



Revista Latina de Comunicación Social

E-ISSN: 1138-5820

jpablos@ull.es

Laboratorio de Tecnologías de la Información
y Nuevos Análisis de Comunicación Social
España

Sánchez-Pita, F; Alonso-Berrocal, JL

Los sitios Web de centros de investigación biosanitaria de Castilla y León. Un análisis cibermétrico

Revista Latina de Comunicación Social, núm. 68, 2013, pp. 383-419

Laboratorio de Tecnologías de la Información y Nuevos Análisis de Comunicación Social

Canarias, España

Disponible en: <http://www.redalyc.org/articulo.oa?id=81928785016>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Los sitios Web de centros de investigación biosanitaria de Castilla y León. Un análisis cibermétrico

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

F Sánchez-Pita [\[CV\]](#) [\[ORCID\]](#) [\[GS\]](#) Instituto de Estudios de la Ciencia y la Tecnología. Universidad de Salamanca - fernandosp@usal.es

JL Alonso-Berrocal [\[CV\]](#) [\[ORCID\]](#) [\[GS\]](#) Departamento de Informática y Automática. Universidad de Salamanca - berrocal@usal.es

Abstracts

[ES] Introducción. Los sitios web institucionales son fuentes de contenidos donde entidad y usuario cruzan intereses en busca de objetivos comunes. Desde la presente investigación se aborda el análisis de este tipo de recursos digitales con una visión estructural de los mismos, más allá de su diseño o de la calidad de los contenidos que alojan. **Metodología.** Para ello se realiza un estudio desde el análisis de redes sociales, apoyado en la cibermetría, para conocer la composición de los sitios web de cada centro de investigación biosanitaria de Castilla y León. Se presentan diferentes propiedades de las estructuras de red, así como medidas topológicas con las que conocer las páginas más importantes de cada sitio web en función de los flujos de información, lo cual repercutirá en la visibilidad de los contenidos que se pretendan difundir. **Resultados.** Los resultados obtenidos indican que existen ciertos puntos de mejora en la estructura de los sitios web analizados, al detectarse exigüas tramas de relaciones informativas o páginas de inicio con poca relevancia sobre el conjunto. **Conclusiones.** Finalmente, tras la evaluación de resultados, se ha concluido que existen carencias en los sitios web analizados. Ello supone deficiencias que afectarán a los objetivos en base a los cuales se conciben este tipo de sitios web, dificultando, desde la propia estructura, la difusión de contenidos y la visibilidad de los mismos.

[EN] Introduction. Institutional websites are sources of content and spaces that bring together users and institutions with common interests and objectives. This article offers an analysis of this type of digital resources from a structural perspective that goes beyond their design or the quality of their content. **Methods.** To this end, this article offers a social network analysis, supported by cybermetrics, of the structure of the websites of the biomedical research centres of Castile and León. This study presents the different properties of the network structures and the topological measures that allow the identification of the most important pages of each website according to the flows of information, which influences the visibility of the contents that these websites want to disseminate. **Results.** The results indicate that there are certain areas that can be improved in the structure of the sample of websites. In particular, there the study detected a lack of information links and homepages that had little relevance in comparison to the rest of the website. **Conclusions.** Finally, after the evaluation of results, this study revealed deficiencies in the structure of the sample of websites, which will affect their objectives and will complicate the dissemination of contents and their visibility.

Keywords

[ES] Cibermetría; Internet; Análisis Web; Comunicación institucional.

[EN] Cybermetrics; Internet; Web analysis; Institutional communication.

Forma de citar este trabajo / Referencia normalizada

F Sánchez-Pita, JL Alonso-Berrocal (2013): "Los sitios Web de centros de investigación biosanitaria de Castilla y León. Un análisis cibermétrico", en *Revista Latina de Comunicacion Social*, 68. La Laguna (Tenerife): Universidad de La Laguna, páginas 383 a 419, recuperado el __ de __ de 2 ___, de http://www.revistalatinacs.org/068/paper/982_Salamanca/16_Sanchez.html
DOI: [10.4185/RLCS-2013-982/CrossRef link](https://doi.org/10.4185/RLCS-2013-982)

Contents

[ES] 1. Introducción. 2. Método. 3. Resultados. 4. Conclusiones. 5. Bibliografía

[EN] 1. introduction. 2. Methods. 3. Results. 4. Conclusions. 5. List of references.

Traducción de **CA Martínez Arcos** (Universidad Autónoma de Tamaulipas)

1. Introducción

Las últimas décadas han supuesto un cambio fundamental en el sistema español de ciencia y tecnología, debido, en gran medida, a las decisiones estratégicas para la introducción e implantación de las tecnologías de la información y la comunicación (Zamora, Agullo, Ortega, Granadino, 2007).

Este nuevo escenario, en el que los contenidos científicos son más accesibles que nunca, dejando atrás disponibilidades de horarios y de instalaciones, se facilitó el posicionamiento de los recursos Web como parte de lo que la FECYT ha denominado como e-Ciencia, es decir, actividades científicas realizadas mediante recursos distribuidos accesibles a través de Internet. Así mismo, la distribución online de contenidos plantea un nuevo escenario, donde han de conjugar los medios y actores tradicionales (Izquierdo, Castillo, 2012), junto con una nueva categoría de actores, considerados nativos digitales.

La nueva realidad planteada revive la idea de Infoesfera, del ensayista futurista de los 80 Alvin Toffler, en la que la principal criatura viva del ciberespacio es la información.

Gracias a Internet, la comunicación y la divulgación se convierten en elementos de inestimable valor, debido a las grandes posibilidades que brinda el nuevo entorno de trabajo a los profesionales de la información. Como base de tales posibilidades, se encuentran las páginas Web o los portales Web (Fernández, Aguirre-Gotia, Boix, 2011), siendo de interés para el presente estudio las webs institucionales de los centros públicos de investigación biosanitaria de Castilla y León.

Para los centros de investigación, Internet se ha convertido en un motor de cambio en su propia evolución. Se ha permitido el desarrollo de nuevos modelos de comunicación y docencia, en el caso de aquellos centros que cuenten con programas docentes. Ello posibilita desarrollar programas de formación a distancia y da pie a que el centro pueda desarrollarse más allá de su ámbito geográfico cercano. De este modo, las noticias de interés de centros de investigación, así como sus ofertas de servicios, salvan las limitaciones geográficas, extendiendo su campo de actuación y, por ende, sus capacidades.

Teniendo en cuenta la dimensión investigadora, algo propio de los centros de investigación, Internet ha favorecido la difusión de resultados, mostrando la capacidad investigadora de las entidades, a través de herramientas tales como los sitios web institucionales, los repositorios institucionales o las páginas personales, entre otros. Esto ha de sumarse a las variables de la direccionalidad del mensaje, ya que gracias al uso de herramientas digitales como las mencionadas se ha posibilitado el cambio de rol en los diferentes actores que forman los procesos de comunicación (Martínez-Sanz, 2012).

Todo ello ha supuesto muestras de prestigio, lo cual permite atraer nuevos recursos tanto económicos como humanos.

Esta muestra de capacidades y de difusión de contenidos a través de Internet, ha incrementado la competitividad entre instituciones de diverso tipo, no siendo ajeno a ello el caso de los centros de investigación, que pugnan por atraer nuevos recursos, siendo esta actividad objeto de evaluación en multitud de estudios y múltiples disciplinas, destacando especialmente a la cibermetría, desde la que se han orientado actividades hacia la explicación y descripción de la presencia investigadora en el Web, así como la posibilidad de evaluar el esfuerzo científico y educativo que se lleva a cabo en Internet (Thelwall, 2004).

El estudio de la presencia investigadora en Internet queda reflejado en trabajos como los realizados por el Laboratorio de Cibermetría del Consejo Superior de Investigaciones Científicas, que elabora cuatro rankings mundiales sobre la presencia en el Web de centros de investigación, hospitales, universidades, repositorios y escuelas de negocios, siendo de especial interés para la presente investigación los dos primeros. Los rankings elaborados por el Laboratorio de Cibermetría se actualizan dos veces al año, normalmente a finales de los meses de enero y julio, por lo que los datos reflejados en el presente estudio son los referentes las clasificaciones elaboradas en julio de 2012.

Hay que destacar que estas clasificaciones se ordenan en función de parámetros cibermétricos, relacionados con la presencia en el Web, y no por cuestiones relativas a la calidad del servicio o trabajo de las entidades. Si bien es cierto, en los casos en los que el rendimiento Web de una institución se encuentre por debajo de lo esperado, en función a la excelencia en otros ámbitos de la misma, los dirigentes de tales entidades deberían reconsiderar su política Web, lo que puede derivar hacia un incremento en el volumen y la calidad de las publicaciones accesibles a través de Internet.

Según los resultados vertidos de la última actualización, tanto en el caso de los centros de investigación, los primeros puestos están ocupados por entidades estadounidenses. En el caso de los centros de investigación, elaborando un top 100, la primera entidad española en aparecer es el CSIC, en el puesto 11. Le seguirán la Biblioteca Virtual Miguel de Cervantes, situada en el puesto 35, el Instituto de Salud Carlos III, en el 45, la Xarxa Telemática Educativa de Catalunya, en el 51, y el Instituto de Estudios Catalanes, ocupando el puesto 94.

Los datos expuestos dan cuenta de la escasa presencia Web de los centros de investigación españoles. Pero además, por el interés del presente estudio, hay que destacar que en las clasificaciones elaboradas, con 7532 centros de investigación, solamente se han encontrado tres centros de los incluidos en la presente investigación, ocupando en los tres caso posiciones muy bajas. Así, el Instituto de Oftalmología Aplicada se encuentra en el puesto 4003, el Instituto de Biología y Genética Molecular, en el 5459 y el Instituto de Biotecnología de León en el 7153.

Por ello, esta investigación pretende aportar nuevos datos sobre la situación real de la presencia Web de los centros de investigación biosanitaria, de perfil público, de Castilla y León. Ello se abordará desde una base cibermétrica, aplicando análisis de redes sociales, aportando así resultados clarificadores sobre los aspectos cuantitativos en la construcción de los sitios Web, así como las tendencias en la construcción de las mismas, apoyado todo ello en representaciones gráficas con un claro elemento visual como elemento central.

1.1. Justificación

Esta investigación parte de la premisa básica de conocer determinados aspectos cuantitativos de los sitios Web pertenecientes a los centros de investigación biosanitaria de Castilla y León de carácter

público. Los centros responsables de las redes analizadas, por su vocación de servicio público, tienen un deber implícito, desde el momento de su creación, de comunicar, de difundir que ocurre de puertas hacia dentro, tanto en clave divulgativa como científica. El valor añadido de la comunicación, no siempre atendido con la dedicación que se debiera, parece estar más vinculado al concepto de beneficio de la institución privada que al servicio desempeñado por la entidad pública, lo cual es algo totalmente erróneo.

A ello se ha de sumar el nuevo ecosistema digital en el que conviven este tipo de instituciones con sus públicos, es decir, lo que se conoce como Web 2.0, tal y como recogen Túñez y Sixto (2011: 2): “Las redes sociales y la filosofía 2.0 se han integrado en todo tipo de organizaciones –con o sin ánimo de lucro, públicas o privadas– de manera que Web 2.0 es también empresa 2.0, política 2.0 o Gobierno 2.0”. Dicha convivencia con los públicos está muy marcada por el peso de las redes sociales, que han determinado las nuevas vías de difusión de productos o servicios (Herrero, Álvarez, López, 2011).

Por ello, partiendo de una más que recomendable presencia en Internet, un análisis cuantitativo, basado en la cibermetría y con el apoyo del análisis de redes sociales, se convierte en un estudio de ciertas tendencias en la construcción de las estructuras sobre las que se basa el tejido comunicacional de entidades con necesidades informativas en los flujos de información hacia el exterior. De este modo, de la presente investigación se desprenderán conclusiones que buscarán aportar nuevas mejoras a las bases comunicación digital, desde un perfil métrico, aproximándose así desde un punto de vista eminentemente cuantitativo, a las realidades y necesidades de los sitios Web de los centros de investigación biosanitaria de Castilla y León.

1.2 Objeto y objetivos de la investigación

Los sitios Web analizados, como ya se ha dicho, son los pertenecientes a los centros de investigación biosanitaria de Castilla y León. A continuación se muestra la relación de las entidades con cuyos sitios Web se realizó la investigación:

- Instituto de Biología Funcional y Genómica
- Instituto de Biología Molecular, Genómica y Proteómica
- Instituto de Biotecnología
- Instituto de Investigación de Endocrinología y Nutrición Clínica
- Instituto de Estudios de Alcohol y Drogas
- Instituto de Farmacoepidemiología
- Instituto de Neurociencias de Castilla y León
- Instituto de Oftalmobiología Aplicada
- Instituto de Biología Molecular y Celular del Cáncer
- Instituto de Biomedicina

Sobre dicho objeto de estudio, se planteó un objetivo principal, entorno al que giraría el grueso de la investigación, planteado del siguiente modo: Conocer, desde una aproximación cibermétrica y mediante el uso de análisis de redes sociales, la composición de los sitios Web de los principales centros de investigación biosanitaria de carácter público de Castilla y León.

Para ello, el estudio se basará en la obtención de las medidas topológicas de cada una de las redes, así como a representación gráfica de las mismas, lo que facilitará la comprensión de los resultados.

Este objetivo planteado, busca conocer, desde un punto de vista cuantitativo, ciertas aptitudes comunicativas de los centros públicos de investigación biosanitaria de Castilla y León, basándose en el panorama digital de la comunicación institucional de este tipo de entidades, donde el entorno es de suma importancia.

En relación con el objetivo principal de la investigación, se plantean las siguientes hipótesis, estando todas ellas en relación con la obtención de medidas e índices de las redes analizadas en el estudio:

- Hipótesis 1: Se trata de redes compuestas por conjuntos de nodos fuertemente conectados.
- Hipótesis 2: La página de inicio no es la más importante de red.
- Hipótesis 3: Los sitios Web de los centros públicos de investigación biosanitaria de Castilla y León, que dependen de una misma universidad, no comparten patrones estructurales entre sí.
- Hipótesis 4: Las páginas de noticias y actualidad, en caso de existir, cuentan con un PageRank alto con respecto al conjunto de la red.
- Hipótesis 5: Las redes analizadas cuentan con buenos niveles de conectividad.
- Hipótesis 6: Los sitios Web analizados cuentan con exigidas tramas de relaciones.

1.3. Cibermetría y antecedentes

En origen, la cibermetría tenía dos objetivos básicos: El estudio de la evolución del tamaño de la web y la descripción de los primeros motores de búsqueda. A estas dos bases, se le han sumado nuevos aspectos, impulsados por los aportes teóricos de los distintos especialistas de la materia. Los orígenes de la cibermetría se pueden situar a mediados de los noventa (Aguillo, 2000), con la llegada de las primeras teorías al respecto (Bossy, 1995), (Abraham, 1997) y el creciente interés en la disciplina, que conllevó a la aparición de la revista electrónica Cybermetrics, en el año 1997, que fue presentada en el primer seminario sobre la diseminación de resultados del análisis cuantitativo de Internet (Aguillo, 1997) en el marco de las conferencias de la International Society for Scientometrics and Informetrics (ISSI).

Según Alonso, Figuerola y Zazo (2004, 74) se debe entender la cibermetría como la medida, el estudio y el análisis cuantitativo de todas las clases de información y de los medios de información que existen y que funcionan dentro del ciberespacio empleando técnicas bibliométricas, cienciométricas e informétricas.

Para Aguillo y Granadino (2006, 69), “La cibermetría es una disciplina emergente que, a partir de las técnicas y el modelo bibliométrico, pretende extender la aplicación de los métodos cuantitativos a la

descripción de los procesos de comunicación científica en Internet, a la determinación del volumen y de la tipología de los contenidos académicos en la Web, y a tratar de desentrañar las interrelaciones sociales y el consumo de información por los usuarios. Otros aspectos también susceptibles de estudio cibermétrico son la descripción de las herramientas de búsqueda en la Web, la llamada Internet invisible o las particularidades de los servicios basados en el correo electrónico y en los foros personales”.

Björneborn (2004) considera que la cibermetría es el estudio de los aspectos cuantitativos de la construcción y uso de los recursos de información, estructuras y tecnologías en Internet, desde perspectivas bibliométricas e infométricas. No se debe confundir este término con el de webmetría, que, a pesar de ser prácticamente lo mismo, se diferencia en el objeto de estudio, el Web, entendiendo esta disciplina como “el estudio de los aspectos cuantitativos de la construcción y el uso de la información, recursos, estructuras y tecnologías Web, a partir de enfoques bibliométricos e informétricos”.

En inicio, para designar a esta nueva disciplina, se utilizaron diversos términos, como señala Björneborn (2004): Netometrics (Bossy, 1995), Webmetry (Abraham, 1997), Internetometrics (Almind, Ingwersen, 1997), Web Bibliometry (Chakrabarti et al. 1999). Finalmente los términos adoptados fueron cybermetrics y webometrics, que se tradujeron al castellano como cibermetría y webometría. Estos dos términos, a menudo, se consideran sinónimos, algo incorrecto y que se puede aclarar fácilmente mediante el siguiente gráfico:

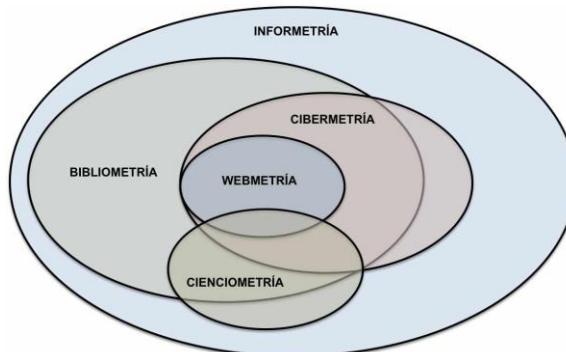


Figura N° 1: Cibermetría. Adaptado de Björneborn (2004)

Según las definiciones de Björneborn (2004) para cibermetría y webmetría, cabe destacar que ambas disciplinas tienen más parecidos que diferencias, ya que ambas trabajan sobre aspectos cuantitativos, se centran en el uso de recursos de información, estructuras y tecnologías, así como en la construcción de los mismos y están basadas en enfoques bibliométricos e informétricos.

Hoy, la cibermetría se postula como una posible solución para abordar el análisis de la explosión documental, basándose en el uso del hipertexto, en los enlaces entre páginas, los cuáles recuperan el papel de la cita en la literatura clásica. De tal modo se pueden localizar los documentos considerados más importantes, en función del número de enlaces que tenga.

A pesas de tratarse de una disciplina relativamente joven, la utilidad de la cibermetría ha quedado demostrada en diversos estudios previos, como los de Berrocal (1999, 2002), Cothey (2004) y Ortega y Aguillo (2008, 2009).

1.3.1. Creación de indicadores

La creación y diseño de indicadores es uno de los campos de la cibermetría que cuenta con más producción literaria científica. Ello es debido, principalmente, a la gran importancia de los índices de impacto. Al igual que sucedió con los sistemas clásicos de citas literarias y la aplicación a entornos digitales, los indicadores tradicionales han pasado a llevar el apellido Web. Así, se comenzaron a generar índices de impacto aplicables a entornos Web.

Ingwersen (1998, 237), definió el concepto de índice de impacto de la información Web como “la suma lógica del número de enlaces externos y autoenlaces de páginas Web que apuntan a un determinado sitio Web en Internet, dividido por el número de páginas que pueden encontrar en dicho sitio Web en momento en concreto. El indicador resultante es, por tanto, el número de páginas que enlazan, no el número de enlaces que se pueden encontrar”.

Otra definición de indicador Web, esta vez orientado hacia en análisis del sistema europeo de ciencia, tecnología y economía en Internet, es la ofrecida por el V Programa Marco de Investigación y Desarrollo de la Comisión Europea (EICSTES – European Indicators, Cyberspace and the Science-Technology-Economy System), indicando que se trata de “una medida que cuantifica los aspectos de creación, difusión y aplicación de la ciencia y la tecnología en la medida en que estas se encuentran representadas en Internet”. El mismo V Programa Marco de Investigación y Desarrollo de la Comisión Europea define 10 de los indicadores más característicos del estudio en Internet:

- Profundidad: Número de niveles de la estructura de un sitio Web, siendo la raíz en nivel uno.
- Densidad: Número total de enlaces por página, incluyendo enlaces salientes, internos, externos y los enlaces dentro la misma página.
- Conectividad: Número de enlaces diferentes en un sitio Web, incluyendo los enlaces salientes, internos y externos, pero excluyendo los enlaces dentro de la misma página.
- Navegabilidad: Densidad de los enlaces salientes internos de un sitio Web, teniendo en cuenta los enlaces repetidos.
- Endogamia: Porcentaje de enlaces salientes internos diferentes con respecto al número total de enlaces salientes diferentes.
- Luminosidad: Número de enlaces salientes externos, es decir, el número de enlaces desde un sitio Web a otros diferentes.
- Dispersión: Tipología y frecuencia de los enlaces salientes de un sitio Web, de acuerdo a diferentes criterios de distribución.
- Visibilidad: Número de enlaces externos recibidos por un sitio Web.
- Popularidad: Número de visitas recibidas por un sitio Web.
- Diversidad: Tipología y frecuencia de enlaces recibidos por un sitio Web, según distintos criterios de distribución.

1.3.2. Recuperación automatizada de información

La recuperación automatizada de la información es una disciplina relacionada directamente con la documentación y la informática, que tiene como objetivo, principalmente, la recuperación de registros y la discriminación del grado de ajuste de un documento en función de los términos de consulta formulados o las necesidades informativas del usuario (Salton y McGill, 1983).

Según Ricardo Baeza-Yates (1999), es la parte de la informática que estudia la recuperación de la información de una colección de documentos escritos, pudiendo estos satisfacer la necesidad de información del usuario, que en la mayor parte de los casos utiliza el lenguaje natural para concretar los términos de búsqueda.

Korfhage (1997) define la disciplina como la localización y presentación al usuario de una necesidad de información expresa en forma de pregunta.

En base a las definiciones recogidas, la recuperación de información se puede definir como la ordenación de documentos por relevancia en función de las condiciones de la consulta realizada y el conjunto de documentos total sobre el que se realiza la búsqueda.

La recuperación de la información se basa en una serie de modelos, de los que cabe resaltar los modelos clásicos, entre los que se encuentran el modelo booleano, el modelo vectorial y el modelo probabilístico.

El modelo booleano, que actualmente se encuentra en desuso, muestra los documentos como un conjunto de términos y las preguntas como expresiones booleanas.

El modelo vectorial es el más usado en las operaciones de recuperación de la información, así como en las de categorización automática y filtrado de la información, entre otras. En este modelo se eligen las palabras útiles, es decir, todas las palabras de los textos excepto las palabras vacías, y se enriquece mediante técnicas como la tematización y el etiquetado (Jaimes, Vega, 2010).

El modelo probabilístico indica que dada una consulta q y un documento d_j , se trata de estimar la probabilidad de que el usuario encuentre dicho documento d_j considerado relevante. De ese modo, se pretende observar la distribución de los documentos en el conjunto de la colección. El modelo supone que la probabilidad de relevancia sólo depende de la consulta y de la propia representación del documento.

Hoy, la recuperación de la información está ligada a los grandes motores de búsqueda accesibles a través de Internet, para poder recuperar información almacenada en el Web.

La Web es una colección de billones de documentos con vínculos entre si, y los crawlers son las herramientas gracias a las cuales se puede explorar la Web. Según (Arroyo et al. 2005), el uso de los crawlers es óptimo para realizar estudios en sitios Web o sedes Web concretas (nivel micro), sin embargo para el estudio de grandes cantidades de información (nivel macro) es necesario valerse de los motores de búsqueda.

Debido a las cantidades ingentes de información que se pueden encontrar en Internet, gran parte de la responsabilidad en la búsqueda y la recuperación de la información existente en la red le concierne a los motores de búsqueda.

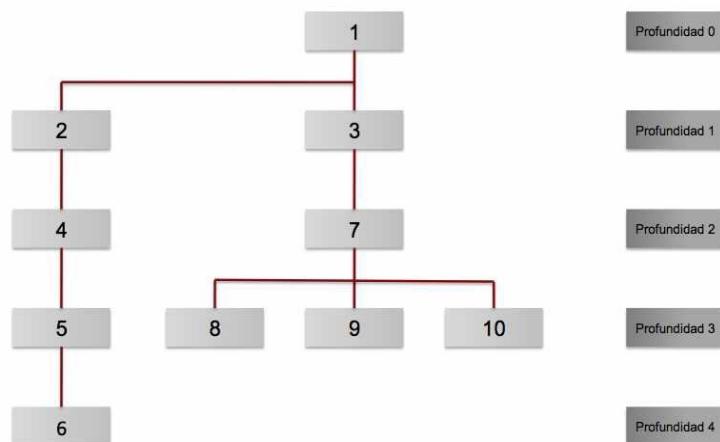


Figura N° 2: Crawling

Lo que se conoce como crawler, spider, wanderer, robot o bot, es una aplicación informática con la capacidad de recuperar páginas web extrayendo de estas sus redes de enlaces, para, a su vez, recorrerlas. Esta acción recibe el término genérico crawling, que puede darse de tres formas diferentes: Recorrido en anchura (breadth-first), Recorrido en profundidad (depth-first), El mejor posible (best-first).

A través de los crawlers, y otras aplicaciones específicas, es posible llevar a cabo la recuperación de la información, área de la ciencia que permite obtener información sobre una materia en concreto sobre una colección de datos mayor. Hay que destacar que la recuperación de información no debe ser confundida con la recuperación de datos, consistente esta última en determinar qué documentos, de una colección, contienen las palabras que el usuario utiliza para la búsqueda.

Actualmente existen multitud de crawlers con los que llevar a cabo un estudio cibermétrico, como el WebBot, del World Wide Web Consortium (W3C), SocSciBot 3 y SocSciBotTools, del Statistical Cybermetrics Research Group de la University of Wolverhampton, WIRE crawler, del Centro de Investigación de la Web de la Universidad de Chile o Webvac Spider, del Standford Infolab de la Standford University.

Con todos ellos se pueden llevar a cabo, de forma satisfactoria, estudios cibermétricos a nivel local. Para el presente estudio se ha optado por el software Sacarino, creado en la Universidad de Salamanca, y que está explicado en profundidad en el apartado metodológico de esta investigación.

2. Método

2.1. Planteamiento y metodología

La metodología utilizada para la consecución de la investigación ha sido de carácter cuantitativo, al girar en torno a la obtención de medidas e índice y a la representación gráfica de los datos recogidos, todo ello estrechamente vinculado a las técnicas de análisis de redes sociales, también denominado análisis estructural, o SNA (Social Network Analysis). Esta técnica de análisis se ha desarrollado como una herramienta de medición y análisis de estructuras que emergen de las relaciones entre actores de diverso tipo, en este caso nodos o páginas Web.

Para la obtención de las medidas y los índices se ha utilizado software específico, así como para la representación gráfica de las redes estudiadas.

Sacarino: Se ha realizado la recogida de datos, con los criterios técnicos que se exponen a continuación y las fechas de recogida, previamente planificadas en un calendario diseñado explícitamente para el presente trabajo de investigación.

EloisaBot Tools: Conversión de las recogidas realizadas con Sacarino de cara a su posterior uso en Gephi y Pajek, y obtención del PageRank de Google interno para los sitios Web.

Pajek: Obtención de medidas (Densidad, diámetro y centralidad de la red) y representación gráfica de las redes mediante la teoría Bow-Tie.

Gephi: Representación gráfica de las redes para los grados de entrada, salida, intermediación y cercanía. También se representó el PageRank calculado con EloisaBot Tools

Tras un estudio inicial, de carácter exploratorio, se cifraron 10 sitios Web de entidades con un perfil adecuado para ser incluidas en la investigación. Se ha trabajado con 10 sitios Web, por lo que el proceso de recogida de datos y posterior representación de los mismos se ha repetido en 10 ocasiones para poder concluir el estudio completo.

A continuación se indican las entidades seleccionadas: Instituto Universitario de Biomedicina, Instituto de Biotecnología, Instituto de Biología Molecular, Genómica y Proteómica, Instituto de Neurociencias de Castilla y León, Instituto de Biología Funcional y Genómica, Instituto Universitario de Biología Molecular y Celular del Cáncer, Instituto de Farmacoepidemiología, Instituto de Endocrinología y Nutrición, Instituto de Estudios de Alcohol y Drogas, Instituto de Oftalmobiología Aplicada.

La selección de los centros de investigación se realizó en base a su naturaleza pública y su ámbito de actuación: el marco biosanitario.

Los estudios de las redes se realizaron a través de medidas propias del análisis de redes sociales, gracias a las cuales se pueden conocer las principales características estructurales de una red en concreto. Mediante la aplicación de este tipo de análisis, fue posible estudiar las estructuras sociales que surgen de la recurrencia de las relaciones, en este caso, entre los nodos que forman las redes.

Partiendo de la asunción básica del análisis de redes sociales, la explicación de los fenómenos representados y su presentación gráfica mejoraría las propias estructuras en cambios venideros y nuevas construcciones que se decida realizar en los sitios Web analizados.

2.2. Software utilizado

Sacarino

En la presente investigación el software seleccionado, para realizar la recogida de datos de cada una de las 10 redes seleccionadas, fue Sacarino (Sonda Automática para la Recuperación de Información en la Web), desarrollado bajo la dirección de José Luis Alonso Berrocal, profesor del Departamento de Informática y Automática de la Universidad de Salamanca.

El origen de Sacarino se remonta a 1994, a partir del software denominado Sonda Ciberdocumental, empleado para realizar cálculos de tipo cuantitativo (Alonso Berrocal, 1996). Esta versión fue

implementándose con los años, hasta que en 2003 se decidió reprogramar por completo el software, con un lenguaje de programación nuevo a la vez que se optimizaron las estructuras de datos, lo que otorgó al programa más potencia, rapidez y flexibilidad. De este modo, y aplicando toda la experiencia adquirida con los trabajos de la Sonda Ciberdocumental, surgió Sacarino, favoreciéndose también de los proyectos fin de carrera de los alumnos María del Carmen Montejo Villa y Faustino Frechilla Daza, ambos de la titulación Ingeniería en Informática de Sistemas de la Universidad de Salamanca.

Para llevar a cabo una correcta recogida de datos, óptima para el posterior tratamiento de los mismos, se utilizaron los siguientes parámetros en la configuración de Sacarino:

- Rapid fire de 250ms
- Time Out: 1000 ms
- Número de hilos: 20
- Límite de la exploración: Host especificado y Host y directorio especificado, variando en función de las características de la URL de partida.
- Limitar nivel de profundidad: Sin límite
- Páginas dinámicas: Recorrer todas las páginas
- Priorización de URLs: Balanceo de servidores

Durante el proceso de recogida de datos hubo que tener en cuenta ciertas particularidades de los sitios Web analizados, ya que en unos casos contaban con URLs propias, mientras que en otros se disponía de directorios dependientes de las URLs principales. Además, en algunos casos hubo que aplicar restricciones de búsqueda en la configuración de Sacarino para impedir la exploración de ciertos elementos, tales como los calendarios que, propios de las secciones “agenda” o “noticias”, como en el caso del Instituto de Estudios de Ciencias de la Salud de Castilla y León.

Eloisa

EloisaBoot Tools es un paquete de herramientas de gran utilidad para el procesamiento de datos. Utiliza una interfaz de tipo MDI (Multiple document interface) que permite mostrar varias ventanas bajo el marco de la ventana principal, lo cual facilita el uso de la aplicación en un entorno multitarea. El proyecto, dirigido por José Luis Alonso Berrocal, profesor del Departamento de Informática y Automática de la Universidad de Salamanca, surge de la necesidad de agrupar diversos programas utilizados durante procesos de trabajo como el presente bajo un único entorno.

Las funciones básicas de EloisaBot Tools son el procesamiento de recogidas, el tratamiento de los grafos generados, el cálculo de algoritmos y el cálculo de índices y medidas. Gran parte de las rutinas están elaboradas para Matlab, debido a la potencia de cálculo que ofrece.

A diferencia de Sacarino, con el programa EloisaBot Tools no fue necesario realizar una configuración exhaustiva para llevar a cabo las actividades de la investigación.

En este caso se cargaron las respectivas recogidas de datos para realizar su conversión al formato de lectura en Pajek, que también serviría para su interpretación en Gephi. Tras ello, se obtuvo el PageRank de Google, que se facilita a través de tablas de datos de las que se debe extraer la información.

Pajek

Pajek (Araña en esloveno) es un programa para sistemas operativos Windows que permite el análisis y la visualización de redes. Se trata de software gratuito, desarrollado por Vladimir Batagelj y Andrej Mrvar, ambos profesores de la Universidad de Ljubljana, Eslovenia, que cuenta con una gran comunidad que permite su continuidad en el tiempo, así como la constante inclusión de mejoras.

Los principales objetivos, en base a los que se desarrolló el software fueron la posibilidad de contar con una potente herramienta de visualización, la implementación de algoritmos eficientes para la representación de grandes redes y la factorización de grandes redes en otras redes de menor tamaño, para ser tratadas mediante métodos más sofisticados (Batagelj, Mrvar, 1998).

Permite la obtención de gran cantidad de medidas, además de alta personalización de la representación gráfica de las redes, aunque con una estética 2D. Mediante el uso de Pajek se obtuvieron las siguientes medidas, índices y representaciones:

- Densidad
- Diámetro
- Grado de centralidad de entrada
- Grado de centralidad de salida
- Relación de nodos en función de la teoría Bow-Tie
- Representación gráfica de la estructura Bow-Tie

El proceso de trabajo con Pajek consistió en la carga de cada uno de los ficheros convertidos con EloisaBot Tools, la obtención de datos mediante tablas generadas por el propio programa y la representación gráfica de las redes.

Para el proceso de representación gráfica se utilizó la presentación dispuesta por el algoritmo Fruchterman Reingold, en su versión 2D. Dicha representación de las redes se hizo en función de la teoría Bow-Tie, por lo que se utilizó la siguiente relación de colores, de cara a la diferenciación de los distintos tipos de nodos que conforman las redes:

- LSSC (Largest Strongly Connected Component): Rojo
- IN: Verde
- OUT: Rosa
- TUBES: Amarillo
- TENDRILS: Azul
- OTHERS: Naranja

Tras la representación gráfica de cada una de las 11 redes, se exportaron los grafos generados en los formatos SVG y BMP, siendo las primeras óptimas para futuras consultas dadas particularidades del formato vectorizado, –el cual que permite una gran ampliación para la consulta en detalle–, y las segundas para ser incrustadas en el presente documento.

Gephi

Gephi es un programa open source multiplataforma (Windows, Linux, Mac OS) para analizar redes y crear grafos derivados del análisis de las mismas. Basa su módulo de visualización en un motor 3D capaz de renderizar los grafos en tiempo real, es decir, el usuario puede observar el movimiento de los nodos, los arcos y las aristas en función de la disposición que decida. Para ello, se fundamenta en el uso de tarjetas gráficas, tal como sucede en los editores de video y fotografía más avanzados, que permiten al equipo informático con el que se desarrolla el trabajo liberar memoria para otras tareas.

En cuestiones de diseño se trata de un paquete de software que ofrece múltiples posibilidades de presentación de los grafos, ya que cuenta con varios algoritmos configurables con los que distribuir los nodos, así como opciones de color y forma.

Además cuenta con un módulo de exploración de las redes en modo interactivo, en el que poder observar los resultados de las distintas mediciones de forma gráfica.

Al igual que sucede con Pajek, un aspecto destacable de Gephi es la posibilidad de exportar las representaciones gráficas en formato SVG, además de formato PDF. Con Gephi, las medidas obtenidas fueron:

- Nodo con mayor grado de entrada
- Nodo con mayor grado de salida
- Nodo con mayor grado de intermediación
- Nodo/s con mayor grado de cercanía

Además para cada uno de los resultados vertidos de dichos cálculos, se generó un grafo para una correcta visualización del nodo en el conjunto de la red. Para la obtención de las medidas anteriormente expuestas, se activaron las siguientes estadísticas en la interfaz de Gephi:

- Average Degree
- Network Diameter
- PageRank

Tras ello, se hizo la discriminación por tamaño y forma de los nodos, en función del valor de los mismos. En el caso del color se partió del amarillo para los valores más bajo, hasta el rojo para los valores más altos, lo que, debido a las posibilidades de mezcla disponibles en Gephi, provocó que los valores intermedios resultaran de color naranja.

En el aspecto del tamaño, se partió de un tamaño 10, para los nodos con menor valor, y 30 para los de mayor valor.

Para la obtención de los grafos de las redes, se optó por la disposición generada por el algoritmo Fruchterman-Reingold. Una vez finalizada la colocación de los nodos según las reglas de atracción y repulsión propuestas por el algoritmo.

Para facilitar el proceso de interpretación de los grafos se modificó la representación en base a la disposición Default Curved, a lo que hubo que añadirle nuevos colores, debido a las grandes cantidades de enlaces existentes.

2.3. Índices y medidas aplicables

2.3.1. Propiedades de la red

Densidad

A través de la densidad, se mide la proporción de enlaces que existen entre las relaciones posibles de una red en concreto. Gracias a este indicador se puede apreciar la intensidad de las relaciones en el conjunto de la red, pudiendo variar el resultado entre 0 y 1, siendo los mejores aquellos resultados más próximos al 1.

La fórmula para realizar el cálculo puede variar, en función de las características de la red. El presente estudio se basa en redes dirigidas, por lo que la densidad se obtiene con la siguiente función:

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

N: nº de nodos r: nº de enlaces

En el supuesto de tratarse de redes no dirigidas, la fórmula variaría de este modo:

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

N: nº de nodos r: nº de enlaces

Diámetro

Se trata de la distancia geodésica más larga que se puede encontrar en el grafo obtenido del análisis de una red en concreto. Como distancia se debe entender el esfuerzo necesario que se debe realizar para que un nodo alcance a otro, o lo que es lo mismo, el número de relaciones existentes en el camino más corto entre nodos.

Del mismo modo, también se puede decir que el diámetro se corresponde con el número de clicks necesario que se deben dar para llegar de un nodo en concreto al nodo más alejado.

2.3.2. Medidas de centralidad

Las medidas de centralidad permiten conocer determinados grados de importancia entre los nodos que conforman una red, en función de su centralidad, su prestigio o su poder. Freeman (1979), en su revisión sobre la literatura de la centralidad, señaló que el concepto de centralidad se podía definir, así como hacer operativo, en base a tres formas:

Grado

El grado es la medida que muestra el número de enlaces directos que tiene un nodo que, en el caso de los grafos dirigidos, tiene dos variantes: Grado de entrada y grado de salida.

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

V: Conjunto de nodos A_{ij} : Matriz de adyacencia

Grado de entrada o indegree

El grado de entrada, o indegree, es la medida que indica el número de relaciones entrantes de las que dispone un nodo en concreto, es decir, indica el número de nodos que apuntan hacia dicho nodo.

Mediante el cálculo del grado de entrada, es posible conocer cuáles son los nodos con más prestigio de una red. Esto también se traduce en poder sobre el conjunto de los nodos que conforman la red.

Tras la pertinente normalización de la medida obtenida tras el cálculo, esta puede redefinirse bajo el concepto de visibilidad.

Grado de salida o outdegree

El grado de salida, o outdegree, da cuenta del número de relaciones que se inician desde un nodo en particular. Esta medida, además de recoger el número de enlaces salientes de un nodo hacia otros nodos de la red, es un reflejo de la actividad de dicho nodo con respecto al conjunto, mostrando su capacidad para acceder a distintos lugares de la red.

Al igual que sucede con el grado de entrada, tras el proceso de normalización de las medidas obtenidas, el grado de salida puede redefinirse como luminosidad.

Grado de intermediación o betweenness

Esta medida indica hasta qué punto en concreto un nodo de la red se encuentra en una posición intermedia con respecto al conjunto de los nodos que conforman la red. Los nodos con un grado de intermediación mayor serán poderosos, ya que en ellos está el mayor control de los flujos de información. El grado de intermediación mide el número de veces que un nodo aparece en las rutas existentes en la red, lo que, desde un punto de vista cibermétrico, permite detectar pasarelas que conectan diferentes tramos de la red (Ingwersen, 1998).

Para la obtención del grado de intermediación, o betweenness, la ecuación a utilizar sería la siguiente:

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

g_{ij} : Número de pasos necesarios para llegar del nodo i al nodo j

g_{ikj} : Número de los pasos de g_{ij} que atraviesan el nodo k

Grado de cercanía o closeness

Muestra la distancia media de cada nodo con el conjunto de nodos que conforman la red. En este caso, los nodos que obtienen resultados más altos tienen una mayor facilidad de acceso al resto de nodos que forman la red. Este hecho otorga a los nodos con un mayor grado de cercanía una mayor capacidad para el envío y la recepción de información.

$$C_i = \frac{1}{|V|} \sum_{j=1}^{|V|} D_{ij}$$

D_{ij} : Número de pasos necesarios para llegar del nodo i al nodo j

$|V|$: Tamaño del conjunto de los nodos

2.3.3. Posicionamiento

El posicionamiento de los distintos sitios Web se mide mediante el Pagerank, el cual determina que la valoración de una página web está influenciada por el número de enlaces que recibe desde otras páginas y a su vez matizada por la importancia de las páginas de las que provienen dichos enlaces. La fórmula básica para obtener dicho cálculo es la siguiente:

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

x_i : Importancia de la página i

$j \in B_i$: Páginas j que enlazan a la página i

N: Número de enlaces salientes desde la página j

x_j : Importancia de la página j

2.3.4. Bow-Tie

La teoría de la pajarita o bow-tie, nace a partir de estudios del Web realizados por (Broder et al. 2000), en los que extrajeron datos almacenados en Altavista y procesaron 200 millones de páginas junto con 1,5 billones de enlaces. Tras ello, detectaron un núcleo de páginas fuertemente conectadas, unos lazos a ambos lados del núcleo –con enlaces de salida y entrada al mismo–, hilos de salida y entrada en cada uno de los lazos –de salida y entrada no definida–, y una serie de componentes no conectados.

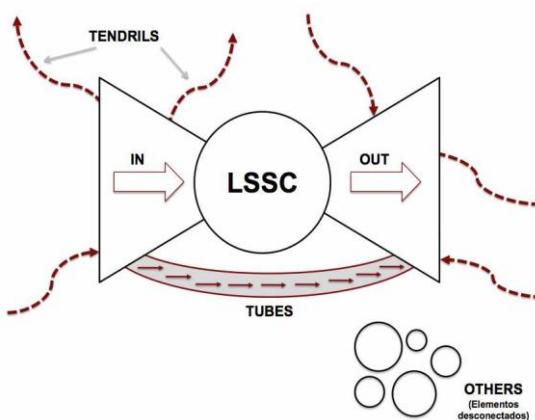


Figura N° 3: Bow-Tie. Adaptado de Broder et al. (2000)

La parte central la forma un grupo de páginas fuertemente conectadas (LSSC o Largest Strongly Connected Component) y todas sus páginas pueden trazar rutas directas de unas a otras. Una de las partes contiguas al LSSC es el grupo de páginas de salida (OUT), a los que se puede acceder desde el núcleo, pero sin poder realizar el camino inverso. En el caso del grupo de entrada (IN) se permite acceder directamente al núcleo, pero no salir.

Entre al grupo un y el grupo out se encuentra una tercera agrupación denominada Tubes, que conecta la zona de in con la de out, sin que tengan que pasar las rutas por el núcleo.

Ajenos a las rutas explicadas se encuentra un grupo de elementos desconectados, también denominado islas y, con cierta conexión con las tres zonas principales del modelo presentados, se encuentran los Tendrils, que pueden ser de salida o entrada y que toman rutas en diferentes direcciones (Alonso et al., 2008).

2.4. Representación gráfica de redes

Un grafo es un conjunto de líneas y vértices, que nos permite representar la estructura de una red. En dicha representación, los vértices, llamados nodos, están comunicados mediante líneas, las cuales pueden ser de dos tipos: Arcos o aristas.

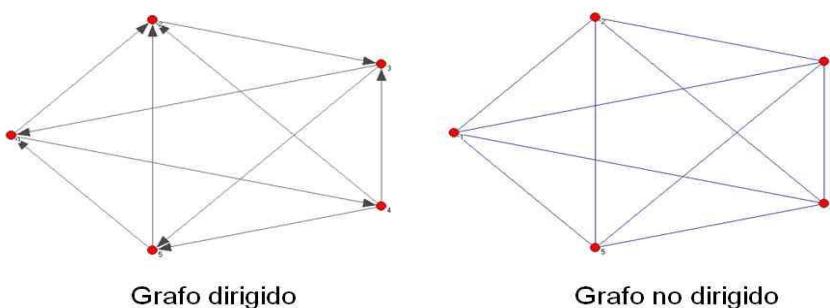


Figura N° 4: Grafos dirigidos y no dirigidos

Los arcos son líneas dirigidas, con un único sentido en la conexión entre nodos, mientras que las aristas son líneas no dirigidas, dando así lugar a conexiones entre nodos de carácter bidireccional. Esto repercute notoriamente en la composición de los grafos, catalogándolos como grafos dirigidos, también llamados dígrafos, y grafos no dirigidos, en función de su composición.

La teoría de grafos es un aporte muy valioso al campo de la cibermetría, debido a sus algoritmos de extracción de datos y búsqueda e identificación de datos (Arroyo et al., 2005).

Las representaciones gráficas derivadas de la teoría grafos, favorecen la visualización de grandes cantidades de información, algo que se ha desarrollado con éxito durante muchos años en investigaciones previas (Batagelj, 1998) (Shannon, 2003) (Adar, 2006). De este modo, mediante la visualización gráfica se maximizan las habilidades humanas para extraer información de las características de las redes y de los propios datos. Sin embargo, este proceso requiere una estrategia de exploración por su dificultad.

Además de esta representación gráfica, una red también se puede visualizar por medio de matrices de adyacencia. Por ejemplo, en una matriz n-por-n cuyas entradas en la fila i y la columna j dan el número de arcos desde el nodo i-ésimo al j-ésimo.

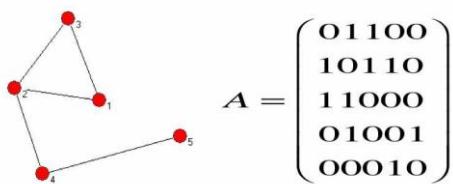


Figura N° 5: Matriz de adyacencia

Formalmente, la matriz de adyacencia se define como $V = \{v_1, v_2, v_3, \dots, v_n\}$, de modo que:

$$M = \begin{cases} 1 \text{ si } (v_i, v_j) \in G \\ 0 \text{ en otro caso} \end{cases}$$

Las estructuras de enlaces, tras la representación de las redes, ya sea a través de grafos o matrices de adyacencia, permiten advertir los patrones estructurales del sitio Web en cuestión, a través de índices y medidas de gran utilidad. Así, dependiendo de las funciones de un sitio Web, su estructura de enlaces estará marcada por unas tendencias de construcción u otras, lo que permite, entre otras cosas, la comparación entre sitios Web con objetivos similares. Además, de las líneas generales de la red, se pueden obtener índices y medidas referentes únicamente a nodos de forma individual.

Para la realización de los grafos, independientemente de sus características, se debe utilizar software específico, de cara a una correcta representación final, mediante programas como Pajek, Gephi, Graphviz o Touchgraph. En la presente investigación se utilizaron Pajek y Gephi, aunque esto se explicará en el apartado metodológico.

Los programas cuentan con diversas herramientas de representación, acompañados de diferentes opciones de diseño en la visualización final, como Fruchterman & Reingold (Fruchterman & Reingold, 1991), Kamada & Kawai (Kamada & Kawai, 1989), o diseños basados en la expansión, la contracción o simplemente aleatorios.

De entre todas las opciones disponibles en cada uno de los paquetes de software antes mencionados, fue la propuesta por el algoritmo Fruchterman & Reingold la que produjo los gráficos más claros y más comprensibles, por lo que se desestimaron las demás representaciones.

El algoritmo de Fruchterman & Reingold es un algoritmo para la colocación de los nodos basado en la fuerza-dirigida, lo que produce dibujos de grafos en dos dimensiones, mediante la simulación simplificada de sistemas físicos.

El método propuesto por Fruchterman & Reingold compara el grafo con una colección de anillos cargados eléctricamente que se encuentran conectados mediante enlaces. El sistema de funcionamiento se basa en que cada dos nodos se produce un rechazo entre si, mediante una fuerza repulsiva, y los nodos adyacentes, aquellos que están conectados por un enlace, son atraídos entre si, en este caso por una fuerza atractiva.

Tras este funcionamiento, se dan una serie de iteraciones y se vuelven a calcular las fuerzas que modelan cada uno de los enlaces, mientras que los nodos se mueven para reducir dichas fuerzas.

El cálculo de la distancia óptima entre vértices, que luego revertirá en la representación gráfica de las redes, se obtiene mediante la aplicación de la siguiente fórmula:

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

K: Distancia óptima entre vértices

C: Constante C, recurso experimental

A: Área

N: Número de vértices

A continuación se muestra la representación gráfica del sitio Web del Instituto de Farmacoepidemiología, mediante los algoritmos Fruchterman & Reingold y Kamada & Kawai, donde se pueden apreciar las diferencias que llevaron a la elección final del algoritmo Fruchterman & Reingold para la representación:

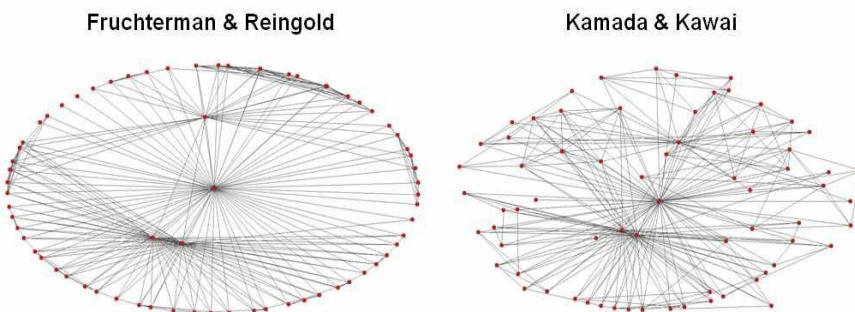


Figura N° 6: Algoritmos

3. Resultados

3.1. Propiedades de las redes

Como se puede apreciar, las tres redes con una mayor densidad son las del Instituto de Biología Molecular Genómica y Proteómica, con un 0,71, el Instituto de Biotecnología, con un 0,38 y el Instituto de Biología Funcional y Genómica, con un 0,33.

El conjunto de redes cuenta con unos niveles de densidad muy bajos, entre el 0,11 y el 0,01, lo cual es significativo recordando que los resultados al calcular la densidad de una red van del 0 al 1. Salvo en el caso del Instituto de Biología Molecular Genómica y Proteómica, que cuenta con una densidad del 0,71, en el resto de los casos sería recomendable hacer modificaciones de cara a reforzar las líneas de comunicación entre los distintos nodos que conforman las redes, lo que mejoraría sustancialmente este parámetro.

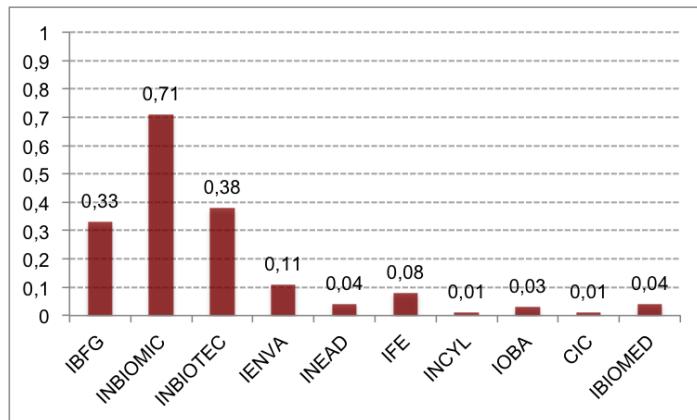


Figura N° 7: Densidad

Con respecto al diámetro de las redes estudiadas, el conjunto muestra un rango de resultados entre el 2 del Instituto de Biología Molecular, Genómica y Proteómica, y el 25 del Instituto de Oftalmobiología Aplicada. En base a sus diámetros, se trata de redes que requieren varios pasos para que puedan ser recorridas. En este caso, el sitio Web del Instituto de Biología Molecular, Genómica y Proteómica cuenta con un diámetro Web más favorable que el resto. Frente a una media de 6,9 del conjunto de las redes, cabe destacar que hasta seis redes se encuentran por encima de esta, siendo de especial interés el caso del Instituto de Oftalmobiología Aplicada, con un diámetro de 25, seguido por el Instituto Universitario de Biología Molecular y Celular del Cáncer.

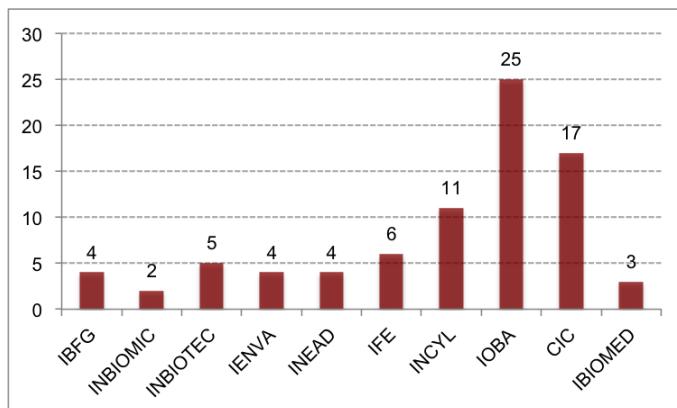


Figura N° 8: Diámetro

3.2. Centralidad de las redes

Respecto a la centralidad de la red, partiendo de la base que indica que los valores óptimos serán aquellos que se aproximen los más posible al 0, se puede observar que en las redes estudiadas, solamente cuatro redes están por debajo del 0,5: El Instituto de Estudios de Alcohol y Drogas, con un 0,04, el Instituto de Biomedicina con un 0,07, el Instituto de Biología Molecular, Genómica y Proteómica, con un 0,25, y por último el Instituto de Biología Funcional y Genómica con un 0,47.

Por el contrario, el resto de las redes superan el 0,5, destacando los casos del Instituto de Neurociencias de Castilla y León, con un 0,99 y el Instituto de Oftalmobiología Aplicada, con un 0,97.

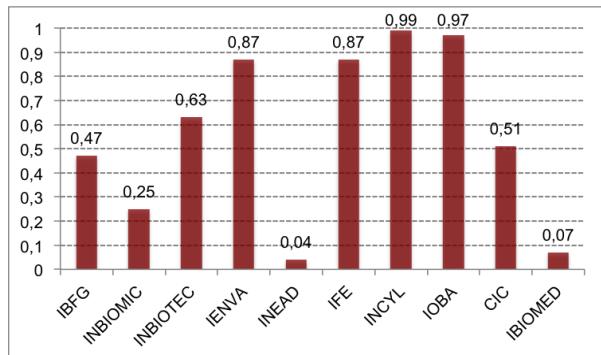


Figura N° 9: Centralidad IN

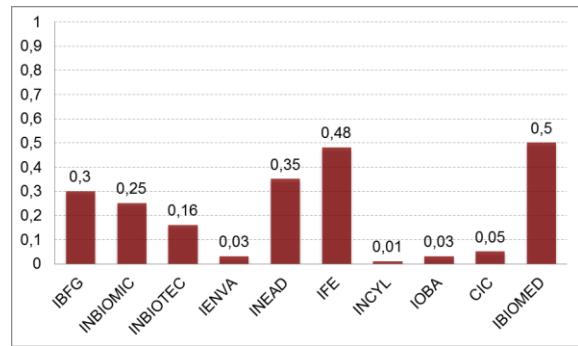
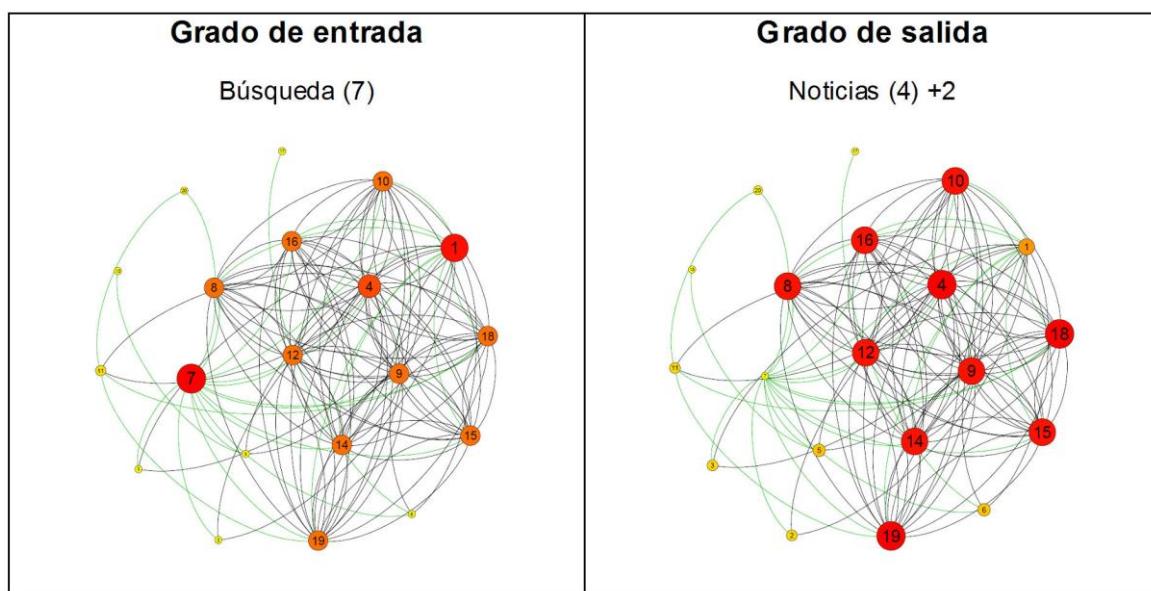


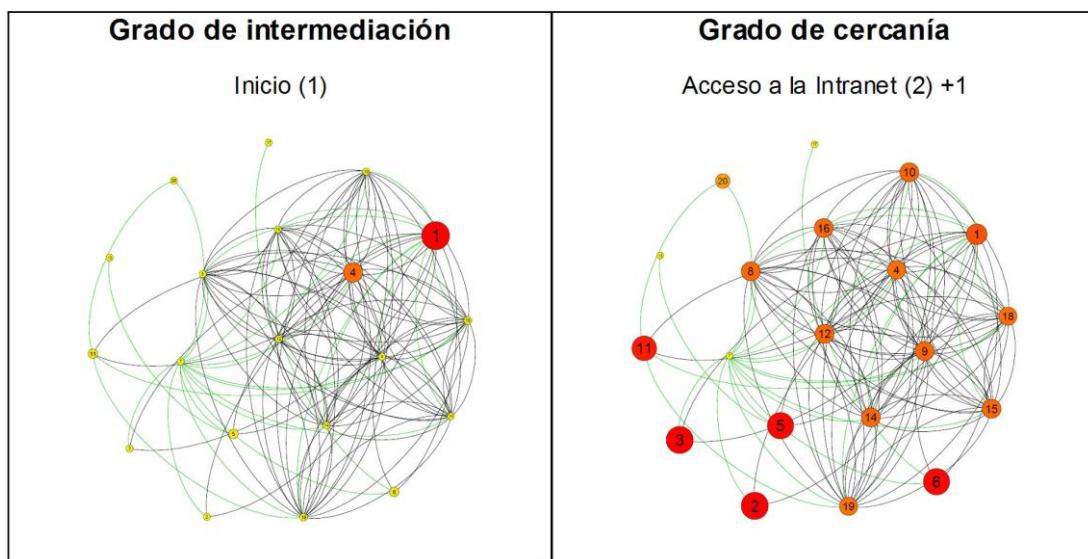
Figura N° 10: Centralidad OUT

3.3. Representación gráfica de las redes

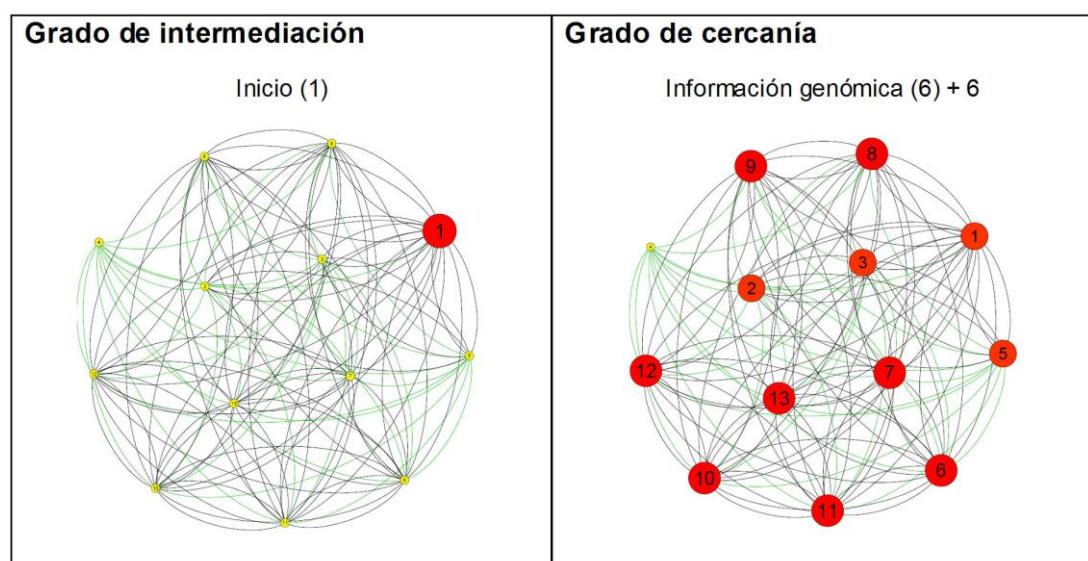
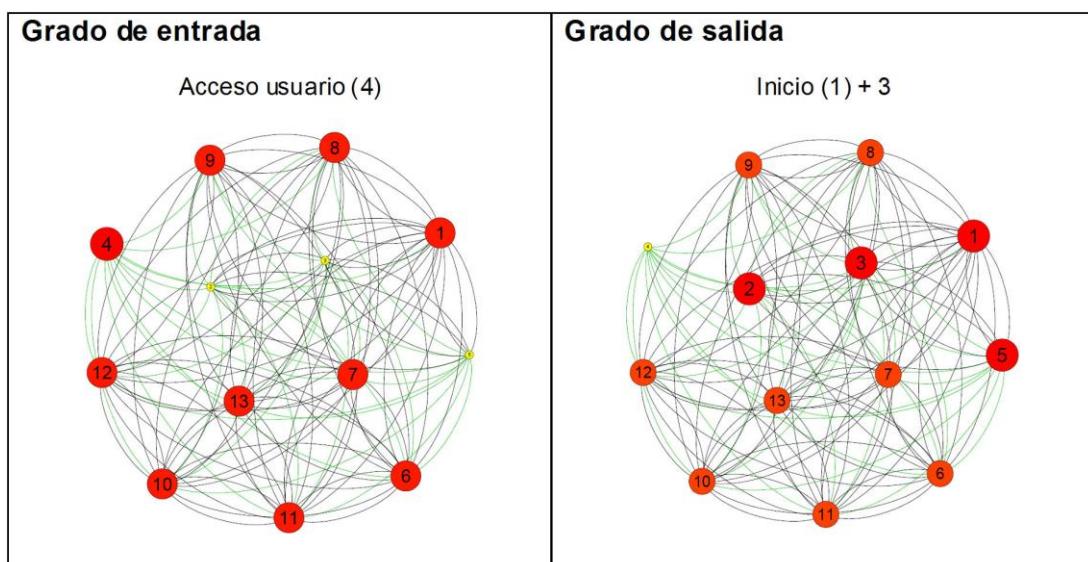
A continuación se muestra la relación de sitios Web y sus respectivas redes, destacando en cada caso los nodos con mayor peso en cada una de las medidas propuestas en la elaboración del estudio. Para ello, se presentan los gráficos agrupados bajo cada uno de los centros, de cara a facilitar la comprensión de los mismos.

Grafos del Instituto de Biología Funcional y Genómica

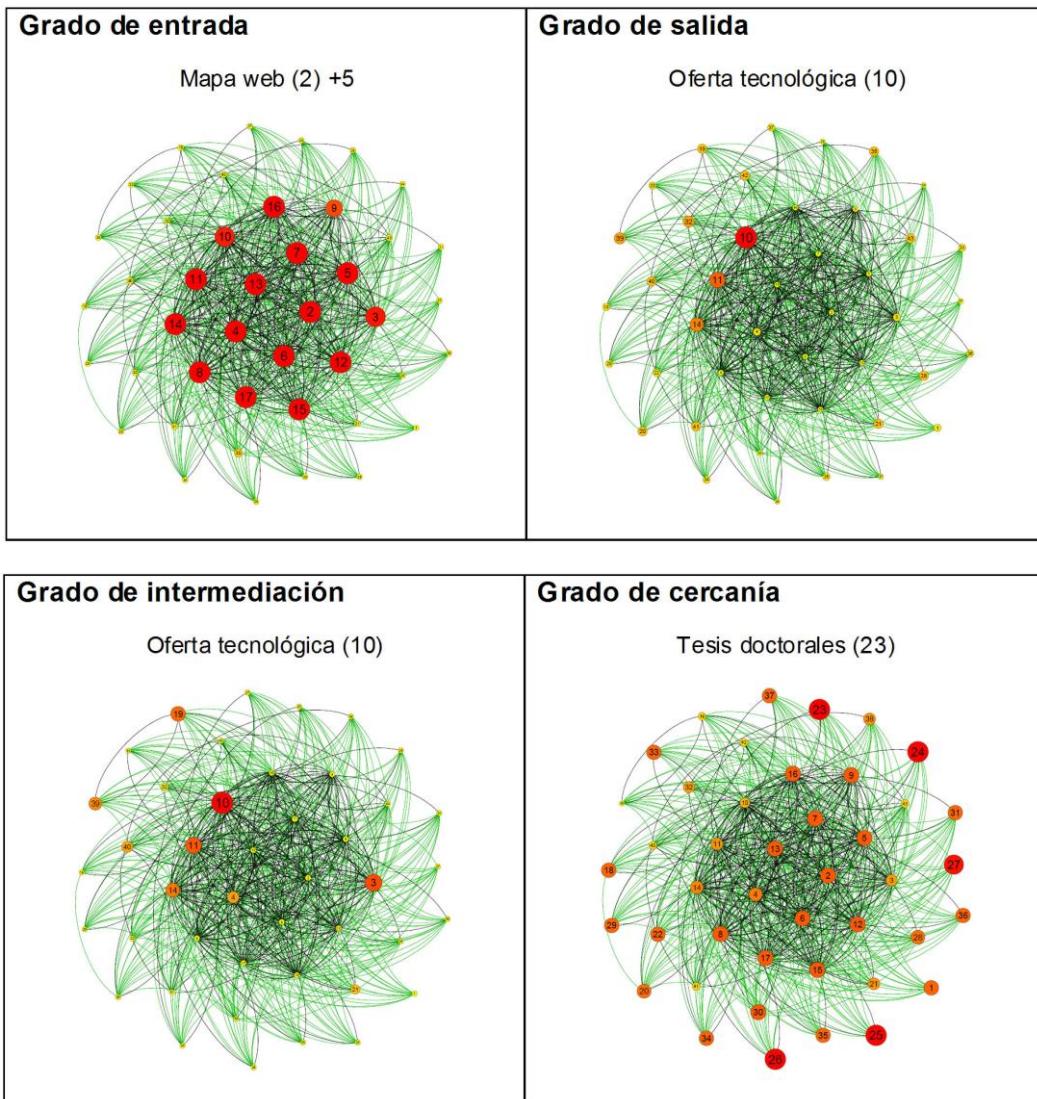




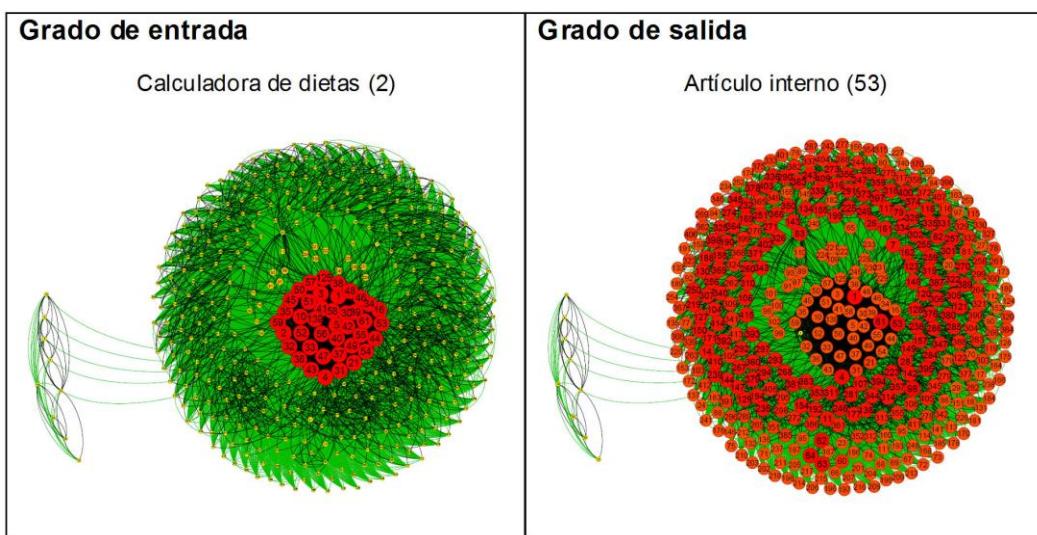
Grafos del Instituto de Biología Molecular Genómica y Proteómica

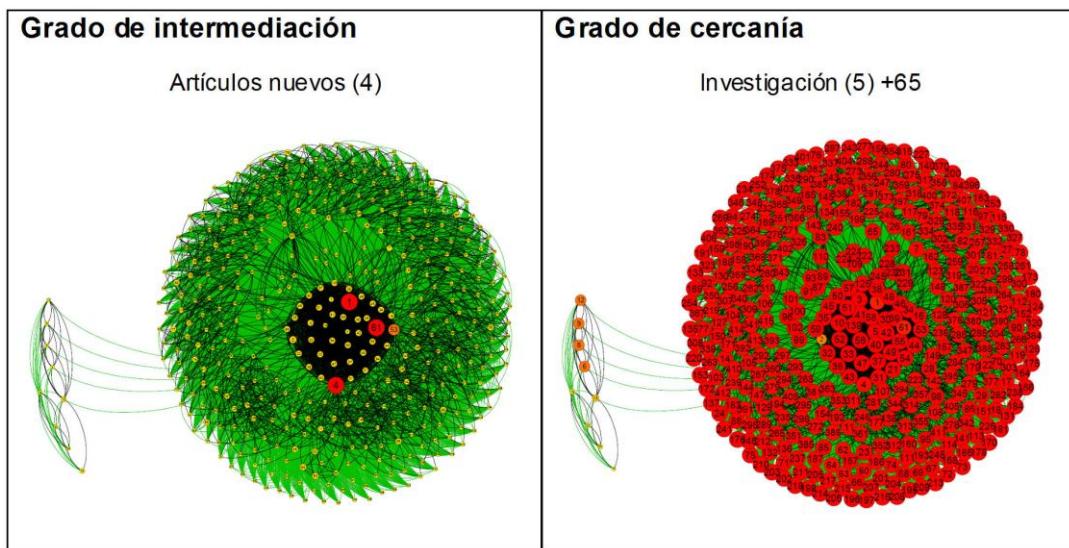


Grafos del Instituto de Biotecnología

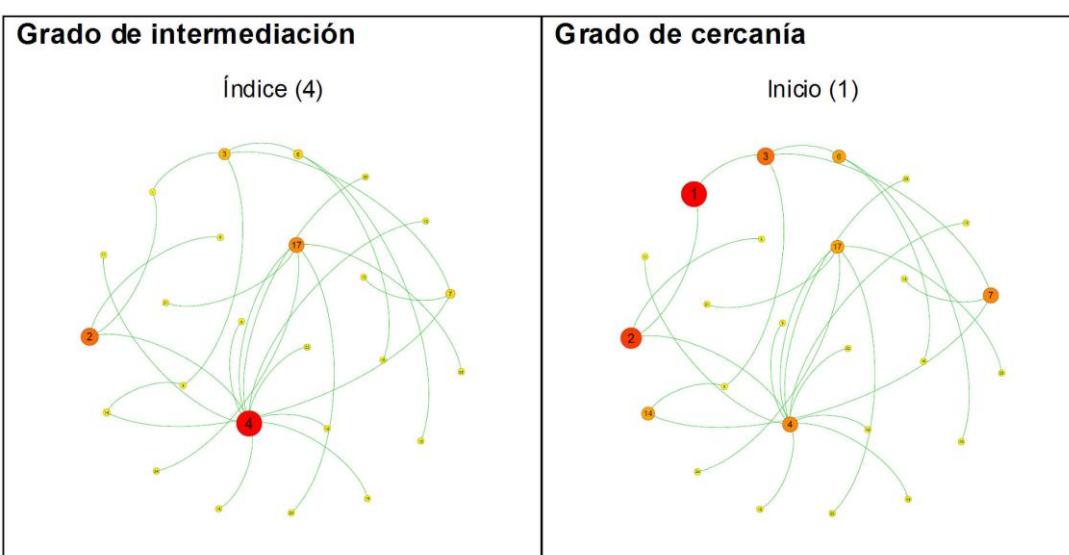
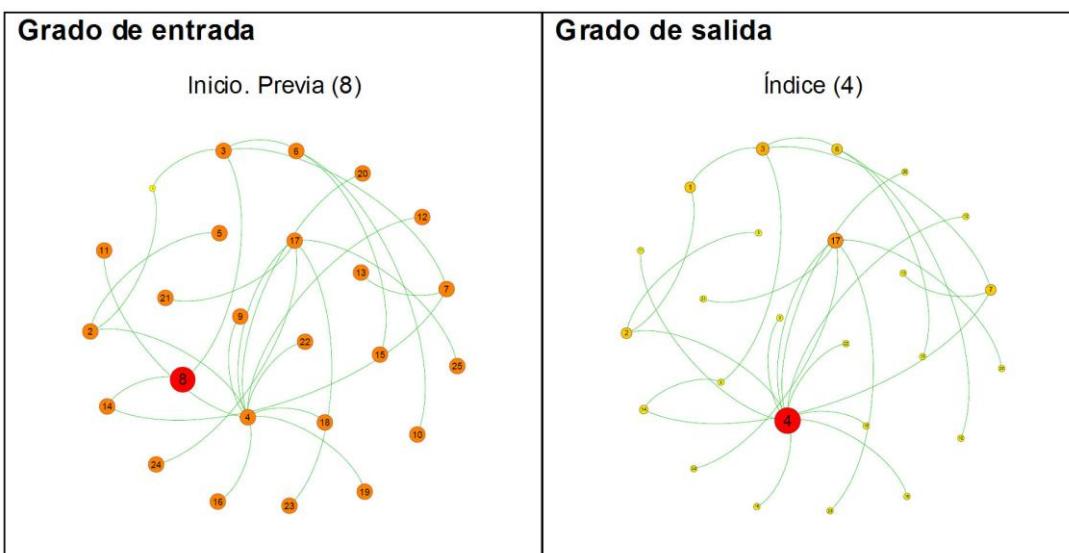


Grafos del Instituto de Endocrinología y Nutrición

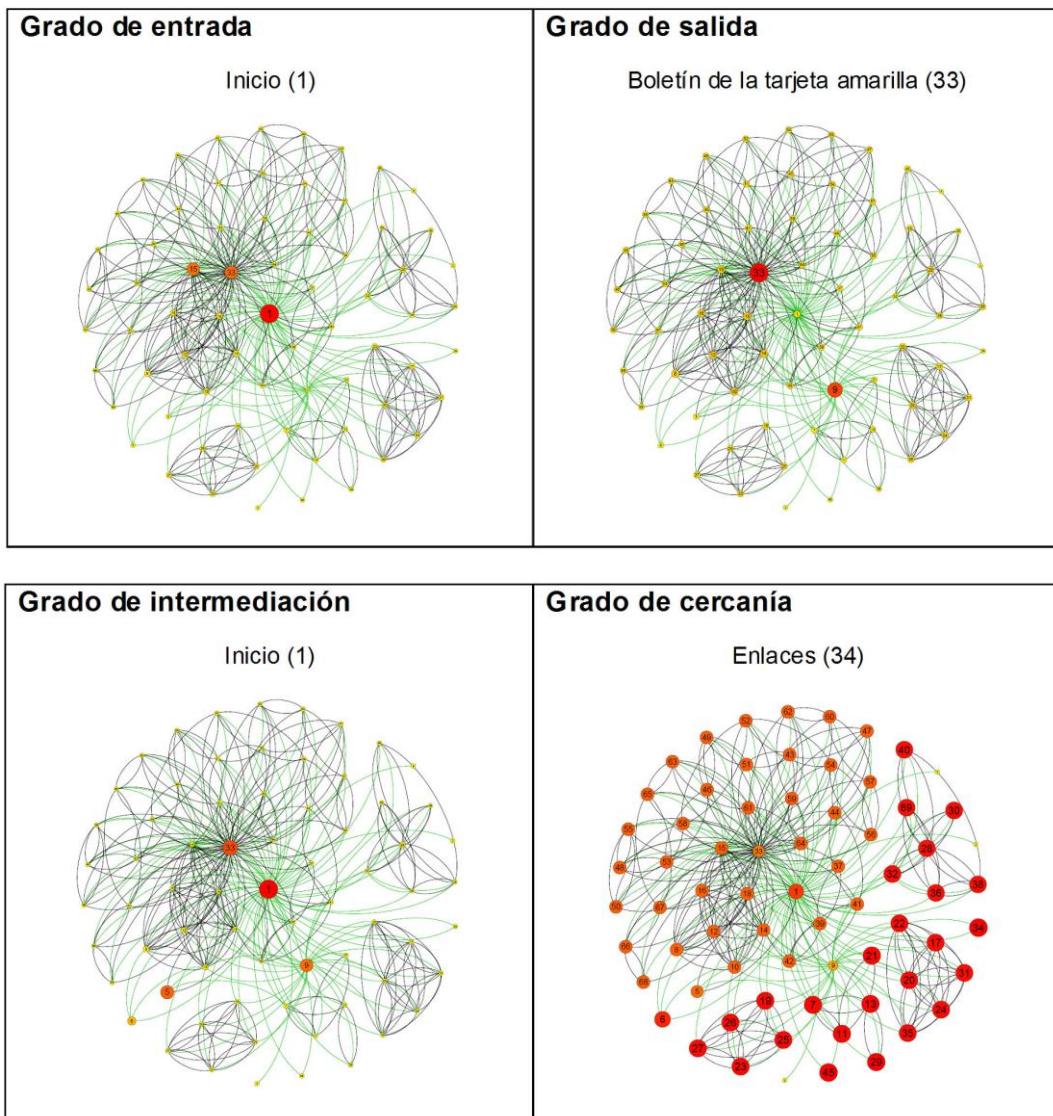




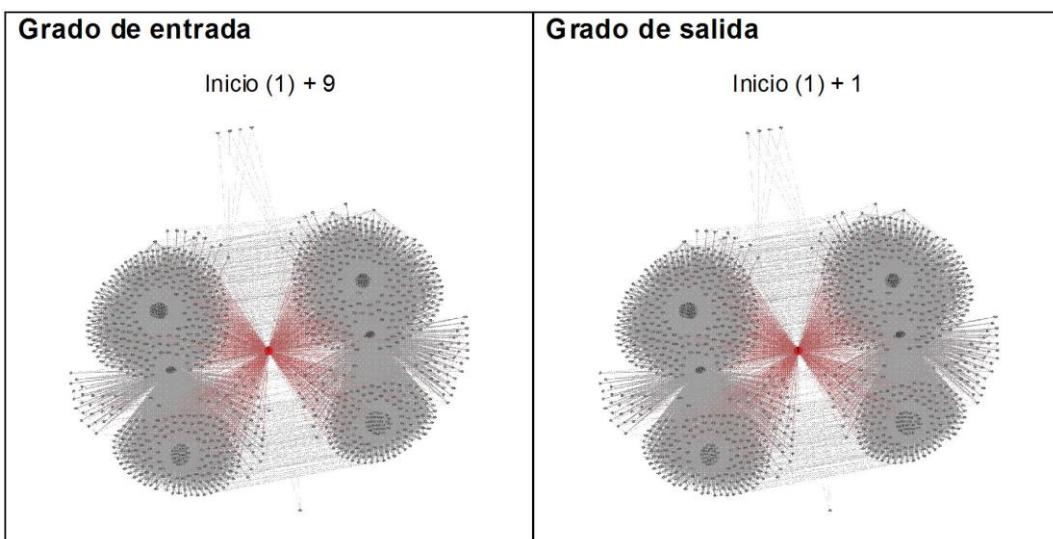
Grafos del Instituto de Estudios de Alcohol y Drogas

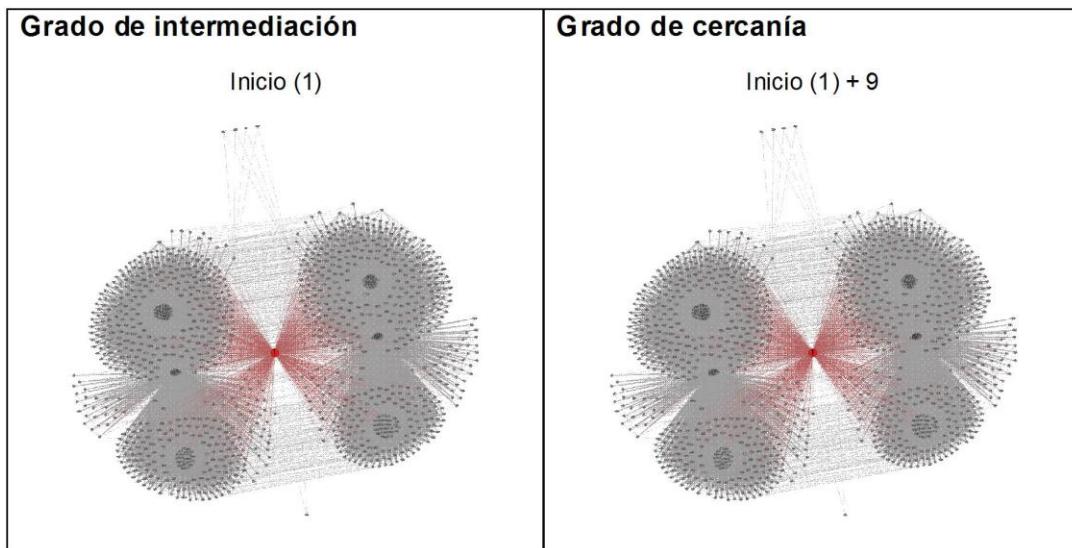


Grafos de la red del Instituto de Farmacoepidemiología

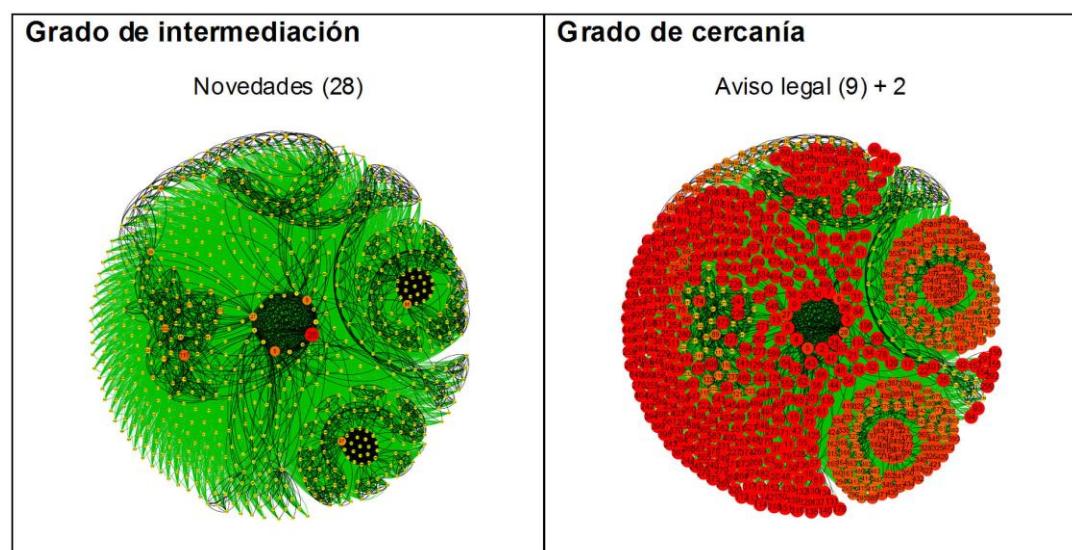
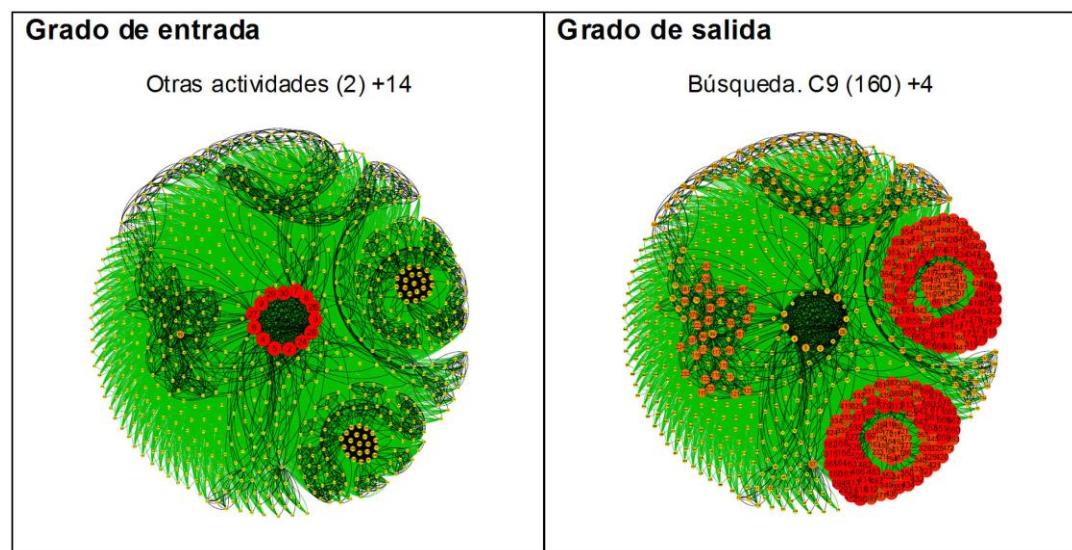


Grafos del Instituto de Neurociencias de Castilla y León

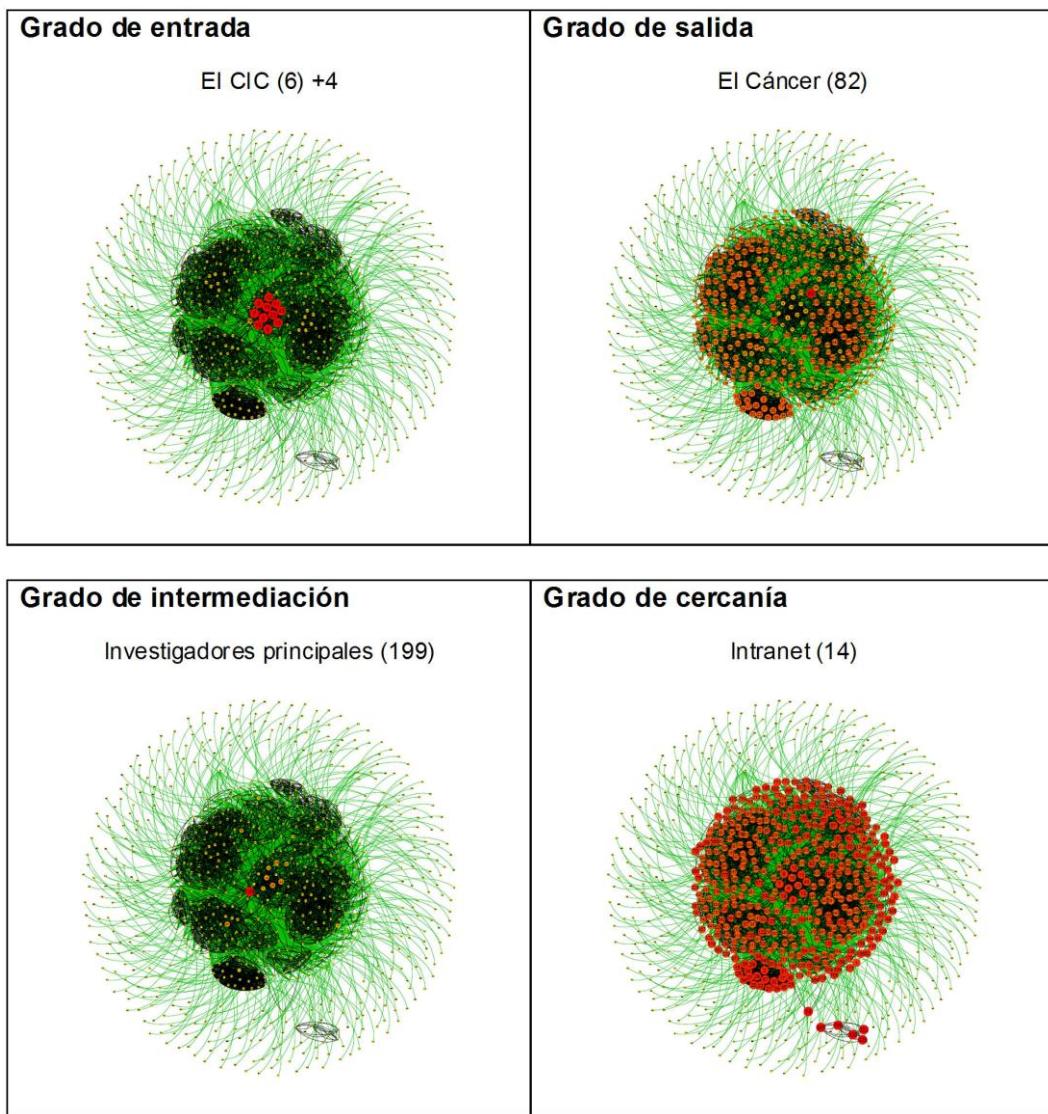




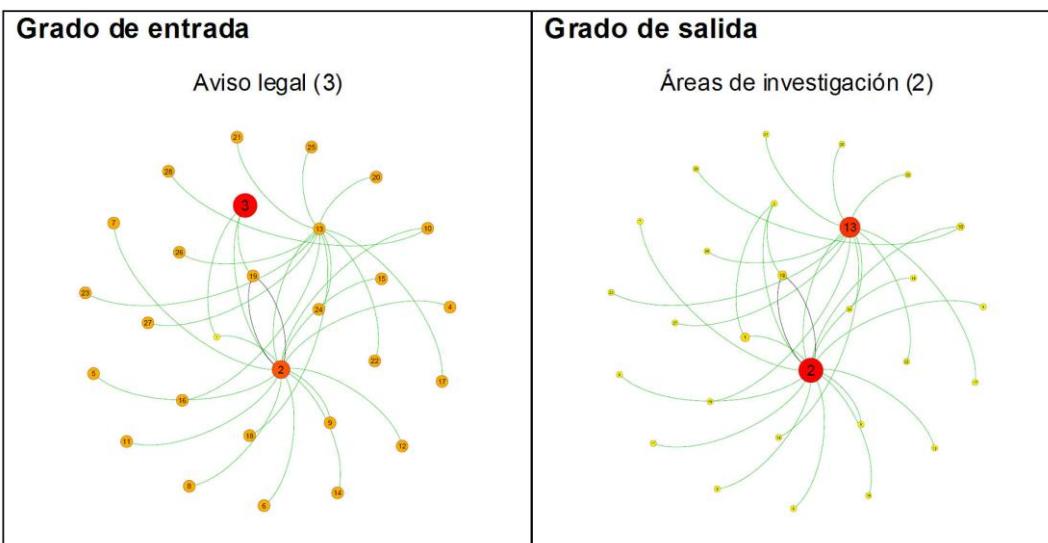
Grafos del Instituto de Oftalmobiología Aplicada

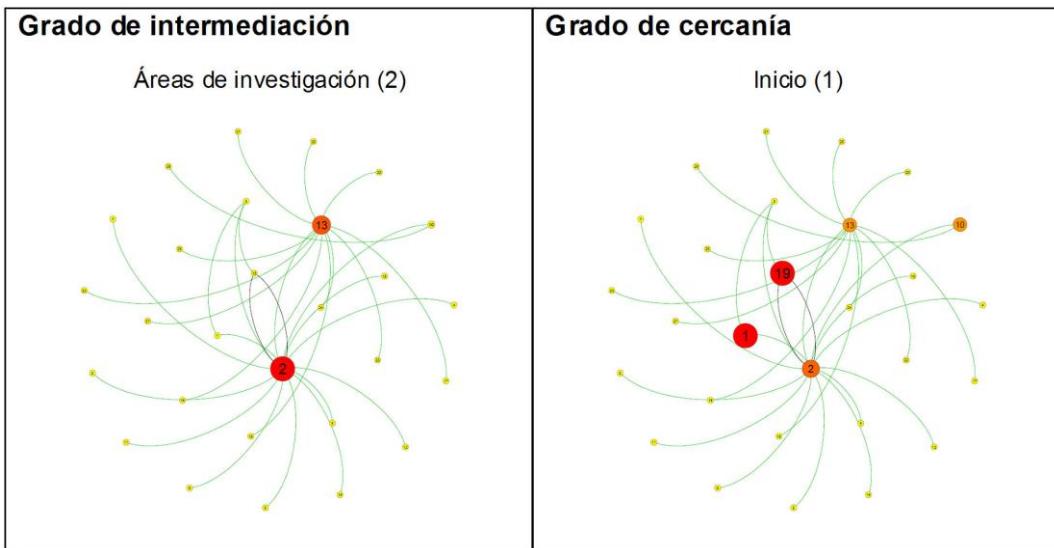


Grafos del Instituto Universitario de Biología Molecular y Celular del Cáncer



Grafos del Instituto Universitario de Biomedicina



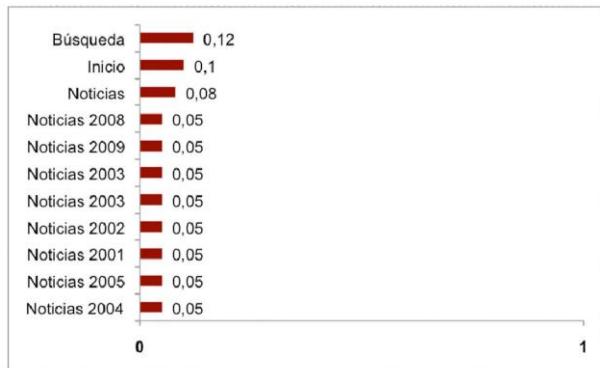


3.4. Posicionamiento de las redes

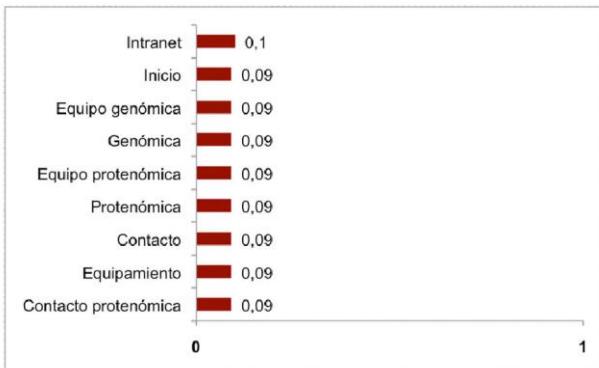
A continuación se muestran los datos del cálculo del PageRank de Google, para cada uno de los sitios Web. Se trata de una excelente visión sobre la importancia que tienen los diferentes nodos de las redes, lo cual es muy significativo, ya que da cuenta de cuáles son las páginas más valoradas desde el exterior. Esto es un asunto de suma importancia si desde la entidad que sustenta el sitio Web se pretende una presencia adecuada en los resultados de los motores de búsqueda.

De todos modos, pese a su importancia, se ha de tener en cuenta que el PageRank es un sistema de medición en bruto, es decir, que éste no garantiza que el contenido de un sitio web sea de mayor o de menor calidad, pero sí que se trata de un sitio importante en Internet. Ello ha de tenerse muy en cuenta, ya que no sirve de nada tener un PageRank alto si no se optimiza el sitio web.

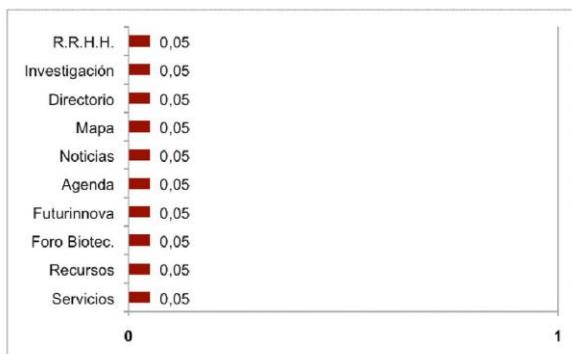
Instituto de Biología Funcional
y Genómica



Instituto de Biología Molecular Genómica
y Proteómica



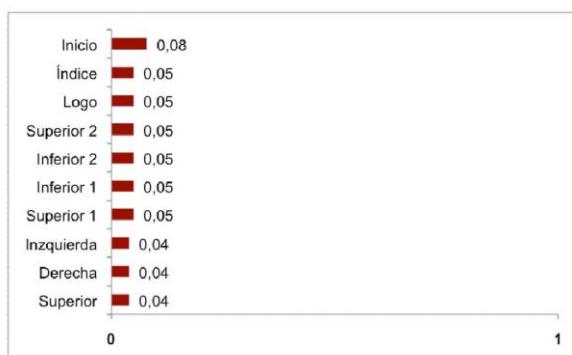
Instituto de Biotecnología



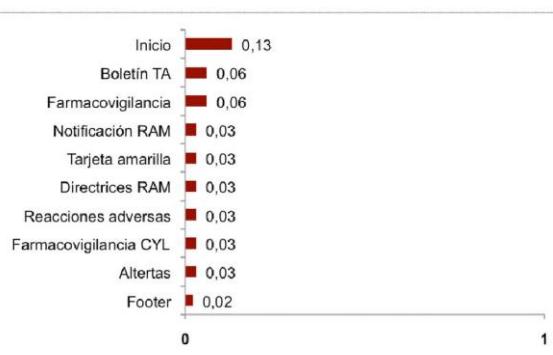
Instituto de Endocrinología y Nutrición



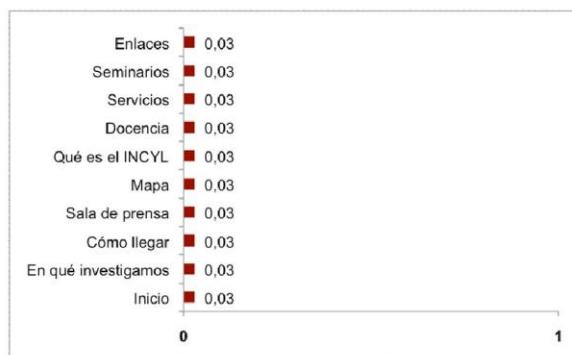
Instituto de Estudios de Alcohol y Drogas



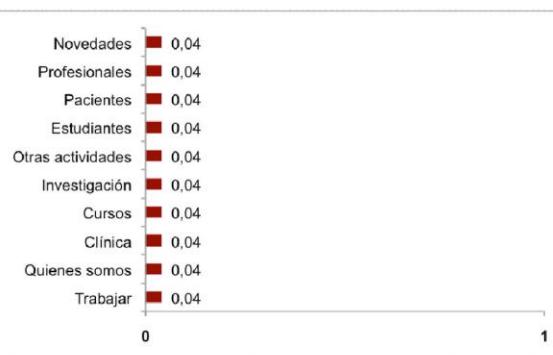
Instituto de Farmacoepidemiología



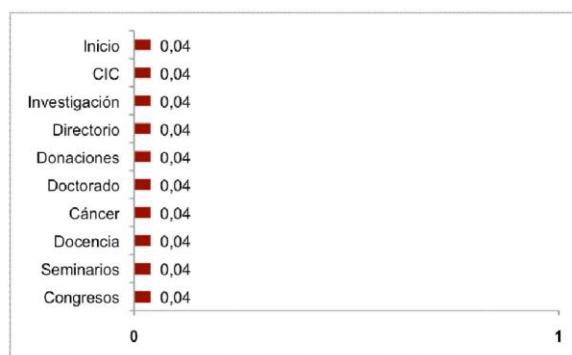
Instituto de Neurociencias de Castilla y León



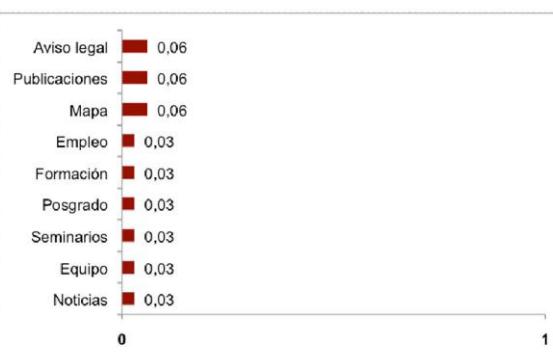
Instituto de Oftalmobiología Aplicada



Instituto Universitario de Biología Molecular y Celular del Cáncer



Instituto Universitario de Biomedicina



3.5. Bow-Tie de las redes

Con el fin de identificar los grupos de nodos, según la tipología propuesta por la teoría Bow-Tie, se presentan a continuación los datos obtenidos tras el análisis de las redes estudiadas. A través de las siguientes tablas se recoge la relación de tipos de nodos existentes en la red, su número y el porcentaje que supone al conjunto de la red.

Instituto de Biología Funcional y Genómica

Grupo	Nodos	Porcentaje
LSSC	16	80%
IN	-	-
OUT	4	20%
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

Instituto de Biología Molecular Genómica y Proteómica

Grupo	Nodos	Porcentaje
LSSC	12	92,31%
IN	-	-
OUT	1	7,69%
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

Instituto de Biotecnología

Grupo	Nodos	Porcentaje
LSSC	42	97,67%
IN	1	2,33%
OUT	-	-
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

Instituto de Endocrinología y Nutrición

Grupo	Nodos	Porcentaje
LSSC	405	97,59%
IN	10	2,41%
OUT	-	-
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

Instituto de Estudios de Alcohol y Drogas

Grupo	Nodos	Porcentaje
LSSC	1	4%
IN	4	16%
OUT	-	-
TUBES	-	-
TENDRILS	20	80%
OTHERS	-	-

Instituto de Farmacoepidemiología

Grupo	Nodos	Porcentaje
LSSC	66	95,65%
IN	-	-
OUT	3	4,35%
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

Instituto de Neurociencias de Castilla y León

Grupo	Nodos	Porcentaje
LSSC	2623	99,96%
IN	-	-
OUT	1	0,04%
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

Instituto de Oftalmobiología Aplicada

Grupo	Nodos	Porcentaje
LSSC	737	99,86%
IN	1	0,14%
OUT	-	-
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

Instituto de Neurociencias de Castilla y León

Grupo	Nodos	Porcentaje
LSSC	2623	99,96%
IN	-	-
OUT	1	0,04%
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

Instituto de Oftalmobiología Aplicada

Grupo	Nodos	Porcentaje
LSSC	737	99,86%
IN	1	0,14%
OUT	-	-
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

Instituto Universitario de Biología Molecular y Celular del Cáncer

Grupo	Nodos	Porcentaje
LSSC	409	52,98%
IN	-	-
OUT	363	47,02%
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

Instituto Universitario de Biomedicina

Grupo	Nodos	Porcentaje
LSSC	2	7,14%
IN	1	3,57%
OUT	25	89,29%
TUBES	-	-
TENDRILS	-	-
OTHERS	-	-

3.6. Conclusiones

Tras la consecución del estudio cibermétrico, con base en el análisis de redes sociales, de los sitios Web de los centros de investigación biosanitaria de Castilla y León, el siguiente paso es valorar el cumplimiento del objetivo general de la investigación, así como la interpretación de los resultados obtenidos.

La consecución del estudio ha desprendido una enorme cantidad de información, relativa a las tendencias de construcción de este tipo de sitios Web desde un punto de vista eminentemente cuantitativo, lo cual resulta de interés para la presente investigación y también puede ser tomado como referencia para estudios venideros. Todo ello, tras el pertinente tratamiento con las herramientas y técnicas anteriormente descritas, ha permitido conocer la composición de las redes analizadas, permitiendo así aseverar el cumplimiento del objetivo.

Por otra parte, las hipótesis planteadas, en torno al objetivo de la investigación, han sido verificadas, tal y como se expone a continuación:

Hipótesis 1: Se trata de redes compuestas por conjuntos de nodos fuertemente conectados.

De las 10 redes analizadas, 9 presentan valores por encima del 50% en la composición de sus respectivos LSSC (Largest Strongly Connected Component), lo que indica una predominancia de sitios Web compuestos por nodos fuertemente conectados. Las redes que cuentan con valores por debajo del 50% son las pertenecientes al Instituto de Estudios de Alcohol y Drogas (INEAD), con un 4% y al Instituto Universitario de Biomedicina (IBIOMED), con un 7,14%. Estos valores destacan de una forma llamativa entre el conjunto analizado si se tiene en cuenta que el rango, en el resto de los casos, parte del 52,98% del Instituto Universitario de Biología Molecular y Celular del Cáncer, al 99,96% del Instituto de Neurociencias de Castilla y León.

Hipótesis 2: La página de inicio no es la más importante de red.

La importancia de una página en concreto, dentro del conjunto de una red, se puede valorar en función de varias medidas, por lo que sería una desidia la fundamentación de la misma en base a un único valor, como sucede, en muchas ocasiones, con la obtención del PageRank de Google. Partiendo de este hecho, se ha concluido en afirmar la hipótesis planteada, tras la medición de la importancia de las páginas e inicio, en función de su grado de entrada, grado de salida, grado de intermediación, grado de cercanía y PageRank interno.

De los cinco valores planteados, solamente en uno, el grado de entrada (visibilidad), se observa una predominancia de redes con la página de inicio como nodo más destacado en la medición, tres (INEAD, IFE, INCYL) frente a ocho.

Para el grado de salida (luminosidad), se presentan dos redes con la página de inicio como nodo más importante (INBIOMIC, INEAD), frente a nueve en la que dicho puesto lo ocupa otra página.

En valor del grado intermediación, la relación es de dos sitios Web con la página de inicio como nodo más importante (INBIOMIC; IFE), frente a nueve.

En el caso del grado de cercanía, son dos las redes con la página de inicio a la cabeza de la medición (INEAD, IBIOMED) y nueve las que cuentan con otras páginas del sitio Web.

Por último, en la medición del PageRank, se observa una mayoría de sitios web con páginas de inicio poco valoradas. En este caso, cuatro sitios web cuentan la página de inicio como la página más valorada (INEAD, IFE, INCYL, CIC) frente a seis en los que no es así.

Estos resultados evidencian que la página de inicio no siempre es la más importante de una red, a pesar de la dedicación que supone el diseño, la construcción y el cuidado de la misma. Esto es un aspecto de suma importancia en aquellos casos en los que la página con un mayor PageRank no sea la deseada, ya que puede suponer algún tipo de trastorno comunicacional en la imagen virtual de la entidad titular del sitio web. El conjunto de sitios analizados se sitúa en general en posiciones medias y bajas, lo cual no ha de interpretarse como que estos son mejorables a nivel de funcionamiento o diseño, sino que su influencia es media en el entorno global que es Internet. Como ya se ha comentado, parte de la valoración del PageRank se obtiene a través de los enlaces que salen y entran de los sitios web, pero sin llegar a ser una cuestión meramente cuantitativa, sino que dichos enlaces han de ser relevantes, es decir, a su vez han de estar bien valorados. Ello evoca una de las bases de Internet, la conectividad y la actividad relacional que se da a través de ella, donde un sitio influyente puede ayudar a que otro sitio menos influyente gane peso en la red, dejando parte del éxito en los conocidos como “factores off the page”, aquellos ajenos al sitio web en cuestión.

Hipótesis 3: Los sitios web que dependen de una misma universidad no comparten patrones estructurales entre sí.

A través del estudio morfológico de los grafos, procesados para la investigación, es posible comparar unas redes con otras. Asimismo, tras el pertinente análisis, es posible apreciar las grandes diferencias que existen entre cada una de las redes, pudiendo, de este modo, afirmar la hipótesis planteada.

Del mismo modo que la representación gráfica muestra una desigualdad entre los tres grupos de redes, los datos extraídos de la medición de las mismas, corroboran la afirmación de la hipótesis, algo evidente, ya que la representación gráfica de las redes, además de basarse en el algoritmo Fruchterman-Reingold, recoge aspectos referentes a las medidas de centralidad, así como las propiedades de la red. Si bien es cierto, cabe anotar que se dan algunos casos menores de similitud.

El Instituto Universitario de Biología Molecular y Celular del Cáncer (CIC) y el Instituto de Neurociencias de Castilla y León (INCYL), ambos vinculados a la Universidad de Salamanca tienen una densidad de 0,01, el Instituto de Endocrinología y Nutrición (IENVA) y el Instituto de Estudios de Alcohol y Drogas (INEAD), de la Universidad de Valladolid, comparten un valor en sus diámetros de 4 y por último, el Instituto de Endocrinología y Nutrición y el Instituto de Farmacoepidemiología (IFE), ambos de la Universidad de Valladolid, cuentan con una centralidad de entrada de un 0,87.

Hipótesis 4: Las páginas de noticias y actualidad, en caso de existir, cuentan con un PageRank alto con respecto al conjunto de la red.

Del total de redes analizadas, seis, frente a cuatro, tienen sección de comunicación. Independientemente del indudable valor de la comunicación en las instituciones, la valoración interna, de carácter cuantitativo, obtenida tras el cálculo del PageRank interno de cada una de las redes, demuestra que las secciones de actualidad, catalogadas como sala de prensa, noticias, gabinete de prensa o actualidad, entre otras nomenclaturas, suele contar con un buen PageRank sobre el conjunto de cada una de las respectivas redes, por lo que se puede confirmar la hipótesis. En este caso, de los sitios web que cuentan con sección de comunicación, cuatro de ellos cuentan entre sus rankings internos con dicha sección, en un cómputo realizado sobre las 10 páginas con más valor.

Cabe destacar que de los siete sitios web que cuentan con sección de comunicación, pero que ésta no aparece entre las 10 páginas más importantes, dos de ellas son secciones vacías, es decir, tras acceder a ellas no se encuentra ningún tipo de contenido. Esto sucede en el caso del Instituto de Estudios de Alcohol y Drogas

Hipótesis 5: Las redes analizadas cuentan con buenos niveles de conectividad.

En base a los resultados obtenidos, la conectividad de las redes estudiadas es buena, pero existen ciertos casos que deberían cambiar sustancialmente de cara a mejorar el valor analizado. Partiendo de un diámetro de valor seis, como medio y aceptable, en el presente estudio se ha concluido que seis de ellas tienen una conectividad buena, al encontrarse valores menores a seis, una de ellas se mantiene en el límite aceptable y cuatro de ellas cuentan con valores que señalan carencias en materia de conectividad. Resultan sorprendentes los valores extraídos del análisis de las redes del Instituto de Oftalmobiología Aplicada (IOBA) y del Instituto Universitario de Biología Molecular y Celular del Cáncer (CIC), con valores de 25 y 17 respectivamente, lo que denota una falta de cuidado notable en la generación de las redes.

Hipótesis 6: Los sitios Web analizados cuentan con exigua tramas de relaciones.

Con la medición de la densidad, de cada una de las redes, se ha podido afirmar la hipótesis planteada, observando que solamente una red supera el valor 0,5, el Instituto de Biología Molecular Genómica y Proteómica (INBIOMIC) obteniendo un 0,71. El rango, en el resto de los casos, va desde el 0,38 del Instituto de Biotecnología (INBIOTEC), al 0,01, con el que cuentan el Instituto de Neurociencias de Castilla y León (INCYL) y el Instituto Universitario de Biología Molecular y Celular del Cáncer (CIC).

En conjunto, los valores presentados son bajos, presentando el total de las redes un valor medio de 0,1.

Estas conclusiones evidencian ciertas carencias de los sitios web analizados, lo que puede derivar en otro tipo de deficiencias que afecten a los aspectos comunicativos en Internet de las entidades seleccionadas para el estudio. Por ello, la información aquí recogida se presenta como un recurso de utilidad para la mejora de los sitios web institucionales actuales, así como un punto de referencia para la construcción y el diseño estructural de los sitios web en desarrollo.

Tal y como se recoge en el estudio, muchos de los aspectos presentados son vitales para la exploración de los recursos web. Ello es útil para la experiencia del usuario, algo de indudable valor en el desarrollo web actual, pero también es de suma importancia para aspectos cuantitativos de Internet, tales como el posicionamiento, del que dependen los resultados ofrecidos por los grandes buscadores como Google, Bing o Yahoo!, entre otros, y que supondrá parte del éxito o el fracaso de determinados contenidos en el ecosistema digital.

- Este artículo se ha desarrollado al amparo de la ayuda para la contratación de personal investigador de reciente titulación universitaria, convocada por la Consejería de Educación de la Junta de Castilla y León, en el marco de la Estrategia Regional de Investigación Científica, Desarrollo, Tecnológico e Innovación 2007-2013, cofinanciadas por el Fondo Social Europeo. Orden EDU/537/2010 de 20 de abril (BOCyL, 28 de abril de 2010) ([Referencia](#))

4. Bibliografía

- RH Abraham (1997): “Webometry: measuring the complexity of the World Wide Web”. *World Futures* 12, Philadelphia, pp. 785-791.
- E Guess Adar (2006): “A language and interface for graph exploration”. *Computer Human Interaction*, Vancouver, pp. 791-800.
- IF Agullo, 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 Agullo, 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): *Cibermetría: 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 cibermetría”. *Ibersid* n. 2, pp. 69-78.
- N Arroyo, JL Ortega, VM Pareja, JA Prieto (2005): *Cibermetría. Estado de la cuestión*. IX Jornadas Españolas de Documentación, Madrid, 14 – 15 de abril de 2005, pp. 14-15.
- V Batagelj, A Mrvar (1998): “Pajek-program for large network 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. Disponible en: <http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d02/2bossy.html> [Consultado el 20/01/2012].
- A Broder, R Kumar, F Maghoul, P Raghavan (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 Networks 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 Aguirreitia-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. Disponible en: http://www.revistalatinacs.org/067/art/946_Alicante/02_Lola.html [Consultado el 16/01/2012]

LC Freeman (1979): “Centrality in social networks conceptual clarification”. *Social Networks* 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", en Revista Latina de Comunicación Social, 66. La Laguna (Tenerife): Universidad de La Laguna, páginas 526 a 548 recuperado el 3 de enero de 2013 de 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. Disponible en: http://www.revistalatinacs.org/11/art/930_Santiago/09_Tunez.html [Consultado el 23/01/2012]

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.

CÓMO CITAR ESTE ARTÍCULO EN BIBLIOGRAFÍAS – HOW TO CITE THIS ARTICLE IN BIBLIOGRAPHIES / REFERENCES:

F Sánchez-Pita, JL Alonso-Berrocal (2013): “Los sitios Web de centros de investigación biosanitaria de Castilla y León. Un análisis cibermétrico”, en *Revista Latina de Comunicación Social*, 68. La Laguna (Tenerife): Universidad de La Laguna, páginas 383 a 419 recuperado el ____ de ____ de 2_____, de
http://www.revistalatinacs.org/068/paper/982_Salamanca/16_Sanchez.html
DOI: [10.4185/RLCS-2013-982/CrossRef link](https://doi.org/10.4185/RLCS-2013-982)

Artículo recibido el 29 de enero de 2013. Sometido a pre-revisión el 1 de febrero. Enviado a revisores el 3 de febrero. Aceptado el 12 de mayo de 2013. Galeradas telemáticas a disposición de los autores el 19 de mayo de 2013. Visto bueno de los autores: 25 de mayo de 2013. Publicado el 12 de junio de 2013.

Nota: el DOI es **parte de** la referencia bibliográfica y ha de ir cuando se cite este artículo.

Artículos de alguna forma relacionados

[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 ...](#)

[The communication in the organizations of the third sector](#)

[Resonant communicators, successful communicators. Flow and communicator's credibility](#)

[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](#)