Toward a Basic Framework for Webometrics

Lennart Björneborn and Peter Ingwersen
Department of Information Studies, Royal School of Library and Information Science, DK 2300 Copenhagen S–Denmark. E-mail: {lb, pi}@db.dk

In this article, we define webometrics within the framework of informetric studies and bibliometrics, as belonging to library and information science, and as associated with cybermetrics as a generic subfield. We develop a consistent and detailed link typology and terminology and make explicit the distinction among different Web node levels when using the proposed conceptual framework. As a consequence, we propose a novel diagram notation to fully appreciate and investigate link structures between Web nodes in webometric analyses. We warn against taking the analogy between citation analyses and link analyses too far.

Introduction

Library and information science (LIS) and related fields in the sociology of science and science and technology studies have developed a range of theories and methodologies—now including webometrics—concerning quantitative aspects of how different types of information are generated, organized, disseminated and used by different users in different contexts. Historically, this development arose during the first half of the twentieth century from statistical studies of bibliographies and scientific journals (Hertzel, 1987). These early studies revealed bibliometric power laws like *Lotka's law* on productivity distribution among scientists (Lotka, 1926); *Bradford's law* on the scattering of literature on a particular topic over different journals (Bradford, 1934); and *Zipf's law* of word frequencies in texts (Zipf, 1949). Similar power-law distributions have been identified on the Web, for example, the distribution of TLDs (top level domains) on a given topic (Rousseau, 1997) or inlinks per Web site (Adamic & Huberman, 2000, 2001; Albert, Jeong, & Barabási, 1999).

Decisive for the development of bibliometrics and scientometrics was the arrival of citation indexes of scientific literature introduced by Garfield (1955) that enabled analyses of citation networks in science (e.g., Price, 1965). Access to online citation databases catalyzed a wide range of citation studies, especially mapping scientific domains, including growth, diffusion, specialization, collaboration, impact, and obsolescence of literature and concepts. For extensive coverage, see the ARIST chapters by White and McCain (1989) and Borgman and Furner (2002).

The breakthrough of online citation analysis parallels the later avalanche of webometric studies enabled by access to large-scale Web data. In particular, the apparent yet ambiguous resemblance between citation networks and the hypertextual interdocument structures of the Web triggered much interest from the mid-1990s (e.g., Almind & Ingwersen, 1997; Bossy, 1995; Downie, 1996; Ingwersen, 1998; Kuster, 1996; Larson, 1996; McKiernan, 1996; Moulthrop & Kaplan, 1995; Pitkow & Pirolli, 1997; Rousseau, 1997; Spertus, 1997).

Furthermore, the central bibliometric measures of cocitation (Small, 1973) and bibliographic coupling (Kessler, 1963) have been applied to studies of Web clustering, Web growth, and Web searching (e.g., Ding, Zha, He, Husbands, & Simon, 2001; Efe et al., 2000; Larson, 1996; Menczer, 2002; Pitkow & Pirolli, 1997; Weiss et al., 1996).

Since its advent, the Web has been widely used in both formal and informal scholarly communication and collaboration (e.g., Cronin, Snyder, Rosenbaum, Martinson, & Callahan, 1998; Harter & Ford, 2000; Hurd, 2000; Thelwall & Wilkinson, 2003; Wilkinson, Harries, Thelwall, & Price, 2003; Zhang, 2001). Webometrics thus offers potentials for tracking aspects of scientific endeavor traditionally more hidden from bibliometric or scientometric studies, such as the use of research results in teaching and by the general public (Björneborn & Ingwersen, 2001; Cronin, 2001; Thelwall & Wilkinson, 2003; Thelwall, Vaughan, & Björneborn, forthcoming) or the actual use of scientific Web pages.

A range of new terms for the emerging research field were rapidly proposed from the mid-1990s, for example, *netometrics* (Bossy, 1995); *webometry* (Abraham, 1996); *internetometrics* (Almind & Ingwersen, 1996); *webometrics* (Almind & Ingwersen, 1997); *cybermetrics* (journal started 1997 by Isidro Aguillo)[1]; *Web bibliometry* (Chakrabarti, Joshi, Punera, & Pennock, 2002). This and similar more

Accepted January 23, 2004

 ● Published online 13 August 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20077

[1]http://www.cindoc.csic.es/cybermetrics/

specific conceptual diversity and development often made (and make) it difficult to understand what actually is analyzed in the contributions. The transformation over a year from internetometrics to webometrics by the same authors, Almind and Ingwersen (1996, 1997), is typical of the conceptual confusion.

Tomas C. Almind wanted, originally, to investigate both the communicative and networking aspects of the Internet *and* to analyze the typology, contents, and characteristics of the national Web pages, as in traditional bibliometric publication analyses. But it was unclear where the Internet stopped and the Web started; hence the broad notion of internetometrics in the original CIS Report (1996)[2]. However, because Almind was very careful to distinguish between communication processes and contents, he and Ingwersen decided that the publication analysis-like study published in 1997 were entirely concerned with Web page types and properties—not with communication on the Internet; hence the conception of webometrics in the title of that classic article.

As a consequence of this conceptual variety, the present paper proposes a consistent framework and terminology with which to deal with matters of webometrics. The paper is organized the following way. First, we set webometrics and associated metrics into the LIS framework of informetrics. This is followed by an introduction of basic link terminology and fundamental Web node diagram configurations. The subsequent section is devoted to advanced link terminology and Web node diagrams. The paper ends with a brief discussion section and conclusions.

Webometrics, Bibliometrics, and Informetrics

Being a global document network initially developed for scholarly use (Berners-Lee & Cailliau, 1990) and now inhabited by a diversity of users, the Web constitutes an obvious research field for bibliometrics, scientometrics and informetrics.

Webometrics and cybermetrics are currently the two most widely adopted terms in library and information science for this emerging research field. They are generically related, see Figure 1, but often used as synonyms. In continuation of the Almind case above, the present paper proposes a differentiated terminology distinguishing between studies of the Web and studies of *all* Internet applications. In this framework, webometrics is defined 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. (Björneborn, 2004)

This definition thus covers quantitative aspects of both the construction side and the usage side of the Web embracing four main areas of present webometric research: (1) Web page content analysis; (2) Web link structure analysis; (3) Web usage analysis (including log files of users' searching and browsing behavior); (4) Web technology analysis (including search engine performance). This includes hybrid forms, for example, Pirolli, Pitkow, and Rao (1996) who explored Web analysis techniques for automatic categorization utilizing link graph topology, text content and metadata similarity, as well as usage data. Further, all four main research areas include longitudinal studies of changes on the dynamic Web of, for example, page contents, link structures and usage patterns. So-called Web archaeology (Björneborn & Ingwersen, 2001) could in this webometric context be important for recovering historical Web developments, for example, by means of the Internet Archive (www.archive.org).

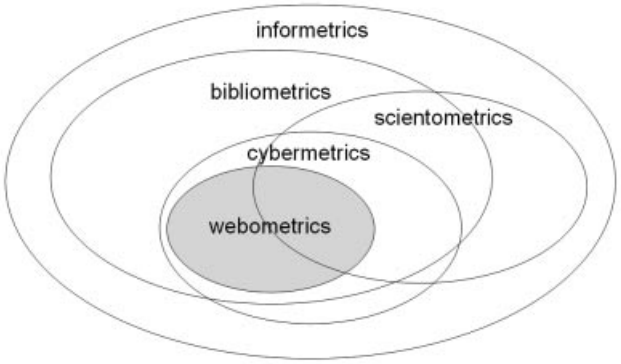

FIG. 1. Relationships between the LIS fields of infor-/biblio-/sciento-/cyber-/webo-/metrics. Sizes of the overlapping ellipses are made for sake of clarity only.

The above definition places webometrics as a LIS specific term in line with bibliometrics and informetrics (also cf., e.g., Cronin, 2001; Björneborn & Ingwersen, 2001). This domain lineage is stressed by the formulation "drawing on bibliometric and informetric approaches" because "drawing on" denotes a heritage without limiting further methodological developments of Web-specific approaches, including the incorporation of approaches of Web studies in computer science, social network analysis, hypertext research, media studies, and so forth.

In the present framework, *cybermetrics* is proposed as a generic term for:

> The study of the quantitative aspects of the construction and use of information resources, structures and technologies on the *whole* Internet drawing on bibliometric and informetric approaches. (Björneborn, 2004)

Cybermetrics thus encompasses statistical studies of discussion groups, mailing lists, and other computer-mediated communication on the Internet (e.g., Bar-Ilan, 1997; Hernández-Borges, Pareras, & Jiménez, 1997; Herring, 2002; Matzat, 1998) *including* the Web. Besides covering all computer-mediated communication using Internet applications, this definition of cybermetrics also covers quantitative measures of the Internet backbone technology, topology, and traffic (cf. Molyneux & Williams, 1999). The breadth

[2]Published by the now closed Centre for Informetric Studies (CIS) at the Royal School of Library and Information Science, Denmark.

of coverage of cybermetrics and webometrics implies large overlaps with proliferating computer-science-based approaches in analyses of Web contents, link structures, Web usage, and Web technologies. A range of such approaches has emerged since the mid-1990s with names like *cyber geography* and *cyber cartography* (e.g., Dodge, 1999; Dodge & Kitchin, 2001, 2002; Girardin, 1995, 1996)[3], *Web ecology* (e.g., Pitkow, 1997; Chi et al., 1998; Huberman, 2001), *Web mining* (e.g., Etzioni, 1996; Cooley, Mobasher, & Srivastava, 1997; Kosala & Blockeel, 2000), *Web graph analysis* (e.g., Broder et al., 2000; Clever Project, 1999; Kleinberg, Kumar, Raghavan, Rajagopalan, & Tomkins, 1999), *Web dynamics* (e.g., Levene & Poulovassilis, 2001), and *Web intelligence* (e.g., Yao, Zhong, Liu, & Ohsuga, 2001).

The raison d'être for using the term *webometrics* in this context could be to denote a close lineage to bibliometrics and informetrics and stress a LIS perspective on Web studies as noted previously. In this context, the earlier mentioned term *Web bibliometry* used by Chakrabarti et al. (2002) is especially interesting because computer scientists thus recognize the heritage in bibliometric research to be drawn on in Web studies. Other computer science approaches to link structure analysis also pay tribute to inspiration from citation studies, for example, Albert and Barabási (2002), Chakrabarti et al. (1999), Efe et al. (2000), Kleinberg (1999), Kosala and Blockeel (2000), Pitkow and Pirolli (1997), Vázquez (2001).

There are different conceptions of informetrics, bibliometrics, and scientometrics. Figure 1 shows the field of informetrics embracing the overlapping fields of bibliometrics and scientometrics following widely adopted definitions by, for example, Brookes (1990), Egghe and Rousseau (1990), and Tague-Sutcliffe (1992). According to Tague-Sutcliffe (1992, p. 1), *informetrics* is "the study of the quantitative aspects of information in any form, not just records or bibliographies, and in any social group, not just scientists." Furthermore, *bibliometrics* is defined as "the study of the quantitative aspects of the production, dissemination and use of recorded information" and *scientometrics* as "the study of the quantitative aspects of science as a discipline or economic activity" (ibid.). In the figure, politico-economical aspects of scientometrics are covered by the part of the scientometric ellipse lying outside the bibliometric one.

The diagram in Figure 1 further shows the field of webometrics entirely encompassed by bibliometrics, because Web documents, whether text or multimedia, are *recorded* information (cf. Tague-Sutcliffe's abovementioned definition of bibliometrics) stored on Web servers. This recording may be temporary only, just as not all paper documents are properly archived. Webometrics is partially covered by scientometrics, as many scholarly activities today are Web-based, while other such activities are even beyond bibliometrics, i.e., nonrecorded, like person-to-person conversation. Furthermore, webometrics is totally included within the field of cybermetrics as defined previously.

[3]Cf. http://www.cybergeography.org/

In the diagram in Figure 1, the field of cybermetrics exceeds the boundaries of bibliometrics, because some activities in cyberspace normally are not recorded but rather communicated synchronously, as in chat rooms. Cybermetric studies of such activities still fit in the generic field of informetrics as the study of the quantitative aspects of information "in any form" and "in any social group" as stated above by Tague-Sutcliffe (1992).

Naturally, the inclusion of webometrics expands the field of bibliometrics, as webometrics inevitably will contribute with further methodological developments of Web-specific approaches. As ideas rooted in bibliometrics, scientometrics, and informetrics contributed to the emergence of webometrics, ideas in webometrics might now contribute to the development of these embracing fields.

Terminology and Web Node Diagrams

The following three subsections deal with terminological issues and forms of diagrams for conceptualizing and illustrating Web structures at different levels of analysis in a consistent way.

Basic Link Terminology

The initial exploratory phases of an emerging field like webometrics inevitably lead to a variety in the terminology used. For example, a link received by a Web node (the network term *node* here denotes a unit of analysis like a Web page, directory, or Web site but could also be an entire top-level domain of a country) has been named, for example, *incoming link, inbound link, inward link, back link,* and *sitation;* the latter term (McKiernan, 1996; Rousseau, 1997) has clear connotations to bibliometric citation analysis. An example of a more problematic terminology is the two opposite meanings of an *external link:* either as a link pointing out of a Web site or a link pointing into a site.

Figure 2 illustrates an attempt to create a consistent basic webometric terminology for link relations between Web nodes (Björneborn, 2004). The figure reflects that the Web may be viewed as a so-called directed graph, using a graph-theoretic term (e.g., Broder et al., 2000; Kleinberg et al., 1999). In such a Web graph, Web nodes are connected by

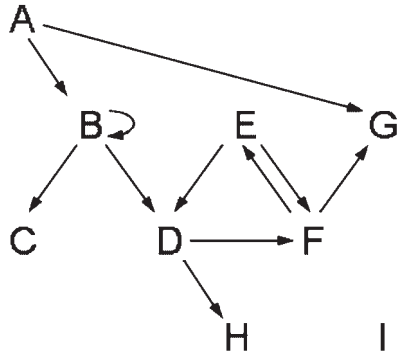

FIG. 2. Basic link relations (Björneborn, 2004). The letters may represent different Web node levels, e.g., Web pages, Web directories, Web sites, or top-level domains of countries or generic sectors. See legend in Table 1.

directed links. In this context, it should be noted that graph theoretic approaches have been used in bibliometrics and scientometrics since the 1960s for analyzing citation networks and other information networks (e.g., Egghe & Rousseau, 1990; Furner, Ellis, & Willett, 1996; Garner, 1967; Hummon & Doreian, 1989; Nance, Korfhage, & Bhat, 1972). Social network analysis (e.g., Scott, 2000; Wasserman & Faust, 1994) makes extensive use of graph theoretical approaches. A review article by Park and Thelwall (2003) compared information science approaches to studying the Web to those from social network analysis. It was found that information science tended to emphasize data validation and the study of methodological issues, whereas social network analysis suggested how its existing theory could transfer to the Web. Otte and Rousseau (2002) give an excellent overview of applications and potentials of social network analysis in the information sciences with regard to studies of, for example, citation and cocitation networks, collaboration structures and other forms of social interaction networks, including the Internet. In a forthcoming ARIST chapter on webometrics by Thelwall, Vaughan, and Björneborn, applications of graph theory and social network analysis in webometrics are further discussed. The proposed basic link terminology in Table 1 has origins in graph theory, social network analysis and bibliometrics.

The terms *outlink* and *inlink* are commonly used in computer-science Web studies (e.g., Broder et al., 2000; Chen, Newman, Newman, & Rada, 1998; Pirolli et al., 1996). The term *outlink* implies that a directed link and its two adjacent nodes are viewed from the source node providing the link, analogous with the use of the term *reference* in bibliometrics. A corresponding analogy exists between *inlink* and *citation* with the target node as the spectator's perspective; compare to Figure 3 (Björneborn, 2004). A link crossing a Web site border, like link *e* in Figure 4, is thus called a *site outlink* or a *site inlink* depending on the perspective of the spectator. Similar considerations of consistent terminology have been put forward in bibliometrics by, for example, Price (1970) who emphasized a conceptual difference between the reference and citation, which matches the difference between outlink and inlink just described.

FIG. 3. Different link terminology for the same link depending on the spectator's perspective as denoted by the eyes (Björneborn, 2004).

The terms *out-neighbor* and *in-neighbor* in the proposed terminology are also used in graph-theoretic Web research (e.g., Chakrabarti et al., 2002). On the Web, *self-links* are used for a wider range of purposes than self-citations in scientific literature. This reflects a special case of the general difference between outlinks/inlinks and references/citations. Page self-links point from one section to another within the same page. Site self-links (also known as internal links) are typically navigational pointers from one page to another within the same Web site.

Because of its dynamic and distributed nature, the Web often demonstrates Web pages reciprocally linking to each other—a case not normally possible in the traditional print-based citation world. *Reciprocal links,* such as those between nodes E and F in Figure 2, is a widespread existing Web term for mutual inlinks and outlinks between two Web nodes. This reciprocity is not necessarily completely symmetrical as there may be more links in one direction between two Web nodes. Sometimes, reciprocal links may be deliberately agreed by two Web site creators for attempting to obtain higher ranking in search engines employing inlink counts in ranking algorithms as in Google (Brin & Page, 1998; also cf. Walker, 2002).

In Figure 2, the *triadically linked* nodes D, E, and F correspond to the social network analytic term *triadic closure* (e.g., Skvoretz & Fararo, 1989), for example, used to denote the probability that nodes D and F are transitively connected if there are already links between D and E, and between E and F. In social networks, such simple *triadic structures* or *triads* are the building blocks of larger social structures

TABLE 1. Basic link terminology (Björneborn, 2004) for link relations in Fig. 2.

- B has an *inlink* from A; B is *inlinked;* A is *inlinking;* A is an *in-neighbor* of B.
- B has an *outlink* to C; B is *outlinking;* C is *outlinked;* C is an *out-neighbor* of B.
- B has a *self-link;* B is *self-linking*.
- A has no inlinks; A is *nonlinked*.
- C has no outlinks; C is *nonlinking*.
- I has neither in- nor outlinks; I is *isolated*.
- E and F have *reciprocal links;* E and F are *reciprocally* linked.
- D, E, and F all have in- or outlinks connecting each other; they are *triadically interlinked*.
- A has a *transversal* outlink to G: functioning as a shortcut.
- H is *reachable* from A by a directed *link path*.
- C and D are *colinked* by B; C and D have *co-inlinks*.
- B and E are *colinking* to D; B and E have *co-outlinks*.
- Co-inlinks and co-outlinks are both cases of *colinks*.

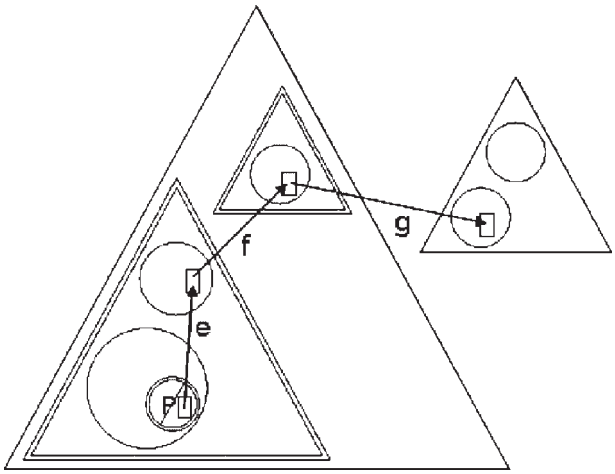

FIG. 4. Simplified Web node diagram illustrating basic Web node levels (Björneborn, 2004).

(e.g., Scott, 2000; Wasserman & Faust, 1994). Milo et al. (2002) use the term *motif* for similar simple triadic building blocks of complex networks in general, for example, in biochemistry, neurobiology, ecology, and engineering.

Most links on the Web connect Web pages containing cognate topics (Davison, 2000). However, some links in a Web node neighborhood may break such typical linkage patterns and connect dissimilar topical domains. Such (loosely defined) *transversal* links (Björneborn, 2001, 2004; Björneborn & Ingwersen, 2001) function as cross-topic shortcuts and may affect so-called small-world phenomena on the Web. Small-world phenomena are concerned with short distances along interconnection paths between nodes in a network graph. For example, short distances between two arbitrary persons through intermediate chains of acquaintances of acquaintances as studied in social network analysis (e.g., Milgram, 1967; Kochen 1989; Pool & Kochen, 1978/1979), and popularized by the notion of "six degrees of separation." Watts and Strogatz (1998) introduced a small-world network model characterized by highly clustered nodes as in regular graphs, yet with short characteristic path lengths between pairs of nodes as in random graphs. In their seminal paper, Watts and Strogatz (1998) showed that a very small percentage of long-range connections is sufficient in a small-world network to function as shortcuts connecting distant nodes of the network.

The concepts of *reachability* and link *paths* as illustrated in Figure 2 are both used in graph theory (e.g., Gross & Yellen, 1999), for example, when describing small-world properties as outlined previously.

The two *colinked* Web nodes C and D in Figure 2 with *co-inlinks* from the same source node are analogous to the bibliometric concept of *cocitation* (Small, 1973). Correspondingly, the two *colinking* nodes B and E having *co-outlinks* to the same target node are analogous to a *bibliographic coupling* (Kessler, 1963). *Colinks* is proposed as a generic term covering both concepts of co-inlinks and co-outlinks. The underlying assumption for the use of both the bibliometric and webometric concepts is that two documents (or two authors/link creators) are more similar, i.e., more semantically related, the higher the frequency of shared outlinks (references) or shared inlinks (citations).

Basic Web Node Terminology and Diagrams

In webometric studies, it may be useful to visualize relations between different units of analysis, for example, in the so-called Alternative Document Model (Thelwall, 2002; Thelwall & Harries, 2003). Figure 4 shows a diagram illustrating some basic building blocks in a consistent Web node framework (Björneborn, 2004). In the diagram, four basic Web node levels are denoted with simple geometrical figures: *quadrangles* (Web pages), *diagonal lines* (Web directories), *circles* (Web sites), and *triangles* (country or generic top level domains, TLDs). Sublevels within each of the four basic node levels are denoted with additional borderlines in the corresponding geometrical figure. For example, a triangle with a double borderline denotes a generic second level domain (SLD), also known as a sub-TLD, assigned by many countries to educational, commercial, governmental, and other sectors of society, for example, .ac.uk, .co.uk, .ac.jp, .edu.au.

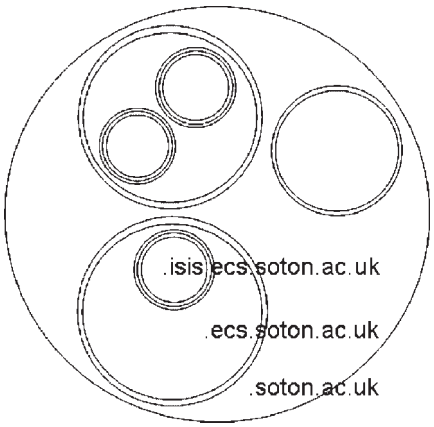

FIG. 5. Simplified Web node diagram of a Web site containing subsites and sub-subsites.

The simplistic Web node diagram in Figure 4 shows a page *P* located in a directory of a subsite in a sub-TLD. The page has a site outlink *e* to a page at a site in the same sub-TLD. The outlinked page in turn is outlinking to a page at a site in another sub-TLD in the same country. The link path *e-f-g* ends at a page at a site in another TLD.

Zooming in on a single Web site, this may comprise several subunits in the shape of subsites, sub-subsites, and so forth, as indicated by hierarchically derivative domain names. For example, as shown in Figure 5, the sub-subsite of The Image, Speech and Intelligent Systems Research Group (isis.ecs.soton.ac.uk) is located within the Department of Electronics and Computer Science (ecs.soton.ac.uk), one of many subsites at the University of Southampton, United Kingdom (soton.ac.uk). Subsites and sub-subsites are denoted as circles with double and triple borderlines, respectively. Subordinate sublevels would logically be denoted with additional number of borderlines. For sake of simplicity, the diagram does not reflect actual numbers and sizes of elements.

Although some Web sites subdivide into derivative domain names, as shown previously, other Web sites locate the same type of subunits into folder directories in their Web site file hierarchy. Obviously, such diverse allocation and naming practices complicate comparability in webometric studies. In Figures 6A and 6B, one or more diagonal lines (resembling URL slashes and reflecting the number of directory levels

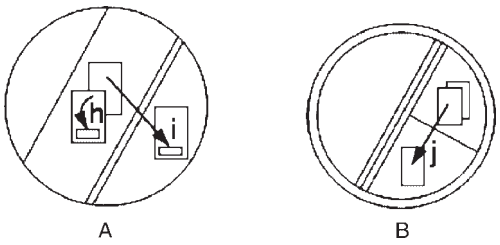

FIG. 6. Simplified Web node diagrams of a Web site and a subsite with links between different directory levels including page subelements.

below the URL root level) denote directories, subdirectories, and so forth.

Web pages may also consist of subelements such as text sections, frames, and so forth. Additional bands illustrate such page subelements as in the targets of the page self-link *h* and the page outlink *i* from the two sibling Web pages in the same directory in Figure 6A. More numerous and complex linkages within a site or subsite, and so forth, can be illustrated by combinations of elements in Figures 6A and 6B, showing links between pages located either at different directory levels (Figure 6A) or in sibling directories at the same level (Figure 6B) in the Web site file hierarchies.

Naturally, any diagrammatic representation of large-scale hypertext structures will get too tangled to be of any practical use or to be interpreted in any quantitative way. However, the proposed Web node diagrams with their simple and intuitive geometrical figures are intended to be used to emphasize and illustrate qualitative differences between investigated Web node levels in a webometric study. Figure 7 shows an example of such a Web node diagram used to illustrate included and excluded Web nodes and links in a connectivity analysis of the UK academic Web space (Björneborn, 2004). Moreover, the diagrams can illustrate actual structural aspects of limited subgraphs of an investigated Web space. Figure 8 gives an example of how the Web node diagrams were used in the above study more specifically concerned with what types of links, pages, and sites function as small-world connectors across dissimilar topical domains in an academic Web space (Björneborn, 2004).

Advanced Link Terminology and Diagrams

The Web can be studied at different granularities employing what here will be called *micro, meso,* and *macro* level perspectives (Björneborn, 2004). Micro level webometrics consists of studies of the construction and use of Web pages, Web directories, and small sub-subsites, and so forth, for example, constituting individual Web territories. Meso level webometrics is correspondingly concerned with quantitative aspects of larger subsites and sites, and macro level webometrics comprises studies of clusters of many sites, or focuses on sub-TLDs or TLDs. Several webometric studies, including classic ones by Larson (1996) and Almind and Ingwersen (1997), have used meso level approaches concerned with site-to-site interconnectivity as well as macro level TLD-to-TLD analysis, primarily applying page level link counts. However, to extract useful information, links may also be aggregated on different node levels as in the earlier mentioned Alternative Document Model (Thelwall, 2002; Thelwall & Harries, 2003).

An adequate terminology for aggregated link relations should capture both the link level under investigation and the reach of each link. Such a terminology should reflect at least three elements: (1) the investigated link level, (2) the highest-level Web node border crossed by the link, and (3) the spectator's perspective (cf. Figure 3). For sake of simplicity, the perspective from the outlinking nodes is chosen in the following examples showing higher and higher link aggregations.

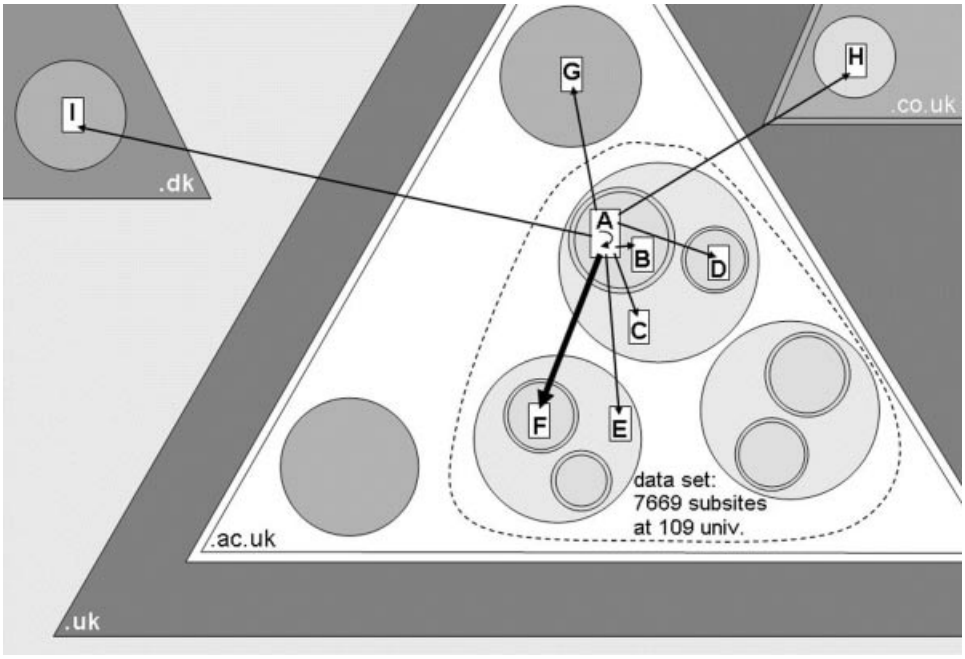

FIG. 7. Example of Web node diagram illustrating qualitative differences between links and Web node levels in a webometric study. The figure illustrates included and excluded Web nodes and links in an analysis of small-world link structures across the UK academic Web space (Björneborn, 2004). The bold link AF symbolizes all included 207,865 page level links between 7,669 subsites at 109 different UK universities in the analysis. All other links were excluded: AA (page self-links); AB (subsite self-links); AC and AD (site self-links); AE (site outlinks to university main sites); AG (site outlinks to ac.uk sites outside data set); AH (sub-TLD outlinks, i.e., links to other UK sub-TLD); and AI (TLD outlinks, i.e., links to other TLD).

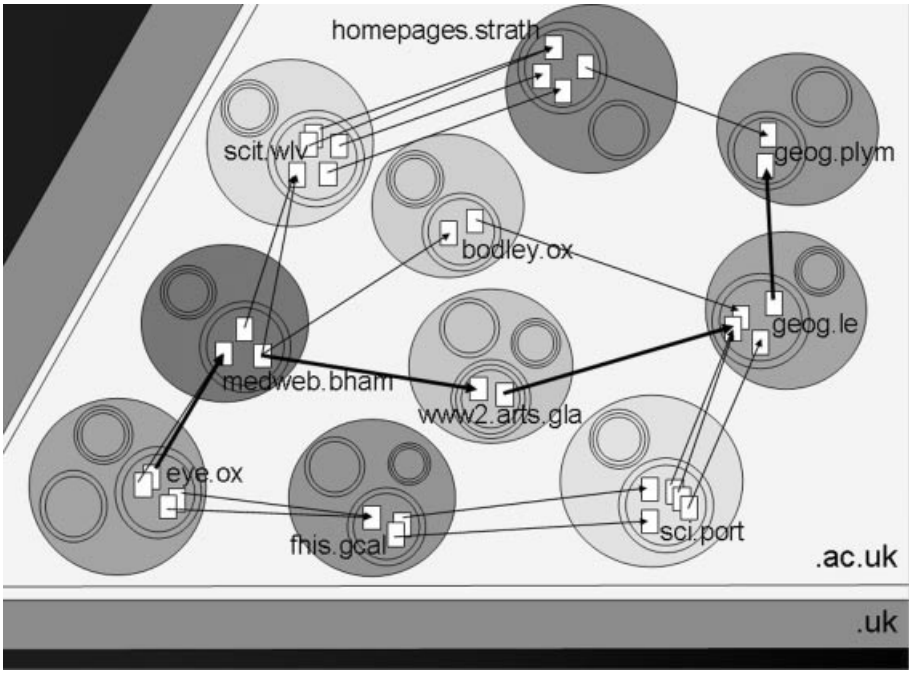

FIG. 8. Example of Web node diagram showing a limited subgraph. It contains an excerpt of shortest link paths (path length 4) between a subsite on eye research (www.eye.ox.ac.uk) and a subsite in geography (www.geog.plym.ac.uk) to identify pages and sites that provide transversal (cross-topic) links across dissimilar topical domains in the UK academic Web space (Björneborn, 2004). Bold links show one example of a shortest link path between the two mentioned subsites. Only links connecting subsites at different UK universities were considered (cf. Figure 7). See Appendix for affiliations.

Figure 9 below shows 14 *page level links* including a page level subsite outlink, k_p (also being a page level site self-link). The subscript in k_p denotes page level. If a webometric study comprises just one level of links, the terminology can be simplified to cover merely the link reach. In such a case, l_p is a site outlink, m_p a sub-TLD outlink, and n_p a TLD outlink.

For sake of simplicity, directory and subsite level links will not be treated here. However, the terminology for these levels would parallel the other levels included.

Figure 10 illustrates 11 site level links. For example, o_s is a site level site outlink aggregating three page level links from Figure 9. Site self-links are denoted with curved arrows.

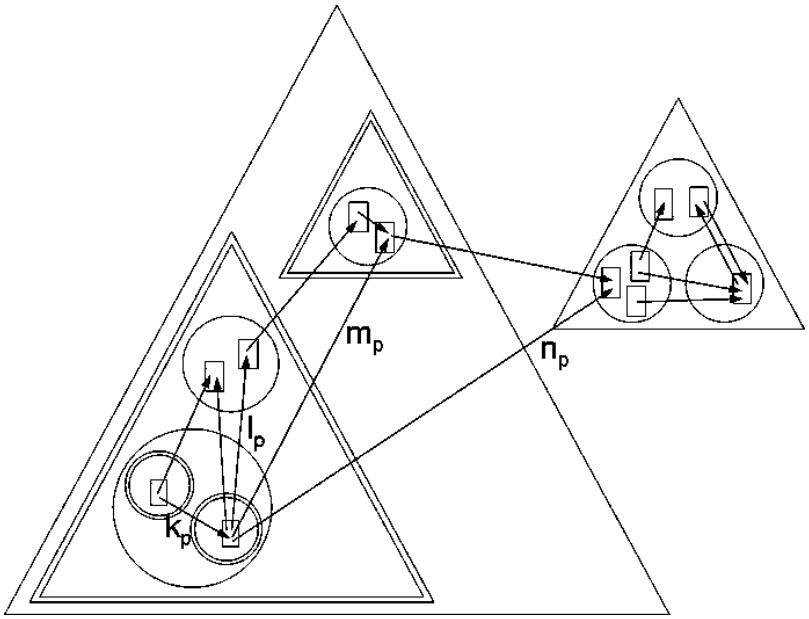

FIG. 9. Web node diagram with page level links (Björneborn, 2004).

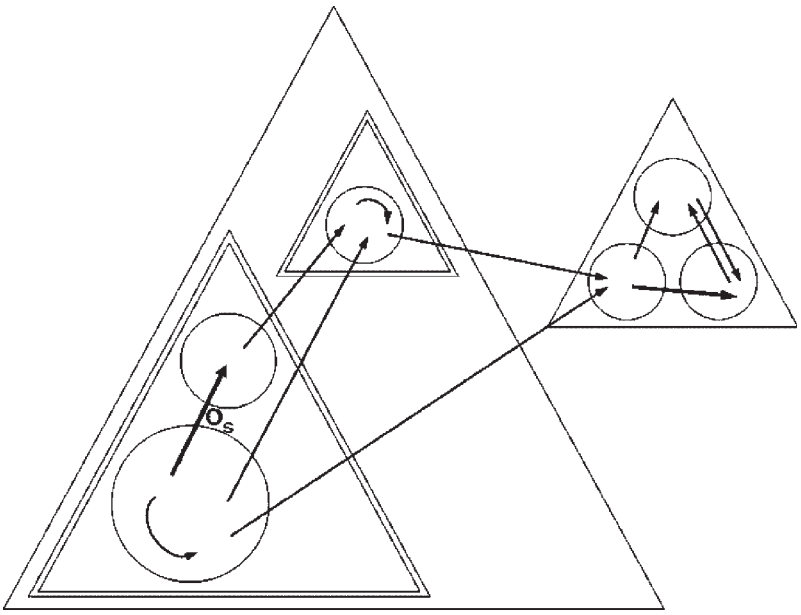

FIG. 10. Web node diagram with site level links.

In this context, it should be noted that a site level link always connects a source site with a target site. Correspondingly, a page level link always connects a source page with a target page; compare to Figures 8 and 9. This point is necessary to make, because a target URL for a Web page may deceivingly look like an URL for a Web site. It is thus common Web practice to stem the target URL of top entry pages of a Web site. For example, instead of writing the full URL www.db.dk/default.htm in a target link pointing to the top entry page of the Royal School of Library and Information Science, it is more convenient to stem the URL to www.db.dk because Web servers automatically look for default pages for stemmed URLs. However, this stemmed URL still denotes a Web page and not a Web site.

This line of higher and higher link aggregations ends with sub-TLD level links as shown in Figure 11 and TLD level links in Figure 12. Terminology for these levels parallel the other levels included.

Discussion and Conclusion

We have demonstrated the relationships between the various metrics associated with library and information science in the framework of its established subfield informetrics. Most basically, we refer webometrics as belonging to cybermetrics and covered by an expanded concept of bibliometrics. We believe that a general consensus exists as to this framework within library and information science.

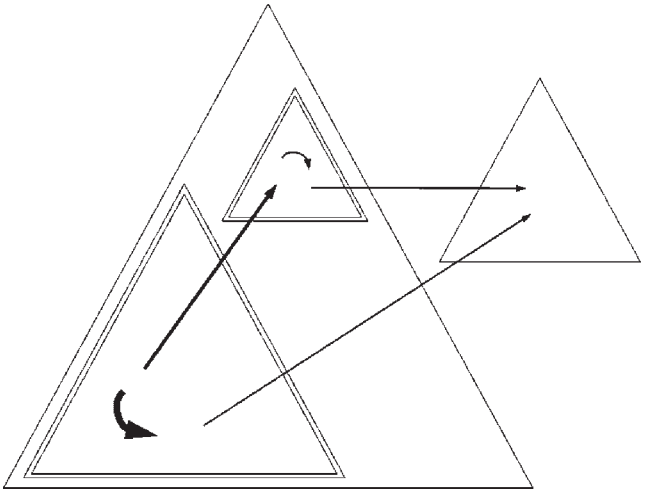

FIG. 11. Web node diagram with sub-TLD level links.

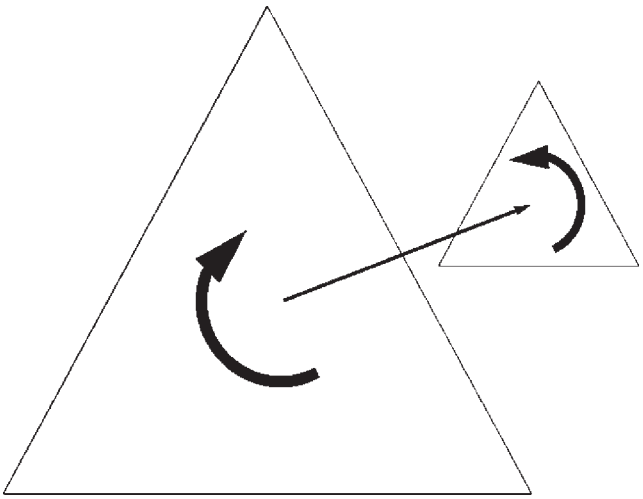

FIG. 12. Web node diagram with TLD level links.

The proposals concerning the basic link terminology are consistent with the increasingly common notation of the most used concepts in the field of webometrics, such as inlink or outlink. However, other notations are obviously required for the additional possible forms of hypertextual associations between Web nodes, for example, reciprocal or transversal links. However, the term *sitation,* introduced by McKiernan (1996) and Rousseau (1997), is not seen as a convenient notation for (in)links. *Sitation* suffers from the same conceptual problem as the term *citation*—namely, that it can be interpreted as synonymous with outlink, i.e., an outgoing reference to other work. Moreover, during oral presentations the distinction between the words *citations* and *sitations* is far from obvious and requires context to be fully understood.

From our perspective, two dimensions of the link terminology are particularly important. First, an analogy exists between references or citations and outlinks or inlinks. Likewise, traditional cocitation or bibliographic coupling is technically similar to colinked or colinking Web nodes, respectively. Nevertheless, it *is* only an analogy, as also stressed by, for example, Björneborn and Ingwersen (2001), Egghe (2000), Meyer (2000), Prime, Bassecoulard, and Zitt (2002), and van Raan (2001). The reasons for giving scholarly references to other scientific work are not fully understood and are different from providing outlinks in the dynamic Web environment (cf. Kim, 2000; Thelwall, 2003; Wilkinson et al., 2003). In many cases, for example, navigational reasons prevail. Operationally, however, one may calculate, analyze, or map the manifestations of such activities. Hence, analogous to citation analyses one must take care when making interpretations of link analyses on different Web spaces.

Second, it is important to be aware of what is measured or counted. For example, there is a rather large difference between counting the *real* number of inlinks to a Web site or page and counting the number of in-neighbors in the shape of Web pages (or sites) inlinking at least once to some Web node. This difference is often overlooked in both calculus and applying terminology. Again, we observe an analogy to citation analysis, when numbers of citations—not only the number of citing articles—are counted. The intellectual and conceptual confusion increases, however, in particular for newcomers in the informetric subfields, when one considers that it is exactly the number of cociting articles, not the actual citations, that commonly are applied to calculating the strength of cocitation.

The distinction among Web node levels, its terminological impact, and the proposal of a consistent diagram notation is necessary for the topology of the Web to be understood and investigated. For example, this distinction is useful when analyzing and illustrating different aggregated Web node levels—nested as Chinese boxes within boxes—as shown in Figures 9–12. There exists a constant possibility of loosing the point of perspective in such analysis, in particular if terminological rigor is lacking.

In conclusion, it should be emphasized that the outlined webometric framework as well as the terminology and diagram notation proposals are seen as conceptual foundations and building blocks by which future discoveries and perspectives of the Web and webometrics hopefully will thrive.

References

Abraham, R.H. (1996). Webometry: Measuring the complexity of the World Wide Web. Visual Math Institute, University of California at Santa Cruz. Retrieved July 9, 2004, from http://www.ralph-abraham.org/vita/redwood/vienna.html

Adamic, L.A., & Huberman, B.A. (2000). Power-law distribution of the World Wide Web. Science, 287, 2115a.

Adamic, L.A., & Huberman, B.A. (2001). The Web's hidden order. Communications of the ACM, 44(9), 55–59.

Albert, R., & Barabási, A.L. (2002). Statistical mechanics of complex networks. Reviews of Modern Physics, 74(1), 47–97.

Albert, R., Jeong, H., & Barabási, A.L. (1999, September 9). Diameter of the World-Wide Web. Nature, 401, 130–131.

Almind, T.C., & Ingwersen, P. (1996). Informetric analysis on the World Wide Web: A methodological approach to "internetometrics." Centre for Informetric Studies, Royal School of Library and Information Science, Copenhagen, Denmark. (CIS Report 2).

Almind, T.C., & Ingwersen, P. (1997). Informetric analyses on the World Wide Web: methodological approaches to "webometrics." Journal of Documentation, 53(4), 404–426.

Bar-Ilan, J. (1997). The "mad cow disease," usenet newsgroups and bibliometric laws. Scientometrics, 39(1), 29–55.

Berners-Lee, T., & Cailliau, R. (1990). World Wide Web: Proposal for a hypertext project. Retrieved July 9, 2004, from http://www.w3.org/Proposal.html

Björneborn, L. (2001). Small-world linkage and co-linkage. Proceedings of the 12th ACM Conference on Hypertext and Hypermedia (pp. 133–134). New York: ACM Press.

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.

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

Borgman, C.L., & Furner, J. (2002). Scholarly communication and bibliometrics. Annual Review of Information Science and Technology, 36, 3–72.

Bossy, M.J. (1995). The last of the litter: "Netometrics." Solaris, 2 (special issue on "Les sciences de l'information: Bibliométrie, scientométrie, infométrie"). Presses Universitaires de Rennes. Retrieved July 9, 2004, from http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d02/2bossy.html

Bradford, S.C. (1934). Sources of information on specific subjects. British Journal of Engineering, 137, 85–86.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems, 30, 1–7.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., et al. (2000). Graph structure in the Web. Computer Networks, 33(1–6), 309–320.

Brookes, B.C. (1990). Biblio-, sciento-, infor-metrics??? What are we talking about? In L. Egghe & R. Rousseau (Eds.), Informetrics 89/90: Selection of papers submitted for the Second International Conference on Bibliometrics, Scientometrics and Informetrics. London, Ontario, Canada, July 5–7, 1989 (pp. 31–43). Amsterdam: Elsevier.

Chakrabarti, S., Dom, B.E., Kumar, S.R., Raghavan, P., Rajagopalan, S., Tomkins, A., et al. (1999). Mining the Web's link structure. IEEE Computer, 32(8), 60–67.

Chakrabarti, S., Joshi, M.M., Punera, K., & Pennock, D.M. (2002). The structure of broad topics on the Web. Proceedings of the WWW2002 Conference. Retrieved July 9, 2004, from http://www2002.org/CDROM/refereed/338/

Chen, C., Newman, J., Newman, R., & Rada, R. (1998). How did university departments interweave the Web: A study of connectivity and underlying factors. Interacting with Computers, 10, 353–373.

Chi, E.H., Pitkow, J., Mackinlay, J., Pirolli, P., Gossweiler, R., & Card, S.K. (1998). Visualizing the evolution of Web ecologies. Proceedings of

Human Factors in Computing Systems (CHI '98, pp. 400–407). ACM Press.
Clever Project, Members of the (1999). Hypersearching the Web. Scientific American, 280(6), 54–60.
Cooley, R., Mobasher, B., & Srivastava, J. (1997). Web mining: Information and pattern discovery on the World Wide Web. Proceedings of the Ninth IEEE International Conference on Tools with Artificial Intelligence (ICTAI '97). Retrieved July 9, 2004, from http://citeseer.nj.nec.com/cooley97Web.html
Cronin, B. (2001). Bibliometrics and beyond: Some thoughts on Web-based citation analysis. Journal of Information Science, 27(1), 1–7.
Cronin, B., Snyder, H.W., Rosenbaum, H., Martinson, A., & Callahan, E. (1998). Invoked on the Web. Journal of the American Society for Information Science, 49(14), 1319–1328.
Davison, B.D. (2000). Topical locality in the Web. Proceedings of the 23rd annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 272–279). New York: ACM Press.
Ding, C., Zha, H., He, X., Husbands, P., & Simon, H. (2002). Link analysis: Hubs and authorities on the World Wide Web. LBNL Tech Report 47847. Retrieved July 9, 2004, from http://crd.lbl.gov/~cding/papers/hits5.pdf
Dodge, M. (1999). The geography of cyberspace. CASA working paper 8. Centre for Advanced Spatial Analysis, University College London. Retrieved July 9, 2004, from http://www.casa.ucl.ac.uk/cyberspace.pdf
Dodge, M., & Kitchin, R. (2001). Mapping cyberspace. London: Routledge.
Dodge, M., & Kitchin, R. (2002). New cartographies to chart Cyberspace. GeoInformatics, 5(April/May), 38–41. Retrieved July 9, 2004, from http://www.casa.ucl.ac.uk/martin/geoinformatics_article.pdf
Downie, J.S. (1996). Informetrics and the World Wide Web: A case study and discussion. Proceedings of the 24th Annual Conference of the Canadian Association for Information Science, Toronto (pp. 130–141).
Efe, K., Raghavan, V., Chu, C.H., Broadwater, A.L., Bolelli, L., & Ertekin, S. (2000). The shape of the Web and its implications for searching the Web. Proceedings of the International Conference on the Advances in Infrastructure for Electronic Business, Science, and Education on the Internet, L'Aquila, Italy, July 31–Aug 6, 2000. Retrieved July 9, 2004, from http://citeseer.nj.nec.com/317732.html
Egghe, L. (2000). New informetric aspects of the Internet: Some reflections—many problems. Journal of Information Science, 26(5), 329– 335.
Egghe, L., & Rousseau, R. (1990). Introduction to informetrics: Quantitative methods in library, documentation and information science. Amsterdam: Elsevier.
Etzioni, O. (1996). The World-Wide Web: Quagmire or gold mine? Communications of the ACM, 39(11), 65–68.
Furner, J., Ellis, D., & Willett, P. (1996). The representation and comparison of hypertext structures using graphs. In M. Agosti & A.F. Smeaton (Eds.), Information retrieval and hypertext (pp. 75–96). Boston: Kluwer.
Garfield, E. (1955 July 15). Citation indexes for science: A new dimension in documentation through association of ideas. Science, 122, 108–111.
Garner, R. (1967). A computer oriented, graph theoretic analysis of citation index structures. In B. Flood (Ed.), Three Drexel information science research studies (pp. 3–46). Drexel Press. Retrieved July 9, 2004, from http://www.garfield.library.upenn.edu/rgarner.pdf
Girardin, L. (1995). Cyberspace geography visualization: Mapping the World-Wide Web to help people find their way in cyberspace. The Graduate Institute of International Studies, Geneva. Retrieved July 9, 2004, from http://www.girardin.org/luc/cgv/report/report.pdf
Girardin, L. (1996). Mapping the virtual geography of the World-Wide Web. Proceedings of the Fifth WWW Conference. Retrieved July 9, 2004, from http://www.girardin.org/luc/cgv/www5/index.html
Gross, J., & Yellen, J. (1999). Graph theory and its applications. Boca Raton, FL: CRC Press.
Harter, S.P., & Ford, C.E. (2000). Web-based analyses of E-journal impact: Approaches, problems, and issues. Journal of the American Society for Information Science, 51(13), 1159–1176.
Hernández-Borges, A.A., Pareras, L.G., & Jiménez, A. (1997). Comparative analysis of pediatric mailing lists on the Internet. Pediatrics, 100(2), e8. Retrieved July 9, 2004, from http://www.pediatrics.org/cgi/content/full/100/2/e8
Herring, S.C. (2002). Computer-mediated communication on the Internet. Annual Review of Information Science and Technology, 36, 109–168.
Hertzel, D.H. (1987). History of the development of ideas in bibliometrics. Encyclopedia of Library and Information Science (Vol. 42, Supplement 7, pp. 144–219). New York: Marcel Dekker.
Huberman, B.A. (2001). The laws of the Web: Patterns in the ecology of information. Cambridge, MA: MIT Press.
Hummon, N.P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. Social Networks, 11, 39–63. Retrieved July 9, 2004, from http://www.garfield.library.upenn.edu/papers/hummondoreian1989.pdf
Hurd, J.M. (2000). The transformation of scientific communication: A model for 2020. Journal of the American Society for Information Science, 51(14), 1279–1283.
Ingwersen, P. (1998). The calculation of Web impact factors. Journal of Documentation, 54(2), 236–243.
Kessler, M.M. (1963). Bibliographic coupling between scientific papers. American Documentation, 14(1), 10–25.
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.M. (1999). Authoritative sources in a hyperlinked environment. Journal of the ACM, 46(5), 604–632.
Kleinberg, J.M., Kumar, R., Raghavan, P., Rajagopalan, S., & Tomkins, A. (1999). The Web as a graph: Measurements, models, and methods. Lecture Notes in Computer Science, 1627, 1–18.
Kochen, M. (Ed.). (1989). The small world. Norwood, NJ: Ablex.
Kosala, R., & Blockeel, H. (2000). Web mining research: A survey. SIGKDD Explorations, 2(1), 1–15.
Kuster, R.J. (1996). A bibliometric study of the remote hypertext links in public library World Wide Web sites. Proceedings of the ASIS Mid-Year Meeting, San Diego, California (pp. 338–343). Medford, NJ: Information Today.
Larson, R.R. (1996). Bibliometrics of the World Wide Web: An exploratory analysis of the intellectual structure of Cyberspace. Proceedings of the 59th ASIS Annual Meeting, Baltimore, Maryland (pp. 71–78). Medford, NJ: Learned Information Inc./ASIS.
Levene, M., & Poulovassilis, A. (2001). Web dynamics. Software Focus, 2(2), 60–67.
Lotka, A.J. (1926, June 19). The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences, 16, 317–323.
Matzat, U. (1998). Informal academic communication and scientific usage of Internet discussion groups. Proceedings IRISS '98 International Conference. Retrieved July 9, 2004, from http://sosig.ac.uk/iriss/papers/paper19.htm
McKiernan, G. (1996). CitedSites(sm): Citation indexing of Web resources. Retrieved July 9, 2004, from http://www.public.iastate.edu/~CYBERSTACKS/Cited.htm
Menczer, F. (2002). Growing and navigating the small world Web by local content. Proceedings of the National Academy of Sciences, 99(22), 14014–14019.
Meyer, E.K. (2000). Web metrics: Too much data, too little analysis. In D. Nicholas & I. Rowlands (Eds.), The Internet: Its impact and evaluation. Proceedings of an international forum held at Cumberland Lodge, Windsor Park, July 16–18, 1999 (pp. 131–144). London: Aslib/IMI.
Milgram, S. (1967). The small-world problem. Psychology Today, 1(1), 60–67.
Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002, October 25). Network motifs: Simple building blocks of complex networks. Science, 298(5594), 824–827.
Molyneux, R.E., & Williams, R.V. (1999). Measuring the Internet. Annual Review of Information Science and Technology, 34, 287–339.
Moulthrop, S., & Kaplan, N. (1995). Citescapes: Supporting knowledge construction on the Web. Poster at the WWW4 Conference. Retrieved July 9, 2004, from http://iat.ubalt.edu/moulthrop/essays/citescapes/citescapes.html
Nance, R.E., Korfhage, R.R., & Bhat, U.N. (1972). Information networks: Definitions and message transfer models. Journal of the American Society for Information Science, 23(4), 237–247.

Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. Journal of Information Science, 28(6), 441–454.

Park, H.W., & Thelwall, M. (2003). Hyperlink analyses of the World Wide Web: A review. Journal of Computer-Mediated Communication, 8(4). Retrieved July 9, 2004, from http://www.ascusc.org/jcmc/vol8/issue4/park.html

Pirolli, P., Pitkow, J., & Rao, R. (1996). Silk from a sow's ear: Extracting usable structures from the Web. CHI 96 Electronic Proceedings. Retrieved July 9, 2004, from http://www.acm.org/sigchi/chi96/proceedings/papers/Pirolli_2/pp2.html

Pitkow, J.E. (1997). Characterizing World Wide Web ecologies. Doctoral dissertation. Georgia Institute of Technology. Retrieved July 9, 2004, from http://www.pitkow.com/docs/1997-Pitkow-Dissertation.pdf

Pitkow, J.E., & Pirolli, P. (1997). Life, death, and lawfulness on the electronic frontier. CHI 97 Electronic Publications. Retrieved July 9, 2004, from http://www.acm.org/sigchi/chi97/proceedings/paper/jp-www.htm

Pool, I. de S., & Kochen, M. (1978/1979). Contacts and influence. Social Networks, 1, 5–51.

Price, D. de Solla. (1965). Networks of scientific papers. In M. Kochen (Ed.), The growth of knowledge: Readings on organization and retrieval of information (pp. 145–155). New York: Wiley.

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

Prime, C., Bassecoulard, E., & Zitt, M. (2002). Co-citations and co-sitations: A cautionary view on an analogy. Scientometrics, 54(2), 291–308.

Rousseau, R. (1997). Sitations: An exploratory study. Cybermetrics, 1(1). Retrieved July 9, 2004, from http://www.cindoc.csic.es/cybermetrics/articles/v1i1p1.html

Scott, J. (2000). Social network analysis: A handbook (2nd ed.). Thousand Oaks, CA: Sage.

Skvoretz, J., & Fararo, T.J. (1989). Connectivity and the small world problem. In M. Kochen (Ed.), The small world (pp. 296–326). Norwood, NJ: Ablex.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. Journal of the American Society for Information Science, 24(4), 265–269.

Spertus, E. (1997). ParaSite: Mining structural information on the Web. WWW6 Conference. Retrieved July 9, 2004, from http://decWeb.ethz.ch/WWW6/Technical/Paper206/Paper206.html

Tague-Sutcliffe, J. (1992). An introduction to informetrics. Information Processing & Management, 28(1), 1–3.

Thelwall, M. (2002). Conceptualizing documentation on the Web: An evaluation of different heuristic-based models for counting links between university Web sites. Journal of the American Society for Information Science and Technology, 53(12), 995–1005.

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 no. 151. Retrieved July 9, 2004, from http://informationr.net/ir/8–3/paper151.html

Thelwall, M., & Harries, G. (2003). The connection between the research of a university and counts of links to its Web pages: An investigation based on 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(7), 594–602.

Thelwall, M., Vaughan, L., & Björneborn, L. (forthcoming). Webometrics. Annual Review of Information Science and Technology, 39.

Thelwall, M., & Wilkinson, D. (2003). Three target document range metrics for university Web sites. Journal of the American Society for Information Science and Technology, 54(6), 490–497.

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

Vázquez, A. (2001). Knowing a network by walking on it: Emergence of scaling. Europhysics Letters, 54, 430–435.

Walker, J. (2002). Links and power: The political economy of linking on the Web. Proceedings of Hypertext 2002 (pp. 78–79). Baltimore: ACM.

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications. Cambridge: Cambridge University Press.

Watts, D.J., & Strogatz, S.H. (1998, June 4). Collective dynamics of "small-world" networks. Nature, 393, 440–442.

Weiss, R., Vélez, B., Sheldon, M.A., Namprempre, C., Szilagyi, P., Duda, A., et al. (1996). HyPursuit: A hierarchical network search engine that exploits content-link hypertext clustering. Proceedings of the Seventh ACM Conference on Hypertext and Hypermedia (pp. 180–193). New York: ACM Press.

White, H.D., & McCain, K.W. (1989). Bibliometrics. Annual Review of Information Science and Technology, 24, 119–186.

Wilkinson, D., Harries, G., Thelwall, M., & 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), 49–56.

Yao, Y.Y., Zhong, N., Liu, J., & Ohsuga, S. (2001). Web intelligence (WI): Research challenges and trends in the new information age. Lecture Notes in Artificial Intelligence, 2198, 1–17.

Zhang, Y. (2001). Scholarly use of Internet-based electronic resources. Journal of the American Society for Information Science and Technology, 52(8), 628–654.

Zipf, G.K. (1949). Human behavior and the principle of least effort: An introduction to human ecology. Cambridge, MA: Addison-Wesley.

Appendix

Figure A1 shows a so-called path net consisting of all shortest link paths (path length 4) between two subsites, www.eye.ox.ac.uk and www.geog.plym.ac.uk, in a study of small-world link structures across the UK academic Web space (Björneborn, 2004). Only links connecting subsites at different UK universities were considered in the study. ID numbers refer to 7669 investigated subsites. Counts of page level links between subsites are shown. White nodes denote subsites included in the path net excerpt shown in Figure 8. The affiliations of the subsites in the path net are listed in Table A1.

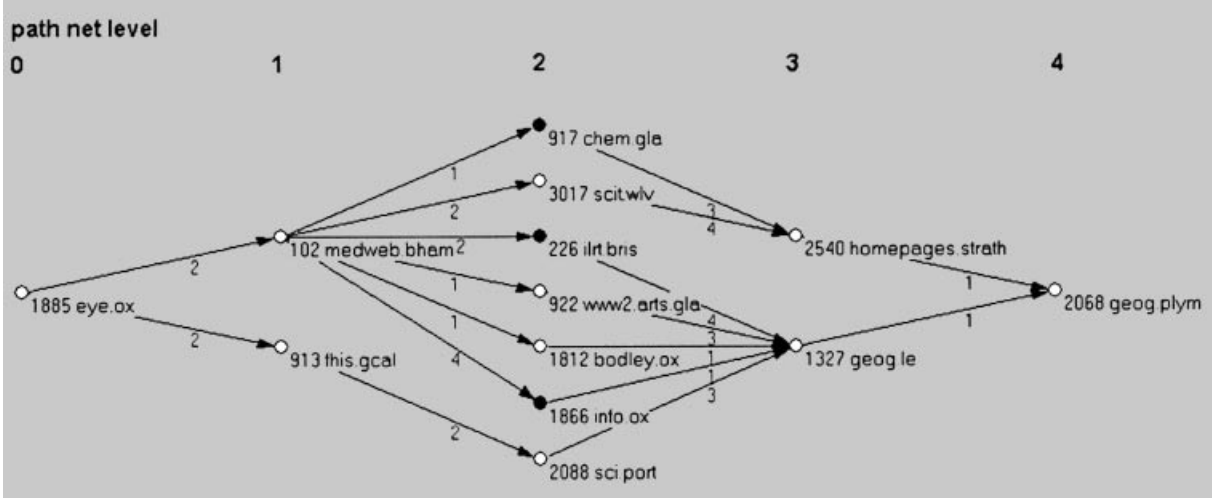

FIG. A1. Path net consisting of all shortest link paths between two subsites.

TABLE A1. The affiliations of the subsites in the path net.

Path net level	Id	Short domain name	Affiliation
0	1885	eye.ox.ac.uk	Dept of Ophthalmology, Univ. of Oxford
1	102	medweb.bham.ac.uk	School of Medicine, Univ. of Birmingham
1	913	fhis.gcal.ac.uk	Faculty of Health, Glasgow Caledonian University
2	226	ilrt.bris.ac.uk	Institute for Learning and Research Technology, Univ. of Bristol
2	917	chem.gla.ac.uk	Dept of Chemistry, Univ. of Glasgow
2	922	www2.arts.gla.ac.uk	Faculty of Arts, Univ. of Glasgow
2	1812	bodley.ox.ac.uk	Bodleian Library, Univ. of Oxford
2	1866	info.ox.ac.uk	Official Oxford University web pages
2	2088	sci.port.ac.uk	Faculty of Science, Univ. of Portsmouth
2	3017	scit.wlv.ac.uk	School of Computing and Information Technology, Univ. of Wolverhampton
3	1327	geog.le.ac.uk	Dept of Geography, Univ. of Leicester
3	2540	homepages.strath.ac.uk	Personal web pages, Univ. of Strathclyde
4	2068	geog.plym.ac.uk	Dept of Geographical Sciences, Univ. of Plymouth