

A Web Link Structure of the American Library & Information Science Field: A Pilot Study

Soohyung Joo
 School of Information Studies
 University of Wisconsin-Milwaukee,
 P.O. Box 413, Milwaukee, WI
 +1-414-793-1748
 sjoo@uwm.edu

Keywords

Web link analysis, Social network analysis, Webometrics, Library and Information Science Field

1. INTRODUCTION

Web link analysis has become one of the methods to delineate a communication structure among different scholars or organizations with the advent of the Web. This study represents a pilot investigation of the Web space in the Library and Information Science (LIS) field using social network analysis techniques. By relying on collected in- and out-link data using a hyperlink crawler and analyzing the directed hyperlink data between Web sites in the LIS field using different techniques, this study tried to explore the structure of the Web space in the field of LIS. Also, this study investigated which Web sites play a central role in communication in the Web space, along with the dependency between Web sites. As an exploratory study, the entire breadth of LIS-related Web sites was not covered.

2. RESEARCH QUESTIONS

This study attempts to address two research questions in relation to Web space analysis of the LIS field:

- 1) Which sites will be located in the central region of the LIS Web space, as measured by link analysis?
- 2) What types of dependency patterns emerge among the Web sites?

3. METHODOLOGY

3.1 Data Collection

Twenty-four Web sites were selected, which include relevant scholarly and professional organizations, government institutions, and American LIS schools. For scholarly and professional organizations, representative organizations were chosen such as the American Library Association (ALA), the American Society for Information Science & Technology (ASIS&T), and the International Federation of Library Associations and Institutions (IFLA). Fourteen LIS schools accredited by ALA were also selected. Table 1 shows the 24 research objects in this study, and the ID (code) used in an italic font to represent the Web site of the institution.

Table 1: The list of 24 research objects

Type	Institution	ID (Code)
Organization	American Library Association	<i>ala</i>
	The International Federation of Library Associations and Institutions	<i>ifla</i>
	Library of Congress	<i>loc</i>
	Online Computer Library Center	<i>oclc</i>
	Special Libraries Association	<i>sla</i>
	Institute of Museum and Library Service	<i>imls</i>
	The American Association of Law Libraries	<i>aall</i>
	The National Archives and Records Administration	<i>archives</i>
	The American Society for Information Science & Technology	<i>asis&t</i>
	The Society of American Archivists	<i>saa</i>
School	School of Library and Information Science, The Catholic University of America	<i>slis.cua</i>
	School of Library and Information Science, Indiana University Bloomington	<i>slis.indiana</i>
	School of Library and Information Science, Kent State University	<i>slis.kent</i>
	School of Information, University of Michigan Ann Arbor	<i>si.umich</i>
	School of Information Sciences, University of Pittsburgh	<i>is.pitt</i>
	School of Communication, Information and Library Studies, Rutgers	<i>scils.rutgers</i>
	Graduate School of Library and Information Science, Simmons University	<i>lis.simmons</i>
	School of Information, The University of Texas Austin	<i>is.utexas</i>
	Department of Information Studies, UCLA	<i>is.ucla</i>
	Graduate School of Library and Information Science, UIUC	<i>lis.uiuc</i>
	School of Information and Library Science, UNC	<i>sil.sunc</i>
	School of Information Sciences, University of Tennessee Knoxville	<i>sis.utk</i>
	College of Information Science, Florida State University	<i>ci.fsu</i>
	Information School, University of Washington	<i>is.washington</i>

In- and out-link data for the 24 Web sites were collected using a Web crawler that was designed to collect the hyperlinks to external domains. The data collection was conducted in February 2009. The crawling depth was set to the fourth level in the study. Even though the fourth level of depth did not necessarily reflect the whole structure of Web site, it would be acceptable for an initial investigation of this topic.

3.2 Analysis Techniques

For this pilot study, two approaches were identified.

1) *In-link and Out-link Visualization*: First, to understand the overall Web space structure of the LIS field, the in- and out-link data were visualized using Netdraw software (<http://www.analytictech.com/Netdraw/netdraw.htm>), which was developed for social network studies.

2) *Degree Centrality Test*: For more detailed investigation of the characteristics of each node, a degree centrality test was included. Centrality in network analysis can be used to quantify an individual node's prominence, influence or dependency using in-degree and out-degree measures. In-degree refers to the number of in-links, whereas out-degree refers to the number of out-links. In network analysis, in-degree represents how prestigious a node is because it indicates how many links a node receives from other nodes[1].

4. RESULTS

An in- and out-link matrix was creating using the external link crawler. The total number of hyperlinks (in-links and out-links) among the twenty four nodes up to the fourth level of each site was 7,898. Table 2 summarizes the descriptive statistics of the asymmetric in- and out-link matrix. The average number of hyperlinks between two nodes was 14.31 with a standard deviation of 41.39. The *ala* node had the largest number of in-links and out-links, totaling 2,156 and 1,147 links, respectively. This reveals that the ALA site is the most frequently linked with others in the LIS Web space. The second largest node was *loc*, with 1,247 in-links and 1,122 out-links. Third was *asis&t* with a total of 1,096 in- and out links. The matrix revealed that the average number of in- and out-links was larger in organizational nodes rather than in school nodes. The average number of links of each organizational node was 474.13, whereas the average links of each school node was 212.65. From this we can deduce that the organizational nodes have more in-links (6,683) than out-links (4,222), while the school nodes have more out-links (3,676) than in-links (1,215).

Table 2: Descriptive statistics of in- and out-link asymmetric matrix data

		(A) in-links from other 23 nodes		(B) out-links to other 23 nodes		(C) Total (A+B)	
		Counts of (A)	Ratio	Counts of (B)	Ratio	Counts of (A+B)	Ratio
org/.gov	<i>ala</i>	2156	27.3%	1147	14.5%	3303	20.9%
	<i>asis&t</i>	405	5.1%	691	8.7%	1096	6.9%
	<i>ifla</i>	412	5.2%	213	2.7%	625	4.0%
	<i>loc</i>	1247	15.8%	1122	14.2%	2369	15.0%
	<i>oclc</i>	820	10.4%	37	0.5%	857	5.4%
	<i>sla</i>	540	6.8%	190	2.4%	730	4.6%
	<i>imls</i>	360	4.6%	44	0.6%	404	2.6%
	<i>aall</i>	150	1.9%	280	3.5%	430	2.7%
	<i>archives</i>	514	6.5%	122	1.5%	636	4.0%
	<i>saa</i>	79	1.0%	376	4.8%	455	2.9%
	Sub-total	6683	84.6%	4222	53.5%	10905	69.0%
.edu	<i>ci.fsu</i>	32	0.4%	259	3.3%	291	1.8%
	<i>slis.cua</i>	11	0.1%	205	2.6%	216	1.4%
	<i>slis.indiana</i>	93	1.2%	800	10.1%	893	5.7%
	<i>slis.kent</i>	118	1.5%	381	4.8%	499	3.2%
	<i>si.umich</i>	111	1.4%	63	0.8%	174	1.1%
	<i>is.pitt</i>	49	0.6%	144	1.8%	193	1.2%
	<i>scils.rutgers</i>	84	1.1%	137	1.7%	221	1.4%
	<i>lis.simmons</i>	12	0.2%	235	3.0%	247	1.6%
	<i>is.utexas</i>	106	1.3%	198	2.5%	304	1.9%
	<i>is.ucla</i>	87	1.1%	341	4.3%	428	2.7%
	<i>lis.uiuc</i>	140	1.8%	609	7.7%	749	4.7%
	<i>slis.unc</i>	294	3.7%	98	1.2%	392	2.5%
	<i>sis.utk</i>	13	0.2%	144	1.8%	157	1.0%
	<i>is.washington</i>	65	0.8%	62	0.8%	127	0.8%
Sub-total	1215	15.4%	3676	46.5%	4891	31.0%	
Total	Sum	7898	100.0%	7898	100.0%	15796	100.0%
	Mean	329.08	-	329.08	-	658.17	-
	SD	488.98	-	318.00	-	732.20	-

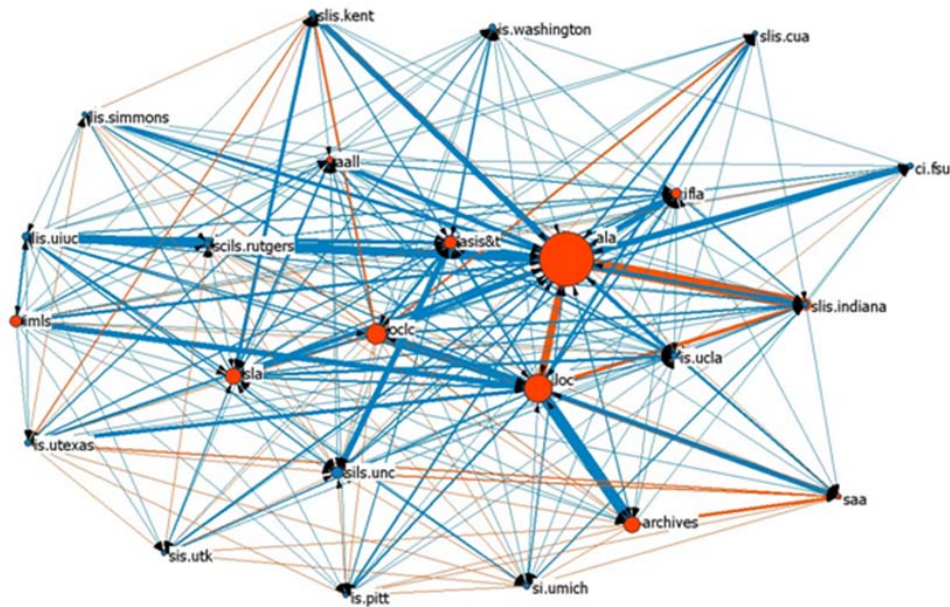


Figure 1: A linkage visualization of a Web structure of the LIS field in America

Figure 1 shows the visualization of the links among the twenty-four nodes. For the layout of nodes, the spring embedding algorithm, which provides an easily legible layout based on node repulsion concept, was applied with equal edge lengths. Since the spring embedding algorithm method considers distance as dissimilarity between nodes, we could interpret similarity among nodes from the output. In this figure, the red node indicates the organizational node and the blue node indicates the school node, and the size of node suggests the number of in-links. The arrow line indicates the link and its direction, and the width of the line reflects the frequency of links between two nodes.

The most notable finding was that most organizational nodes such as *ala*, *asis&t*, *oclc*, *loc* and *ifla* were located at the center of the space, whereas most school sites such as *slis.cua*, *ci.fsu*, *slis.kent*, *lis.simmons*, *sis.utk* and *is.pitt* were located at the periphery. A relatively distinct core/peripheral structure was apparent, where the core area consisted of organizational sites and the peripheral area consisted of school sites. Then the most prominent node was *ala* which had the largest size of node and the most frequent connections with other nodes. This suggests the ALA site serves as a core communicator in the LIS Web space. In addition, we can see the *ala* node sends and receives links to all other nodes, and the widths of those lines are relatively thicker. Furthermore, the four representative LIS related organizations, *ala*, *asis&t*, *oclc* and *loc*, located in close proximity to one another. Also, we can observe that *archives* and *saa*, which relate to the archives field, are located adjacently.

Table 3: Freeman's degree centrality measures

	(A)InDegree	(B)OutDegree	(C)NrmInDeg	(D)NrmOutDeg	(A) - (B)
<i>ala</i>	2156	1147	22.808	12.134	1009
<i>loc</i>	1247	1122	13.192	11.869	125
<i>oclc</i>	820	37	8.674	0.391	783

<i>sla</i>	540	190	5.712	2.01	350
<i>archives</i>	514	122	5.437	1.291	392
<i>ifla</i>	412	213	4.358	2.253	199
<i>asis&t</i>	405	691	4.284	7.31	-286
<i>imls</i>	360	44	3.808	0.465	316
<i>slis.unc</i>	294	98	3.11	1.037	196
<i>aall</i>	150	280	1.587	2.962	-130
<i>lis.uiuc</i>	140	609	1.481	6.442	-469
<i>slis.kent</i>	118	381	1.248	4.03	-263
<i>si.umich</i>	111	63	1.174	0.666	48
<i>is.utexas</i>	106	198	1.121	2.095	-92
<i>slis.indiana</i>	93	800	0.984	8.463	-707
<i>is.ucla</i>	87	341	0.92	3.607	-254
<i>scils.rutgers</i>	84	137	0.889	1.449	-53
<i>saa</i>	79	376	0.836	3.978	-297
<i>is.washington</i>	65	62	0.688	0.656	3
<i>is.pitt</i>	49	144	0.518	1.523	-95
<i>ci.fsu</i>	32	259	0.339	2.74	-227
<i>sis.utk</i>	13	144	0.138	1.523	-131
<i>lis.simmons</i>	12	235	0.127	2.486	-223
<i>slis.cua</i>	11	205	0.116	2.169	-194

To investigate the dependency patterns in the LIS Web space, the Freeman's centrality degree index values were calculated for each node. The higher in-degree centrality indicates a node gets higher attention by other nodes, whereas the higher out-degree centrality implies a node is more like to be dependent on other nodes in a Web space [1]. Table 3 shows the degree centrality in the LIS Web space. Regarding the in-degree centrality, the top eight nodes turned out to be the organizational sites. Similar to the outputs of the descriptive analyses, we see that the organizational sites receive higher numbers of in-links from other sites. Also, we find that most organizational nodes have higher in-degree centrality than out-degree centrality. Conversely, we find that most of the school sites have higher out-degree centrality than in-degree centrality. This pattern implies that the school sites are dependent

on other sites in the Web space. In particular, the *slis.indiana* node ranks third in out-degree centrality but fifteenth in the in-degree centrality, which indicates high dependency. From this degree centrality analysis, we are able to confirm the dependency propensity of school sites on organizational sites.

5. CONCLUSION

This study investigated the Web space of the LIS field using social network analysis techniques. The findings show that the organizational sites have more in- and out-links than school sites. The visualized network diagram of the LIS Web space imply most organizational sites are located near the center region, whereas most school sites are positioned in peripheral areas. Also the centrality analyses reveal that higher prominence of organizational sites than school sites and higher dependency propensity of school sites than organizational sites.

The findings of this study yield some insights in Web site evaluations, especially how to incorporate Web space structure in a specific field to evaluate the importance of Web sites in a specific field. How a specific Web site is situated in a Web space suggests a dynamic approach to evaluate the importance or influence of Web sites in relation to other sites. Also, the application of centrality measures such as degree, closeness and betweenness centrality utilized in social network analyses is useful for evaluating Web sites.

This study also has its limitations. First, even though the sample represents a selection of organizations and schools from the LIS field, the number of research objects could be enlarged to provide a more complete picture of LIS Web space. Second, the data did not represent the entirety of each site since the limitation of the crawling level. For more complete analysis, the depth of crawling could be deeper to reflect the full content of Web sites. Based on this pilot study, further research to implement Web space analysis with a larger dataset is anticipated.

6. REFERENCE

[1] S. Wasserman, and K. Faust. Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press, 1994.