems to be more linked to private institutions than to the services performed by public institutions, which is totally wrong.

This study also takes into consideration the new digital ecosystem where these types of institutions interact with their public, the so-called web 2.0. As Túñez and Sixto (2011: 2) point out, "the social websites and the philosophy 2.0 have been integrated in all kinds of organisations –profit and non-profit, public and private – so that the Web 2.0 can also be business 2.0, politics 2.0, and government 2.0". Such coexistence with the public is very marked by the importance of the social websites, which have determined the new dissemination routes for products and services (Herrero, Álvarez, López, 2011).

Therefore, based on web presence quantitative analysis, supported by Social Network Analysis, cybermetrics is the study of the trends in the construction of the communication platforms of centres with external communication needs. Thus, based on cybermetrics and an exclusively-quantitative point of view, this study will offer conclusions on the situation and needs of the websites of the biomedical research centres from Castile and León and recommendations to improve these digital communication platforms.

1.2. Research objectives

As mentioned, the websites under analysis are those belonging to the biomedical research centres from the Community of Castile and León. The centres whose websites are examined in this study are:

- Institute of Functional Biology and Genomics (*Instituto de Biología Funcional y Genómica*)

- Institute of Molecular Biology, Genomics and Proteomics (*Instituto de Biología Molecular, Genómica y Proteómica*)
- Institute of Biotechnology (*Instituto de Biotecnología*)
- Research Institute of Endocrinology and Clinical Nutrition (*Instituto de Endocrinología y Nutrición*)
- Institute of Alcohol and Drug Studies (*Instituto de Estudios de Alcohol y Drogas*)
- Institute of Pharmacoepidemiology (*Instituto de Farmacoepidemiología*)
- Institute of Neuroscience of Castile and León (*Instituto de Neurociencias de Castilla y León*)
- Institute of Applied Ophthalmobiology (*Instituto de Oftalmobiología Aplicada*)
- University Institute of Molecular Biology and Cancer Cell (*Instituto Universitario de Biología Molecular y Celular del Cáncer*)
- University institute of Biomedicine (*Instituto Universitario de Biomedicina*)

The main research objective guiding this study is to examine, from a cybermetric approach supported by Social Network Analysis, the composition of the websites of the major public biomedical research centres from Castile and León.

To this end, the study was based on the topological examination of each of the websites, as well as on the graphical representation of this analysis in order to facilitate the understanding of the results.

Thus this study seeks to identify, from a quantitative point of view, the communicative skills of the public biomedical research centres from Castile and León, based on the institutional communication they carry out in the digital landscape, where the use of websites is very important.

With regards to the main objective, there are a series of hypotheses about the measurements and indices that will be obtained from the websites under analysis:

- Hypothesis 1: These websites are composed of sets of strongly-connected nodes.
- Hypothesis 2: The homepage is not the most important page in the website.
- Hypothesis 3: The websites of the public biomedical research centres from Castile and León, which depend on the same university, do not use the same structural patterns.
- Hypothesis 4: The news and updates sections, when they exist, have a high PageRank with respect to the rest of the website.
- Hypothesis 5: The websites under analysis have good levels of connectivity.
- Hypothesis 6: The websites under analysis have low numbers of links per page.

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

1.3. Cybermetrics and background

Cybermetrics has two basic objectives: the study of the evolution of the size of the web and the description of the first search engines. To these objectives, new aspects have been added, driven by the theoretical contributions of different specialists. The origins of cybermetrics date back to the mid-1990s (Aguillo, 2000), when the first related theories were proposed (Bossy, 1995; Abraham, 1997) and the increasing interest in the discipline, which led to the emergence of the electronic journal Cybermetrics, in 1997, which was presented at the International Society for Scientometrics and Informetrics (ISSI), where the results of the quantitative analysis of the Internet were presented (Aguillo1997).

According to Alonso, Figuerola and Zazo (2004, 74) cybermetrics should be understood as the study and quantitative analysis of all the kinds of information and media that exist and function in cyberspace, by means of bibliometric, scientometric and informetric techniques.

For Aguillo and Granadino (2006, 69), "cybermetrics is an emerging discipline that, based on bibliometric techniques and model, aims to extend the application of quantitative methods to the description of the processes of scientific communication in the Internet, to the determination of the volume and the typology of the academic content posted on the Web, and tries to unravel the social interrelationships and information consumption by users. Other objectives of cybermetrics are the description of search tools on the Web, the so-called invisible Internet and the particularities of email-based services and personal forums".

Björneborn (2004) considers that cybermetrics is the study of the quantitative aspects of the construction and use of the information resources, structures and technologies in the Internet, from bibliometric and informetric perspectives. Cybermetrics should not be confused with webometrics which, despite being practically the same, differs in the object of study, the Web: this discipline should be understood as "the study of the quantitative aspects of the construction and use of the Web's information, resources, structures and technologies, from bibliometric and informetric perspectives".

According to Björneborn (2004), at first, different terms were used to refer to this new discipline: Netometrics (Bossy, 1995), Webmetry (Abraham, 1997), Internetometrics (Almind, Ingwersen, 1997), Web Bibliometry (Chakrabarti et al., 1999). The terms that were eventually adopted were cybermetrics and webometrics, which are often considered synonyms, which is wrong as we can see in the following figure.

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

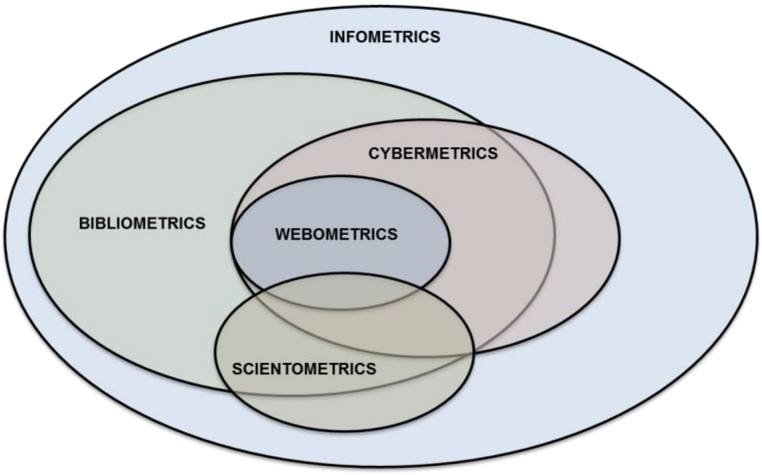


Figure 1: Cybermetrics. Adapted from Björneborn (2004)

According to Björneborn's definitions for cybermetrics and webometrics (2004), both disciplines have more similarities than differences, since both are based on quantitative aspects, focus on the use and construction of information resources, structures and technologies, and are based on bibliometric and informetric approaches.

Today, cybermetrics is postulated as a possible solution to address the analysis of the documentary explosion, based on the use of hypertext, links between pages, which recover the role of citation in classical literature. In this way, the importance of documents depends on the number of links it has.

Despite being a relatively young discipline, the usefulness of cybermetrics has been demonstrated in various studies, such as those carried out by Berrocal (1999, 2002), Cothey (2004) and Ortega and Aguillo (2008, 2009).

1.3.1. Development of indicators

The creation and design of indicators is one of the fields of cybermetrics with more scientific literary production. This is due, mainly, to the great importance of the impact indices. As it happened with the traditional literary citation schemes and the application to digital environments, traditional indicators have adopted the Web surname. This is how the impact indices applicable to web environments emerged.

Ingwersen (1998, 237) defined the concept of web impact factors as "the logic sum of the number of external links and intrapage links of web pages that link to a given website, divided by the number of pages that can be found in this website in a over certain specified time period. The resulting indicator is, therefore, the number of linking pages, not the number of links that can be found".

Another definition of the internet-based indicator is provided by the Fifth Framework programme of R&D of the European Commission (EICSTES - European Indicators, Cyberspace and the Science-Technology-Economy System), which indicates that an indicator is "a measure that quantifies the creation, dissemination and application of science and technology on the Internet". This programme also defines 10 of the most characteristic indicators of the study on the Internet:

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

- Depth: Number of levels in a website, where level 0 is the home page.
- Density: Total number of links per page, including external and internal outlinks and intrapage links.
- Connectivity: Different number of links in a website, including internal and external outlinks, but not intrapage links.
- Navigability: Density of internal outlinks in a website, taking into account even repeated links.
- Endogamy: Percentage of internal different outlinks to the total number of different outlinks.
- Luminosity: Number of external outlinks, i.e., links from a website to other different ones.
- Dispersion: Typology and frequency of the outlinks in a website according to several distribution criteria.
- Visibility: Number of external links received by a website.
- Popularity: Number of different visits received by a website.
- Diversity: Typology and frequency of the links received by a website according to several distribution criteria.

1.3.2. Automatic information retrieval

Automatic information retrieval is a discipline directly related to documentation and Informatics, which is aimed, mainly, to the retrieval of information resources and the discrimination of the degree of adjustment of a document depending on the terms of reference formulated or the information needs of the user (Salton and McGill, 1983).

According to Ricardo Baeza-Yates (1999), it is the part of computer science that studies the retrieval of information from a collection of written documents that can meet the needs for information of the user, who most of the times uses natural language to define the search terms.

Korfhage (1997) defines this discipline as the location and presentation to the user of a need for information expressed in the form of question.

Based on the previous definitions, information retrieval can be defined as the management of documents by relevance depending on the conditions of the consultation carried out and the full set of documents on which the search is performed.

Information retrieval is based on a series of models, like the classical models, which include the Boolean model, the vector model and the probabilistic model.

The Boolean model, which is currently in disuse, shows the documents as a set of terms and questions as Boolean expressions.

The vector model is the most widely used system for information retrieval, as well as automatic categorisation and filtering of information, among others. In this model the useful words are chosen,

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

i.e., all the words of the texts except the empty words, and is enriched by techniques such as lemmatization and labelling (Jaimes, Vega, 2010).

On the other hand, the probabilistic model indicates that given a query *q* and a document d_j, the objective is to estimate the probability of the user to find the document d_j considered to be relevant. In this way, the objective is to observe the distribution of the documents in the entire collection. The model assumes that the probability of relevance depends only on the query and the representation of the document.

Today, information retrieval is linked to major online search engines that allow retrieving information stored in the web.

The web is a collection of billions of documents linked with each other, and the web crawlers are the tools to systematically browse the web. According to Arroyo et al., (2005), the use of the web crawlers is optimal for conducting studies on websites or specific webpages (micro-level), however for the study of large amounts of information (macro-level) it is necessary to make use of the search engines.

Due to the massive amounts of information that can be found on the Internet, the search engines are greatly used for the search and retrieval of the existing information in the web.

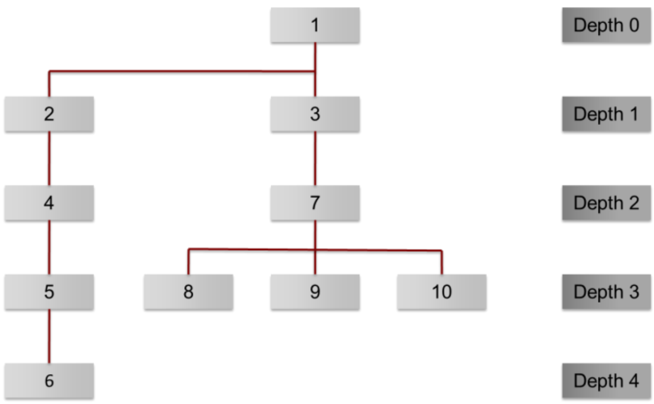


Figure 2: Crawling

A crawler, spider, wanderer, or a bot, is a computer program with the ability to retrieve information from websites by identifying all the hyperlinks inside of them and moving through them. This action receives the generic term crawling, which can occur in three different ways: breadth-first, depth-first and best-first.

Crawlers and other specific applications allow information retrieval, which is an area of science focused on retrieving information about a particular subject contained in a collection of data. It should be noted that the information retrieval should not be confused with data recovery, which consist in determining which documents in a collection, contain certain words a user is searching for.

Currently there are many crawlers to carry out a cybermetric study: the WebBot created by the World Wide Web Consortium (W3C); SocSciBot 3 and SocSciBotTools, created by the Statistical Cybermetrics Research Group of the University of Wolverhampton; WIRE crawler, created by the

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

Web Research Centre of the University of Chile; and Webvac Spider, created by the Stanford Infolab of Stanford University.

All of them can satisfactorily carry out local cybermetric studies. This study, however, will use the SACARINO computer program created at the University of Salamanca. This software is explained in the following section.

2. Methods

2.1. Approach and methods

This study is based on a quantitative method, which revolves around the calculation of measures and indices and the graphical representation of the collected data. All this is closely related to the techniques of Social Network Analysis, which is a tool for measuring and analysing the structures that emerge from the relations between actors of different types, in this case nodes or web pages.

The measurements and indices, as well as the graphical representation of the studied networks have been obtained with specific software.

SACARINO: Used for data collection, based on technical criteria that are explained below and a purpose-created scheduled.

EloisaBot Tools: Used for the conversion of the data collected with SACARINO, in preparation for their subsequent use in Gephi and Pajek, and the estimation of Google's internal PageRank for websites.

Pajek: Use for obtaining measurements (density, diameter and centrality of the website) and graphical representation of networks through the Bow-Tie theory.

Gephi: Graphical representation of the networks' input, output, intermediation and closeness degrees, and the PageRank calculated with EloisaBot Tools

After an initial exploratory study, the websites of ten research centres were analysed to be included in the study. Thus, the process of data collection and subsequent graphical representation has been repeated ten times in order to complete the full study.

The selected centres are:

- Institute of Functional Biology and Genomics (*Instituto de Biología Funcional y Genómica*)
- Institute of Molecular Biology, Genomics and Proteomics (*Instituto de Biología Molecular, Genómica y Proteómica*)
- Institute of Biotechnology (*Instituto de Biotecnología*)
- Research Institute of Endocrinology and Clinical Nutrition (*Instituto de Endocrinología y Nutrición*)
- Institute of Alcohol and Drug Studies (*Instituto de Estudios de Alcohol y Drogas*)
- Institute of Pharmacoepidemiology (*Instituto de Farmacoepidemiología*)

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

- Institute of Neuroscience of Castile and León (*Instituto de Neurociencias de Castilla y León*)
- Institute of Applied Ophthalmobiology (*Instituto de Oftalmobiología Aplicada*)
- University Institute of Molecular Biology and Cancer Cell (*Instituto Universitario de Biología Molecular y Celular del Cáncer*)
- University institute of Biomedicine (*Instituto Universitario de Biomedicina*)

The selection criteria for the research centres were: they had to be public institutions and be part of the biomedical sector.

The analysis of the websites was carried out with Social Network Analysis techniques, which allow the identification of the main structural features of a particular network. The application of this type of analysis allowed us to study the social structures that arise from the recurrence of relations, in this case, between the nodes that form the networks.

Based on the premises of social network analysis, the explanation of the websites and their graphical presentation can be used to improve the structures and design of the analysed websites.

2.2. Software used

SACARINO (*Sonda Automática para la Recuperación de Información en la Web*) is the software used in this study for the collection of data from each of the ten selected websites. This software was developed under the direction of Professor José Luis Alonso Berrocal, from the Department of Computer Science and Automation of the University of Salamanca.

The origin of SACARINO dates back to 1994, when a computer program called *Sonda Ciberdocumental* was created to perform quantitative calculations (Alonso Berrocal, 1996). This program was improved over the years and in 2003 it was completely reprogrammed with a new programming language and its data structures were optimised to make it more powerful, faster and more flexible. This is how SACARINO emerged, with the contribution of two students from the Engineering programme in Computer Systems of the University of Salamanca: María del Carmen Montejo Villa and Faustino Frechilla Daza.

To carry out the data collection correctly, which is necessary for the subsequent analysis, the following parameters were used in the configuration of SACARINO:

- Rapid fire of 250ms
- Time Out: 1000 ms
- Number of threads: 20
- Scanning limits: specified Host, Host, and specified directory, varying depending on the starting URL.
- Depth limit: none

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

- Dynamic pages: scan all the pages
- Prioritization of URLs: balancing of servers

During the data collection, certain peculiarities of the analysed websites were taken into account, since some had their own URLs, while others were dependent on main URLs. In addition, search restrictions were applied in some cases in the configuration of SACARINO to prevent the scanning of certain elements, such as calendars, typical of the "agenda" and "news" sections, as in the case of the Institute of Health Sciences Studies of Castile and León.

ELOISA

EloisaBoot Tools is a package of tools for data processing. Uses a Multiple Document interface which allows the display of multiple windows under the frame of the main window, which facilitates multitasking. The project, directed by Professor José Luis Alonso Berrocal, from the Department of Computer Science and Automation of the University of Salamanca, was motivated by the need of using different programs during work processes like this study under a single environment.

The basic functions of EloisaBot Tools are the processing of collected data, treatment of generated graphs, and the calculation of algorithms, indices and measures. Most of the routines are made for Matlab, due to the its processing power.

Unlike SACARINO, with EloisaBot Tools it was not necessary to undertake a comprehensive configuration to carry out the research activities.

In this case, the respective data collections were loaded for their conversion to the reading format used in Pajek, which was also use for their interpretation in Gephi. Afterwards, we obtained Google’s PageRank, which is provided through data tables from which the information must be extracted.

Pajek

Pajek (spider in Slovenian) is a program for Windows operating systems that enables the analysis and visualisation of websites. It is free software developed by Vladimir Batagelj and Andrej Mrvar, both professors at the University of Ljubljana, Slovenia. This software has a large community of developers which allows constant improvements and updates.

The main objectives that motivated the development of the software were to create a powerful visualisation tool, to implement efficient algorithms for the representation of large networks and the factorisation of large networks in smaller networks for their treatment with more sophisticated methods (Batagelj, Mrvar, 1998).

This program allows the obtaining of many indicators and a high customization of the graphical representation of networks, although in 2D. The following indicators, indices and representations were obtained with Pajek:

- Density
- Diameter
- Degree of incoming centrality
- Degree of outgoing centrality

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

- Relation of nodes according to Bow-Tie theory
- Graphical representation of the Bow-Tie structure

The use of Pajek consisted in the loading of each of the files converted with EloisaBot Tools, the calculation of data through tables generated by this program and the graphical representation of the networks.

The graphical representation was carried out with the Fruchterman Reingold algorithm, in its 2D version. This representation of networks was based on the Bow-Tie theory, and for this reason the following list of colours was used to differentiate the different types of nodes that make up the networks:

- LSSC (Largest Strongly Connected Component): Red
- IN: Green
- OUT: Pink
- TUBES: yellow
- TENDRILS: Blue
- OTHERS: Orange

After the graphical representation of each of the 11 networks, we exported the generated graphs in SVG and BMP formats. The first format more suitable for future reference due to the particularities of the vectorised format (which allows a great expansion for detailed queries), while the second are more suitable to be presented in this article.

Gephi

Gephi is an open source cross-platform (Windows, Linux, Mac OS) program to analyse and graphically represent networks. Its display module is based on a 3D engine capable of rendering graphs in real time, i.e. the user can observe the movement of nodes, arcs and the edges according to the desired settings. It relies on the use of graphics cards, as the most advanced video and photography editors, which allows the computer to free up memory space for other tasks.

With regards to design, this software package offers multiple possibilities in terms of presentation of graphs, since it has various configurable algorithms colour and shape options to distribute and represent the nodes.

It also has a module to interactively explore the networks, which allows us to observe the results of the different measurements in a graphic form.

Like Pajek, a remarkable aspect of Gephi is the possibility of exporting the graphic representations in SVG format, as well as in PDF format. The indicators obtained with Gephi are:

- Node with higher indegree
- Node with higher output degree
- Node with higher intermediation degree

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

- Node(s) with a greater degree of closeness

In addition, a graph was generated for each of these calculations, for the correct visualisation of the node within the network. To obtain the previous indicators we activated the following statistics in Gephi:

- Average Degree
- Network Diameter
- PageRank

This was followed by the discrimination based on the size and shape of the nodes, according to their value. In the case of colour, due to the mixing possibilities available in Gephi, shades of yellow were used for the lowest values and shades of red for the highest values, so the intermediate values were represented with orange. In terms of size, it started in 10 with the lowest value and ended in 30 with the highest value.

The graphs of the networks were generated with the Fructhterman-Reingold algorithm. The nodes were placed according to the rules of attraction and repulsion proposed by the algorithm.

To facilitate the process of interpretation of the graphs, we modified the representation according to the Default Curved style, to which we added new colours, due to the large amounts of links.

2.3. Indices and measures

2.3.1. Network's properties

Density

This indicator refers to the proportion of links that exist between potential relations of a particular network. Thanks to this indicator one can appreciate the intensity of relations in the whole of the network, whose result can vary from 0 and 1, where the values closest to 1 are the best.

The formula for this calculation may vary, depending on the characteristics of the network. This study is based on directed networks, whose density is obtained with the following formula:

$$D = \frac{r}{N(N-1)}$$

N: nº of nodes r: nº of links

In the case undirected networks, the formula would vary in this way:

$$D = \frac{r}{N(N-1)/2}$$

N: nº de nodes r: nº of links

Diameter

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

This indicator refers to the longest geodesic distance that can be found in the graph obtained from the analysis of a given network. Distance must be understood as the effort required from one node to reach another, or the number of relations existing in the shortest path between nodes. Diameter can also be described as the number of clicks that are needed to go from a particular node to the farthest node.

2.3.2. Indicators of centrality

Centrality indicators allow us to know the degrees of importance existing among the nodes that make up a network, according to their centrality, prestige or power. Freeman (1979), in his review of the literature on centrality, has pointed out that the concept of centrality could be also defined, and made operational, according to three indicators:

1) Degree

This indicator shows the number of links that has a node. In the case of directed graphs, there are two variants: indegree and outdegree.

$$d_i = \sum_{j \in V} A_{ij}$$

V: Set of nodes A_{ij}: Adjacency matrix

a) Indegree

The indegree is the measure that indicates the number of incoming relations of a node in particular, i.e. the number of nodes that point to that node. With the calculation of the indegree it is possible to identify the most prestigious nodes within a network. This also translates into power over the set of nodes that make up the network. After the standardisation of the obtained measure, this can be redefined under the concept of visibility.

b) Outdegree

The outdegree accounts for the number of relations that are initiated from a particular node. Apart from indicating the number of outlinks from one node to other nodes on the network, this measure reflects the activity of that node in relation to whole set, showing its ability to access the different parts of the network. Like the indegree, after the standardisation, the outdegree can be redefined as luminosity.

2) Degree of betweenness

This indicator refers to the extent to what a particular node of the network is in an intermediate position with respect to the set of nodes that make up the network. Nodes with a higher degree of betweenness are powerful because they have greater control over the information flows. The degree of betweenness measures the number of times a node appears in the network's existing routes which, from a cybermetric point of view, allows detecting the gateways that connect different sections of the network (Ingwersen, 1998).

The equation to obtain the degree of betweenness is the following:

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

$$g_k = \sum_{i \subset k \subset j} \frac{g_{ikj}}{g_{ij}} \qquad \forall_i \in V$$

g_{ij}: number of steps required to get from node i to node j

g_{ikf}: number of steps of g_{ij} that go through the k node

3) Closeness degree

It shows the average distance of each node with respect to the set of nodes that make up the network. In this case, the nodes that get higher values have greater ease of access to the rest of nodes that make up the website. This fact gives nodes with a greater degree of closeness a greater capacity to send and receive information.

$$c_i = \frac{1}{\sum_{j=1}^{|V|} D_{ij}} \qquad \forall_i \in V$$

D_{ij}: number of steps required to get from node i to node j

|V|: Size of the set of nodes

2.3.3. Positioning

The positioning of the websites is measured through the PageRank, which determines that the value of a web page based on the number of hyperlinks it receives from other pages and nuanced in turn by the importance of the pages from which these links come. The basic formula to obtain the PageRank is as follows:

$$x_i = \sum_{j \in B_i} \frac{1}{N} x_j$$

xi: importance of page i

j Bi: pages j linking to page i

N: number of outlinks from page j

xj: importance of page j

2.3.4. Bow-Tie theory

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

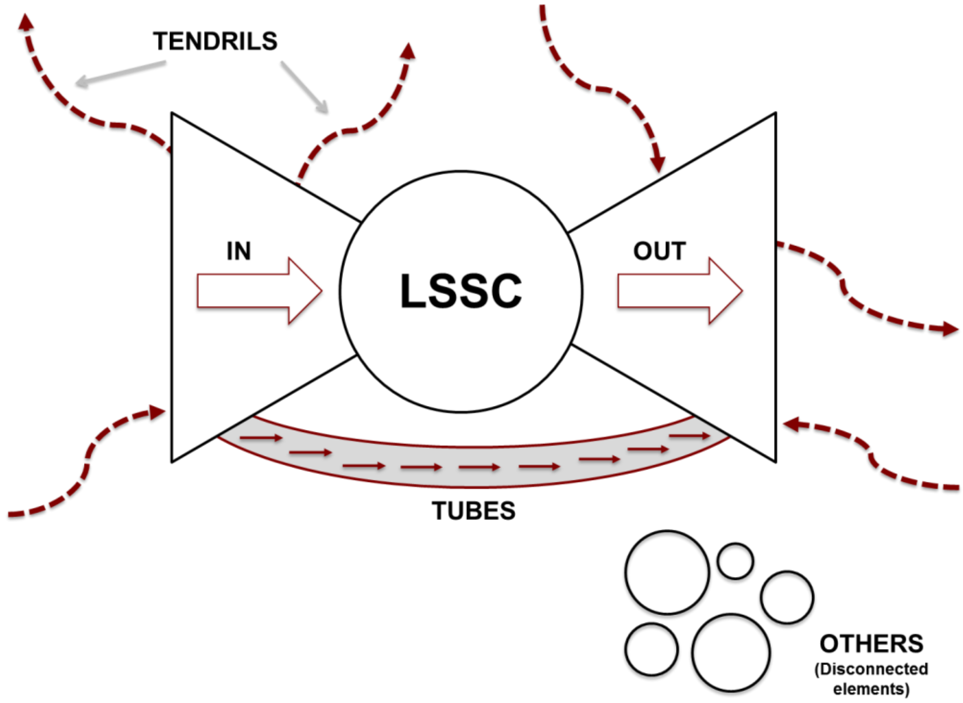


Figure 3: Bow-Tie model, adapted from Broder *et al*. (2000)

The bow-tie theory is derived from the Web studies conducted by Broder *et al.* (2000), who extracted data stored in Altavista and processed 200 million pages and 1.5 billion links. Based on that analysis, Broder *et al.* designed a model to represent the Web. In this model there is the Largest Strongly Connected Component (LSCC), named core, in which all of its pages can trace direct routes to each other; then there are two groups of pages at both sides of the core, which are named IN and OUT groups (the OUT group can be accessed from the core, but does not link back to the core, while the IN group links to the core, but the core does not links to go back to it); there are snake-like regions, called tendrils, that hang off IN and OUT of the three main components; and there are others groups, called tubes, that connect the IN and OUT groups without passing through the core; and a series of disconnected components (Alonso et al., 2008).

2.4. Graphical representation of websites

A graph is a set of lines and vertices that allow us to represent the structure of a network. In this representation, the vertices, also called nodes or points, are connected by two types of lines: arcs or edges.

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

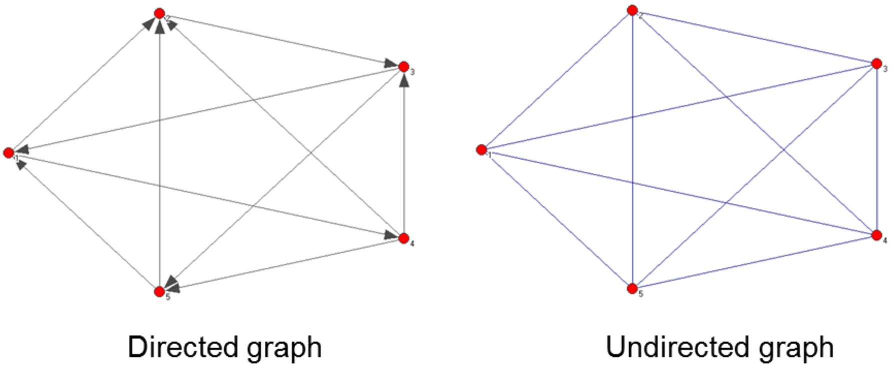


Figure 4: Directed and undirected graphs

Arcs are directed lines, with only one direction on the connection between nodes, while edges are undirected lines that give rise to bi-directional connections between nodes. This significantly affects the composition of graphs, which can be divided in directed and undirected.

Graph theory is a very valuable contribution to the field of cybermetrics, due to its algorithms for data search, extraction and identification (Brook et al., 2005).

Graphic representations derived from graph theory, facilitate the visualisation of large amounts of information, which has been successfully done for many years in previous research (Batagelj, 1998; Shannon, 2003; Adar, 2006). Graphic visualisation enhances human abilities to extract information on the characteristics of networks and the data. However, this difficulty of process requires the design of an exploration strategy.

In addition to this graphic representation, a network can also be visualised by means of an adjacency matrix. For example, in an n-by-n matrix whose entries in row i and column j give the number of arcs from ith node to the jth node.

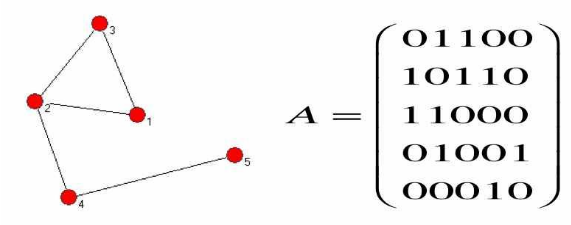


Figure 5: Adjacency matrix

Formally, the adjacency matrix is defined as V = {v_1, v_2, v_3,..., v_n}, so that:

$$M = \begin{cases} 1 & if\ (vi, vj) \in G \\ 0 & otherwise \end{cases}$$

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

The visualisation of the networks' structures of links, either through graphs or adjacency matrices, allows the identification of the structural patterns of a given website, through indices and useful indicators. So, depending on the features of a website, its links structure of links will be marked by certain construction trends, which allows, among other things, the comparison between websites with similar goals. In addition, of the general lines of a website, indices and measures about to nodes can be obtained only individually.

To produce the graph, regardless of its features, and obtain a correct final representation we must use specific software such as Pajek, Gephi, Graphviz and Touchgraph. In this study we used Pajek and Gephi, which will be further explained in the methods section.

These programs have various representation tools, options for visualisation designs, such as the Fruchterman & Reingold and Kamada & Kawai (1989) algorithms, as well as expansion, contraction or simply random designs.

From the options available in each of the previous software packages, the Fruchterman & Reingold algorithm produced the clearest and more understandable graphics, and for this reason we dismissed the use of the other representations.

The Fruchterman & Reingold algorithm offers a force-directed layout, which produces graphs in two dimensions through the simplified simulation of physical systems. This method compares the graph with a collection of electrically charged rings which are connected by means of links. The operating system is based on the premise that each two nodes reject each other, by means of a repulsive force, while the adjacent nodes, which are connected by a link, are attracted to each other, in this case by an attractive force.

After this operation, there is a series of iterations and the forces that shape each of the links are calculated once again, while the nodes move to reduce these forces.

The optimal distance between vertices, which will then affect the graphical representation of the networks, is obtained with the following formula:

$$K = C\sqrt{\left(\frac{A}{N}\right)}$$

K: Optimal distance between vertices

C: C constant, experimental resource

A: Area

N: number of vertices

Below are the graphical representations of the website of the Institute of Pharmacoepidemiology, obtained with the Fruchterman & Reingold and Kamada & Kawai algorithms, which show the differences that led us to choose the Fruchterman & Reingold algorithm for the graphical representation of the websites:

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

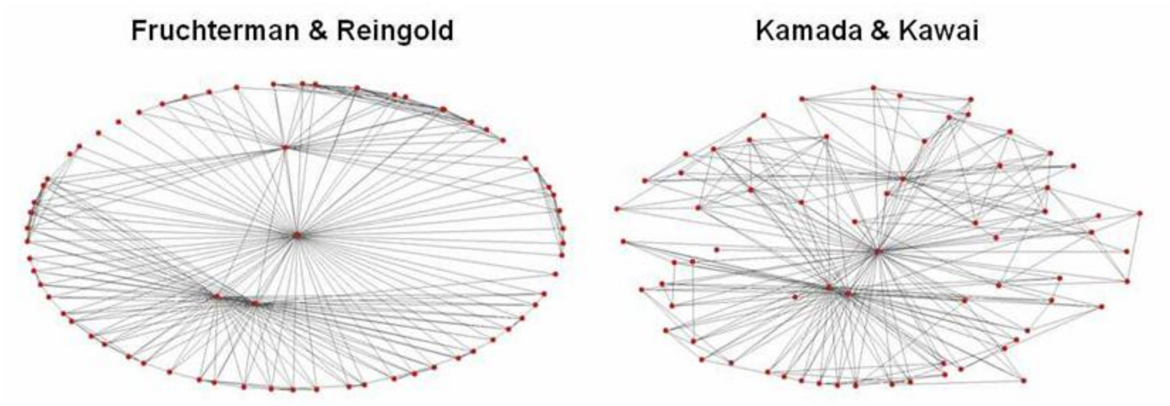


Figure 6: Algorithms

3. Results

3.1. Properties of networks

As we can see, the three networks with the greatest density are those from the Institute of Molecular Biology, Genomics and Proteomics (INBIOMIC), with 0.71, the Institute of Biotechnology (INBIOTEC), with 0.38 and the Institute of Functional Biology and Genomics (IBFG) with 0.33.

The set of networks has very low levels of density, between 0.11 and 0.01, which is significant considering that the density of a network ranges from 0 to 1. With the exception of the Institute of Molecular Biology, Genomics and Proteomics, which has a density of 0.71, in the rest of the cases it is advisable to carry out some changes in order to strengthen the lines of communication between the different nodes that make up the networks, which would substantially improve this parameter.

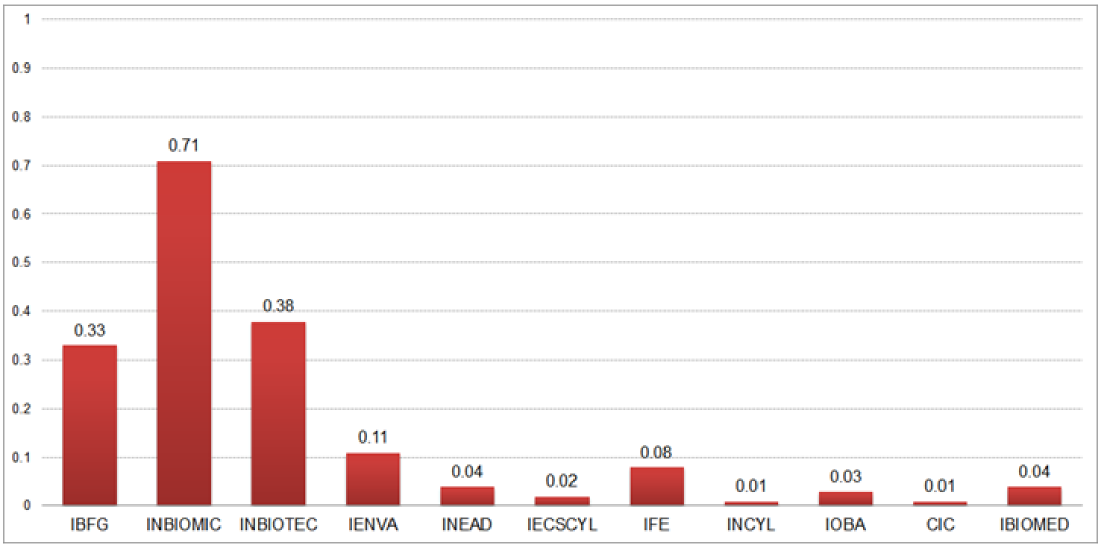


Figure 7: Density

With regards to the diameter of the sample of networks, it ranges from 2, like in the case of Institute of Molecular Biology, Genomics and Proteomics, and 25, like in the case of the Institute of Applied Ophthalmobiology. Based on their diameters, these networks require several steps to be explored. In

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

this case, the web diameter of the website of the Institute of Molecular Biology, Genomics and Proteomics is more favourably than the rest. Here is important to remark that six networks are above the average of 6.9, being of special interest the case of the Institute of Applied Ophthalmobiology, with a diameter of 25, followed by the University Institute of Molecular Biology and Cancer Cell.

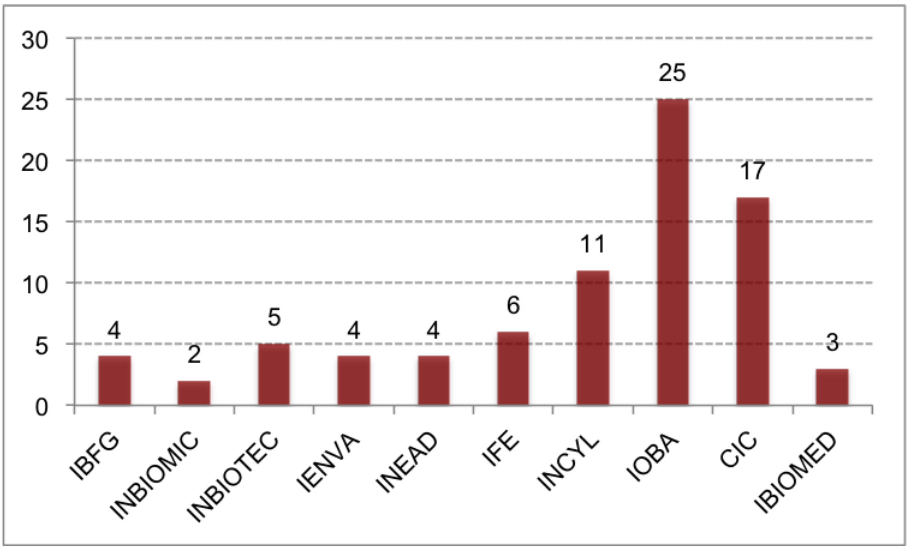
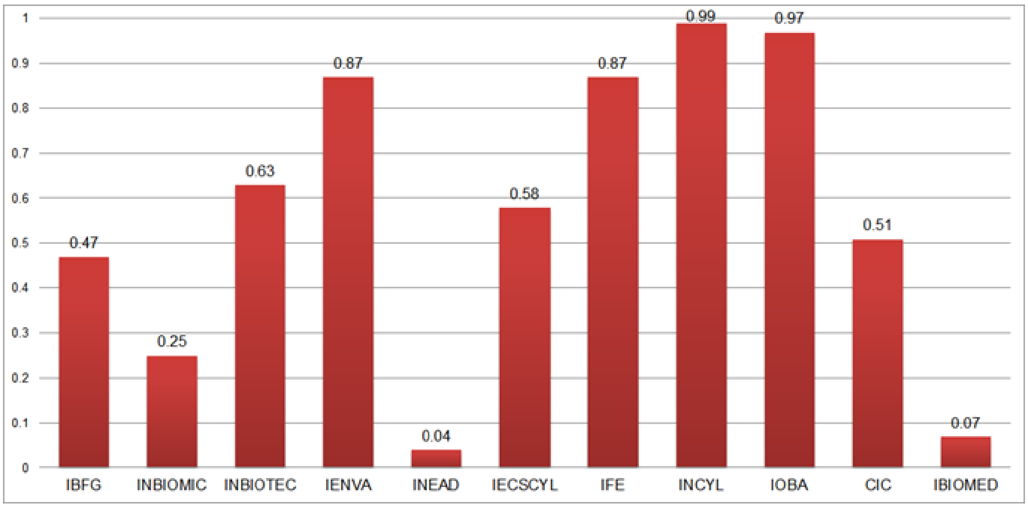


Figure 8: Diameter

3.2. Centrality of networks

Regarding the networks' centrality, whose optimal values are those closest to 0, only four networks are below 0.5: the Institute of Alcohol and Drug Studies, with 0.04, the Institute of Biomedicine with a 0.07, the Institute of Molecular Biology, Genomics and Proteomics, with 0.25, and finally the Institute of Functional Biology and Genomics with 0.47.

On the other hand, the rest of the networks exceeds 0.5, standing out the cases of the Institute of Neuroscience of Castile and León, with 0.99 and the Institute of Applied Ophthalmobiology, with 0.97.



Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

Figure 9: Centrality IN

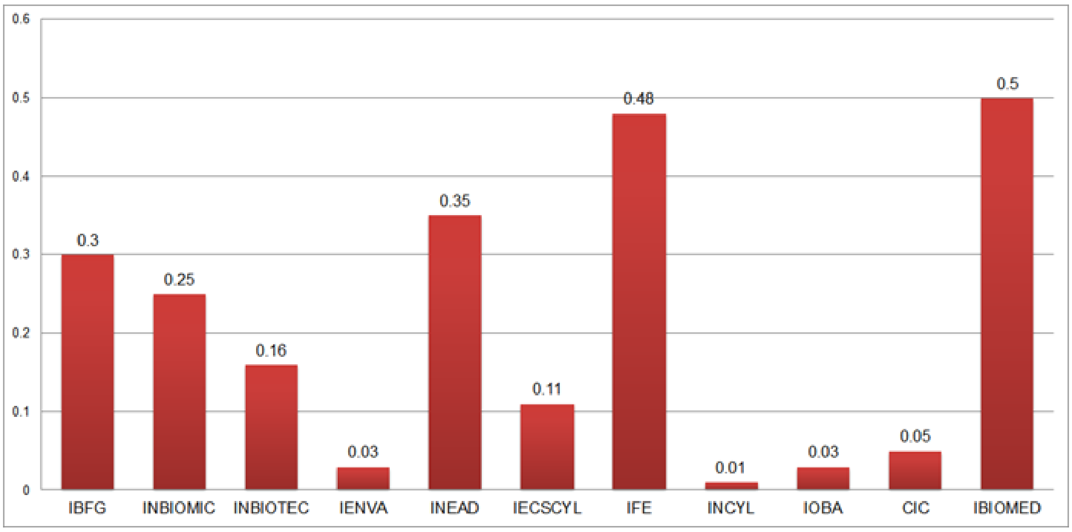
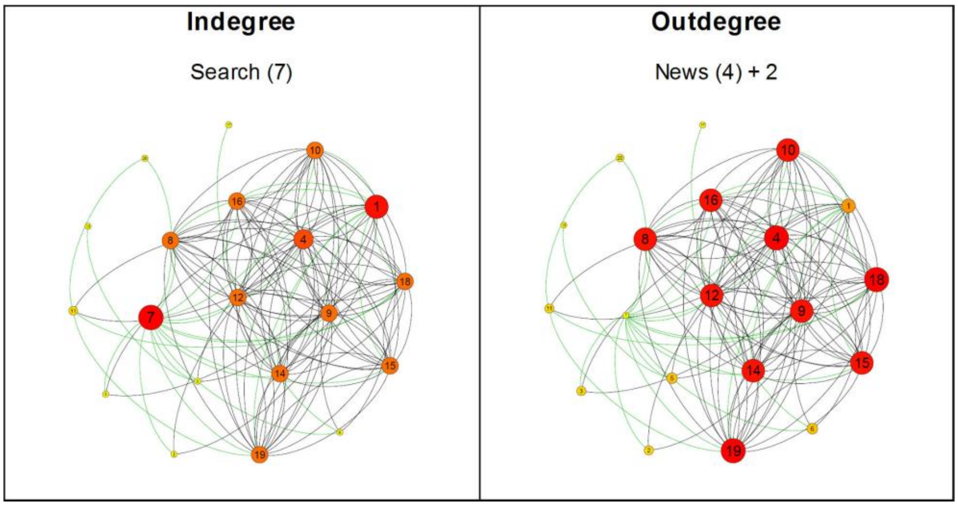


Figure 10: Centrality OUT

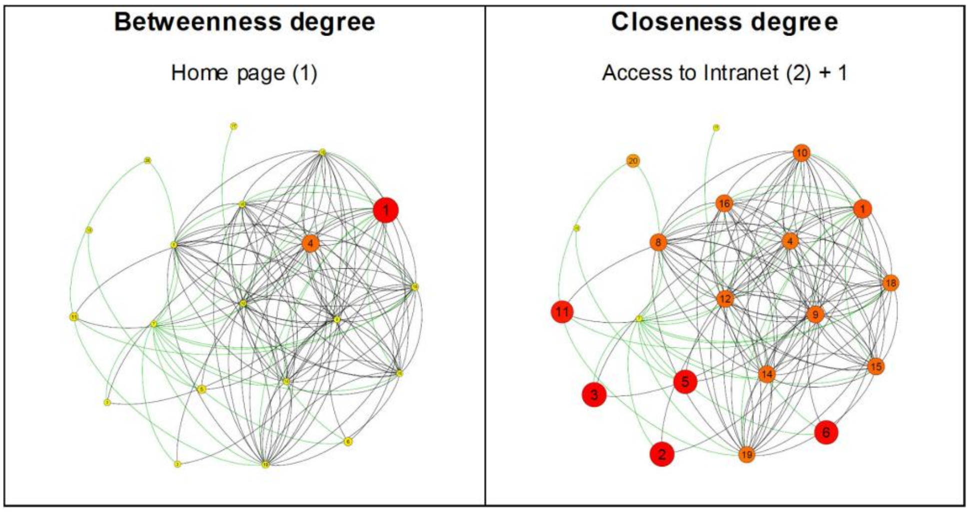
3.3. Graphical representation of networks

Below is the list of websites and their respective networks, standing out in each case the nodes with the heaviest weight in each of the measures proposed in the study. Four graphs are presented for each of the centres.

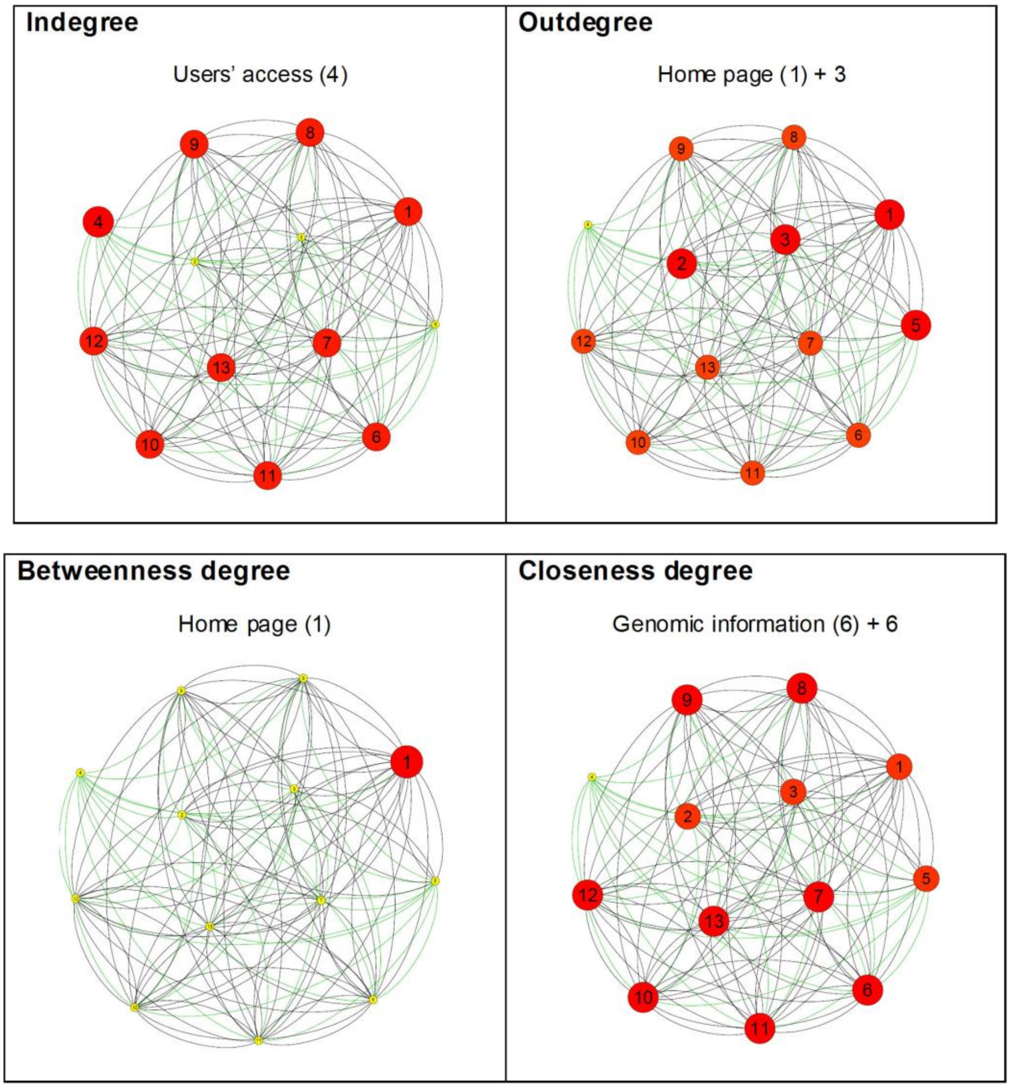
Graphs of the Institute of Functional Biology and Genomics



Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

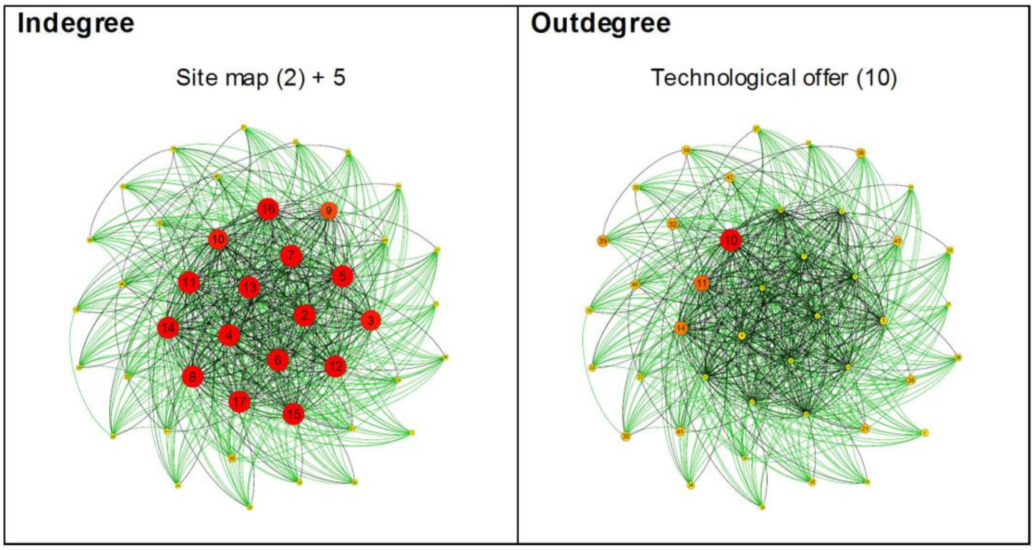


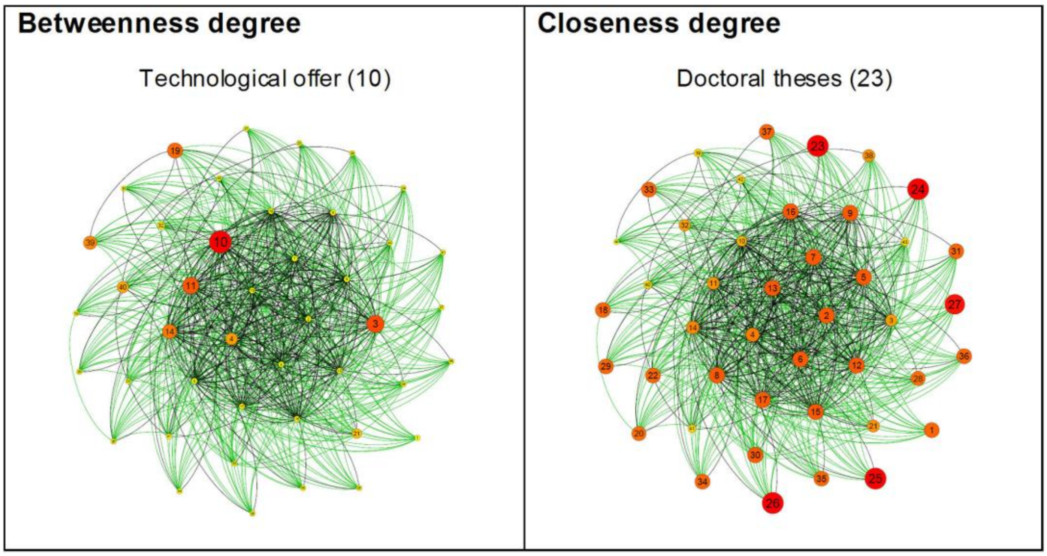
Graphs of the Institute of Molecular Biology, Genomics and Proteomics



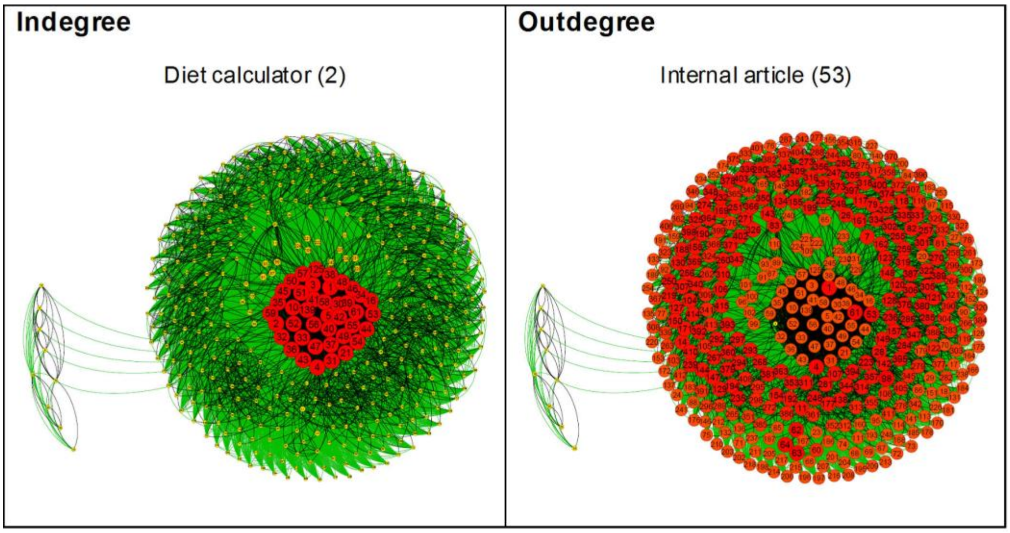
Graphs of the Institute of Biotechnology

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

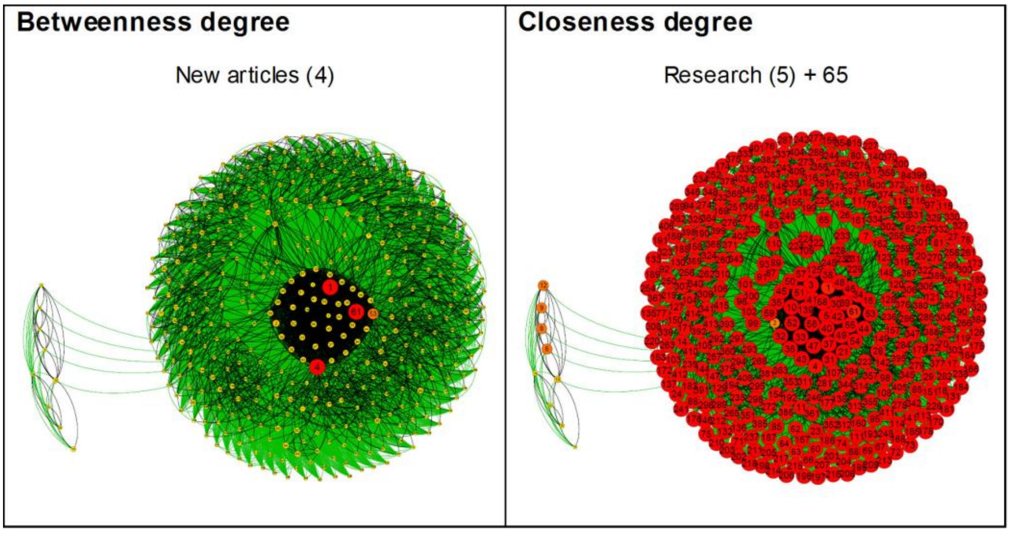




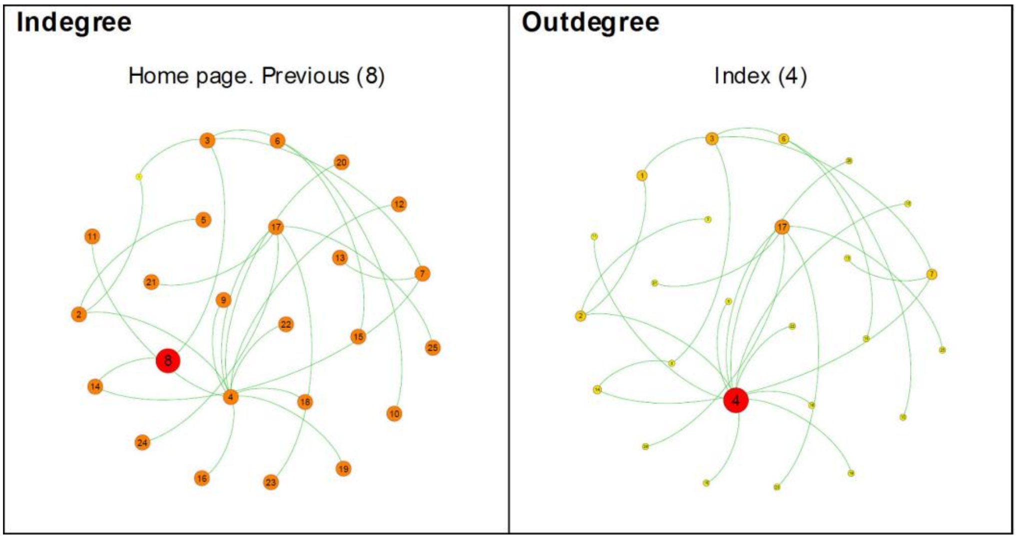
Graphs of the Institute of Endocrinology and Nutrition

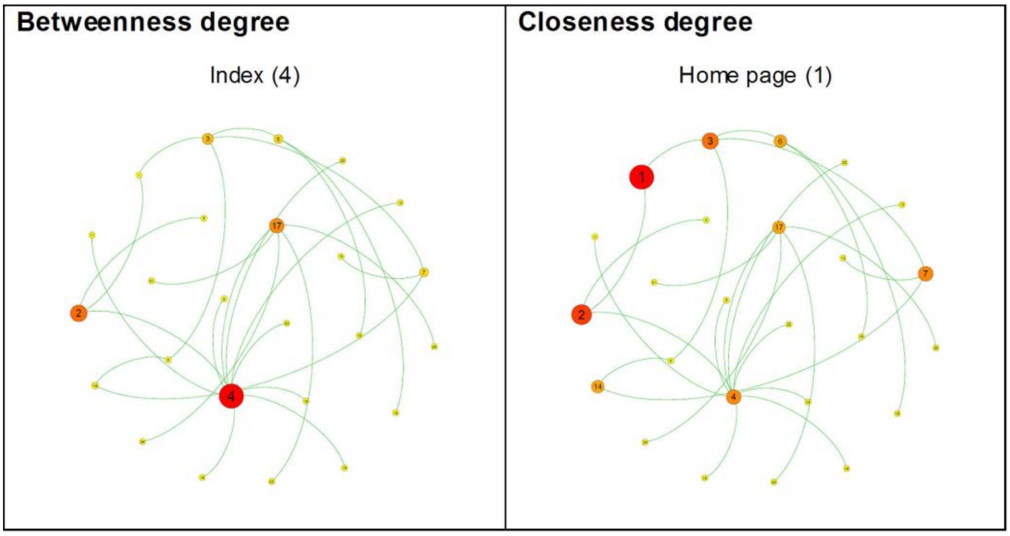


Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013



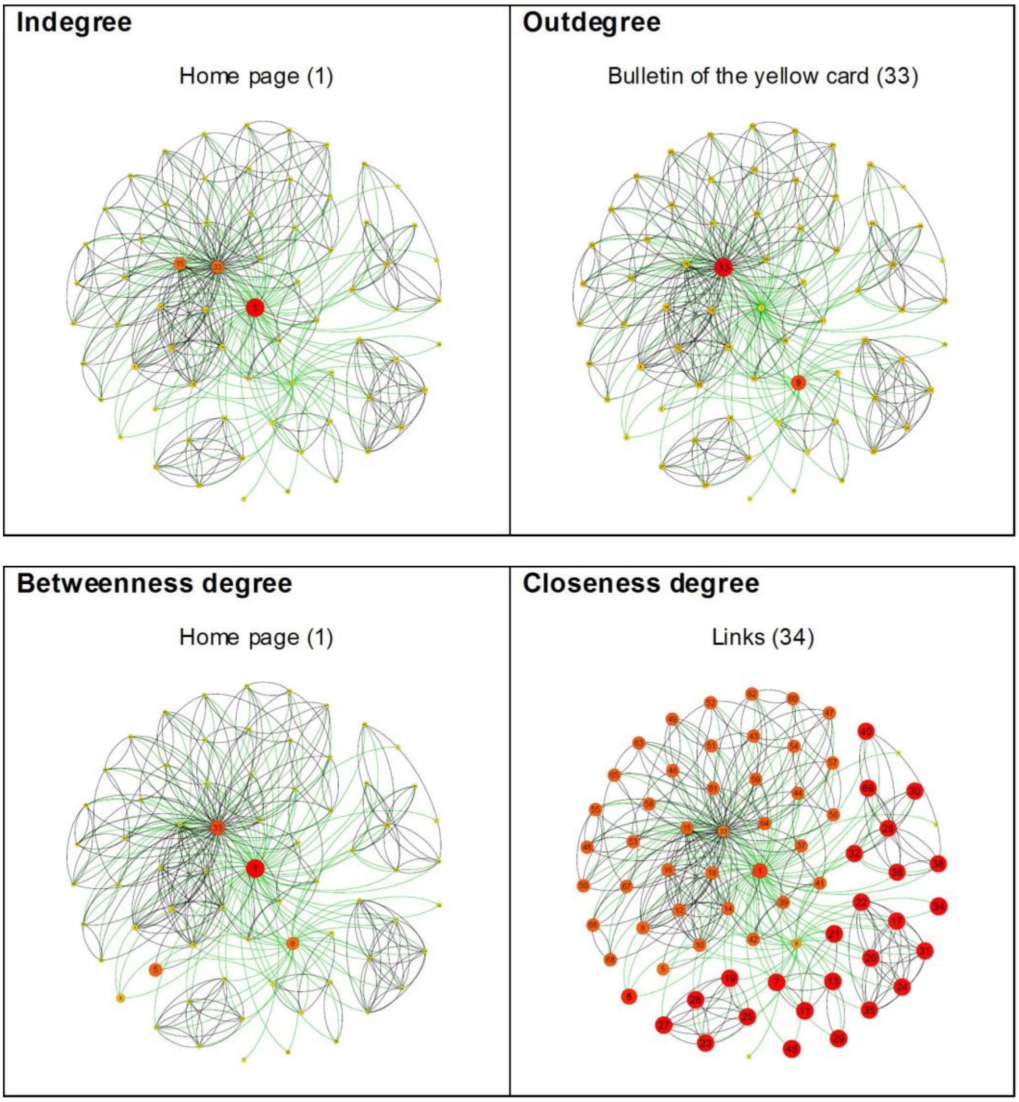
Graph of the Institute of Alcohol and Drug Studies



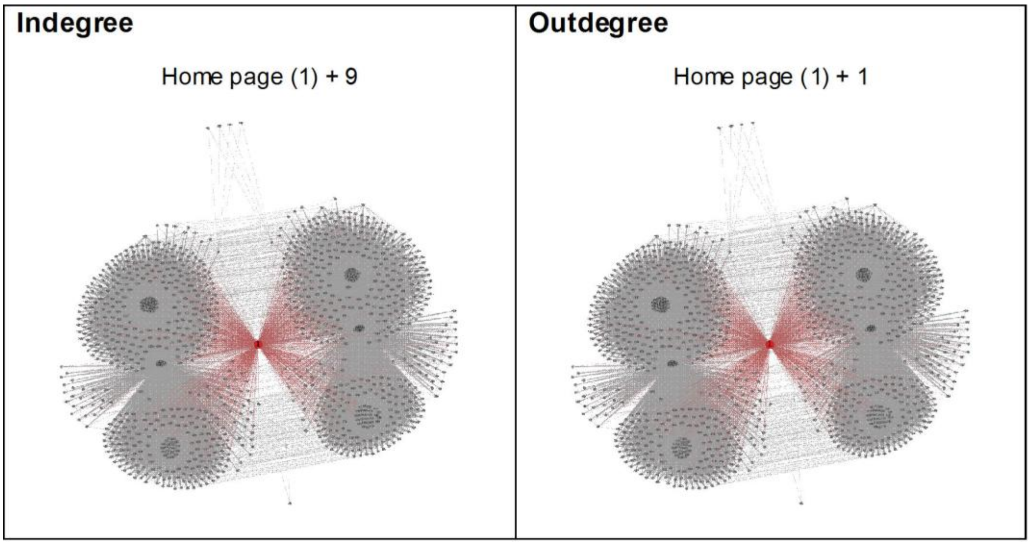


Revista Latina de Comunicación Social # 068 – Pages 383 to 419
Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

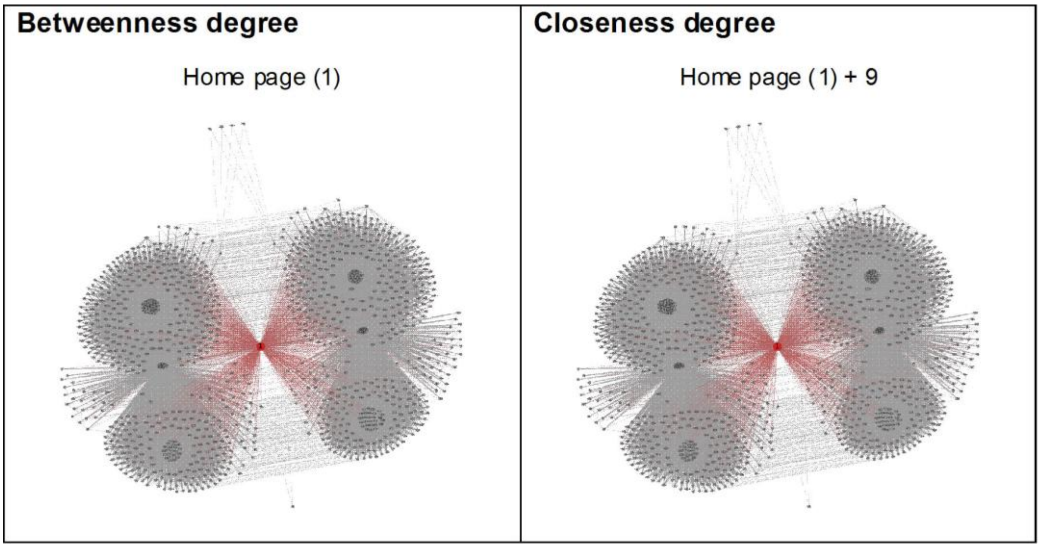
Graphs of the Institute of Pharmacoepidemiology



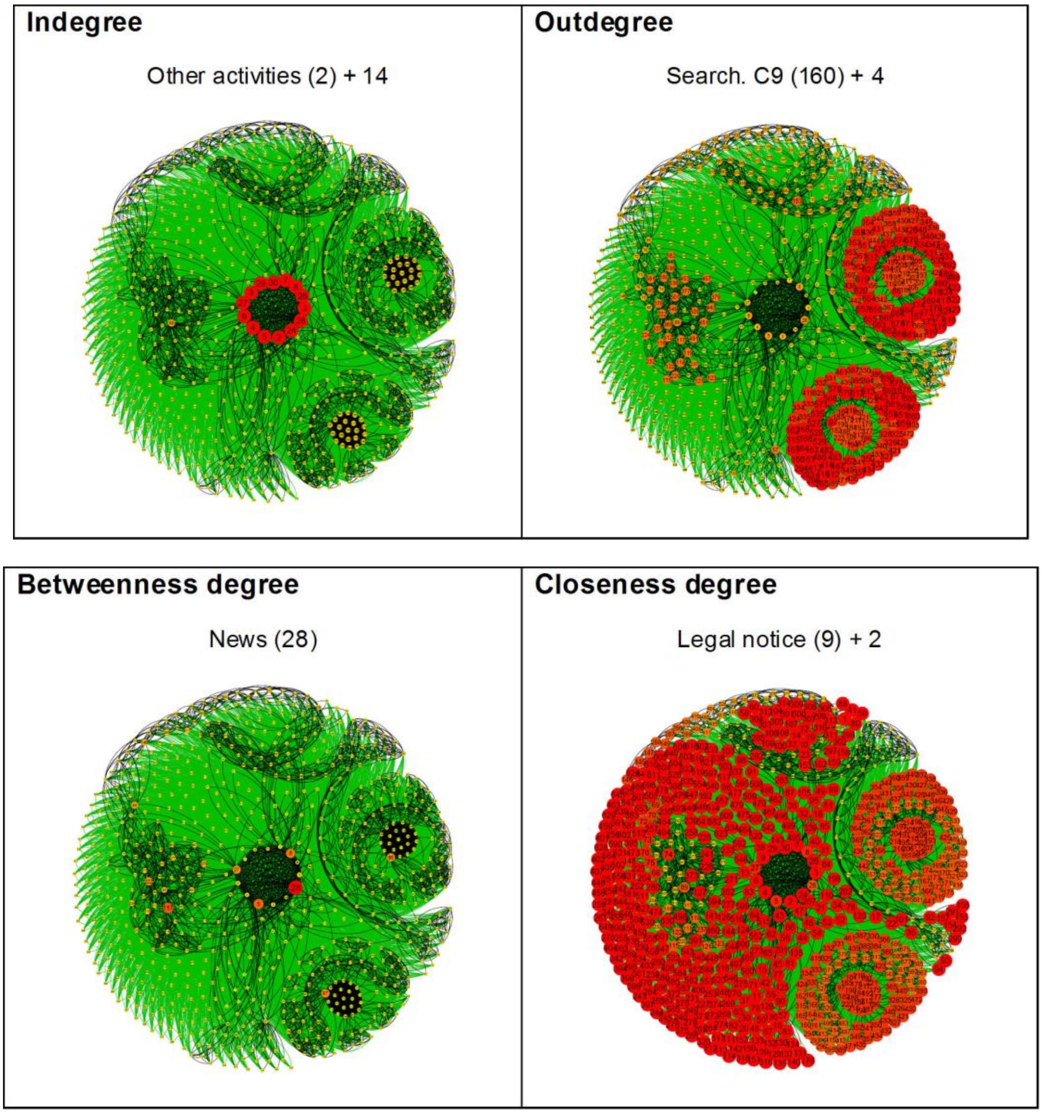
Graphs of the Institute of Neuroscience of Castile and León



Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

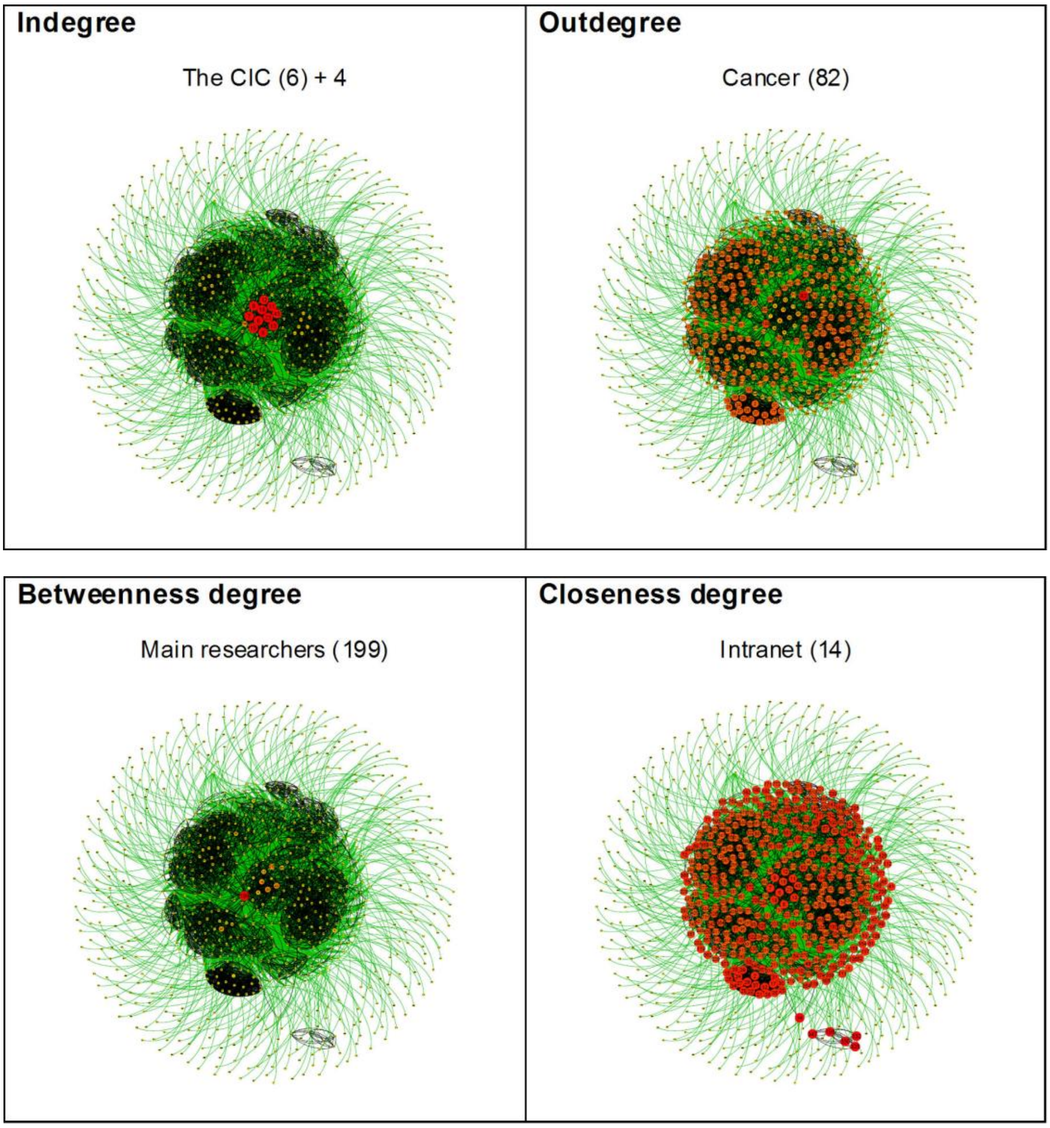


Graphs of the Institute of Applied Ophthalmobiology

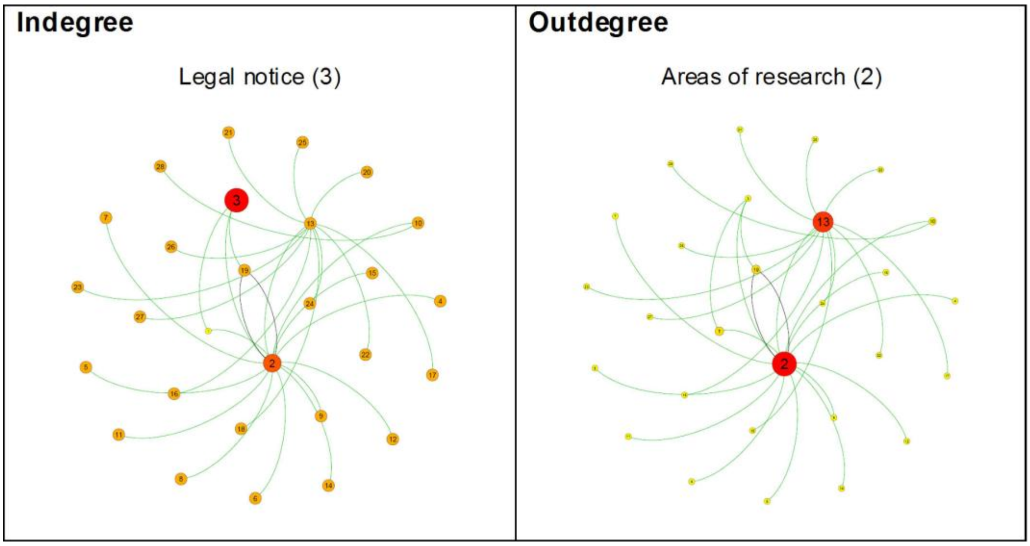


Graphs of the University Institute of Molecular Biology and Cancer Cell

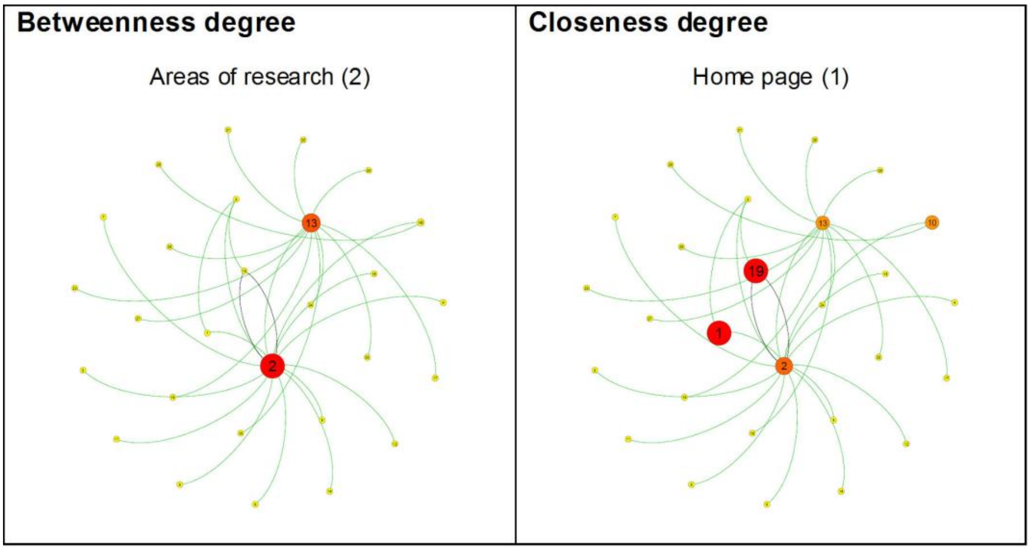
Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013



Graphs of the University Institute of Biomedicine



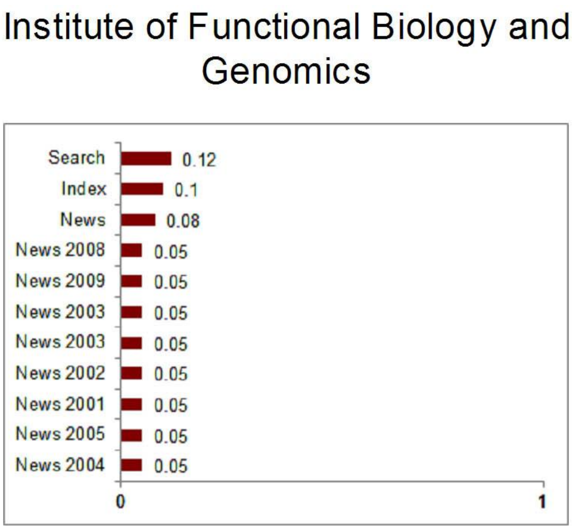
Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

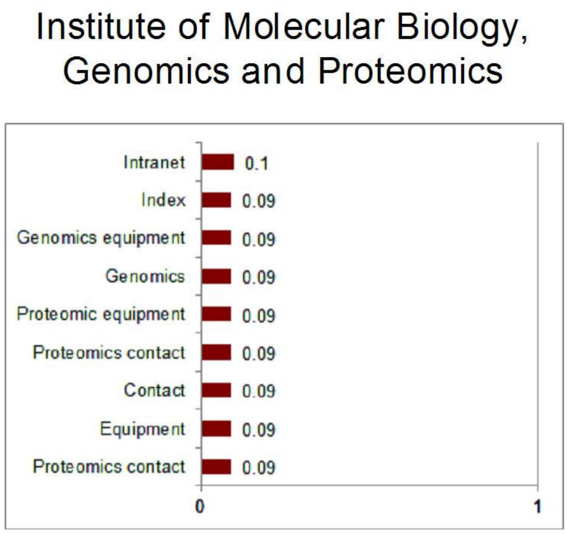


3.4. Positioning of networks

This section presents Google's PageRank for each of the websites. This is an excellent description of the importance of the different nodes of the networks, which is very significant since it shows the most valued pages according to the links received from other pages. This is very important if the institution that runs the website aims to gain a satisfactory presence in search engine results.

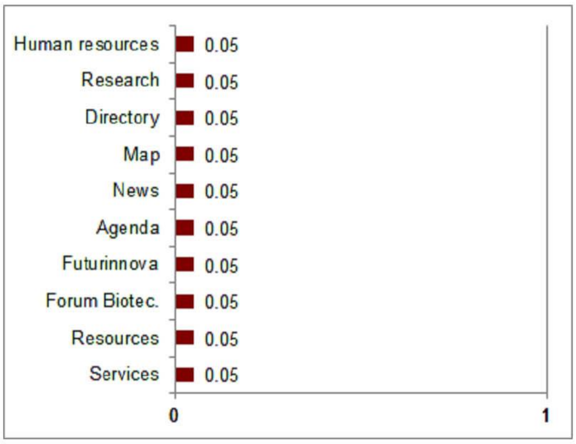
However, despite its importance, it has to be remarked that PageRank is a gross measurement system that does not measure the quality of the content of a website, but only its importance on the Internet. Therefore is important to realise that having a high PageRank is useless if the website is not optimised.





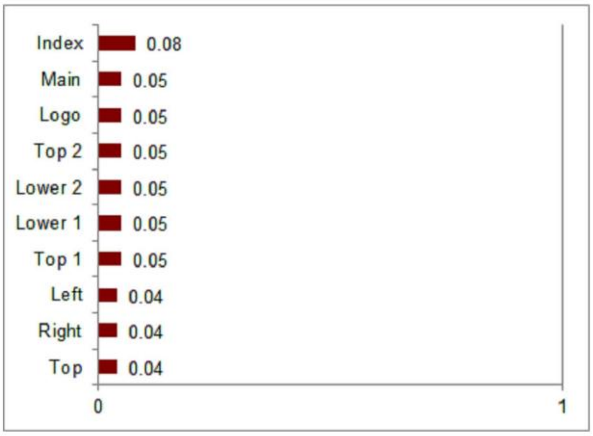
Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

Institute of Biotechnology

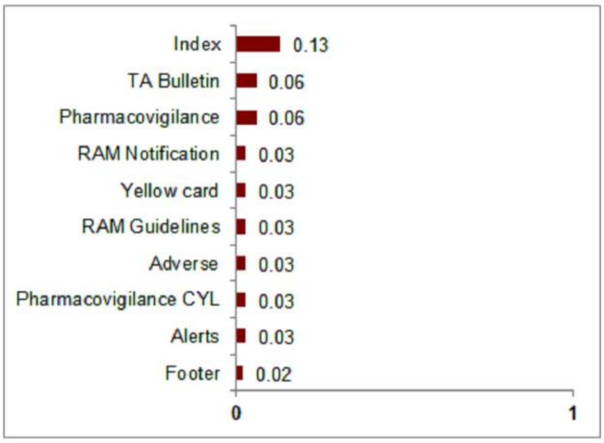


Institute of Endocrinology and Nutrition

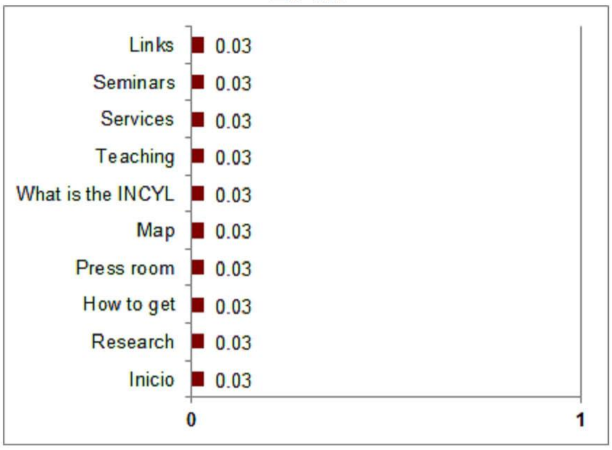
Institute of Alcohol and Drug Studies



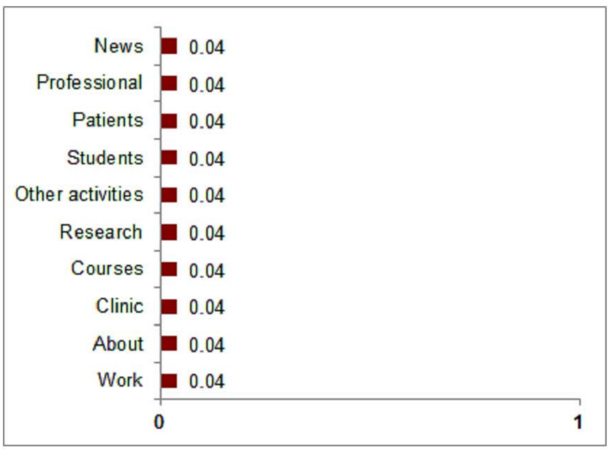
Institute of Pharmacoepidemiology



Institute of Neuroscience of Castile and León

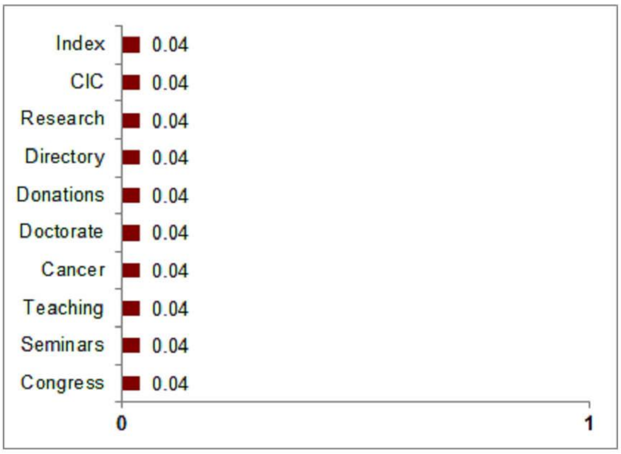


Institute of Applied Ophthalmobiology

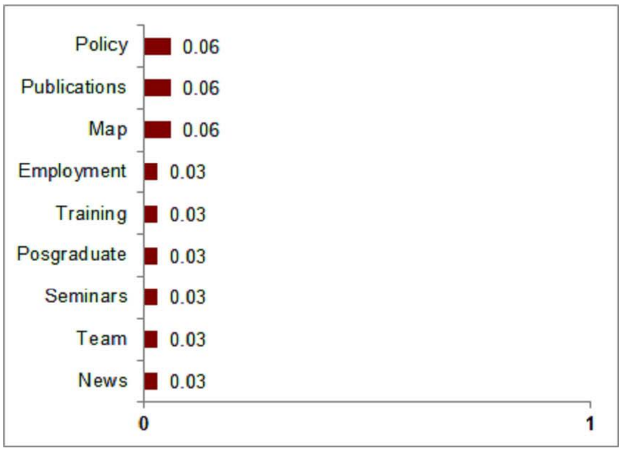


Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

University Institute of Molecular Biology and Cancer Cell



University Institute of Biomedicine



3.5. Bow-Tie analysis of the networks

In order to identify groups of nodes, according to the typology proposed by the Bow-Tie theory, this section presents the data obtained from the analysis of the sample of networks. The following tables present the relation of the types of nodes existing in the network, their number and the percentage they represent within the whole network.

Institute of Functional Biology and Genomics

Group	Nodes	Percentage
LSSC	16	80 %
IN	-	-
OUT	4	20 %
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

Institute of Molecular Biology, Genomics and Proteomics

Group	Nodes	Percentage
LSSC	12	92.31 %
IN	-	-
OUT	1	7.69 %
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

Institute of Biotechnology

Group	Nodes	Percentage
LSSC	42	97.67 %
IN	1	2.33 %
OUT	-	-
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

Institute of Endocrinology and Nutrition

Group	Nodes	Percentage
LSSC	405	97.59 %
IN	10	2.41 %
OUT	-	-
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

Institute of Alcohol and drug studies

Group	Nodes	Percentage
LSSC	1	4 %
IN	4	16 %
OUT	-	-
TUBES	-	-
TENDRILS	20	80 %
OTHERS	-	-

Institute of Pharmacoepidemiology

Group	Nodes	Percentage
LSSC	66	95.65 %
IN	-	-
OUT	3	4.35 %
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

Institute of neuroscience of Castile and León

Group	Nodes	Percentage
LSSC	2623	99.96 %
IN	-	-
OUT	1	0.04 %
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

Institute of Applied Ophthalmobiology

Group	Nodes	Percentage
LSSC	737	99.86 %
IN	1	0.14 %
OUT	-	-
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

University Institute of Molecular Biology and Cancer Cell

Group	Nodes	Percentage
LSSC	409	52.98 %
IN	-	-
OUT	363	47.02 %
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

University Institute of Biomedicine

Group	Nodes	Percentage
LSSC	2	7.14 %
IN	1	3.57 %
OUT	25	89.29 %
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

4. Conclusions

After the cybermetric study, based on social network analysis, of the websites of the biomedical research centres of Castile and León, the next step is to assess compliance with the overall objective of the research, as well as the interpretation of the results obtained.

The study produced a huge amount of information about the construction trends of this type of websites from an eminently quantitative point of view, which be a reference for future studies. The appropriate use of the tools and techniques previously described has allowed us to identify the composition of the networks under analysis, which was the objective of this research.

Moreover, the research hypotheses have been verified, as outlined below:

Hypothesis 1: These networks are composed of groups of strongly connected nodes.

The composition of the LSSC (Largest Strongly Connected Component) of nine of the 10 websites under analysis is over 50%, which indicates that there is a predominance of websites composed by strongly connected nodes. The networks with values below 50% are those belonging to the Institute of Alcohol and Drug Studies (INEAD), with 4%, and the University Institute of Biomedicine (IBIOMED), with 7.14%. These values remarkably stand out among the group of websites considering that the values range from 52.98%, like in the case of the University Institute of Molecular Biology and Cancer Cell, to 99.96%, like in the case of the Institute of Neuroscience of Castile and León.

Hypothesis 2: The home page is not the most important space in the website.

The importance of a page in particular, within the whole network, can be evaluated according to various aspects, so it would be wrong to evaluate it based on a single value, as it happens, on many occasions, with the calculation of Google's PageRank. Based on this fact, we decide to test the hypothesis by measuring the importance of the homepage and the other pages based on their indegree, outdegree, betweenness degree, closeness degree and internal PageRank.

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

Of the five proposed values, only in the indegree (visibility) there was a predominance of networks with the home page as the most outstanding node in the measurement, three (INEAD, IFE, INCYL) against eight.

In the outdegree (luminosity), two networks had the home page as the most important node (INBIOMIC, INEAD), against nine networks in which this post was occupied by another page.

In the betweenness degree, the ratio is two websites with the home page as the most important node (INBIOMIC; IFE), against nine.

In the case of the closeness degree, two networks had the home page as the most important node (INEAD, IBIOMED), while nine networks had other pages as the most important node.

Finally, the PageRank measurement showed that in the majority of websites the importance of the home page was low. In this case, in four websites the home page was the most important node (INEAD, IFE, INCYL, CIC), while this was not the case in six websites.

These results show that the home page is not always the most important part of a website, despite the dedication involved in its design, construction and maintenance. When the page with the highest PageRank is not the expected page the virtual communication and image strategy of the institution managing the website is obviously wrong. The websites under analysis are generally located in middle and low positions, however, this does not mean that their functioning or design has to be improved, but rather that their influence is average in the global environment that is Internet. As mentioned, the calculation of the PageRank does not only take into account the website's in and out links, but also the relevance of these links. This reflects some of the main features of the Internet: connectivity and relational activity, where an influential site can help other less influential sites to gain importance in the Web, so part of the success depends on off the page factors, i.e., those factors outside the control of the website in question.

Hypothesis 3: the websites that depend on a University's network do not share structural patterns

The morphological study of the graphs, produced for this study, allowed us to compare the sample of networks and to notice the great differences that exist between them, which confirm this hypothesis.

Like the graphical representation showed the differences between the three groups of networks, the data extracted from the measurement of these websites also corroborate the hypothesis, which is self-evident, since the graphical representation of the networks, in addition to relying on the Fruchterman-Reingold algorithm, includes aspects relating to measures of centrality, as well as the properties of the network.

However, it is important to remark that there were some minor similarities among the websites. Both the University Institute of Molecular Biology and Cancer Cell (CIC) and the Institute of Neuroscience of Castile and León (INCYL), both linked to the University of Salamanca, have a density of 0.01; the diameter of the Institute of Endocrinology and Nutrition (IENVA) and the Institute of Alcohol and Drug Studies (INEAD), both from the University of Valladolid, is 4 is both cases; and finally, the degree of incoming centrality of the Institute of Endocrinology and Nutrition and the Institute of Pharmacoepidemiology(IFE), both from the University of Valladolid, is 0.87 in both cases.

Hypothesis 4: The news pages, when they are included in the websites, have a high PageRank with respect to the whole website.

Of the 10 analysed networks, six have a communication section. Regardless of the unquestionable value of institutional communication, the internal quantitative analysis, based on the calculation of the internal PageRank, of each of the networks shows that the news sections (named "press room", "news" or "today"), tend to have a good PageRank in comparison to the whole website, which confirms this hypothesis. In this case, in four of the seven websites that have a communication section, this section was in their top ten most valued internal pages.

It should be noted that in two of the seven websites that have a communication section, this section was not in the top 10 most important pages because they were empty sections. This is the case of the Institute of Alcohol and Drug Studies

Hypothesis 5: The networks under analysis have good levels of connectivity.

The obtained results indicate that the studied networks indeed have good degree of connectivity, but there are certain networks that should undertake substantial changes in order to improve their connectivity value. Considering a diameter of six as an acceptable average, the present study concluded that six of the websites have a good connectivity (one of them in the acceptable limit), while four of them have connectivity problems. The values obtained by the networks of the Institute of Applied Ophthalmobiology (IOBA) and the University Institute of Molecular Biology and Cancer Cell (CIC) are outstanding, 25 and 17 respectively, which reflects the poor attention paid in the creation of these networks.

Hypothesis 6: The websites under analysis have low numbers of links per page.

The analysis of the density of the networks has confirmed this hypothesis, since only one website exceeded 0.5 in this indicator: the Institute of Molecular Biology, Genomics and Proteomics (INBIOMIC), whose density was 0.71. The density of the other websites ranged from 0.38 (like the website of the Institute of Biotechnology) to 0.01, which include the Institute of Neuroscience of Castile and León and the University Institute of Molecular Biology and Cancer Cell.

As a whole, these values are low since the average value was 0.1.

These findings reveal that the analysed networks have certain problems, which can lead to other types of deficiencies that affect the internet communication of the research centres selected for the study. Therefore, the information presented in this article can be useful to improve the institutional websites, and can serve as reference for the development and structural design of new websites.

As previously shown, many of the aspects examined in this study are vital in the exploration of web-based resources. This information is very useful to improve users' experience, which is very important in the current web development and particularly to achieve high quantitative values in the Internet, such as positioning, which determine the results offered by the large search engines like Google, Bing and Yahoo! and are part of the success or failure of certain contents in the digital ecosystem.

- This research project has been financed by the Government of Castile and León and the European Social Fund through a work contract (EDU/537/2010) (Reference)

5. List of References

Financed Research | DOI: 10.4185/RLCS-2013-982 | ISSN 1138-5820 | Year 2013

RH Abraham (1997): “Webometry: measuring the complexity of the World Wide Web”. *World Futures* 12, Philadelphia, pp. 785-791.

EG Adar (2006): “A language and interface for graph exploration”. *Computer Human Interaction*, Vancouver, pp. 791-800.

IF Aguillo, B Granadino (2006): “Indicadores web para medir la presencia de las universidades en la Red”. *Revista de universidad y sociedad del conocimiento* 3, pp. 68-75.

IF Aguillo, J Bar-Ilan, M Levene, JL Ortega (2010): “Comparing university rankings”. *Scientometrics* vol. 85, n. 1, pp. 243-256.

TC Almind, P Ingwersen (1997): Informetric analyses on the World Wide Web: methodological approaches to Webometrics. *Journal of documentation* vol. 53, n. 4, pp. 404-426.

JL Alonso, C García, A Zazo (2004): *Cybermetrics: nuevas técnicas de estudio aplicables al web*. Gijón: Trea.

JL Alonso, C García, A Zazo (2006): “Sacarino (Sonda AutomátiCA para la Recuperación de INformaciÓn en la Web): un robot para recorrer y procesar la Web”. *Scire* vol. 12, n. 1, pp. 211-224.

JL Alonso, C García, A Zazo (2008): “Recuperación de información web: 10 años de cybermetrics”. *Ibersid* n° 2, pp. 69-78.

N Arroyo, JL Ortega, VM Pareja, JA Prieto (2005): *Cybermetrics. Estado de la cuestión*. IX Jornadas Españolas de Documentación, Madrid, 14–15 April 2005, pp. 14-15.

V Batagelj, A Mrvar (1998): “Pajek-program for large website analysis”. *Connections* vol. 21, n. 2, pp. 47-57.

L Björneborn (2004): *Small-World Link Structures across an Academic Web Space: A Library and Information Science Approach*. København: Royal School of Library and Information Science, Department of Information Studies.

MJ Bossy (1995): “The Last of the Litter: Netometrics”. *Solaris* n. 2. Retrieved on 20/01/2012 from http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d02/2bossy.html.

Broder, A.; Kumar, R.; Maghoul, F.; Raghavan, P. (2000): “Graph structure in the web”. *Computer networks* vol. 33, n. 1, pp. 309-320.

S Chakrabarti, BE Dom, SR Kumar, P Raghavan (1999): “Mining the Web's link structure”. *Computer Websites and ISDN Systems* vol. 32, n. 8, pp. 60-67.

V Cothey (2004): “Web crawling reliability”. *Journal of the American Society for Information Science and Technology* vol. 55, n. 14, pp. 1228-1238.

MD Fernández Poyatos, A Aguirregoitia Martínez, B Boix Martínez (2011): "Camino de Santiago y Xacobeo 2010 en los portales turísticos de las Comunidades Autónomas". *Revista Latina de Comunicación Social* 67. La Laguna (Tenerife): Universidad de La Laguna, pp. 23 – 46. Retrieved on 16/01/2012 from http://www.revistalatinacs.org/067/art/946_Alicante/02_Lola.html.

LC Freeman (1979): “Centrality in social networks conceptual clarification”. *Social Websites* vol. 1, n. 3, pp. 215-239.

FJ Herrero Gutiérrez, A Álvarez Nobell, M López Ornelas (2011): "Revista Latina de Comunicación Social, en la red social Facebook". *Revista Latina de Comunicación Social*, 66. La Laguna (Tenerife): Universidad de La Laguna, pp. 526-548. Retrieved on 3/01/2013 from http://www.revistalatinacs.org/11/art/944_Salamanca/23_Javier.html. DOI: 10.4185/RLCS-66-2011-944-526-548 / CrossRef link

P Ingwersen (1998): "The calculation of web impact factors". *Journal of documentation*, vol. 54, n. 2, pp. 236-243.

J Izquierdo Castillo (2012): "Distribución online de contenidos audiovisuales: análisis de 3 modelos de negocio". *El Profesional de la Información* vol. 21, n. 4, pp. 385-390.

LG Jaimes, F Vega (2010): "Modelos clásicos de recuperación de la información". *Revista Integración* vol. 23, n. 1, pp. 17-26.

K Järvelin, P Ingwersen (2004): "Information seeking research needs extension towards tasks and technology". *Information Research* vol. 10, n. 1, pp. 10-11.

R Martínez Sanz (2012): "Estrategia comunicativa digital en el museo". *El Profesional de la Información* vol. 21, n. 4, pp. 391-395.

P Morville (2005): *Ambient findability: What we find changes who we become*. Sebastopol: O'Reilly Media Inc.

JL Ortega, I Aguillo (2008): "Análisis estructural de una red social en línea: la red española de Flickr. *El profesional de la información* vol. 17, n. 6, pp. 603-610.

JL Ortega, I Aguillo (2009): "Análisis estructural de la web académica iberoamericana". *Revista española de documentación científica* vol. 32, n.3, pp. 51-65.

P Shannon, A Markiel, O Ozier, NS Baliga (2003): "Cytoscape: a software environment for integrated models of biomolecular interaction networks". *Genome research* vol. 13, n. 11, p. 2498-2504.

M Thelwall (2008): "Bibliometrics to webometrics". *Journal of Information Science* vol. 34, n. 4, pp. 605-621.

JM Túñez, J Sixto (2011): "Redes sociales, política y Compromiso 2.0: "La comunicación de los diputados españoles en Facebook". *Revista Latina de Comunicación Social* 66. La Laguna (Tenerife): Universidad de La Laguna, pp. 210– 246. Retrieved on 23/01/2012 from http://www.revistalatinacs.org/11/art/930_Santiago/09_Tunez.html.

H Zamora, I Aguillo, JL Ortega, B Granadino (2007): "Calidad formal, impacto y visibilidad de las revistas electrónicas universitarias españolas". *El profesional de la información* vol. 16, n. 1, pp. 13-23.

HOW TO CITE THIS ARTICLE IN BIBLIOGRAHIES / REFERENCES:

F Sánchez-Pita, JL Alonso-Berrocal (2013): "The websites of the biomedical research centres of the Autonomous Community of Castile and León. A cybermetric analysis", at *Revista Latina de*

Comunicación Social, 68. La Laguna (Tenerife): La Laguna University, pages 383 to 419 retrieved on ___ de ___th of ____ of 2_______,
DOI: 10.4185/RLCS-2013-982

Article received on 29 January 2013. Submitted to pre-review on 1 February. Sent to reviewers on 3 February. Accepted on 12 May 2013. Galley proofs made available to the authoress on 19 May 2013. Approved by authoress on: 25 May 2013. Published on 12 June 2013.

Related papers:

Social networks as journalistic paradigm. Spanish media on Facebook
Estrategias locales en tiempos de globalización
A methodological proposal for the analysis of participation mechanisms in online media
Localization of international news in a global world
Comunicación en instituciones y organizaciones: una aproximación teórico-analítica a su diversidad conceptual
The communication in the organizations of the third sector
Los valores del periodismo en la convergencia digital: civic journalism y quinto poder
Analysis of Institutional Press Releases and its Visibility in the Press
La tiranía de la apariencia en la sociedad de las representaciones
Economic, political and communicative power in the neoliberal societies
Media representations of social networks: a case study
El valor social de la información de proximidad
The scientific productivity in Communication across the journal *Zer*

Communication Research Bibliometric analysis of the most-cited ISI-indexed Journals

Nuevos soportes para un periodismo necesitado de reflexión crítica
Presencia del formato weblog en los cibermedios: una aproximación a sus usos y funciones
El modelo de propaganda de Noam Chomsky, medios mainstream y control del pensamiento
Periodista digital: de MacLuhan a Negroponte
Las incertidumbres del periodismo en Internet