

Mapping the Intellectual Structure of Open Access Field Through Co-citations

Guleda Duzyol, Zehra Taskin and Yasar Tonta

Hacettepe University, Department of Information Management, Beytepe, Ankara, Turkey

Abstract

Open access is 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 2000-2010. Using bibliometric and co-citation analyses, co-citation patterns of papers are visualized through a number of co-citation maps. CiteSpace is used to analyze and visualize co-citation maps. Maps show major areas, prominent articles, major knowledge producers and journals in 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 following most highly cited papers in the network. "JASIS/JASIST" is the most 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. Rob Stevan Harnad is mostly co-cited author, Alma Swan, Steven Lawrence and Peter Suber follows. 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, information visualization, mapping, Cite Space

1. Introduction

The internet has rapidly become a world-wide publishing platform, and open access to these publications has been a hot topic for scholars, librarians and publishers over last few years. According to Budapest Open Access Initiative, open access provides free availability on the public internet, permits any users to read, download, copy, distribute, print, search or links to the full texts of articles, crawls them for indexing, passes them as data to software, or uses them for any other lawful purpose, without financial, legal, or technical barriers (Budapest Open Access Initiative, 2004).

Open access is an important development 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 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 argued in the library and publishing communities (Zhang, 2006). In parallel with these arguments, the correlation between citations and open access has been major subject in 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 open access field, there have been many works in literature about open access in recent years. It creates major knowledge producers, significant journals and prominent articles in this area. From this point forth, the main aim of this study is to evaluate open access field using scientific visualization techniques. This study attempts 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 the most cited?
- Which keywords are used mostly in the open access field?

2. Data and Methods

A topical search on Web of Science (WoS) database using the term “open access” was performed to identify papers on open access that appeared in the literature after 2000 (2000–2010) and a total of 281 journal articles under the

subject category of “Information Science & Library Science” were identified (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 which can be called as “knowledge domain visualization” aims to create a picture of how science grows and evolves over time (Chen, 2004; Dell, 2004).

Four co-citation networks (document co-citation network, author co-citation network, journal co-citation network, network of keywords and noun phrases) were generated to analyze open access field.

3. Findings and Discussion

Table 1 provides descriptive statistics about 281 articles on open access that appeared in the literature between 2000 and 2010. During this period these articles were cited 730 times. On the average, 26 articles published annually (SD = 22). While the number of papers was very few at the beginnings of 2000s, they have increased considerably then. The increase has slowed down after 2007.

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

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

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 the 245 papers on the network. Citations made in 2000-2003 are shown in blue rings, 2004-2007 in green rings and 2008-2010 in yellow and orange. The colors of co-citation links represents the first year the connection between two documents was made (Chen, C., Song, I. Y., Yuan, X. and Zhang, J., 2008). Colors of the network depict that studies on open access has mainly started in 2002 (light blue colors on the left-hand part of the network), there is not dark blue colors (colors of 2000 and 2001 time slices) on the map.

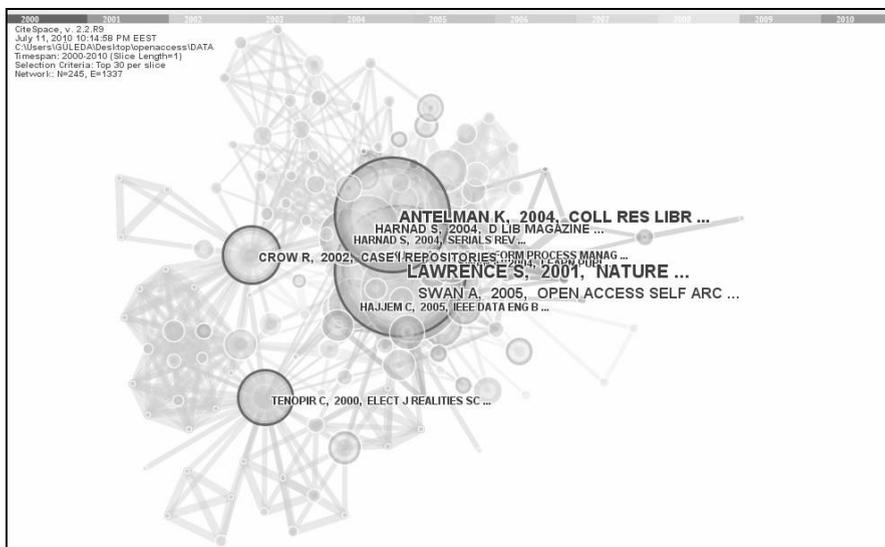


Figure 1. Document co-citation network of open access, 2000-2010

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 Carol Tenopir’s book (“Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers”, 2000) are also seen on the document co-citation networks to be the prominent articles of open access field. These prominent articles were not started to get cited soon after their publication, they have started to get cited about 2-3 years after their publication as the inner ring color is not same with the color of publication year. For example Lawrence’s letter was published in 2001 but started to be cited after 2004, similarly Antelman’s article that published in 2004 started to be cited after 2006. These prominent papers seen on the network still continue to be cited today, as the outer orange rings indicate. This indicates that open access is an evolving field and the studies on open access continue.

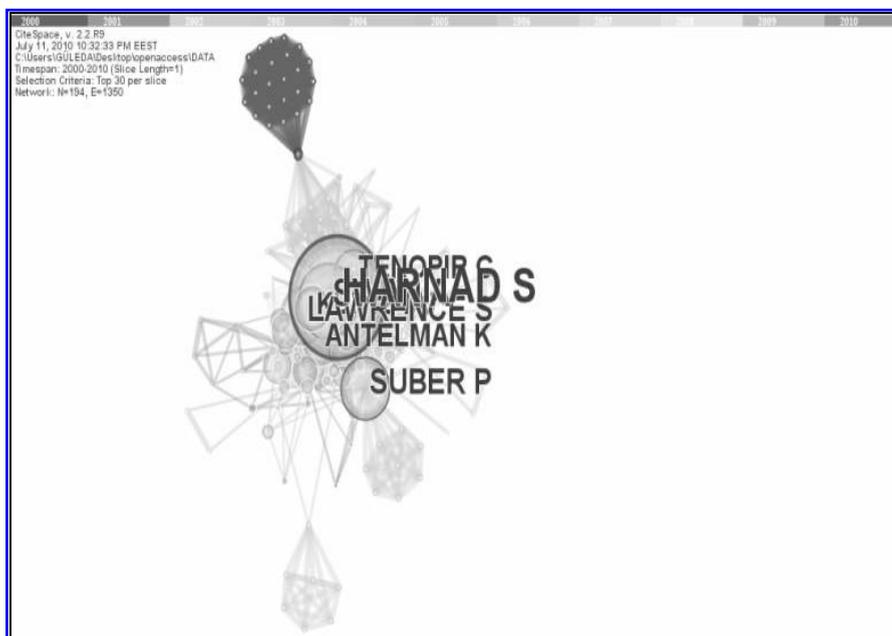


Figure 2. Author co-citation network of open access, 2000-2010

Figure 2 shows the results of author co-citation analysis that consists of authors contributing to open access literature. Author co-citation network contains 194 mostly cited authors by 281 articles in our data set and 1350 co-citation links between these mostly cited 194 authors. On an author co-citation map, the size of a node is proportional to the number of open access articles the writer on that node has published. In Figure 2, Steven Harnad has the largest

citation circle. The colors of nodes give information about the citation patterns of an author, in which years years he/she was cited most/least for example. 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 (Chen et al., 2008). Most of the citations to the prominent articles were made after 2005. The colour of outermost ring of almost all the authors are orange, indicate that many of these authors continue to publish on open access that they continue to be cited.

Mostly cited journals by the 281 open access articles can be seen on the journal co-citation network (Figure 3). Figure 3 consists of 170 journals along with 1395 co-citation links among them. The “JASIS/JASIS” is the most highly cited journal by the writers of open-access. Learned Publishing, Nature and D-Lib Magazine follows. The journal “Learned Publishing” has the highest centrality and have been cited since 2004 in open access articles. The ring color of the mostly cited journals are heavily yellow and orange, indicates that these journal started to get cited in last 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. “Open access” noun phrase is a pivotal node that has a purple ring. “Open access” is also mostly used keyword in open access

References

- Brody, T. (2004). Citation analyses in the open access world. *Interactive Media International*. Retrieved 25 July, 2010 from http://eprints.ecs.soton.ac.uk/10000/1/tim_oa.pdf
- Budapest Open Access Initiative. (2004). Guide to business planning for converting a subscription-based Journal to open access. Retrieved July 25, 2010 from http://www.soros.org/openaccess/oajguides/business_converting.pdf
- Chen, C., Song, I. Y., Yuan, X. and Zhang, J. (2008). The thematic and citation landscape of Data and Knowledge Engineering (1985-2007). *Data and Knowledge Engineering*, 67, 234-259.
- Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences of the USA* (PNAS), 101 (Suppl.(1)), 5303-5310. Retrieved July, 13, 2010 from <http://www.pnas.org/content/101/suppl.1/5303.full.pdf>
- Craig, I., Plume, A., McVeigh, M., Pringle, J., Amin, M. (2007). Do open access articles have greater citation impact?: critical review of the literature. *Journal of Informetrics*, 1(3): 239-248
- Dell, H. (2004). Mapping intellectual milestones. BioMedNet: Special Report. Retrieved July 13, 2010 from <http://cluster.cis.drexel.edu/~echen/citespace/doc/biomednet.pdf>
- Eysenbach, G. (2006). Citation advantage of open access articles. *Plos Biology*, 4(5): 692-698
- Kousha, K., Thelwall, M. (2006). Motivations for URL citations to open access library and information science articles. *Scientometrics*, 68(3): 501-517
- 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
- Turk, N. (2008). Citation impact of open access journals. *New Library World*, 109(1-2): 65-74
- 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.
- Lawrance, 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)

