

Mapping the Intellectual Structure of the Open Access Field Through Co-citation Analysis

Guleda Duzyol, Zehra Taskin and Yasar Tonta

Hacettepe University, Department of Information Management
06800 Beytepe, Ankara, Turkey
{gduzyol, ztaskin, tonta}@hacettepe.edu.tr

Abstract: Open access has been one of the major research trends and hottest topics in electronic publishing. This paper aims to assess the evolution of open access as a research field using bibliometric and scientific visualization techniques. It maps the intellectual structure of open access based on 281 articles that appeared in professional literature on the topic between 2000 and 2010. Using bibliometric and co-citation analyses, co-citation patterns of papers are visualized through a number of co-citation maps. CiteSpace was used to analyze and visualize co-citation maps. Maps show major areas of research, prominent articles, major knowledge producers and journals in the field of open access. The letter written by Steven Lawrence (“Free online availability substantially increases a paper’s impact”, 2001) appears to be the most prominent source as it was cited the most. The journal article by Kristin Antelman (“Do open Access articles have a greater research impact”, 2004) and the report by Alma Swan and Sheridan Brown (“Open access self-archiving: An author study”, 2005) are the second most highly cited papers in the network. *JASIS / JASIST* is the most frequently cited journal by the authors writing on open access. The most recent research topics appear to be institutional repositories, open access publishing/open access journals and scientific communication. Stevan Harnad is most frequently co-cited author, followed by Alma Swan, Steven Lawrence and Peter Suber. The preliminary findings show that open access is an emerging research field. Findings of this study can be used to identify landmark papers along with their impact in terms of providing different perspectives and engendering new research areas.

Keywords: open access, co-citation analysis, information visualization, mapping, CiteSpace

1. Introduction

The Internet has rapidly become a world-wide publishing platform, and open access to publications available through the Internet has been a hot topic for scholars, librarians and publishers over the last decade or so. According to the Budapest Open Access Initiative, open access provides free access to the full texts of articles, permits any users to read, download, copy, distribute, print, search or links to them, allows crawlers to index and pass them as data to software, and entitles anyone to use them for any other lawful purpose, without financial, legal, or technical barriers (Budapest Open Access Initiative, 2004).

Open access is an important development in the field of bibliometrics, too. Citation databases allow the literature to be navigated backwards and forwards in time, following citations to and from any article, guided also by co-citation analysis in order to find related papers. Citation analysis can be used to find emerging fields, to map the time-course and direction of research in progress, and to identify synergies between different disciplines. For current users it will at first be just a pleasant surprise to find that the citation links within an article can retrieve the full texts of the articles it cites; yet this is just one of the many rich scientometric possibilities that will be provided by open access (Brody, 2004).

Open access journals have established a new paradigm of scholarly communication and their scholarly impact has been discussed in the library and publishing communities (Zhang, 2006). In parallel with these discussions, the correlation between citations and open access has been a major subject of many studies (Mukherjee, 2009a; Mukherjee, 2009b; Craig, Plume, McVeigh, Pringle, Amin, 2007; Turk, 2008; Kousha, Thelwall, 2006; Eysenbach, 2006).

In conjunction with the developments in the field of open access, there have been many works in the literature about open access in recent years. These works can be used to find out the major knowledge producers, significant journals and prominent articles in this field. The main objective of this study is to evaluate the open access field using scientific visualization techniques and to answer the following research questions:

- What are the prominent articles in the open access field?
- Which authors are major knowledge producers?
- Which journals are cited the most frequently?
- Which keywords are used most often in the open access field?

2. Data and Methods

A topical search on the Web of Science (WoS) database using the term “open access” was performed to identify papers on open access that appeared in the literature between 2000 and 2010. Searches were carried out on July 6, 2010. A total of 281 journal articles were identified under the subject category of “Information Science & Library Science” (proceedings, book reviews, editorials, letters and other document types were excluded). The full bibliographic records including authors, titles, abstracts and reference lists for 281 articles were downloaded.

CiteSpace was used to produce co-citation networks. CiteSpace is a visualization tool developed by Chaomi Chen from Drexel University (<http://cluster.cis.drexel.edu/~cchen/citespace/>). CiteSpace facilitates the analysis of emerging trends in a knowledge domain. CiteSpace is a “knowledge domain visualization” tool that paints a picture of how science grows and evolves over time (Chen, 2004; Dell, 2004).

Co-citation analysis studies investigate the structures of scientific research based upon citations and co-citations. It enables us to identify the groups of scientists and their publications from which conclusions can be drawn about the inner structure of research disciplines, schools or paradigms (Small, 1980). There are many co-citation networks in the literature to identify co-citations and four of them (document co-citation network, author co-citation network, journal co-citation network and the network of keywords and noun phrases) were generated to analyze the open access field in this study. Document co-citation networks produce maps of highly cited and co-cited documents in clusters representing interrelated research specialties (Small, Sweeney, & Greenlee, 1985). Author co-citation networks provide insights about underlying schools of thought in scientific discourse (White and Griffith, 1981). Journal co-citation networks can help the researcher to study the structure of scholarly disciplines through the published literature (McCain, 1991). The network of keywords is the idea of the co-occurrence of keywords that describes the contents of documents in a file (Callon, Courtial & Laville, 1991).

3. Findings and Discussion

Table 1 provides descriptive statistics about 281 articles on open access that appeared in the literature between 2000 and 2010. These articles were cited 730 times during this period. On the average, 26 articles were published on open access annually ($SD = 22$). The number of papers has increased considerably over the years, although it has slowed down after 2007. The Budapest Open Access Initiative in 2002, the “Bethesda Statement of Open Access Publishing” and “Berlin Declaration of Open Access” announced in 2003 seem to have an impact on the number of publications on open access.

Table 1. Number of articles on open access (2000-2010)

Year	# of articles	# of times cited
2000	2	14
2001	0	0
2002	7	20
2003	11	48
2004	13	136
2005	27	108
2006	36	128
2007	62	148
2008	47	91
2009	55	33
2010	21	5
Total	281	730

Figure 1 shows a document co-citation network derived from the citing behavior of authors writing on open access. This network is the result of merging 11 one-year document co-citation networks generated by the WoS dataset (2000-2010). The document co-citation network consists of 245 papers (not only articles, all type of papers – letters, editorials, conference proceedings etc.) that have been cited by 281 open access articles in our dataset and there are 1337 co-citation links between these 245 papers on the network. Citations between 2000 and 2003 are shown in blue rings, 2004 and 2007 in green rings and 2008 and 2010 in yellow and orange rings.

The colors of co-citation links represent the first year the connection between the two documents was made (Chen, Song, Yuan, & Zhang, 2008). Colors of the network show that studies on open access have mainly started in the year 2002 (light blue colors on the left-hand part of the network) as there is no dark blue colors on the map representing the time slices of the years 2000 and 2001.

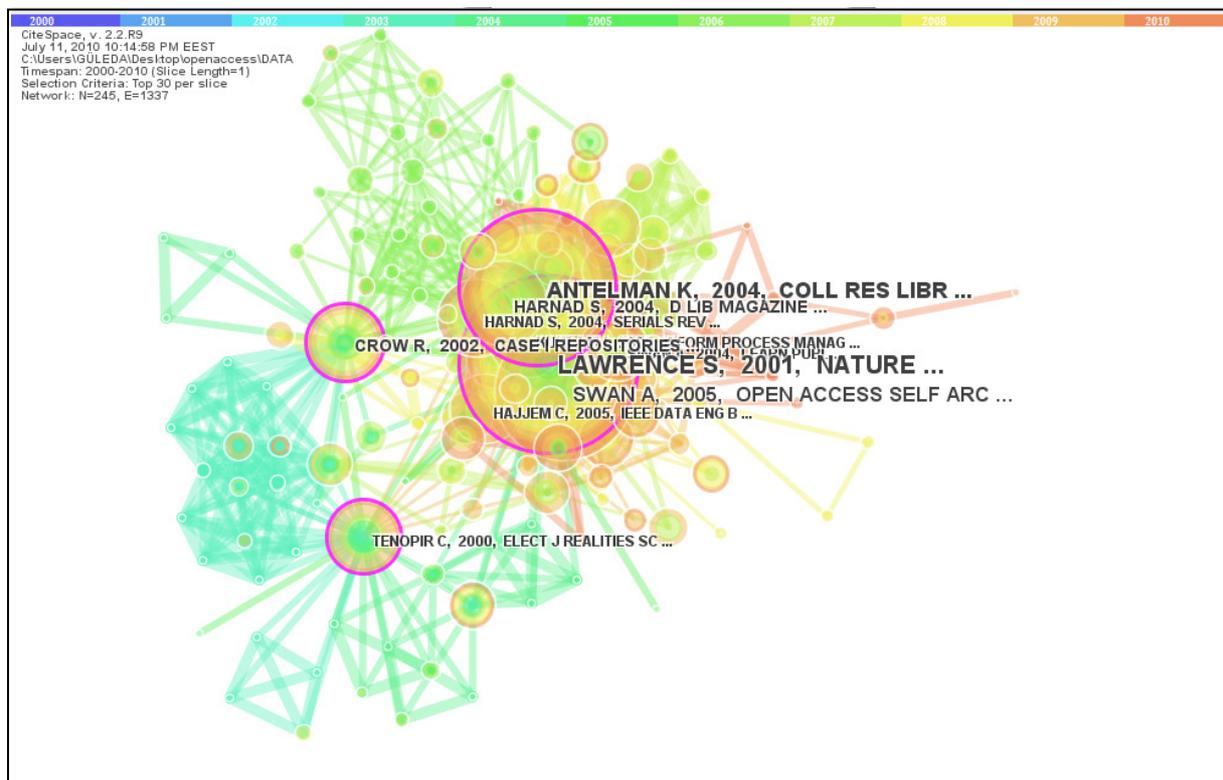


Figure 1. Document co-citation network of open access, 2000-2010

Structurally strategic papers that are most frequently cited by 281 articles on open access can easily be identified in Figure 1. The letter written by Steven Lawrence (“Free online availability substantially increases a paper’s impact”, 2001) appears to be the most prominent source as it was cited the most. The journal article by Kristin Antelman (“Do open Access articles have a greater research impact”, 2004) and the report by Alma Swan and Sheridan Brown (“Open access self-archiving: An author study”, 2005) are the following most highly cited papers in the network. Papers by Stevan Harnad (“The access/impact problem and the green and gold roads to open access”, 2004; “Comparing the impact of open access (OA) vs. non-OA articles in the same journals”, 2004), Raym Crow (“The case for institutional repositories: a SPARC position paper”, 2002), Michael J. Kurtz (“The effect of use and access on citations”, 2005), Chawki Hajjem (“Ten-year cross-disciplinary comparison of the growth of open access and how it increases research citation impact”, 2005) and the book written by Carol Tenopir and Donald King (“Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers”, 2000) are also seen on the document co-citation networks to be the prominent works in the open access field. There appears to be a couple of years’ worth of time lag between publication and citation years of papers, which can easily be followed by different colors of the publication year and of the inner ring that represents the year a paper was first cited. For example, Lawrence’s letter was published in 2001 but it started generating citations after 2004. Similarly, Antelman’s article was published in 2004 but started to be cited after 2006. These prominent papers continue to be cited today, as the outer rings in orange on the map indicate. This further reinforces the fact that open access is still an evolving field.

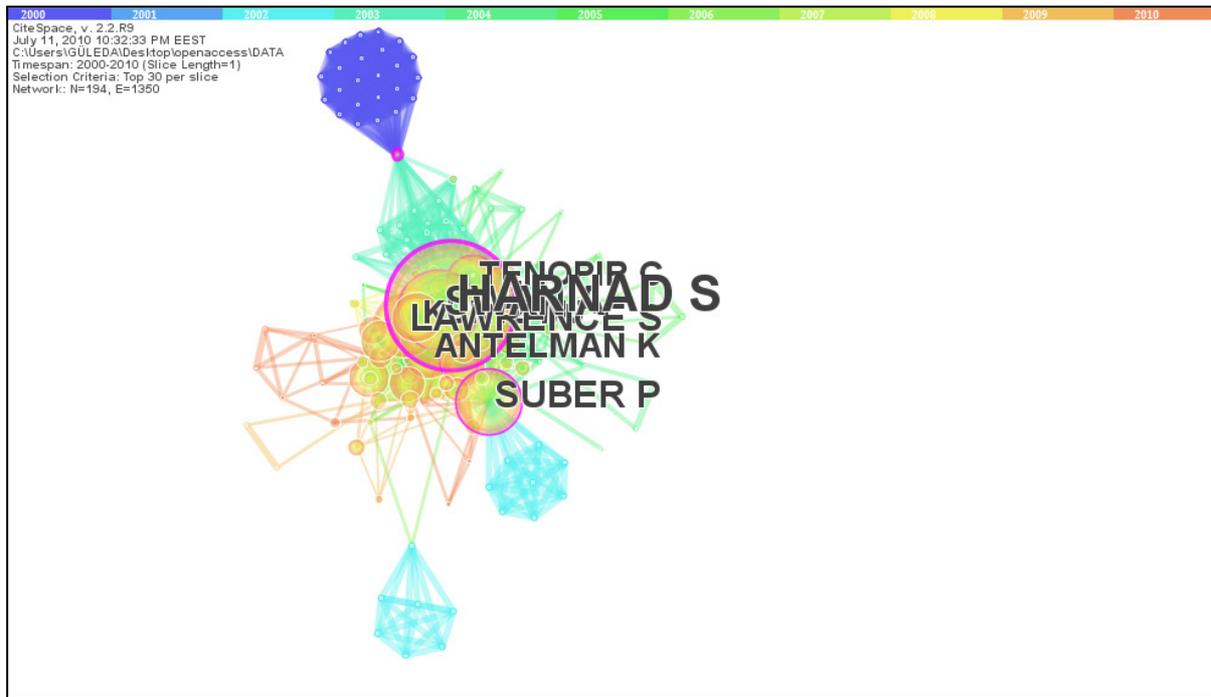


Figure 2. Author co-citation network of open access, 2000-2010

Figure 2 shows the results of the author co-citation analysis that consists of authors contributing to the open access literature. The author co-citation network contains the most frequently cited 194 authors (cited by 281 articles in our data set) and 1350 co-citation links between them. On an author co-citation map, the size of a node is proportional to the number of open access articles the author on that node has published. In Figure 2, Steven Harnad has the largest circle. The colors of nodes provide information about the citation patterns of an author, i.e., the year(s) he/she was cited the most/the least. The nodes of authors with purple rings around them (Steven Harnad, Peter Suber, Carol Tenopir, for example) are strategically important in pulling other nodes together; they have the highest betweenness centrality which is an indicator of a node's ability to make connections to other nodes in a network (Chen et al., 2008; Sentinel Visualizer, 2009). Most of the citations to the prominent articles were made after 2005. The colors of the outermost ring of almost all the authors are orange, indicating that many of these authors continue to publish papers on open access which continue to be cited.

Figure 3 shows the journal co-citation network of the most frequently cited journals by the 281 open access articles. The journal co-citation network consists of 170 journals along with 1395 co-citation links among them. The *JASIS / JASIST* (Journal of the American Society for Information Science and Technology) is the most highly cited journal by the authors of open access papers. *Learned Publishing*, *Nature* and *D-Lib Magazine* follow. The journal *Learned Publishing* has the highest centrality ratio and papers appeared in it have been cited since 2004 by the authors of open access articles. The ring colors of the highly cited journals are heavily yellow and orange, indicating that these journals started to get cited in the last few years and continue to be cited.



Figure 3. Journal co-citation network of open access, 2000-2010

Figure 4 shows a hybrid network of keywords as circles with black labels and noun phrases as triangles. Keywords and noun phrases were extracted from titles and abstracts of papers. The noun phrase of “open access” is a pivotal node that has a purple ring. “Open access” also is the most frequently used keyword in open access papers. The most recent research topics appear to be institutional repositories, open access publishing/open access journals and scientific communication. The lower part of the Figure 4 (not shown) has keywords and noun phrases on electronic publishing that can be seen on the map when threshold values are reduced, which show that open access has evolved as a sub-topic of electronic publishing.

- Mukherjee, B. (2009a). Do open-access journals in library and information science have any scholarly impact?: a bibliometric study of selected open-access journals using Google Scholar. *Journal of the American Society for Information Science and Technology*, 60(3): 581-594
- Mukherjee, B. (2009b). Scholarly research in LIS open access electronic journals: a bibliometric study. *Scientometrics*, 80(1): 167-194
- Sentinel Visualizer. (2009). Social network analysis (SNA). Retrieved July 27, 2010 from <http://www.fmsasg.com/SocialNetworkAnalysis/>
- Small, H. (1980). Co-citation context analysis and the structure of paradigms. *Journal of Documentation*, 36(3), 183-196
- Turk, N. (2008). Citation impact of open access journals. *New Library World*, 109(1-2): 65-74
- Small, H., Sweeney, E. & Greenlee, E. (1985). Clustering the science citation index using co-citations: II. mapping science. *Scientometrics*, 8(5-6): 321-340
- White, H.D. & Griffith, B.C. (1982). Authors as markers of intellectual space: co-citation in studies of science technology and society. *Journal of Documentation*, 38, 255-272.
- Zhang, Y. (2006). The effect of open access on citation impact: a comparison study based on web citation analysis. *Libri*, 56, 145-156.

Appendix: Papers depicted in the network clusters and mentioned in the text

- Antelman, K. (2004). Do open access articles have a greater research impact? *College and Research Libraries*, 65 (5): 372-382.
- Brody, T. and Harnad, S. (2004) Comparing the impact of open access (OA) vs. non-OA articles in the same journals. *D-Lib Magazine*, 10 (6).
- Hajjem, C., Harnad, S. and Gingras, Y. (2005). Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. *IEEE Data Engineering Bulletin*, 28(4): 39-47.
- Crow, R. (2002). The case for institutional repositories: a SPARC position paper.
- Harnad, S., Brody, T., Vallières, F., Carr, L., Hitchcock, S., Gingras, Y., Oppenheim, C., Stamerjohanns, H. and Hilf, E. R. (2004). The access/impact problem and the green and gold roads to open access. *Serials Review*, 30 (4): 310-314.
- Kurtz, M. J., Eichhorn, G., Accomazzi, A. Grant, C., Demleitner, M. Henneken, E. and Murray, S. S. (2005). The effect of use and access on citations. *Information Processing and Management*. 41(6): 1395-1402.
- Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature*, 411, p. 521
- Swan, A. and Brown, S. (2005) Open access self-archiving: An author study. Technical Report , External Collaborators, Key Perspectives Inc..
- Tenopir, C. and King, D. W. (2000). *Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers.* Special Libraries Assoc., Washington DC