Comparative Analysis of Webometric Measurements in Thematic Environments

Cristina Faba-Pérez, Felipe Zapico-Alonso, and Vicente P. Guerrero-Bote
Library and Information Science Faculty, University of Extremadura, Alcazaba de Badajoz (Antiguo Hospital Militar), 06071 Badajoz, Spain. E-mail: {cfabper, fzapalo, Vicente}@alcazaba.unex.es

Félix de Moya-Anegón
Library and Information Science Faculty, University of Granada, Campus Cartuja, Colegio Máximo, 18071 Granada, Spain. E-mail: felix@goliat.ugr.es

There have been many attempts to evaluate Web spaces on the basis of the information that they provide, their form or functionality, or even the importance given to each of them by the Web itself. The indicators that have been developed for this purpose fall into two groups: those based on the study of a Web space's formal characteristics, and those related to its link structure. In this study we examine most of the webometric indicators that have been proposed in the literature together with others of our own design by applying them to a set of thematically related Web spaces and analyzing the relationships between the different indicators.

Introduction

The current rate of growth of electronic resources available on the Internet—Google (2004) claims to have 4.28 U.S. billion Web pages indexed—makes it important to set up mechanisms to examine this information and to analyze the formal characteristics of "Web spaces"—an expression used by Smith (1999b)—that contain this information (Kelly, 1999; Shneiderman, 1997; Strain & Berry, 1996; Trochim, 1996). This was the context in the mid-1990s in which informetric indicators and methods began to be applied to the Internet. It led to the rise of two new disciplines: the first variously denominated Cybermetrics (1997), Netometrics (Bossy, 1995), Cyberinformetrics (Aguillo, 2000b) or Internetmetrics (Shiri, 1998), and the second Webometrics (Almind & Ingwersen, 1997).

Research on the development and application of informetric indicators to the world of the Internet has basically opted for the study of two quite different groups of indicators: those based on the analysis of the characteristics of the electronic information available on the Internet, and those that use the Web's network of hypertext links to investigate the process of search and retrieval of information or to understand the behavior and social dynamics of Web builders and users.

In the former case, some workers have proposed theoretical models for the analysis of the characteristics and Web resources that would be applicable to any field on the Internet, i.e., these are theoretical studies of general criteria for the analysis of the information contained in Web spaces (Codina, 2000; Correa-Uribe, 1999; Jiménez-Piano, 2001; Smith, 1996, 1997; Tillman, 2000; Zhang & Dran, 2000; Zimmerman et al., 1997). Others, however, have proposed applications of general evaluation for such diverse specific thematic areas as medicine (Aguillo, 2000a; Chu & Chan, 1998; Garrison, 1998; Pendergrass, Nosek, & Holcomb, 2001), chemistry (Yates, 2000), government Web spaces (Smith, 2001), and specialized libraries (Osorio, 2001), inter alia. Many characteristics have been put forward for analysis, both relating to content such as "depth or level of thematic detail" (Smith, 1996), and relating to formal aspects such as "appropriate contrast between text/background, images/text, and images/background on the Web" (Codina, 2000, p. 141). From our perspective, most of the characteristics analyzed in these theoretical models are lacking in precision and are difficult to quantify. To overcome this problem, certain Web space evaluation models have been developed based exclusively on the analysis of formal characteristics that can be extracted automatically, thereby ensuring greater objectivity in the results, as in the cases of Woodruff et al. (1996), Nicholas, Huntington, and Williams (2002), and Faba-Pérez (2003).

In the latter case, i.e., the group of link-based indicators, other workers have proposed the use of citation analysis techniques similar to those applied to printed scientific publications to reveal the structure of relationships on the Web and to evaluate the quality of the content. Assuming that the outlinks and inlinks of Web spaces are analogous to references and

Received October 16, 2003; revised April 12, 2004; accepted May 17, 2004

 • Published online 24 March 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20161

citations (Price, 1970) of traditional scientific publications, one can perform the process of "websiting" (Rousseau, 1997), following the links received by a given Web space and analyzing the "sitations." This analogy has found opposing opinions in the literature, with examples in favor (Almind & Ingwersen, 1997; Björneborn & Ingwersen, 2001; Cronin, 2001; Faba-Pérez, Guerrero-Bote, & Moya-Anegón, 2003a; Larson, 1996; McKiernan, 1996; Vreeland, 2000), and against (Egghe, 2000; Harter & Ford, 2000; Kim, 2000; Van Raan, 2001).

In principle, all the links connecting Web documents are equivalent to each other, i.e., the Web in itself expresses no preference for one or another link. However, the connectivity of the network of links that is formed between Web spaces contains implicit information about the comparative importance of certain links with respect to others. Hence, documents of greater quality—which contain clear, precise, and useful information—are linked to many other documents, so that a document's total number of inlinks may be considered as a measure of its importance (Heylighen, 2000). Unlike such "traditional" search engines as AltaVista, Infoseek, HotBot, Lycos, and Excite that use statistical IR techniques with some added heuristics for weighting to determine the ranking of the retrieved pages, other link-based search engines and prototypes have been developed that have obtained greater relevance in their responses to queries. Examples are the Automatic Resource Compiler (ARC) prototype developed in the framework of the Clever Project and which implements the HITS algorithm (Kleinberg, 1998), the Google search engine that implements the PageRank algorithm (Brin & Page, 1998; Page et al., 1998), and the WebQuery system that displays the end results using a prototype of a visualization tool named VANISH (Carrière-Nortel & Kazman, 1999).

Besides its contribution to the development of search services, analysis of the Web's link structure (Faba-Pérez, Guerrero-Bote, & Moya-Anegón 2003a, 2003b, 2004; Thelwall, 2001b; Thelwall & Smith, 2002) also forms part of the field of Social Networks that is centered on the identification and analysis of structures starting from the analysis of the relationships that exist between determined elements, independently of the attributes or characteristics of those elements (Wassermann & Faust, 1994). This field of study assumes that those structures exert some type of influence on the behavior of the elements that make up the system (Molina, Muñoz, & Losego, 2000). One of the ideas that guided the first studies in network analysis was that of the "centrality" of the different social actors forming part of the networks, giving rise to various "centrality measures." In this sense, although the concept of "centrality" is mostly associated with the comparative positions of the nodes of a graph, it occasionally also alludes to another completely different problem—that of the degree of centralization of the graph as a whole. Thus, Freeman (1979) distinguishes between "point centrality" and "graph centrality," and Scott (1991) identifies these two concepts as "centrality" and "centralization," respectively.

In general terms, two concepts in Social Network Analysis—*centrality* and *prestige*—refer to the weight that a given node has in a network. Centrality and centralization are appropriate indicators of both directed and undirected networks. Prestige indicators only have meaning in directed networks. The latter lie outside the scope of the present work, although they will be dealt with in subsequent studies.

In the present work we compare the behavior of different webometric indicators in a generic thematic environment—the case of Extremadura (a region in Spain). The indicators proposed and calculated for the Web spaces of this environment belong to the aforementioned two principal categories. On the one hand, we have developed and applied indicators based on formal characteristics—number of pages, and logarithm of the number of pages; weight of Web spaces according to the weighted sum of their formal characteristics; weight plus logarithm of the number of pages. On the other hand, we have developed and applied webometric measures based on a Web space's links—web impact factor; inlinks received by the Web spaces; betweenness; PageRank provided by Google and calculated mathematically. Whenever possible in this latter group of measurements, we introduced another variable: the source of the inlinks—internal when they came from the Web site's own environment (even though they might be external to the site itself), and external when they came from the rest of the Internet.

Our starting hypothesis is thus that all these indicators of the formal quality and importance or centrality of Web spaces should be correlated (since, though in principle the two concepts—quality and centrality—are not equivalent, one would assume that if the formal quality of Web spaces is high then other spaces will recognize their value by means of links), and that, although they all measure certain aspects, those that are the most correlated with the rest will be the most representative of the set of indicators studied.

The novelty of the present work is hence not in defining new indicators, but in the study of their application a set of thematically related Web spaces and analyzing the relationships between the different indicators on the basis of how they each evolve. We thereby determine: (a) which are the specialized indicators that measure particular aspects and whose values are unrelated to the rest, and (b) which are the indicators that are related to most of the others and therefore take into account in an overall sense all of the aspects measured by the other indicators.

Material and Method (Data and Measurements)

The data used in the study corresponded to 1,180 "Web spaces" with their relationship to the Extremadura region in common. We located this population by looking for a source that gave us a set of Web spaces specializing in Extremadura from which we could retrieve new Web spaces on Extremadura by following the outlinks. Taking as the principal evaluation criterion the "authority" of the source, we selected "Extremadura on Internet," the Web server of the Junta of Extremadura (the local government of the Autonomous Community), which compiles URLs of and about the region of Extremadura (http://www.juntaex.es/todoweb). On the basis of this compilation and of the links obtained from the Web spaces themselves, we generated a database of 1,180 elements related to Extremadura. The database consists of three categories of URL: Web sites, Web pages including personal pages, and sets of Web pages hosted on other servers. We considered the three categories conjointly under the generic term of "Web spaces" from/about Extremadura, at least for their retrieval.

The indicators studied were the following:

Web Impact Factor (WIF). This is a relative indicator and has been one of the most studied in cybermetric analyses (Ciolek, 1997; Ingwersen, 1998; Rodríguez i Gairín, 1997; Smith, 1999a; Thelwall, 2001a; Thomas & Willet, 2000). It can be expressed by the general formula:

$$WIF = \frac{\text{Number of the web space's inlinks}}{\text{Number of the web space's pages}}$$

where the numerator represents the number of links entering a given Web space (i.e., its number of sitations). Ingwersen defines it as: "The logical sum of the number of external and self-link web pages pointing to a given country or web site divided by the number of pages found in that country or web site, at a given point in time" (1998, p. 237), where the external Web pages are those that do not belong to that given country or Web site and that link, at least once, to it, and the internal Web pages are those that belong to that country or Web site and that link, at least once, to it.

In our case, using the retrieval possibilities of AltaVista and our own database with the pages and outlinks of the set of 1,180 pages, we calculated two types of *WIF* that we denominate *Internal- and External-WIF*:

- *Internal-WIF*: relates the number of inlinks of each of the 1,180 URLs of our population that come from the population itself to the number of internal pages of that population.
- *External-WIF*: relates the number of inlinks of each of the 1,180 URLs of our population that come from the rest of the Internet to the number of internal pages of the population.

Internal Inlinks and External Inlinks. These are linear indicators that coincide, respectively, with the numerators of the *Internal-WIF* (number of inlinks of each of the 1,180 URLs of our population that come from the population itself) and of the *External-WIF* (number of inlinks of each of the 1,180 URLs of our population that come from the rest of the Internet).

Betweenness. This is a measure of "centrality" used in Social Networks Analysis that determines to what degree a node acts as intermediary between other nodes by being located on the path between them. In a Social Network, the nodes represent different social actors and an arc represents a union that has been established between two of them. In the present work, the nodes are the 1,180 Extremadura Web spaces of the study population and the arcs are the links that exist between them. The betweenness is a count of how often a node appears on the geodesic paths (the shortest routes between each pair of nodes of the network).

Google PageRank (G.PR.). Google is a search engine that uses the link structure of the Web itself to retrieve relevant results for a query. Google's results are ranked according to their popularity index in terms of links (Cornella, 2000). The model implements the PageRank algorithm which determines each Web page's individual rank in the retrieval according not only to the number of other Web pages that link to it (sitors), but also to the PageRank and the total number of outlinks of the sitor pages (Brin & Page, 1998; Page et al., 1998). In the present study, the value of this indicator is the PageRank displayed in Google's toolbar for each of the 1180 Web spaces of our study population.

PR1. Applying the PageRank formula (Brin & Page, 1998), we calculated *PR*1:

$$PR1(A) = (1 - d) + d \cdot (PR1(T1)/C1(T1) + \cdots + PR1(Tn)/C1(Tn))$$

where

- $PR1(A)$ it is the PageRank of A (which will correspond to each of the 1,180 Extremadura Web spaces).
- $T1, \ldots, Tn$ are the Web spaces of the study population that are sitors of A (i.e., that have links to A).
- $C1(Tn)$ is the number of outlinks of Tn whose target is one of the 1,180 Web spaces of the study population.
- d is a damping factor which can be set between 0 and 1. According to Brin & Page (1998), it is usually set to 0.85, which is therefore the value we shall use.

PR2. Applying the PageRank formula (Brin & Page, 1998), we also calculated *PR*2:

$$PR2(A) = (1 - d) + d \cdot (PR2(T1)/C2(T1) + \cdots + PR2(Tn)/C2(Tn))$$

where

- $PR2(A)$ is the PageRank of A (corresponding to each of the 1,180 Extremadura Web spaces).
- $T1, \ldots, Tn$ are the Web spaces of the study population that are sitors of A (i.e., that have links to A).
- $C2(Tn)$ is the total number of outlinks of Tn.
- d is a damping factor which can be set between 0 and 1. Following Brin & Page (1998), we set d to 0.85.

Pages and Log(Pages). As indicated by their names, these indicators measure, respectively, the number of internal pages associated with each of the 1,180 Extremadura Web spaces, and the logarithm of that number.

Weight. There have been many studies aimed at evaluating Web spaces on the basis both of the information that they provide and of their form. The ultimate goal of such studies is to identify the characteristics of Web spaces that best satisfy the requirements of users who browse through them (Ma, 2002). Following the idea mentioned above, be more objective, this indicator is based exclusively on the formal characteristics of Web spaces, in particular, all those that we have been able to determine automatically for each page. For these characteristics to be equivalent for purposes of comparison (i.e., equiparable), it was first necessary to assign them a weight according to their relevance for retrieval of any requested information. To this end, we chose a range of weights from −0.5 to 0.3, with, positive weights corresponding to those that were considered desirable characteristics and negative weights to those considered undesirable characteristics, and varied the values to make the results equiparable (a complete justification of the values is given in Faba-Pérez, 2003 and Faba-Pérez, Guerrero-Bote, & Moya-Anegón, 2004, 2005). The characteristics considered are (Faba-Pérez et al., 2005):

New Pages. Pages that have changed in the last month.

External Links. Links from a given Web space to another Web space different from that of the origin (external outlinks). The greater or lesser number of external outlinks will determine the "brightness" of the originating Web space (Bray, 1996; Codina, 2000).

Metadata. Indicates the presence of different types of META tags.

"Dublin Core" Metadata. The DCMES (Dublin Core Metadata Element Set) subset of META tags, consisting of 15 elements relating to content (Coverage, Description, Type, Relation, Source, Subject, Title), to intellectual property (Contributor, Creator, Publisher, Rights), and to specific aspects of the document's configuration (Date, Format, Identifier, Language).

Main Metadata. This characteristic includes a subset of Metadata considered of special relevance for access to the information (Author, Creator, Keywords, Subject, Description, Title, Date).

Hypertextual Density. Number of HTML links per page of the Web space.

Multimedia Density. Number of multimedia objects per page of the Web space.

Small Pages. Pages smaller than 3 KB.

Broken Links (External and Internal). Broken outlinks in a Web space. The possible causes include: File Not Found, Cannot Connect, Host Not Found, and Time Out.

Pages With No Title. Pages with a blank (null or white space) in the Title HTML tag or with no such tag.

Slow Pages. Pages (generally larger than 50 KB) that take more than 20 or 30 seconds to load.

Broken Anchors. Broken anchors in Web spaces, i.e., problems of hypertext linking to places within a document.

Links With Non-Critical Problems. Outlinks in Web spaces that do not function correctly because of some type of problem considered of minor importance (e.g., Temporary or Permanent Redirect).

Pages With No Image Attributes. Pages with images that have no Height, Width, or Alt attribute assignments.

Old Pages. Pages that have been unchanged for the last 6 months.

Deep Pages. Pages that require more than four clicks to get to/from the front page.

Table 1 shows their corresponding weights.

The weight (*Ws*) obtained by each of the 1,180 Web spaces of our study is calculated as the weighted sum of its formal characteristics (Faba-Pérez, 2003):

$$W_s = \sum_{i=1}^{n} w_i x_i$$

where

TABLE 1. Formal characteristics and associated weights.

Positive characteristics	Weight	Negative characteristics	Weight
New pages	0.3	Broken external links	−0.5
External links	0.2	Broken internal links	−0.3
Metadata	0.2	Pages with no title	−0.3
"Dublin Core" metadata	0.2	Slow pages	−0.2
Main metadata[a]	0.2	Broken anchors	−0.1
Hypertextual density	0.1	Links with non-critical problems	−0.1
Multimedia density	0.1	Pages with no image attributes	−0.1
Small pages	0.1	Old pages	−0.03
		Deep pages	−0.02

[a] Main metadata includes a subset of metadata or meta tags considered of special relevance for access to the information (Author, Creator, Keywords, Subject, Description, Title, and Date).

- w_i is the weight assigned to the *i-th* characteristic (with *n* being the number of characteristics of each Web space).
- x_i is the number of occurrences of the *i-th* characteristic.

Weight + Logarithmic Factor (LF). This indicator is the above weight with the addition of a size factor, implemented as a logarithmic factor of the number of pages in the Web site (Faba-Pérez, 2003). Specifically, the additional factor is the logarithm of the number of pages multiplied by 10 to match the scale of the two terms.

Results and Discussion

After calculating the webometric indicators described above for each of the 1,180 Web spaces of our data set, we then generated a 12 × 12 Pearson correlation matrix in which the rows and columns represent the 12 indicators and each element is the Pearson correlation coefficient between the corresponding variables (indicators). Because of the great number of cases used, all these correlations are significant at the two-tailed 0.01 level, even correlations of 0.09 in value. Table 2 presents the essential part of the resulting matrix, for the sake of facilitating the interpretation including only the cases where the correlation is very strong (greater than 0.700), substantial (between 0.500 and 0.699), or moderate (between 0.300 and 0.499) (Ferreiro-Aléez, 1993). The related pairs of indices corresponding to these three classes of correlation are as follows:

1. Pairs of indicators with a very strong correlation (>0.700): *PR*1 − *PR*2; *PR*1 − *Int. Inlinks*; *PR*2 − *Int. Inlinks*; *Log(Pages)* − *Weight* + *LF*; *Log(Pages)* − *Weight*; and *Weight* + *LF* − *Weight*.
2. Pairs of indicators with a substantial correlation (between 0.500 and 0.699): *Betweenness* − *PR*1; *Betweenness* − *PR*2; *Betweenness* − *Int. Inlinks*; *PR*1 − *Ext. Inlinks*; *PR*2 − *Ext. Inlinks*; and *Int. Inlink* − *Ext. Inlinks*.
3. Pairs of indicators with a moderate correlation (between 0.300 and 0.499): *Ext. Inlinks* − *Pages*; *Ext. Inlinks* − *Log(Pages)*; *Ext. Inlinks* − *Weight* + *LF*; *Pages* − *Log(Pages)*; and *Pages* − *Weight* + *LF*.

TABLE 2. Correlation matrix between the webometric indicators.

	Int. *WIF*	Ext. *WIF*	Between	*PR*1	*PR*2	Int. Inlinks	Ext. Inlinks	*G.PR*	Pages	Log(pages)	Weight + *LF*	Weight
Int. *WIF*	1.00											
Ext. *WIF*		1.00										
Between			1.00	0.59	0.57	0.55	0.36					
*PR*1			0.59	1.00	0.87	0.90	0.60					
*PR*2			0.57	0.87	1.00	0.76	0.55					
Int. Inlinks			0.55	0.90	0.76	1.00	0.52					
Ext. Inlinks			0.36	0.60	0.55	0.52	1.00		0.46	0.31	0.30	
G.PR								1.00				
Pages							0.46		1.00	0.39	0.37	
Log(pages)							0.31		0.39	1.00	0.98	0.80
Weight + *LF*							0.30		0.37	0.98	1.00	0.91
Weight										0.80	0.91	1.00

One sees that the strongest correlations are found between measurements that belong to the same group of indicators, so that one can classify the indicators into three groups:

1. Indicators that measure formal characteristics.
2. Indicators based on the links established between Web spaces. Within this group there was a greater correlation between measurements based on the inlinks from the same set of the 1,180 Web spaces.
3. Within the group of link-based indicators, there was a set of webometric indicators that remained independent, and consequently had no linear correlation with any other indicator. These were *Int. WIF*, *Ext. WIF*, and *G.PR*.

The (substantial) correlation between the centrality measure (*Betweenness*) on the one hand and the *Inlinks* (*External* and *Internal*) and the two *PR* indicators based on the internal link structure on the other was not necessarily to be expected because, although both measures take all the links into account, the former also includes such other factors as the relationships between originating and destination nodes.

The case of the *Ext. Inlinks* also stands out as being the indicator that is related to other indicators belonging to both the groups that were studied (except for the *WIFs*, *G.PR.*, and *Weight*), thereby lending cohesion to the set of indicators.

The results of applying the Multidimensional Scaling (MDS) technique to the entire correlation matrix of the 12 indicators are shown in Figure 1. They correspond to a stress of 0.15648. In the interpretation, we followed the four basic elements proposed by White and Griffith (1981):

1. *Center/Periphery*. For White and Griffith, in the MDS of an Author Cocitation Analysis (ACA), the most important authors of a discipline are found in the center of the diagram. In our case, we found that the *Betweenness* and the *Ext. Inlinks* are quasi-centered, which indicates that they occupy a privileged position on the map. In a certain sense, the measures located in the center are those that are most related to the group as a whole, and that could therefore be regarded as a compendium of the rest. The greatest centrality is that of the indicator *Pages*, reflecting the influence of size (number of pages) on the other indicators. This result can only be attributed to the rest of the correlations whose value is less than 0.3 and that are not given in Table 2.
2. *Clustering*. Measurements that are more related to each other should appear together in the space of Figure 1. Indeed, one sees in the overview provided by the figure that there seem to be four clusters of indicators: the central (as mentioned above), those related to formal characteristics (*Weight*, *Weight + LF*, and *Log(Pages)*), the link-related measures (*PR*1, *PR*2, and *Int. Inlinks*), and the independent indicators (*WIFs* and *G.PR.*).
3. *Identification of the axes*. The horizontal axis starts from the indicators that have very strong relationships and form part of the most clearly defined clusters (formal and link-based indicators), then passes through those that have more general relationships without these being very strong (*Ext. Inlinks*, *Betweenness*, and *Pages*), and ends in the independent measurements (*G.PR.*, *WIFs*). The

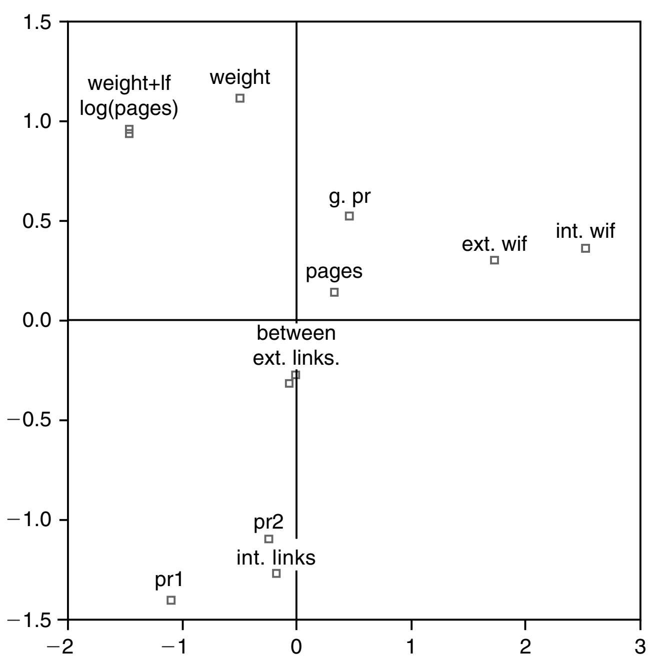

FIG. 1. Multidimensional scaling of the indicators on the basis of the correlation matrix (two-dimensional space).

vertical axis goes from the measures related to the link structure (*PR*1, *Int. Inlinks*, *PR*2, *Ext. Inlinks*, and *Betweenness*) to the independent indicators, and finally to measures based on the formal characteristics.

4. *Relationships between nearby elements*. The closer two indicators are in the MDS, the greater is the relationship between them. In this sense, the relationships between the indicators *Betweenness* and *Ext. Inlinks* on the one hand, and *Weight* + *LF* and *Log(Pages)* on the other merit particular attention (the latter relationship is quite logical considering the logarithmic factor contained in the first of the two indicators). One observes that in both cases these are pairs belonging to the same groups of indicators.

Conclusions

We have used a thematic Web environment to analyze the relationships between several webometric measures belonging to two groups of indicators (one based on the formal characteristics of Web spaces and the other on their link structure). The results lead to the following conclusions:

1. Each indicator is more closely related to those which are clustered within the same group, as was to be expected (since they were calculated in terms of the same or similar characteristics).
2. The *G.PR.* indicator remained independent of the remainder of the measures, including *PR*1 and *PR*2 which are PageRanks calculated by means of the original formula.
3. Neither were the *Int. WIF* or *Ext. WIF* related to the rest of the measures, although we found that the linear indicators forming their respective numerators and denominator (*Int.* and *Ext. Inlinks*–numerator—and *Pages*–denominator) were correlated with other measures. We therefore consider that it is the relative character of the *WIFs* that triggers their independence.
4. The *Ext. Inlinks*, a measure of the visibility of Web spaces (Bray, 1996; Codina, 2000), was the indicator that correlated with most other indicators.
5. We have demonstrated the linear dependence of the *Betweenness* of Web spaces in this closed environment and their importance with respect to the sitations that they receive, since this indicator presents a substantial correlation with *PR*1, *PR*2, and *Int. Inlinks*, and a moderate correlation with *Ext. Inlinks*.
6. The *Ext. Inlinks* and the *Betweenness*, as well as being very intensely correlated with each other, are two of the most centrally positioned indicators, and therefore most related to the rest.
7. The size of the Web spaces in number of pages was found to be more representative than was expected beforehand.

In sum, the indicators that were found to be the most representative were those occupying a central position in the map of Figure 1, because they have a greater correlation in the set of all the indicators considered than the indicators on the periphery. The peripheral indicators could be said to be more specialized, while the central indicators provide a more balanced representation of the set of characteristics.

Acknowledgments

This work was financed by the Junta de Extremadura—Consejería de Educación Ciencia & Tecnología—and the European Social Fund, as part of the “Programa de Ayudas para la Realización de Proyectos de Aplicación de las Tecnologías de la Información y la Comunicación” (published in D.O.E. no. 120, 17 October 2000).

References

Aguillo, I.F. (2000a). A new generation of tools for search, recovery and quality evaluation of World Wide Web medical resources. Journal of Management in Medicine, 14(3–4), 240–248.

Aguillo, I.F. (2000b). Indicadores hacia una evaluación no objetiva (cuantitativa) de sedes web. In VII Jornadas Españolas de Documentación (Bilbao, October 19–21, 2000) (pp. 233–248). Bilbao: Servicio Editorial de la Universidad del País Vasco.

Almind, T.C., & Ingwersen, P. (1997). Informetric analyses on the World Wide Web: Methodological approaches to ‘Webometrics.’ Journal of Documentation, 53(4), 404–426.

Björneborn, L., & Ingwersen, P. (2001). Perspectives of webometrics. Scientometrics, 50(1), 65–82.

Bossy, M.J. (1995). The last of the litter: Netometrics. Solaris, 2. Retrieved December 3, 1999, from http://www.info.unicaen.fr/bnum/jelec/Solaris/d02/2bossy.htm

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems, 30, 107–117.

Bray, T. (1996). Measuring the Web. The World Wide Web Journal, 1(3). Retrieved July 30, 2000, from http://www5conf.inria.fr/fich_html/papers/P9/Overview.html

Carrière-Nortel, J., & Kazman, R. (1999). WebQuery: Searching and visualizing the web through connectivity. Computer Networks and ISDN Systems, 29(8), 1257.

Chu, L.F., & Chan, B.K. (1998). Evolution of web site design: Implications for medical education on the Internet. Computers in Biology and Medicine, 28(5), 459–472.

Ciolek, T.M. (1997). The size, content and geography of Asian Cyberspace: An initial measurement. The Journal of East Asian Libraries. Retrieved July 19, 2001, from http://www.ciolek.com/PAPERS/AsianCyberspace-97.html

Codina, L. (2000). Parámetros e indicadores de calidad para la evaluación de recursos digitales. In VII Jornadas Españolas de Documentación (Bilbao, October 19–21, 2000) (pp. 135–144). Bilbao: Servicio Editorial de la Universidad del País Vasco.

Cornella, A. (2000). Desarrollos en buscadores. El Profesional de la Información, 9(1–2), 32–34.

Correa-Uribe, G. (1999). Colombia conectada al mundo: Sitios web colombianos. Antioquía (Colombia): Universidad.

Cronin, B. (2001). Bibliometrics and beyond: Some thoughts on web-based citation analysis. Journal of Information Science, 27(1), 1–7.

Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics (1997). Retrieved August 11, 2002, from http://www.cindoc.csic.es/cybermetrics/

Egghe, L. (2000). New informetric aspects of the Internet: Some reflections—Many problems. Journal of Information Science, 26(5), 329–335.

Faba-Pérez, C. (2003). Análisis cibermétrico de la información WEB: El caso de Extremadura en Internet. Unpublished doctoral dissertation, University of Granada, Spain.

Faba-Pérez, C., Guerrero-Bote, V.P., & Moya-Anegón, F. (2003a). “Sitation” distributions and Bradford’s Law in a closed web space. Journal of Documentation, 59(5), 558–580.

Faba-Pérez, C., Guerrero-Bote, V.P., & Moya-Anegón, F. (2003b). A study of data-mining in a closed web environment. Scientometrics, 58(3), 623–640.

Faba-Pérez, C., Guerrero-Bote, V.P., & Moya-Anegón, F. (2004). Fundamentos Y Técnicas Cibermétricas. Mérida: Junta de Extremadura.

Faba-Pérez, C., Guerrero-Bote, V.P., & Moya-Anegón, F. (2005). Self-organizing maps of web spaces based on formal characteristics. Information Processing and Management, 41(2), 331–346.

Ferreiro-Aláez, L. (1993). Bibliometría: análisis bivariante. Madrid: Espasa.

Freeman, L.C. (1979). Centrality in social networks: Conceptual clarification. Social Networks, 1, 215–239.

Garrison, S. (1998). Evaluating health Internet sites: A white paper's criteria. Medical Reference Services Quarterly, 17(3), 41–47.

Google. (2004). Google achieves search milestone with immediate access to more than 6 billion items. Retrieved March 26, 2004, from http://www.google.com/press/pressrel/6billion.html

Harter, S.P., & Ford, C.E. (2000). Web-based analyses of e-journal impact: Approaches, problems, and issues. Library Science With a Slant to Documentation and Information Studies, 51(13), 1159–1176.

Heylighen, F. (2000). Web connectivity analysis. Retrieved June 2, 2001, from http://pespmc1.vub.ac.be/WEBCONAN.html

Ingwersen, P. (1998). The calculation of the Web impact factors. Journal of Documentation, 54(2), 236–243.

Jiménez-Piano, M. (2001). Evaluación de sedes web. Revista Española de Documentación Científica, 24(4), 405–432.

Kelly, B. (1999). Technologies and standards for building Web sites. Online and CD-ROM Review, 23(3), 183–185.

Kim, H.J. (2000). Motivations for hyperlinking in scholarly electronic articles: A qualitative study. Journal of the American Society for Information Science, 51(10), 887–899.

Kleinberg, J. (1998). Authoritative sources in a hyperlinked environment. Retrieved July 20, 2001, from http://www.cs.cornell.edu/home/kleinber/auth.pdf

Larson, R.R. (1996). Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of cyberspace. In S. Hardin (Ed.), Proceedings of the 59th Annual Meeting of the American Society for Information Science (pp. 71–78). Medford, NJ: Information Today.

Ma, Y. (2002). A design analysis model for developing World Wide Web sites. Journal of the American Society for Information Science & Technology, 53(7), 531–535.

McKiernan, G. (1996). CitedSites(sm): Citation indexing of web resources. Retrieved February 6, 2000, from http://www.public.iastate.edu/~CYBERSTACKS/Cited.htm

Molina, J.L., Muñoz, J., & Losego, P. (2000, September). Red y realidad: Aproximación al análisis de las redes científicas. Paper presented at VII Congreso Nacional de Psicología Social, Oviedo, Spain.

Nicholas, D., Huntington, P., & Williams, P. (2002). Evaluating metrics for comparing the use of web sites: A case study of two consumer health web sites. Journal of Information Science, 28(1), 63–75.

Osorio, N.L. (2001, Winter). Web sites of science-engineering libraries: An analysis of content and design. Issues in Science and Technology Librarianship. Retrieved July 27, 2002, from http://www.library.ucsb.edu/istl/01-winter/refereed.html

Page, L., Brin, S., Motwani, R. & Winograd, T. (1998). The PageRank citation ranking: Bringing order to the Web. Retrieved January 27, 2002, from http://www.lans.ece.utexas.edu/ ~skumar/DataMining/repository/PageRank.pdf

Pendergrass, S., Nosek, M.A., & Holcomb, J.B. (2001). Design and evaluation of an Internet site to educate women with disabilities on reproductive health care. Sexuality and Disability, 19(1), 71–83.

Price, D.J. de Solla. (1970). Citation measures of hard science, soft science, technology and non-science. In C.C. Nelson & D.E.Pollock (Eds.), Communication among scientists and engineers (pp. 3–22). Lexington, MA: D.C. Heath and Co.

Rodríguez i Gairin, J.M. (1997). Valoración del impacto de la información en Internet: Altavista, el 'Citation Index' de la red. Revista Española de Documentación Científica, 20(2), 175–181.

Rousseau, R. (1997). Sitations: An exploratory study. Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics, 1(1). Retrieved September 5, 2000, from http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html

Scott, J. (1991). Social network analysis. A handbook. Beverly Hills, CA: Sage.

Shiri, A.A. (1998). Cybermetrics: A new horizon in information research. Retrieved November 13, 2000, from http://www.deck.com/people/pages/shiri/cybermet1.html

Shneiderman, B. (1997). Designing information-abundant web sites: Issues and recommendations. International Journal of Human–Computer Studies, 47(1), 5–29.

Smith, A.G. (1996). Criteria for evaluation of Internet information resources. Retrieved March 3, 2000, from http://www.vuw.ac.nz/~agsmith/evaln/index.htm

Smith, A.G. (1997). Testing the surf: Criteria for evaluating Internet information resources. The Public-Access Computer Systems, 8(3). Retrieved February 28, 2002, from http://info.lib.uh.edu/pr/v8/n3/smit8n3.html

Smith, A.G. (1999a). A tale of two Web spaces: Comparing sites using Web impact factors. Journal of Documentation, 55(5), 577–592.

Smith, A.G. (1999b). The impact of web sites: A comparison between Australasia and Latin America. Retrieved May 14, 2001, from http://www.vuw.ac.nz/~agsmith/publns/austlat/

Smith, A.G. (2001). Applying evaluation criteria to New Zealand government websites. International Journal of Information Management, 21(2), 137–149.

Strain, H.C., & Berry, P.M. (1996). Better page design for the World Wide Web. Online and CD-ROM Review, 20(5), 227–238.

Thelwall, M. (2001a). Results from a web impact factor crawler. Journal of Documentation, 57(2), 177–191.

Thelwall, M. (2001b). Exploring the link structure of the Web with network diagrams. Journal of Information Science, 27(6), 393–401.

Thelwall, M., & Smith, A.G. (2002). Interlinking between Asia-Pacific University Web sites. Scientometrics, 55(3), 363–376.

Thomas, O., & Willett, P. (2000). Webometric analysis of departments of librarianship and information science. Journal of Information Science, 26(6), 421–428.

Tillman, H.N. (2000). Evaluating quality on the Net. Retrieved June 3, 2000, from http://www.tiac.net/users/hope/findqual.html

Trochim, W.M.K. (1996). Evaluating websites. Retrieved January 6, 2001, from http://trochim.human.cornell.edu/webeval/webintro/webintro.htm

Van Raan, A.F.J. (2001). Bibliometrics and Internet: Some observations and expectations. Scientometrics, 50(1), 59–63.

Vreeland, R.C. (2000). Law libraries in hyperspace: A citation analysis of World Wide Web sites. Law Library Journal, 92(1), 9–25.

Wassermann, S., & Faust, K. (1994). Social network analysis: Methods and applications. Cambridge, UK/New York: Cambridge University Press.

White, H.D., & Griffith, B.C. (1981). Author cocitation: A literature measure of intellectual structure. Journal of the American Society for Information Science, 32(3), 163–171.

Woodruff, A., Aoki, P.M., Brewer, E., Gauthier, P., & Rowe, L.A. (1996). An investigation of documents from the World Wide Web. Computer Networks and ISDN Systems, 28(7–11), 963-980.

Yates, P.C. (2000). Use of a World Wide Web site evaluation tool in chemistry. Journal of Science Education and Technology, 9(4), 357–365.

Zhang, P., & Dran, G.M. von. (2000). Satisfiers and dissatisfiers: A two-factor model for website design and evaluation. Journal of the American Society for Information Science, 51(14), 1253–1268.

Zimmerman, D.E., Murasaki, M., Estes, E., & Hallmark, B. (1997). A formative evaluation method for designing WWW sites. In IPCC '97 Proceedings (pp. 223–230). Piscataway, NJ: IEEE.