

Una introducción a la investigación webométrica de empresas

Esteban Romero Frías

Departamento de Economía Financiera y Contabilidad

Universidad de Granada

erf@ugr.es

RESUMEN DEL ARTÍCULO

Palabras clave: webmetría, minería de datos en la web, rendimiento empresarial, ratios.

La Web ha experimentado un enorme desarrollo en los últimos 20 años. La Webmetría se presenta como una nueva disciplina que pretende estudiar desde un punto de vista cuantitativo este nuevo canal de comunicación mediante la adaptación y aplicación de técnicas bibliométricas. Este artículo aborda la relevancia de los hipervínculos como fuente de información y explica brevemente las técnicas más prometedoras que pueden emplearse para obtener datos acerca de fenómenos que se manifiestan *online*. Hasta el momento los estudios en este campo se han centrado en el estudio de espacios académicos, sin embargo esta metodología se puede aplicar igualmente a sitios comerciales, que son los que dominan la Web.

El artículo revisa algunos estudios que han descubierto que el número de enlaces que apuntan a páginas web de empresas se correlacionan significativamente con determinadas medidas de desempeño financiero. Este descubrimiento sugiere que los enlaces entrantes en un sitio web podrían emplearse como un indicador adicional del éxito de las actividades empresariales. Los trabajos realizados hasta el momento se han limitado a un determinado sector empresarial, el de tecnologías de la información, y a un número reducido de países (Estados Unidos, Canadá y China). Es preciso llevar a cabo más investigación de carácter exploratorio con el fin de valorar las posibilidades de futuros desarrollos en esta área de investigación.

ARTICLE SUMMARY

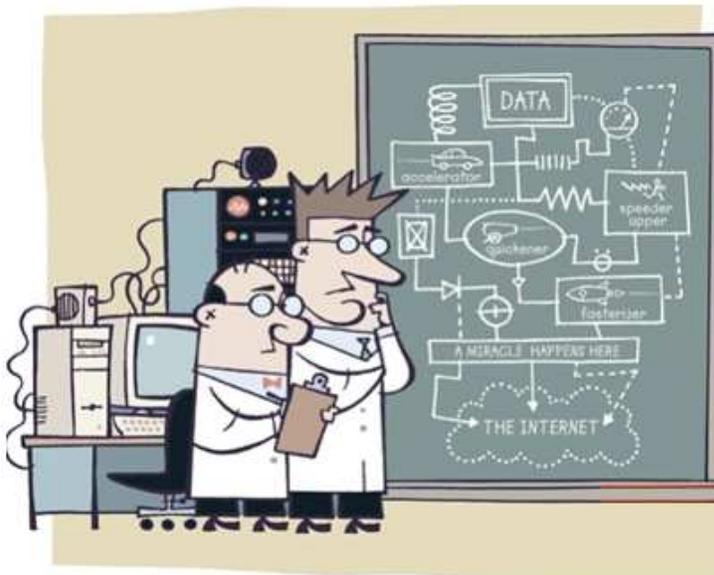
Keywords: webometrics, web data mining, business performance, ratios.

The Web has undergone an enormous development in the last 20 years. Webometrics is a new discipline that applies Bibliometric techniques to study the Web from a quantitative approach. This paper addresses the relevance of hyperlinks as a source of information, and it explains briefly the most promising techniques that can be used to mine data about online phenomena. So far studies in this field have focused on academic spaces, however this methodology is equally applicable to commercial sites which dominate the Web.

This paper reviews studies that found that the number of links pointing to company websites correlates significantly with some company's business performance measures. This finding may imply that links to a website could be used as an additional indicator of business performance. Nevertheless this research has some limitations such as the focus on the IT industry and on a limited amount of countries (e.g. USA, Canada and China). More exploratory research is needed in order to assess possible future developments in this area

1. POR QUÉ ES RELEVANTE EL ESTUDIO DE LA WEB

La variedad de posibilidades de uso de la Web, y de Internet en su conjunto¹, unida a una reducción generalizada de los costes de conexión y de los equipos necesarios para ello, han redundado en que un número creciente de personas se conecte a Internet, empleando distintos tipos de dispositivos y haciéndolo desde más lugares y durante más tiempo. El “poblamiento” de la Red ha ocasionado que una gran mayoría de actividades sociales, económicas, políticas, educativas, de toda índole, que hasta hace poco se desarrollaban únicamente en el mundo físico, ahora encuentren reflejo en Internet o hayan desarrollado manifestaciones online de naturaleza puramente autóctona. Es decir, son realidades cotidianas que sin Internet no existirían. Entre ellos se pueden mencionar el desarrollo de una potente red de blogs a nivel mundial o la profusión de las redes sociales virtuales. De acuerdo con los datos del informe *La Sociedad en Red. Informe anual 2007*, elaborado por el ONTSI² (2008), actualmente existen en España más de 22 millones de usuarios conectados a Internet.



Estas cifras tan elevadas se han visto acompañadas de una significativa presencia de las empresas en la Web. Así, una amplia mayoría de empresas de cierto tamaño (un 52% de empresas de 10 ó más empleados, un 68% de empresas medianas y un 87% de empresas grandes) dispone de página web propia. Se tratan de páginas centradas generalmente en la promoción de la entidad frente a la prestación de servicios online, sin embargo la calidad de la presencia en la Red también se ha incrementado diversificándose el tipo y la cantidad de operaciones comerciales que las organizaciones llevan a cabo entre

sí, con la Administración o con los ciudadanos. Valgan algunos datos (ONTSI, 2008). El comercio electrónico B2C (*Business to Consumer*) en España experimentó en 2007 un crecimiento del 72%. En 2006 se identificaron más de 5 millones de internautas que realizaban compras online. Se estima que el 94% de las empresas españolas (de 10 ó más empleados) disponen de acceso a Internet, con un 93% de ellas con banda ancha contratada, superando la media de la UE. Por último, en relación con los usos que se da a Internet, cabe señalar que la mayoría de empresas busca información (96,1%) y accede a servicios bancarios y financieros (85,8%).

Del auge de Internet y de sus implicaciones comerciales dan buena cuenta las cifras de crecimiento de los ingresos por publicidad en España. En 2007, frente al estancamiento de otros medios, se produjo un incremento del 55% con respecto al año anterior (Fundación Telefónica, 2008), acercándonos a la situación de mercados más maduros de nuestro entorno. Se puede

¹ Internet y la Web no son lo mismo aunque frecuentemente se empleen de forma intercambiable. Internet es un conjunto descentralizado de redes de comunicación distribuidas que se interconectan entre sí empleando la familia de protocolos TCP/IP. Su origen se remonta a la red ARPANET en Estados Unidos a finales de los años 60. La World Wide Web (WWW) o Web es, sin embargo, un servicio específico que funciona en Internet basado en un conjunto de protocolos como es el HTTP. Su origen se sitúa a finales de los años 80 y principios de los 90 a partir de los trabajos llevados a cabo por Tim Berners-Lee en el CERN. El crecimiento y popularización de la Web ha sido tan importante a lo largo de los últimos 20 años que los términos Internet y Web se confunden de forma habitual. Actualmente muchos de los otros servicios de Internet, como son el correo electrónico (basado en el protocolo SMTP), la transmisión de archivos (FTP y P2P), las conversaciones en línea (IRC), la mensajería instantánea, etc., son accesibles desde una interfaz Web, lo que contribuye a una mayor confusión entre ambos conceptos. Este trabajo se centra fundamentalmente en el estudio de la Web, teniendo las menciones que se realizan a Internet en sentido equivalente a ésta, salvo que se especifique lo contrario. Otro término que se emplea como sinónimo es el de Red.

² ONTSI son las siglas del Observatorio Nacional de las Telecomunicaciones y la Sociedad de la Información.

afirmar que las empresas empiezan a tomar conciencia de que Internet no es un canal más, algo complementario, sino que se está convirtiendo en un ámbito tan importante como pueda ser la propia presencia física de la entidad. De hecho la tendencia en los próximos años será a que desaparezca la, en muchos casos, artificiosa distinción entre *online* y *offline*.

Uno de los elementos fundamentales que nos permite entender el porqué millones de personas pasan más tiempo conectados a Internet que viendo la televisión (especialmente entre la gente más joven) es el desarrollo de la conocida como Web 2.0 (O'Reilly, 2004), que ha posibilitado una democratización de los medios de producción de información, eliminando barreras técnicas y económicas que hacían de la industria de medios de comunicación una de las fortalezas más inexpugnables de las pasadas décadas. Las aplicaciones que entran dentro de la 2.0 (blogs, redes sociales, wikis, páginas para compartir vídeos, fotografías o enlaces, etc.) han sabido aprovechar los recursos disponibles, como la disponibilidad de un mayor ancho de banda o la aplicación de avances informáticos, para enriquecer la experiencia en Internet del usuario (Jenkins, 2006). Sin embargo, no es un enriquecimiento que nazca del control de la información, sino por el contrario de facilitar la posibilidad de generarla y difundirla. El ciudadano en red ha abandonado en muchos casos el simple rol de espectador para convertirse también en un creador de contenidos. Los hiperenlaces constituyen un elemento fundamental de variada tipología de información que se genera.

2. LA IMPORTANCIA DE LOS HIPERENLACES

Los hiperenlaces son un elemento básico para comprender las posibilidades de la investigación en la Web y, en concreto, de la investigación basada en la aplicación de técnicas webmétricas. Un hiperenlace es básicamente una referencia que una página web establece a una sección de esa misma página o a otra página web completamente distinta. Los enlaces conforman la estructura de la Web (Berners-Lee, 1999), ya que sin ellos resultaría imposible acceder a los contenidos alojados en páginas, a no ser que conociéramos su URL³. Si bien algunas voces han señalado la naturaleza caótica y desordenada del medio (Keen, 2007), la ausencia de cualquier tipo de orden es aparente, ya que cada enlace constituye potencialmente una fuente de información sobre el modo en que la Web se haya distribuida. Un buen ejemplo de ello es la industria de los buscadores de información (Battelle, 2005; Machill et al., 2008) que tienen al análisis de hiperenlaces como una de las bases para la detección y ordenación de la información. Éste es precisamente el origen del éxito de Google, líder mundial del sector, cuyo sistema de ordenación de la información, *Pagerank* (Brin y Page, 1998) se basa fundamentalmente en dos principios. El primero es que cada enlace que se establece a una página web constituye en cierta forma un voto que se otorga a la misma, una llamada de atención, una cita o referencia. El segundo es que no todos los enlaces tienen la misma repercusión. Los enlaces provenientes de páginas relevantes, con una buena reputación, tendrán más valor que aquellos provenientes de páginas poco importantes.

La investigación webmétrica se sitúa en este contexto y se encarga del estudio cuantitativo de todos aquellos aspectos que pueden cuantificarse en la Web. De ahí que precisamente sean los hiperenlaces su materia prima fundamental.

3. WEBMETRÍA: CONCEPTO Y METODOLOGÍA

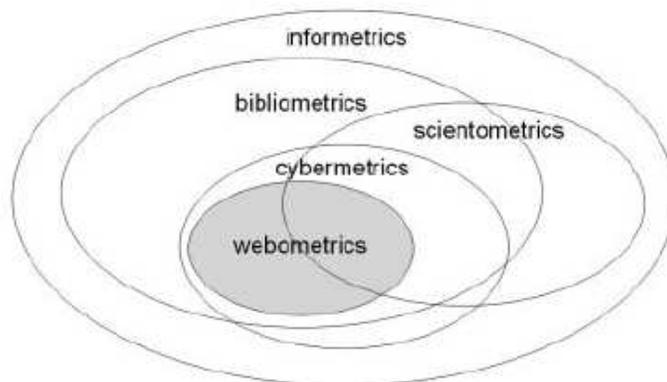
El término inglés *Webometrics*⁴ aparece por primera vez en 1997 de la mano de Almind e Ingwersen. Casi simultáneamente se empieza a utilizar el término Cibermetría (*Cybermetrics*, en inglés), con una especial preferencia en el ámbito hispano. Si bien ambos se han usado en ocasiones de forma indistinta para referirse a un mismo tipo de estudios, en 2004 Björneborn propone delimitar su empleo definiéndolos. De este modo, la Webmetría se define como “el estudio de los aspectos cuantitativos de la construcción y empleo de los recursos, estructuras y

³ URL son las siglas de *Uniform Resource Locator*, cuya traducción es “localizador uniforme de recurso”. Se trata de una secuencia de caracteres con un formato estándar que permite referenciar recursos en Internet a partir de su localización específica.

⁴ Para evitar el componente cacofónico, se prefiere la traducción del término por “Webmetría” en vez de “Webometría”.

tecnologías de la información en la Web empleando una enfoque bibliométrico e informétrico”⁵ (citado en Björneborn e Ingwersen, 2004, p. 1217), mientras que la Cibermetría se define de igual forma pero circunscrita a un ámbito más amplio, englobando tanto la Web como otros servicios de Internet como son los chats, grupos de noticias, correo electrónico, etc. La Figura 1 permite entender mejor el lugar que ocupan estas disciplinas dentro del área de las ciencias de la información, si bien en los últimos años se ha insistido en el carácter interdisciplinar de la materia.

Figura 1: La Webmetría y la Cibermetría en el contexto de las Ciencias de la Información



Fuente: Björneborn y Ingwersen, 2004, p. 1217.

La idea, mencionada anteriormente, de que un enlace a una página Web representa de alguna manera un voto a favor de la misma se encuentra ya en la base de los sistemas de citación y de elaboración de rankings de impacto en el campo de la investigación científica (Garfield, 1979). De ahí que sea desde el campo de ciencias bibliométricas desde donde, de forma natural, se han desarrollado los primeros estudios que se centran en la investigación cuantitativa del fenómeno de la Web. De este manera, se han puesto de manifiesto relaciones entre variables como son los *inlinks* (número de enlaces que apuntan a una determinada URL) y diversas medidas de desempeño de la actividad investigadora. Por ejemplo, se han estudiado las relaciones entre enlaces recibidos por sitios web de universidades con los niveles de investigación alcanzados por las mismas (Thelwall, 2001; Smith y Thelwall, 2002), enlaces a sitios web de facultades con su productividad investigadora (Chu, He y Thelwall, 2002), y enlaces a sitios web de revistas con las calidad de las revistas (Vaughan y Hysen, 2002; Vaughan y Thelwall, 2003).

Antes de examinar las posibilidades de aplicar de forma análoga estos razonamientos al ámbito de las webs empresariales, a continuación se exponen tres tipos fundamentales de investigación webométrica que se basan en el empleo de buscadores comerciales para la obtención de los datos.

En primer lugar, el **Análisis de Impacto en la Web** (*Web Impact Analysis*) consiste en estudiar el impacto de ideas, conceptos, marcas, organizaciones, empresas, etc. en la Web a partir de la medición y análisis de las URL de los resultados proporcionados por los buscadores comerciales para un término concreto de búsqueda. Se trata de un método especialmente útil como procedimiento exploratorio en tanto que permite identificar fenómenos proclives a ser investigados más en profundidad mediante el empleo de análisis complementarios, tanto de tipo cualitativo (por ejemplo, a través del análisis de contenidos) o de tipo cuantitativo. Entre sus principales inconvenientes destaca la falta de precisión, siendo ésta mayor en tanto que más común o polisémico sea el término de búsqueda empleado. Una opción para refinar la búsqueda es incorporar otros términos que acoten el fenómeno que se examina. Una limitación significativa es la relativa a la medición de fenómenos con un gran número de resultados, ya que los buscadores comerciales generalmente permiten acceder a un máximo de 1.000 de manera simultánea, reduciendo las posibilidades de análisis. Resumiendo, este tipo de investigación se basa en

⁵ Traducción propia.

contar y analizar las URL recuperadas por los buscadores en respuesta a un término de búsqueda específico.

Ejemplo práctico: en el caso de querer analizar el impacto en la Web de los dos candidatos a las últimas elecciones estadounidenses analizaríamos los resultados (número total estimado de resultados y tipo de URL) que buscadores como Google, Yahoo! o Live Search (MSN) proporcionan para los términos [Obama] y [McCain]. Dado que el número total estimado es tan elevado, podría ser conveniente afinar la búsqueda en función de los intereses de la investigación. Imaginemos que nos interesa estudiar las veces en que éstos aparecen vinculados con sus candidatos a la vicepresidencia. Las búsquedas que realizaríamos serían: [Obama AND Biden] y [McCain AND Palin].

Para aprovechar mejor las posibilidades de investigación es preciso conocer las opciones de búsqueda avanzada de los buscadores así como el empleo de operadores booleanos.

El **análisis de enlaces a sitios web** (Thelwall, 2004) representa un método especialmente indicado para la investigación ya que permite precisar claramente el objeto de la misma. Consiste en analizar las URL que incluyen un enlace a una página o a un sitio web determinado. Este tipo de investigación se puede enfrentar con algunos problemas como son el empleo de varios dominios por una misma organización o de dominios que redireccionan a otros, así como con las limitaciones impuestas por los buscadores para efectuar este tipo de búsquedas. Se pueden distinguir dos tipos principales de investigación. Por un lado, el **análisis de impacto de enlaces**, que se basa en contar el número de enlaces que recibe una página web con el fin de determinar que su relevancia en relación con otras; y, por otro lado, el **mapeado de relaciones de enlaces**, que se lleva a cabo estudiando un conjunto cerrado de sitios web con el objeto de dibujar el diagrama de relaciones entre ellos medido a través de los enlaces que los interconectan.

Ejemplo práctico: siguiendo con el caso anterior, las búsquedas que realizaríamos en Yahoo!, buscador con la mayor gama de funciones avanzadas, sería: [linkdomain:barackobama.com -site:barackobama.com] y [linkdomain:johnmccain.com -site:johnmccain.com]. Realizando estas peticiones de búsqueda se obtienen enlaces que apuntan a cualquiera de las páginas alojadas dentro del dominio señalado exceptuando los enlaces procedentes de páginas del propio sitio. Este procedimiento permite obtener únicamente enlaces externos que son los que mejor permiten evaluar la relevancia de un sitio, ya que los internos pueden ser fácilmente manipulados por la propia organización.

Por último, la **investigación basada en co-enlaces** se centra en el estudio de páginas web que enlazan al mismo tiempo a dos páginas o sitios que son objeto de la investigación. Se trata de un procedimiento especialmente indicado para el estudio de webs que por su tipología no se enlazan frecuentemente entre sí. Es el caso concreto de las webs de empresas, que evitan enlazar a la competencia con el objeto de no desviar tráfico de visitas hacia ella. Estas relaciones de competencia en un sector pueden sin embargo investigarse a mediante el empleo de co-enlaces, como ya se ha hecho analizando la industria de las telecomunicaciones en el mercado chino y en el mercado global (Vaughan y You, 2006) y el subsector del WIMAX también perteneciente a la industria de las telecomunicaciones (Vaughan y You, 2008).

En virtud de lo señalado, la webmetría es una disciplina que aborda el estudio de la Web desde un punto de vista cuantitativo, sin embargo esta aproximación no permite obtener resultados enteramente satisfactorios si no se complementa con una visión cualitativa del fenómeno que permita comprobar, por ejemplo, la significatividad de los enlaces analizados o las motivaciones existentes a la hora de enlazar.

4. LA INVESTIGACIÓN WEBMÉTRICA APLICADA AL ESTUDIO DE EMPRESAS

El estudio webmétrico de sitios comerciales no ha alcanzado el desarrollo de la investigación de sitios webs académicos, lo cual resulta paradójico si tenemos en cuenta el predominio de los primeros y el incremento continuo del comercio electrónico (Thelwall et al. 2005; Vaughan y Wu,

2004). La explicación reside en la inercia investigadora de una nueva disciplina que deriva de la bibliometría y que en sus primeros desarrollos ha asumido el campo de estudio propio de aquella.

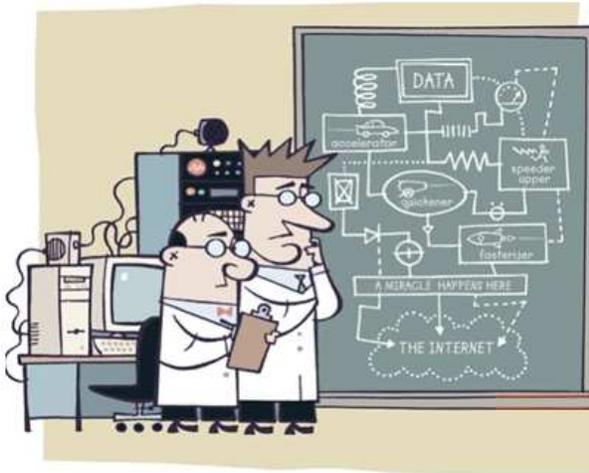
Hasta ahora, los estudios de webs comerciales se han realizado al albur del surgimiento del comercio electrónico, atendiendo fundamentalmente a cuestiones que van desde el diseño general, por ejemplo relativos a la estructura y atractivo del sitio hasta aspectos de índole más técnica, como el empleo de meta-etiquetas para la indexación. Se han llevado a cabo también trabajos de minería de datos a partir de la información recopilada de empresas en la Web, especialmente con el objetivo de obtener información del sector industrial o de la competencia. Sin embargo, pocos han sido los estudios que hayan profundizado en la obtención de información a partir de la estructura de enlaces de la Web. En esta línea destacan algunos trabajos del profesor Mike Thelwall y especialmente de la investigadora Liwen Vaughan.

Han sido las contribuciones de esta profesora las que se han centrado en analizar la posibilidad de emplear el número de enlaces recibido por la web de una empresa como un indicador de desempeño de su actividad. Para ello intenta comprobar si dichos datos están correlacionados con determinadas variables contables, de forma análoga a lo que se ha venido haciendo con las webs académicas. Su investigación se ha centrado fundamentalmente en el sector de las telecomunicaciones y de tecnologías de la información así como en los mercados norteamericano (EE.UU. y Canadá) y chino.

Vaughan y Wu (2004) investiga si existen relaciones entre los enlaces que reciben las 100 primeras empresas de tecnología de la información (grupo 1) y las 100 primeras empresas de titularidad privada (grupo 2) en el mercado chino y algunas variables contables.

La elección del sector de tecnologías de la información (grupo 1) responde fundamentalmente a la propia naturaleza de la actividad que llevan a cabo las empresas, con un componente altamente tecnológico, lo cual asegura una presencia destacada en la Web. Es importante tener en consideración que la presencia de cada sector de actividad parece seguir un patrón de comportamiento distinto, siendo más significativa y fiable la investigación cuanto mayor sea la exposición de la industria en la Web, ya sea por la actividad que desempeña o por los niveles de desarrollo de comercio electrónico. La elección de un segundo grupo de empresas pertenecientes a sectores de actividad distintos tiene por objeto comparar resultados para determinar si la homogeneidad sectorial es un factor significativo a la hora de obtener resultados.

Las variables financieras recabadas para las empresas tecnológicas (grupo 1) son: ingresos brutos, resultado, ingresos por exportaciones y gastos de investigación y desarrollo; mientras que para las empresas de titularidad privada, el único dato disponible es el correspondiente a los ingresos brutos.



La obtención de los datos de enlaces entrantes a la páginas web objeto de estudio (*inlinks*) se realiza empleando distintos buscadores comerciales que permiten esta función. En verano de 2002, momento en que se recaba la información, los buscadores disponibles eran: Google, AltaVista, AllTheWeb y MSN Search.

Una variable a la que se presta especial consideración en el trabajo es la **edad de las páginas web**. Vaughan y Thelwall (2003) muestran que el número de *inlinks* que recibe un sitio está correlacionado con la edad del mismo, es decir, las páginas webs más antiguas reciben mayor número de enlaces por un simple proceso de acumulación a lo largo del tiempo. Por ello, con el fin de controlar este efecto se utiliza la variable *edad* de la página web a partir de los datos

proporcionados por el Internet Archive (www.archive.org).

Los resultados al calcular el coeficiente de correlación de Spearman para las empresas del grupo 1 se muestran en la siguiente tabla:

Coeficiente de correlación Spearman ($p < 0,01$) de de	Ingresos brutos	Resultado	Gasto en I+D
Nº enlaces	0,51	0,3	0,64
Nº enlaces / edad de la web	0,5	0,3	0,63

Los datos muestra una correlación significativa entre los enlaces recibidos y los ingresos brutos de las empresas, siendo especialmente alta en el caso del gasto en investigación y desarrollo. Esto parece indicar que las empresas que invierten más en I+D gozan también de una mejor presencia en la Web y por tanto sus sitios son más visibles y más enlazados. Por el contrario, los resultados para el grupo 2 de empresas indican que no existe correlación entre los ingresos brutos y los enlaces recibidos, lo cual parece indicar que las medidas de desempeño de las empresas de distintos sectores no son directamente comparables, siendo los niveles de presencia en la web también de intensidad distinta.

En sintonía con este trabajo, Vaughan (2004b) investiga si las relaciones examinadas se verifican en países distintos, para lo cual estudia las empresas del sector de tecnología de la información en China y en Estados Unidos. Los resultados se pueden observar en la siguiente tabla:

Coeficiente de correlación de Spearman ($p < 0,01$) de	Enlaces & Ingresos	Enlaces & Resultado	Enlaces/edad de la web & Ingresos	Enlaces/edad de la web & Resultado
China	0,51	0,30	0,50	0,30
Estados Unidos	0,51	0,35	0,58	0,37

Se refuerza pues la evidencia de la relación entre *inlinks* y las magnitudes de ingresos y beneficio, especialmente la primera de ellas. Las conclusiones de este trabajo sugieren que con independencia de grado de desarrollo tecnológico de un país o con independencia del momento en que éste se produjo, la relación entre *inlinks* y variables financieras es consistente.

Por último, un tercer trabajo (Vaughan, 2004a) compara las empresas de tecnologías de la información de Canadá con las de Estados Unidos. En este trabajo se analiza la población completa de empresas del sector, a diferencia de los anteriores trabajos que basaban la selección de empresas en rankings. Con ello se pretende llegar a resultados más generalizables.

Se toma en consideración una nueva variable, el tamaño de las empresas, medido a partir del número de empleados, al considerar que una compañía mayor obtendrá mayores ingresos, permaneciendo el resto de variables iguales. Calculando de nuevo el coeficiente de correlación de Spearman se obtienen los siguientes resultados:

Coeficiente de correlación Spearman ($p < 0,01$) de de	Enlaces & Número de empleados	Enlaces & Ingresos	Enlaces & Ingresos por empleado	Enlaces/edad de la web & Ingresos	Enlaces/edad de la web & Ingresos por empleado
Canada	0,57	0,55	0,35	0,55	0,36
Estados Unidos	0,68	0,71	0,53	0,67	0,51

Al margen de apoyar las conclusiones de los trabajos anteriores, los datos indican que aún incluyendo la variable tamaño sigue existiendo una correlación significativa entre ambas.

Las conclusiones obtenidas a partir de los anteriores trabajos permiten indicar que los enlaces recibidos por páginas webs comerciales podrían emplearse como un indicador complementario del desempeño de las mismas. Dada la correlación positiva existente, se puede afirmar que empresas con un mejor desempeño financiero muestran también una mejor presencia en la Web atrayendo la atención de mayor número de páginas. De acuerdo con lo observado, parece que únicamente sería comparable el desempeño a través de los *inlinks* si el grupo de empresas seleccionado es homogéneo, es decir, pertenecen a un mismo sector de actividad.

Los tres estudios presentados anteriormente son de carácter puramente cuantitativo, sin embargo la realización de análisis cualitativos puede ser muy útil para interpretar y comprender el proceso y las motivaciones por los que se establecen los enlaces (Bar-Ilan, 2005; 2008). Si bien este tipo de investigación sobre las motivaciones para la creación de enlaces tienen una mayor tradición en el ámbito de las webs académicas (Chu, 2005; Thelwall, 2003; Wilkinson et al., 2003), Vaughan et al. (2006) estudiaron las motivaciones para enlazar webs comerciales concluyendo que la inmensa mayoría de enlaces se establecían con motivaciones comerciales siendo el origen más frecuente de los mismos directorios online, listados de empresas y listados de productos.

5. POSIBLES LÍNEAS DE INVESTIGACIÓN

Los caminos marcados para futuras investigaciones sugieren apostar por tener en cuenta las siguientes cuestiones:

- Prestar atención a variables como el tamaño del sitio web de la empresa, ya que sitios web más grandes pueden atraer mayor número de enlaces.
- Analizar el caso de empresas en otras áreas geográficas (por ejemplo, Europa) y en otros sectores de actividad.
- Incluir una variable temporal que permita estudiar la evolución de los enlaces en el tiempo y su relación con el cambio en las variables financieras.
- Emplear otras variables contables y financieras (por ejemplo, vinculadas a la rentabilidad en bolsa para el caso de empresas cotizadas).
- Recabar datos de la web procedentes de fuentes alternativas a los buscadores comerciales, como son por ejemplo los buscadores de blogs o los buscadores de noticias.

6. CONCLUSIÓN

La amplitud y la profundidad del fenómeno de la Web son tan significativas que desde áreas de conocimiento muy distintas se está abordando el desarrollo de nuevas metodologías de investigación (Fielding et al., 2008). En el ámbito de las ciencias sociales las disciplinas que con más interés se han acercado al estudio de esta realidad van desde los estudios culturales, la sociología, la psicología, el análisis de redes sociales, las ciencias de la información, hasta, de forma no menos importante, la economía, el marketing y los estudios de empresa en sentido global. Las técnicas webmétricas son una respuesta más a esta nueva realidad emergente y este gran reto que es para las ciencias sociales, Internet y la Web. Es por esto que la tendencia sea hacia una mayor transversalidad e interdisciplinariedad en la labor investigadora.

BIBLIOGRAFÍA

- Almind, T. C. e Ingwersen, P. (1997) Infometric analyses on the World Wide Web Methodological approaches to "webometrics." *Journal of Documentation*, 53(4), pp. 404-426.
- Bar-Ilan, J. (2005) What do we know about links and linking? A framework for studying links in academic environments. *Information Processing and Management*, 41(4), pp. 966–973.
- Bar-Ilan, J. (2008) Infometrics at the beginning of the 21st century – A review. *Journal of Informetrics*, 2, pp. 1-52.
- Battelle, J. (2005) *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. London: Portfolio.
- Berners-Lee, T. (1999) *Weaving the Web*. San Francisco: Harper.
- Björneborn, L. (2004) *Small-world link structures across an academic Web space: A library and information science approach*. Doctoral dissertation, Royal School of Library and Information Science, Copenhagen, Denmark. Disponible en: <http://vip.db.dk/lb/phd/phd-thesis.pdf> [21 Enero 2009].
- Björneborn, L. e Ingwersen, P. (2004) Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, 55(14), pp. 1216–1227.
- Brin, S. y Page, L. (1998) The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30, pp. 1-7.
- Chu, H., He, S. y Thelwall, M. (2002) Library and information science schools in Canada and USA: A Webometric perspective. *Journal of Education for Library and Information Science* 43, pp. 110-125
- Chu, H. (2005) Taxonomy of inlinked Web entities—What does it imply for webometric research? *Library & Information Science Research*, 27(1), pp. 8–27.
- Fielding, N., Lee, R. M. y Blank, G. (eds.) (2008) *SAGE Handbook of Online Research Methods*. Sage Publications.
- Fundación Telefónica (2008) *La Sociedad de la Información en España 2008*. Ariel. Disponible en: <http://e-libros.fundacion.telefonica.com/sie08/> [17 Enero 2009].
- Garfield, E. (1979) *Citation indexing: Its theory and applications in science, technology and the humanities*. Wiley. Interscience, New York.
- Jenkins, H. (2006) *Convergence Culture: Where Old and New Media Collide*. Cambridge, MA: MIT Press.
- Keen, A. (2007) *The Cult of the Amateur: How the Democratization of the Digital World is Assaulting Our Economy, Our Culture, and Our Values*. New York: Doubleday Currency.
- Machill, M., Beiler, M., y Zenker, M. (2008) Search-engine research: a European-American overview and systematization of an interdisciplinary and international research field. *Media, Culture & Society*, 30(5), pp. 591–608.
- ONTSI (2008) *La sociedad en red. Informe anual 2007*. Observatorio Nacional de las Telecomunicaciones y la Sociedad de la Información (ONTSI) Disponible en: <http://observatorio.red.es/informes-anales/articulos/id/2426/informe-anual-2007.html> [16 Enero 2009].
- O'Reilly, T. (2005) What is Web 2.0. Design Patterns and Business Models for the Next Generation of Software. Disponible en: <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html> [17 Enero 2009].
- Smith, A. y Thelwall, M. (2002) Web Impact Factors for Australasian universities. *Scientometrics*, 54, pp. 363-380.
- Thelwall, M. (2001) Extracting macroscopic information from Web links. *Journal of the American Society for Information Science and Technology*, 52(13), pp. 1157–1168.
- Thelwall, M. (2003) What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation. *Information Research*, 8(3), paper 151. Disponible en: <http://informationr.net/ir/8-3/paper151.html> [17 Enero 2009].
- Thelwall, M. (2004) *Link Analysis: An Information Science Approach*. Amsterdam: Elsevier.
- Thelwall, M., Vaughan, L. y Björneborn, L. (2005) Webometrics. In B. Cronin (Ed.), *Annual review of*

information science and technology. Medford, NJ: Information Today, pp. 81–135.

Vaughan, L. y Wu, G. Z. (2004) Links to commercial websites as a source of business information. *Scientometrics*, 60(3), pp. 487–496.

Vaughan, L. (2004a) Exploring website features for business information. *Scientometrics*, vol. 61, No. 3, pp. 467-477.

Vaughan, L. (2004b) Web hyperlinks reflect business performance—A study of US and Chinese IT companies. *Canadian Journal of Information and Library Science*, 28(1), pp. 17–31.

Vaughan, L. y Hysen, K. (2002) Relationship between links to journal web sites and Impact Factors. *ASLIB Proceedings: New Information Perspectives*, 54, pp. 356-361.

Vaughan, L. y Thelwall, M. (2003) Scholarly use of the Web: What are the key inducers of links to journal web sites? *Journal of the American Society for Information Science and Technology*, 54, pp. 29-38.

Vaughan, L. y You, J. (2006) Comparing business competition positions based on Web co-link data—The global market vs. the Chinese market. *Scientometrics*, 68(3), pp. 611–628.

Vaughan, L. y You, J. (2008) Content assisted web co-link analysis for competitive intelligence. *Scientometrics*, 77 (3), pp. 433-444.

Vaughan, L., Gao, Y. y Kipp, M. (2006) Why are hyperlinks to business Websites created? A content analysis. *Scientometrics*, 67(2), pp. 291–300.

Wilkinson, D., Harries, G., Thelwall, M. y Price, L. (2003) Motivations for academic web site interlinking—Evidence for the Web as a novel source of information on informal scholarly communication. *Journal of Information Science*, 29(1), pp. 49–56.