

Webometrics: A Personal Review

Ronald Rousseau

Associate Professor,

Department Industrial Sciences and Technology, Association K. U. Leuven , Belgium

E-mail: ronald.rousseau@khbo.be

【Abstract】

We present a short and necessarily incomplete review on recent developments in the field of webometrics. Considering the Internet as a large network, we discuss its general structure and possible use for research evaluation. We finally mention its potential for business applications.

Keyword

Webometrics ; Social network theory ; Web 2.0 ; Business applications

Introduction: A Definition

Björneborn (2004), (see also Björneborn & Ingwersen (2001), Ingwersen & Björneborn (2004)), defines webometrics as:

The study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web drawing on bibliometric and informetric approaches.

These authors, and we fully agree with them, consider webometrics as a subfield of cybermetrics and of bibliometrics, which are both considered as subfields of informetrics. Following Tague-Sutcliffe (1992) they define informetrics as:

The study of quantitative aspects of information in any form, not just records or bibliographies, and in any social group, not just scientists.

For some reason terminology *in* and *of* our field was and still is a delicate matter (Brookes, 1990). Many colleagues have strong opinions about the particular use of the terms bibliometrics, informetrics, scientometrics, webometrics, cybermetrics and so on. In our opinion Björneborn and Ingwersen's approach is the most sensible one.

An important part of webometrics deals with Web link analysis (Thelwall, 2004). Link analysis, in general, deals with identifying relationships and associations between objects not apparent from isolated pieces of information (Payne, 2007). Web link analysis, in particular, is carried out in order to understand and extract information from the hyperlink structure of collections of Web documents. The Statistical Cybermetrics Research Group of Wolverhampton University under the direction of Mike Thelwall is the leading group in the field of link

analysis. Using a specialist Web crawler they are able to avoid many (but not all) of the problems related to the use of commercial search engines. Indeed, commercial search engines do not offer the control on harvested data a scientific investigator wants to have. Problems with commercial search engines and their lack of transparency, coverage and reliability were made abundantly clear by Bar-Ilan (2005).

The Web as a Large Network and Some Social Implications

The fact that the Internet is a large network has intensified scientific relations between computer scientists, social network analysts, economists and information scientists. Nowadays social network theory directly influences the way researchers think and formulate ideas on the Web and other network structures such as those shown in enterprise interactions (Raghavan, 2001). Underlying any concrete network lies a graph, a structure studied by mathematicians since Euler solved the problem of the Königsberg bridges in 1735 (published in 1741). Moreover, as many biological interactions and relations between disease genes can be described using graph or network terminology (Barbano, et al., 2007; Goh, et al., 2007). Network studies have become a hot topic in many fields of science, including systems biology and human genetics. We further mention applications of graph theory in computer science and artificial intelligence (neural networks), and recent theories concerning the Web and the free market economy, geography and transport networks. Informetrics researchers study citation networks, co-citation networks, collaboration structures and other forms of social interaction

networks (Melin, Danell, & Persson, 2000; Otte & Rousseau, 2002).

Also other collaborations, such as movie actor collaborations have inspired fellow scientists (Barabási & Albert, 1999). These authors and others link their research to the so-called small-world phenomenon or “six degrees of separation” phenomenon (Björneborn, 2004; Braun, 2004; Guns, 2008; Karinthy, 1929; Milgram, 1967; Newman & Watts, 1999). A small-world network is then characterized as a network exhibiting a high degree of clustering and having at the same time a small average distance between nodes. Moreover, the “hubs” and “authorities” approach (Kleinberg, 1999) is related to the Pinski-Narin influence weight citation measure (Pinski & Narin, 1976) and Google’s PageRank (Langville & Meyer, 2006) and mimics the idea of “highly cited documents” (authorities) and reviews (hubs). Following Kleinberg we note that hubs and authorities exhibit a mutually reinforcing relationship: a good hub will point to many authorities, and a good authority will be pointed at by many hubs.

Trying to avoid the long scrolling list syndrome, link structure plays an important role in rankings obtained in Web information retrieval (Langville & Meyer, 2006; Lempel & Moran, 2001). Clearly, understanding the (social) structure of networks, and in particular of the Internet, is a first step for data mining exercises.

Language is an important element in social, academic and business interactions. This also applies to Web interactions. Thelwall, Tang, & Price(2003) found that English is the dominant language of the Western Europe academic Web, with Greece,

however, as a major exception. Thelwall & Tang (2003) studying academic information exchange in and between Taiwan and Mainland China found that English is not the language of preference for cross-strait links and that Chinese is preferred instead. When it comes to the use of English on the Web, most non-native speakers prefer the American variant above the British one (Rogge & Rousseau, 2007).

Web Links and the Bibliometric Laws

Already in 1997 Rousseau (1997), using a small sample, established that inlinks on the Internet follow a power law relation. This observation was subsequently confirmed on a sample of a much larger scale by M. Faloutsos, P. Faloutsos and C.Faloutsos (1999). Since then many refinements of this general power law model have been proposed. As many link relations on the Web can be described, or at least approximated, by Zipf-type and Lotka-type power laws, the whole theory of so-called Lotkaian informetrics (Egghe, 2005) is applicable to this kind of structural studies.

One of the most interesting recent developments related to modeling networks is the work published by Jackson and Rogers (2007). They provide a model for all types of socially-generated networks. Such networks are characterized by the following features:

- Low distances between nodes (small diameter – small world property).
- The presence of more high and low degree nodes than when links are formed independently at random.
- Exhibiting high clustering of links on a local level.
- Higher degree links tend to be linked to other

higher degree links; and lower degree links tend to be linked to other lower degree links (assortativity = homophily = birds of a feather flock together).

Jackson and Rogers then construct such networks and show that the random/network-based meeting ratio is the essential parameter.

The Web as a Tool for Retrieving Information Used in Research Evaluation Exercises (Rousseau, 2008)

One of the first applications of classical informetrics to the Web was Ingwersen's "Web impact factor" (Ingwersen, 1998). One (loosely stated) definition of a Web impact factor is: the number of external inlinks divided by The number of pages found at the entity of which a Web impact factor is counted (typically this is a country or a university). Yet, it soon turned out that the definition itself was not without problems. What is the correct (or at least an acceptable) way of defining Web pages and hence of calculating Web impact factors? What is their validity in terms of measuring the impact of a particular Web space?

Concentrating on the numerator of the Web impact factor, it was found by Thelwall and Harries (2003) that generally the number of links to Web sites of an academic scholar or institution can indeed be conceived as a measure of prominence. Yet, Thelwall & Tang (2003) studying how the Web has become an important means of academic information exchange discovered that in Taiwan inlinks to university Web sites correlate with research productivity, but that this was not the case for Mainland China.

Using the public Web for research evaluation purposes in a similar way as global or local citation databases (such as Thomson Scientific's Web of Knowledge) is another matter. Important questions to be answered are: How do researchers cite on the Web? What kind and which percentage of the total number of scientific documents are available on the Web? It is clear that the role of pre-print archives, institutional repositories and personal Web pages is of utmost importance for such Web-based evaluation exercises.

How can we study references and relations from the Web to paper-based sources? Vaughan and Shaw (2003; 2005) were the first ones to make a full-scale investigation of this type of relation. They define Web citations as mentions of an article published in a paper-based source in a source on the Web. The term Web-to-print citation for this type of citation has been proposed by Van Impe and Rousseau (2006). Vaughan and Shaw propose an interesting classification of Web references in three categories, according to academic level:

- Research impact, similar to a classical reference.
- Other intellectual impact (reference in a syllabus, a popular science Web site, academic questions & answers, ...).
- Non-intellectual impact (reference in a table of contents, an online bibliography, an author's home page, etc.).

Investigating journal articles in four scientific domains (biology, genetics, medicine and multidisciplinary sciences), Vaughan and Shaw found that about 30% of Web citations belonged to the first category. Vaughan & Shaw (2005)

discovered that, at least in the four domains studied by them, there exists a significant correlation between Web citation counts and classical citation counts. These four domains (biology, genetics, medicine and multidisciplinary sciences) belong to the exact and medical sciences. If their findings could be confirmed in general (for other scientific domains, including the social sciences and the humanities), then the Vaughan-Shaw approach would, in particular, be very useful for citation studies in the humanities, where collecting classical (paper-based) citations is tedious. Web citation counts would then offer a relatively simple way to study the visibility of authors, articles and journals in all, or at least many, humanities fields. A small-scale study by Van Impe and Rousseau (2006) in the fields of general history, history of the book and archaeology, and for articles written mainly in other languages than English (Dutch, French) was only moderately successful. Classical, as well as Web-to-print citation scores were found to be too low to draw significant conclusions.

Björneborn (2004) studied what types of Web links, Web pages and Web sites function as cross-topic connectors in small-world link structures across an academic Web space. In his investigations he found that the structure of the Web can better be compared to a corona, rather than a bow-tie, as suggested by Broder, et al. (2000). Within the academic Web space, computer departments play a special role as connectors between other departments. In terms of social network indicators, it is expressed by the fact that they have a high betweenness centrality in the academic Web.

In relation to research evaluation we remind the reader that outputs of technological and innovation research are in many cases not written up as such but appear as designs, applications, models or know-how (Jansz, 2000). We suggest that network analysis covering not only scientific journal articles, but also trade literature (as an example), may lead to a better understanding of the visibility and practical importance of such technological outputs.

New Challenges: Web 2.0 and the (Social) Semantic Web

The term Web 2.0 refers to the second phase in the development of the World Wide Web. Technically, its main feature is the move from separate Websites to complete platforms of interactive Web applications. Weblogs, wikis, podcasts, RSS-feeds, Web videos and other social software are developments related to Web 2.0 (O'Reilly, 2005). Within the scientific community we can mention scientific blogs, article sharing (such as sharing preprints through the arXiv.org or the E-LIS server) and open scientific data sets as typical Web 2.0 developments.

Web 2.0 can be seen as a step towards the semantic Web, or, in particular, the social semantic Web (Guns, 2008; Mikroyannidis, 2007). The semantic Web (Berners-Lee, Hendler, & Lassila, 2001) is an extension of the original Web, incorporating a vision on Web structure and tools in which not only humans deal with and understand information, but also software. Essential aspects of the semantic Web are the use of metadata and ontologies. The term *social semantic Web* refers to

data on the semantic Web containing social information.

Ontologies providing shared and common domain “theories” are key assets for the semantic Web. The best known definition of an ontology is that Gruber (1993) states that an ontology is a formal, explicit specification of a shared conceptualization. The term *conceptualization* refers to an abstract model of phenomena in the world. This model came to existence by having identified the relevant concepts related to these phenomena. *Explicit* means that the type of concepts used, and the constraints on their use are explicitly defined. *Formal* refers to the fact that the ontology should be machine-readable. The term *shared*, finally, reflects the fact that an ontology should capture consensual knowledge, accepted by communities of users. An ontology hence determines the classes of a field and organizes them within a class structure (a taxonomy). The unfolding of an ontology moreover provides criteria for distinguishing various types of objects (concrete and abstract, existent and non-existent, real and ideal, independent and dependent) and their ties (relations, dependences and predication). In this way, it becomes possible to reason about these relations (Fensel, 2000; Ding, 2001; B. Rousseau & R. Rousseau, 2002).

An important tool for the development of the semantic Web is the RDF-standard, developed by the World Wide Web Consortium (W3C). RDF stands for Resource Description Framework and describes relations between resources, unambiguously identified by a Uniform Resource Identifier (URI), in RDF statements (triples). These triples consist of a subject and an object, linked by a predicate. In order

to control what types of RDF statements are allowed ontology languages such as OWL (Web Ontology language) have been developed.

RDF extends the basic XML (eXtensible Markup Language) model and syntax with the aim of describing resources. For this, RDF uses the “namespace” functionality of XML. Documents, containing multiple markup vocabularies, pose problems of recognition and collision. Software modules need to be able to recognize the tags and attributes which they are designed to process, even in the face of “collisions” occurring when markup intended for some other software package uses the same element type or attribute name. These considerations require that document constructs should have universal names, whose scope extends beyond the document in which they are contained. XML namespaces accomplish this. Formally, a namespace is a formal collection of terms managed according to a policy or algorithm (Duval, Hodgins, Sutton, & Weibel, 2002). An XML namespace, in particular, is a collection of names, identified by a URI reference, which are used in XML documents as element types and attribute names. Namespaces allow RDF to define a uniquely identifiable set of properties. This set is called a schema. It can be accessed via the URI defined in the namespace. Because RDF is defined within XML it inherits all XML properties, such as support for rendering data in several different languages. The main advantage of RDF lies in the fact that resource description groups can concentrate on semantic problems, instead of syntax and structure of metadata (B. Rousseau & R. Rousseau, 2002). As many informetric (webometric) relations can be straightforwardly expressed in RDF,

an RDF is a useful tool for studying the interplay of several relations at once.

Web 2.0 applications such as Flickr (an online image management and sharing application), have already been studied by Webometric researchers. Angus, Thelwall and Stuart (2007) studied tagging behavior in Flickr in order to find out if tagging was useful for people other than the person who uploaded the pictures. This type of research is related to the “taxonomies versus folksonomies“ question. Generally one may say that the larger the target group and the more serendipity plays a role, the more social tagging (folksonomies) becomes important. Focused business applications, however, do best by using well-defined taxonomies.

Intelligent Agents and Applications to Business Information

We are constantly obliged to make decisions without having enough information or experience to make the “best“ or even an ”intelligent“ choice. The amount of information made available via networks and databases is, moreover, still increasing at a high rate. Search engines can not cope with this enormous amount of data and yield only limited support in localizing the information searched for by the user. Intelligent agents can be of help here because they transform passive machines into active personal assistants and counselors (Maes, 1994; Rousseau, 2008).

The semantic Web structures the contents of Web pages and hence, creates an environment where software agents, roaming from page to page, can – almost - autonomously perform complex tasks on behalf of users. Intelligent agents are small software

programs searching the Internet in order to find information that answers the queries of their owners.

Agents are semi-autonomous computer programs assisting the user in handling computer applications of all kinds. Agents do not only use the available semantic infrastructure, but also create and maintain this infrastructure. Good agents help people finding the information they need, allowing them to spend less time in the search process, and more on actually analyzing the information they have found. A good Internet agent is communicative, capable, autonomous and adaptive (Hendler, 2001).

An agent must not only be able to act, but also to make suggestions to its user. In other words: An agent offers advice and services. Good agents can do things on the Web without its user knowing all the details. In this way users can delegate tasks to their agents. Examples of such tasks are: searching, classifying and storing information, but also reading e-mail, making appointments, keeping a diary and scheduling a trip abroad (Maes, 1994), see also (Langville and Meyer, 2006).

In a number of studies Liwen Vaughan investigated the relation between linking to commercial sites and business information. In 2004, she found that the number of inlinks to an information technology (IT) company in the US or China correlated significantly with the company’s revenue and profit, reflecting business performance. Clearly this observation can be used in web mining practice. Continuing these investigations she mapped business competitive positions of a number of telecommunication companies from different regions of the world (Vaughan & You, 2005). In this investigation she used Web co-links, mapped by

multidimensional scaling (MDS). Although competitors usually do not link to one another, they were often co-linked, a fact that was confirmed in yet another study (Vaughan, Gao, & Kipp, 2006).

Conclusion

Clearly, the study of networks (graphs) and of the Internet in particular, as performed in the fields of webometrics and cybermetrics is nowadays a hot topic. It brings together the information sciences, computer science, physics, social networks theory and even the life sciences, uniting them around a common theme. For further information we refer the reader to Newman's review (Newman, 2003) and Noruzi's and Payne's doctoral theses (Noruzi, 2007;

Payne, 2007). Bar-Ilan (2008) reviews webometrics in relation to informetrics in general. Mathematically inclined readers may start their investigations for new applications based on Langville and Meyer's book about Google's PageRank (Langville & Meyer, 2006).

Acknowledgements

The author thanks Yuxian Liu (Tongji University) and Raf Guns (Antwerp University) for helpful suggestions to improve this text.

(收稿日期：2008 年 5 月 28 日)

References

- Angus, E., Thelwall, M., & Stuart, D. (2007). University groups in Flickr: Tagging for purpose or pleasure? In Daniel Torres-Salinas, & Henk F. Moed (Eds.), *Proceedings of ISSI 2007* (pp.824-825). Madrid: CSIC.
- Barabási, A. & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286 (5439), 509-512.
- Barbano, P.E., Spivak, M., Flajolet, M., Nairn, A.C., Greengard, P. & Greengard L. (2007). A mathematical tool for exploring the dynamics of biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 19169-19174.
- Bar-Ilan, J. (2005). Expectations versus reality – search engine features needed for Web research at mid 2005. *Cybermetrics*, 9(1). Retrieved May 28, 2008, from <http://www.cindoc.csic.es/cybermetrics/articles/v9i1p2.pdf>
- Bar-Ilan, J. (2008). Informetrics at the beginning of the 21st century – a review. *Journal of Informetrics*, 2, 1-52.
- Berners-Lee, T., Hendler, J. & Lassila, O. (2001). The semantic Web. *Scientific American*, 284(5), 34-43.
- Björneborn, L. (2004). *Small-world link structures across an academic Web space: a library and information science approach*. Unpublished doctoral dissertation, Royal School of Library and Information Science, Denmark.
- Björneborn, L. & Ingwersen, P. (2001). Perspectives on webometrics. *Scientometrics*, 50(1), 65-82.
- Braun, T. (2004). Hungarian priority in network theory. *Science*, 304(5678), 1745-1746.
- Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., & et al. (2000). Graph structure in the Web. *Computer Networks*, 33(1), 309-320.

- Brookes, B.C. (1990). Biblio-, sciento-, infor-metrics??? What are we talking about? In L. Egghe, & R. Rousseau (Eds.), *Informetrics 89/90* (pp.31-43). Amsterdam: Elsevier.
- Ding, Y. (2001). A review of ontologies with the semantic web in view. *Journal of Information Science*, 27(6), 377-384.
- Duval, E., Hodgins, W., Sutton, S. & Weibel, S. L. (2002). Metadata principles and practicalities. *D-Lib Magazine*, 8(4). Retrieved May 28, 2008, from <http://www.dlib.org/april02/weibel/04weibel.html>
- Egghe, L. (2005). *Power laws in the information production process: Lotkaian informetrics*. Amsterdam: Elsevier.
- Faloutsos, M., Faloutsos, P., & Faloutsos, C. (1999). On power-law relationships of the Internet topology. *Computer Communication Review*, 29, 251-263.
- Fensel, D. (2000). *Ontologies: Silver bullet for knowledge management and electronic commerce*. Berlin: Springer-Verlag.
- Goh, KI., Cusick, M.E., Valle, D., Childs, B., Vidal, M., & Barabási, A.-L. (2007). The human disease network. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 8685-8690.
- Gruber, T.R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5, 199-220.
- Guns, R. (2008). Unevenness in network properties on the social semantic Web. In D. Flejter, S. Grzonkowski, T. Kaczmarek, M. Kowalkiewicz, T. Nagle, & J. Parkes (Eds.), *BIS 2008 Workshop Proceedings*, (pp.27-36). Innsbruck, Austria: Department of Information Systems, Poznań University of Economics.
- Hendler, J. (2001). Agents and the semantic web. *IEEE Intelligent Systems*, 16(2), 30-37.
- Ingwersen, P. (1998). The calculation of web impact factors. *Journal of Documentation*, 54(2), 236-243.
- Ingwersen, P., & Björneborn, L. (2004). Methodological issues of webometric studies. In H. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research* (pp.339-369). Dordrecht: Kluwer Academic Publishers.
- Jackson, M. O., & Rogers, B.W. (2007). Meeting strangers and friends of friends: how random are social networks? *American Economic Review*, 97(3), 890-915.
- Jansz, C.N.M. (2000). Some thoughts on the interaction between scientometrics and science and technology policy. *Scientometrics*, 47(2), 253-264.
- Karinthy, F. (1929). *Minden másképpen van (Ö tvenkét vasárnap)*. Athenaeum, Irodalmi és Nyomdai Rt., Budapest.
- Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46, 604-632.
- Langville, A. N., & Meyer, C. D. (2006). *Google's PageRank and beyond: the science of search engine rankings*. Princeton (NJ): Princeton University Press.
- Lempel, R., & Moran, S. (2001). SALS: the stochastic approach for link-structure analysis. *ACM Transactions on Information Systems*, 19(2), 131-160.
- Maes, P. (1994). Agents that reduce work and information overload. *Communications of the ACM*, 37(7), 31-40 + 146.
- Melin, G., Danell, R., & Persson, O. (2000). A bibliometric mapping of the scientific landscape on Taiwan. *Issues & Studies*, 36(5), 61-82.
- Mikroyannidis, A. (2007). Toward a social semantic Web. *Computer*, 40(11), 113-115.
- Milgram, S. (1967). The small world problem. *Psychology Today*, 2, 60-67.

- Newman, M.E.J. (2003). The structure and function of complex networks. *SIAM Review*, 45(2), 167-256.
- Newman, M.E.J., & Watts, D.J. (1999). Scaling and percolation in the small-world network model. *Physical Review E*, 60, 7332-7342.
- Noruzi, A. (2007). Webométrie – concepts, pratiques, outils et applications: analyses webométriques et étude du facteur d'impact des sites web académiques français. Université Paul Cézanne – Aix-Marseille III. Retrieved May 28, 2008, from http://crrm.u-3mrs.fr/web/IMG/pdf/AlirezaNoruzi_T.pdf
- O'Reilly, T. (2005). What is Web 2.0. Design patterns and business models for the next generation of software. Retrieved May 28, 2008, from <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-Web-20.html>
- Otte, E., & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, 28(6), 441-453.
- Payne, N. (2007). *A longitudinal study of academic Web links: identifying and explaining change*. Unpublished doctoral dissertation, University of Wolverhampton, England (UK).
- Pinski, G., & Narin, F. (1976). Citation influences for journal aggregates of scientific publications: theory, with applications to the literature of physics. *Information Processing and Management*, 12(5), 297-312.
- Raghavan, P. (2001). Social networks on the Web and in the enterprise. In *Web Intelligence: Research & Development* (N. Zhong, Y.Y. Yao, J. Liu, and S. Ohsuga, Eds., pp.58-60). (Lecture Notes in Artificial Intelligence, No.2198). Berlin: Springer.
- Rogge, G., & Rousseau, R. (2007). Is the European Web British or American? In D. Torres-Salinas, & H. F. Moed (Eds.), *Proceedings of ISSI 2007* (pp.928-929). Madrid: CSIC.
- Rousseau, B., & Rousseau, R. (2002). Some ideas concerning the semantic Web. *Library and Information Service*, 8, 39-49.
- Rousseau, R. (1997). Sitations: an exploratory study. *Cybermetrics*, 1(1), paper 1. Retrieved May 28, 2008, from <http://www.cindoc.csis.es/cybermetrics/articles/v1i1p1.html>
- Rousseau, R. (2008). Publication and citation analysis as a tool for information retrieval. In D. Goh, & S. Foo(Eds.), *Social Information Retrieval Systems* (pp.252-268). Hershey: IGI Global.
- Tague-Sutcliffe, J. (1992). An introduction to informetrics. *Information Processing and Management*, 28(1), 1-3.
- Thelwall, M. (2004). *Link analysis: an information science approach*. Boston: Elsevier Academic Press.
- Thelwall, M., & Harries, G. (2003). The connection between research of a university and counts of links to its Web pages: an investigation based upon a classification of the relationships of pages to the research of the host university. *Journal of the American Society for Information Science and Technology*, 54, 594-602.
- Thelwall, M., & Tang, R. (2003). Disciplinary and linguistic considerations for academic Web linking: an exploratory hyperlink mediated study with mainland China and Taiwan. *Scientometrics*, 58(1), 155-181.
- Thelwall, M., Tang, R., & Price, L. (2003). Linguistic patterns of academic Web use in Western Europe. *Scientometrics*, 56(3), 417-432.
- Van Impe, S., & Rousseau, R. (2006) Web-to-Print Citations and the Humanities. *Information - Wissenschaft und Praxis*, 57(8), 422-426.

- Vaughan, L. (2004). Web hyperlinks reflect business performance: a study of US and Chinese IT companies. *The Canadian Journal of Information and Library Science*, 28(1), 17-31.
- Vaughan, L., Gao, Y., & Kipp, M. (2006). Why are hyperlinks to business Websites created? A content analysis. *Scientometrics*, 67(2), 291-300.
- Vaughan, L., & Shaw, D. (2003). Bibliographic and Web citations: what is the difference? *Journal of the American Society for Information Science and Technology*, 54(14), 1313-1322.
- Vaughan, L., & Shaw, D. (2005). Web citation data for impact assessment: a comparison of four science disciplines. *Journal of the American Society for Information Science and Technology*, 56(10), 1075-1087.
- Vaughan, L., & You, J. (2005). Mapping business competitive positions using Web co-link analysis. In P. Ingwersen, & B. Larsen (Eds.), *Proceedings of ISSI 2005* (pp. 534-543). Stockholm: Karolinska University Press.