Scientometrics (2020) 125:2617–2635
https://doi.org/10.1007/s11192-020-03682-4

Intellectual structure evolution of open access research observed through correlation index of keyword centrality

Jane Cho[1]

Received: 19 May 2020 / Published online: 28 August 2020

Abstract

This study captured intellectual structures of open access by time frame using the pathfinder keyword network analysis method. 1998 papers published on Web of Science from 2005 to 2019 were divided into 3-year units, and keyword pathfinder networks were analyzed in five time segments. Thus, this study examined the time series changes of intellectual structure and keyword centrality. In addition, by analyzing the correlation index of keyword centrality between time segments, this study examined how long the similarities of the intellectual structure persisted and how it has changed. As a result, a weak correlation ($r=0.10\sim r=0.49$) was obtained from the observations in 2005 for 9 years; however, the correlation decreased sharply since 2014 ($r=-0.06\sim r=0.00$). New research topics have emerged that have not been highlighted in centrality, such as article processing charge, altmetrics, and research data. The scope of research has changed as subjects such as document delivery that showed high centrality initially, disappeared.

Keywords Intellectual structure · Open access · Pathfinder network analysis · Web of science · Time series changes

Introduction

Since the declaration of the Budapest Open Access Initiative in February 2002, various open access (OA) policies have been implemented in each country. In the United States, United Kingdom, Australia, etc., the obligations that allow public access to public funded research have been actively implemented. In the journal publishing market, many publishers are actively supporting the green OA policy by allowing self-archiving by authors for pre- or post-print. In addition, there is an increasing number of hybrid journals based on the Article Processing Charge (APC) policy of major research support agencies in Europe. As a result, according to the data released by the Open Access Scholarly Publishers Association (OASPA) in July 2019, the number of OA articles published under the CC-BY license in 2018 was 261,621, up 23% from the previous year. As the number of OA papers

✉ Jane Cho
chojane123@naver.com; chojane@inu.ac.kr

[1] Department of Library and Information Science, Incheon National University, 119 Academy-ro, Songdo 1(il)-dong, Yeonsu-gu, Incheon, Republic of Korea

increased, libraries in each country began to insist on a reasonable journal subscription price; in addition, the model of a transformative agreement, which replaces the subscription cost with the editing service required for publishing open access content, is being promoted. Furthermore, with the announcement of PlanS (www.coalition-s.org) to realize complete and immediate open access to publicly funded research results, the open access movement is taking a new step.

As the movement progresses, the number of researches related to this topic is increasing, and research fronts are expanding. More and more studies are being done on the impact of OA papers, altmetrics, and predatory journals, one of the side effects of open access. With this background, this study aims to find out what intellectual structure the OA-related global research shows, how long has it been maintained and at what point the intellectual structure changes using pathfinder keywords network analysis. Keyword network analysis refers to a network that is constructed by extracting keywords from the literature and calculating similarity between keywords through the co-occurrence frequency of each keyword pair. It is an important analytical method of informetrics applying the same concepts as co-word analysis. This study intends to perform five keyword pathfinder network analysis by classifying papers published in international academic journals listed on WOS (Web of Science) from 2005 to 2019 in 3-year segments. Track changes in intellectual structure based centrality changes of keywords. Although many studies have been conducted to identify the intellectual structure on OA (Bailey 2010; Zhao and Wu 2014; Suh and Jeong 2013), few studies have explored how long the intellectual structure has been maintained and how it has changed. This study tries to explore it using a methodology proposed by Chen et al. (2019), which observed changes in intellectual structure through observation of the correlation index of keyword centrality between time gaps.

Research background

Open access development

Around 2000, efforts were made to promote open access, such as the establishment of the public science library PLOS (2000) and the SPARC Europe (1998). Since the mid-2000s, specific policies for realizing open access were announced through the Federal Research Public Access Act and the National Institutes of Health Public Access Policy of the United States. The Office of Science and Technology Policy (2013) mandated open access to results of research conducted with a federal research fund of $ 100 million or more (Office of Science and Technology Policy 2013). At the same time, the UK stipulated that all academic achievements obtained from Research Councils UK (RCUK) funds should be made public within 6–12 months of publication. Unlike the United States, which seeks to realize open access indirectly with a focus on green OA, the European Union has a definite OA policy direction for governments and research funding institutions (McGlashan and Hadley 2019). The European Union announced the principle of PlanS in November 2018 at cOAlition, an initiative to realize full and immediate open access to publicly funded research outcomes, which is still supported by the library community (Cho 2020).

With these developments, the volume of OA publishing has increased, and the method has diversified. An OA method, in addition to the gold method of publishing in OA journals listed in the Directory of Open Access Journals (DOAJ), is a green OA method that involves publication in paid journals but can be opened through OA archives. There is also a hybrid method, in which the publication is in a paid journal, but can be released by an APC (Article Processing Charge). There are also bronze methods that make articles open but do not permit re-use without a license. The number of hybrid OAs has increased significantly since European research sponsors have implemented APC support policies. However, OA papers shared by bronze methods and academic social networks (ASNs) have also increased significantly. As the volume of freely accessible papers has increased in recent years, there is an increase in the number of cases in which the library replaces the subscription cost with the editing service required for OA publishing while claiming to cancel the journal subscription or demand a reasonable price. Recently, the concept of a hybrid journal offset or transformative agreements supporting open access based on the Publish & Read model has evolved. As of 2019, more than 60 transformative agreements have been registered in the registry of Efficiency and Standards for Article Charges (ESAC), and transformative agreements negotiations are underway among 19 publishers in 14 countries (OA2020 2019).

Prior research review

As described above, attempts have been made to analyze the progress being made in open access and to classify or systemize related research areas. Bailey (2010) organized the research subject of open access based on books and journal articles published between 1999 and 2010 through *Open Access: A Bibliography*. Zhao and Wu (2014) searched for major OA research topics in academic journals published in China by conducting co-word and social network analyses. Miguel et al. (2016) also identified major sub-themes related to open access by analyzing the occurrence of specific words in 1179 articles. Similarly, Suh and Jeong (2013) analyzed the intellectual structure of open access based on the analysis of co-words. They analyzed the intellectual structure by extracting keywords from research related to open access in international journals published between 1998 and 2012 and found that there were 18 subject areas related to institutional repositories, OA journals, and academic publishing. Cho (2014) performed keyword network analysis on institutional repositories, which are a part of the basic infrastructure for open access. As a result, four clusters were formed as open access, academic journals, libraries, and services. Lastly, Rodrigues et al. (2016) conducted content analysis on OA papers registered in SCOPUS between 2001 and 2015. Research on open access has been published in earnest since 2005. The focus of the initial study was the scientific community's perception and attitude regarding OA and the citation performance of OA papers. In the 2010s, it was observed that papers on OA economics and publishing market appeared. In 2015, discussions were focused on legal and ethical issues such as copyrights and licenses and quality control. In addition, they classified OA research papers like follow: OA awareness and status/citation and impact measurement of OA papers/economic issues such as APC/technology and systems/quality control and visibility/law and ethics/philosophy and values.

Meanwhile, the studies that analyzed changes in intellectual structure were also conducted as follows. Choi et al. (2018) and Kim (2014) analyzed the changes in intellectual structure in the field of Library and Information Science and Sports Sociology, respectively, by dividing the period and observing changes in the ranking of keyword centrality. Cho (2018) also observed changes in intellectual structure through changes in the ranking of centrality of library-related keywords reported in the media using a similar method. The precedence studies described above have identified changes in the intellectual structure just through observing changes in the ranking of the centrality of keywords. However such analysis is difficult to explain the similarity and dissimilarity of the multi- intellectual structure. Recently, Chen et al. (2019) published a study observing changes in intellectual structure through the keywords correlation index between multi-sections by setting a time gap. By analyzing how the centrality index of keywords calculated in each step is related to each other through correlation coefficients, the similarity and dissimilarity of the networks were explained. A study that analyzed the similarity of the intellectual structure by analyzing the correlation of the keyword centrality index was found in Cho (2018) in addition to Chen et al. (2019). And then, there have been many other studies that use centrality to distinguish multiple network similarities. Szell et al. (2010) used the similarity coefficient to determine the relevance of the multi-network, Bae (2015) also performed a Pearson correlation analysis using the centrality to measure the relevance of input and performance networks in convergence research.

In summary, this study is similar to previous studies (Zhao and Wu 2014; Miguel et al. 2016; Suh and Jeong 2013) in that it observed the centrality of keywords through keyword network analysis in the open access research field. However, there is a difference in that the analysis of the keyword correlation index between multiple time sections was performed to explain the similarity and dissimilarity of the intellectual structure, and the change in the intellectual structure was identified based on this.

Pathfinder keyword network analysis

Scholars from many disciplines are exploring intellectual structures based on keywords or citation information. Intellectual structures have been analyzed in various fields, such as informatics (Ding et al. 1999: Suh and Jeong 2013: Park and Jeong 2013), medicine (Heo and Song 2013), economics and management (Alan and Jack 2008; Kwak and Chung 2012). Those studies used the author co-citation (Kwak and Chung 2012), author bibliographic coupling analysis (Park and Jeong 2013), and co-words analysis (Heo and Song 2013; Chen et al. 2016). In particular, co-words analysis has proven to be an effective technique that reveals the relationship between topics through relationships between keywords (Ronda-Pupo and Guerras-Martin 2012; Rokaya et al. 2008).

In the intellectual structure analysis, the pathfinder network analysis which reflects the weight of co-occurrence, is mainly used. The pathfinder network refers to a network created by removing a path that violates a triangular inequality in a state where all weighted links are generated. When a triangular inequality is violated, it is a case where a path indirectly connected through several short links exists rather than a long path directly connected (Lee 2006a). Social relationships are generally based on binary networks. However, in the field of bibliometrics, weighted network analysis with distinct strengths is more useful. This is because a weighted relationship based on the number of co-occurrence is important in co-citation or co-word analysis (Lee 2013). The Pathfinder Network started with cognitive psychology, but was widely known by White (2003) and Chen (2004). The technique was spread while developing open software that can analyze the bibliographic data (Lee 2006a) and has been used in a number of studies in bibliometrics (Cho 2019; Choi et al. 2018; Suh and Jeong 2013). PFNet is considered to be more advantageous in expressing the detailed structure as well as expressing the overall structure, such as multidimensional scaling or cluster analysis, which is a traditional methodology for analyzing the intellectual structure in the academic field (Lee 2006a). Meanwhile, clusters on a pathfinder network can be effectively represented through a parallel nearest neighbor clustering (PNNC) technique. PFNet guarantees the survival of only the shortest link among the links connected to each node. PNNC works effectively on PFNet because each node connects the nearest neighbors to form a cluster step by step (Lee 2006b).

In the analysis of general social networks, the degree centrality, closeness centrality, and betweenness centrality are commonly used. These centralities do not take into account the link weights between keywords. Therefore, citation analysis commonly uses the nearest neighbor centrality, mean association, mean profile association, and triangular betweenness centrality (which indicates the degree of binding that one node mediates to other nodes) (Lee 2006c).

Research method

As shown in Table 1, this study analyzed 1998 research papers from 2005 in the field of Information Science & Library Science that were searched with term of "open access" in WOS. The searched fields are titles, abstracts, keywords, etc. and retrieved in December 2019. Three keywords were extracted in order to get a total of 3871 keywords. Data was received by metadata export function of WOS.

First, keywords for analysis were grouped every 3 years and divided into 5 year units from 2005, as shown in Table 1. The OA declaration was made in the early 2000s, but the relevant research results were published in earnest from 2005 onwards (Rodrigues et al. 2016). The keywords were extracted from each section of the papers published, and the frequency was calculated to derive the centrality of the keywords that appeared more than 4 times. The centrality of the nodes is the triangle betweenness centrality, which is derived from the results of the pathfinder network analysis. As describe earlier, pathfinder network is created by removing paths that violate triangular inequalities with all weighted links created. The COOC program was used for the co-occurrence matrix for network analysis, and the centrality was calculated using WNET software based on the cosine values calculated between words. The COOC and WNET software used here are both pathfinder network analysis tools developed by Lee (https://cafe.daum.net/wnets). Second, clusters were formed using PNNC based on the extracted keywords. Clusters at each stage were constructed using keywords extracted from each of the five sections, and network and clusters were visualized using Nodexl (https://www.smrfoundation.org/nodexl/). Third, by analyzing the correlation between time segments using the centrality index of the keyword, this study observed how long the formed intellectual structure has been maintained and the extent to which it has changed in a specific period. The method was applied to the methodology of Chen et al. (2019), which set a time gap and observed changes in intellectual structure through changes in the correlation index of keyword centrality between

Table 1 Data to be analyzed

Time segment		Number of documents	Number of keywords
W1	2005.1–2007.12	233	214
W2	2008.1–2010.12	278	455
W3	2011.1–2013.12	323	568
W4	2014.1–2016.12	626	1401
W5	2017.1–2019.1	538	1233
	Sum	1998	3871

time segments. To observe changes in intellectual structure through changes in the ranking of keywords' centrality or frequency is common, but such analysis is difficult to explain the similarity and dissimilarity of the multi-networks. Correlation analysis is a technique used to grasp the relationship between two variables. It has a correlation coefficient indicating the direction and intensity of the relationship between the two variables. Usually 0.0–0.1 is almost irrelevant 0.1–0.2 means weak positive correlation, 0.2–0.4 means normal positive correlation, and from 0.4, it means relatively strong positive correlation. There are a number of studies explaining the similarity through correlation coefficients of centrality in different networks (Bae 2015; Cho 2018; Chen et al. 2019; Szell et al 2010). In this study correlation analysis was performed by pairing each segment from W1 to W5. The change in the calculated correlation index value was observed to identify a point where the correlation may have rapidly dropped. This enabled us to estimate how long the intellectual structure was maintained and to determine the period when the change began on the research front. The Spearman correlation coefficient was used as the correlation index, and SPSS 25 was used for statistical processing.

Analysis result

Intellectual structure analysis results by time segment

The results of analyzing the intellectual structure of OA research in each time segment are described in this section. First, let's take a look at the intellectual structure formed in W1 (2005.1–2007.12), the stage in which we analyzed 233 documents. This period was followed by the announcement of policies supporting OA by research institutes, universities, and academic associations in each country. In the academic communities, research on the concept and philosophy of open access (Yiotis 2005) began, but the concept of open access had not yet been established as an independent research topic. As shown in Table 2, keywords with high centrality in the first period were *electronic publishing* (0.50), *document delivery* (0.47), *world wide web* (0.42) and *digital library* (0.36). For reference, when looking at the frequency of occurrence of keywords, the *digital library* showed the highest frequency of 11 times. In addition, two clusters were formed, an Archive cluster (C1) containing keywords such as *digital storage, archives management, digital libraries, and electronic publishing,* and an Information Resource Sharing cluster (C2) containing keywords such as *document delivery, inter-lending,* and *information resource sharing*. Note that we subjectively assigned the cluster name considering the centrality and frequency of keywords belonging to the cluster. Through this, it is presumed that the concept of open access at that time was discussed as an extension of digital preservation (Park 2007) or was dealt with in relation to the sharing of information resources (Boukacem-Zeghmouri and Schöpfel 2006), one of the major research topics at that time..

W2 (2008.1–2010.12) segment which 278 papers were collected, was the time when the depositing of the granted research papers in PubMed Central was obliged by the Public Access Policy of NIH. In addition, with the creation of the European Commission's OpenAIRE, Confederation of Open Access Repositories, and other repositories, green OA practice was activated. As shown in Table 3, *open access (22)* had a higher frequency of appearance than before and the keywords showing high centrality at this time are *bibliometric* (0.60), *academic library* (0.56), *repository* (0.53). However, the institutional repositories had not yet formed a major subject cluster in the study. Looking at the clusters formed

Table 2 Main keyword centrality, cluster and network map of W1 (2005–2007.12)

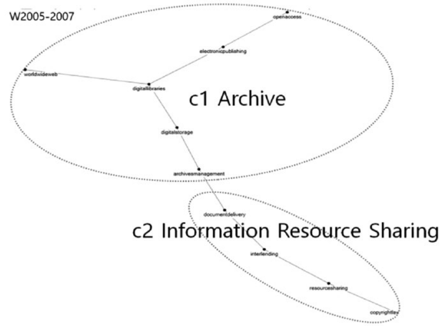
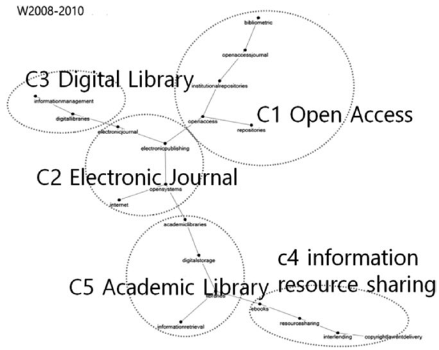
Node	Keyword frequency	Centrality rTBC(0–1)	Cluster		PathFinder Network
Electronic publishing	7	0.50	c1	Archive	
Open access	5	0.50	c2	Information resource sharing	
Document delivery	8	0.47			
World wide web	4	0.42			
Digital libraries	11	0.36			

Table 3 Main keyword centrality, cluster and network map of W2 (2008.1–2010.12)

Node	Keyword frequency	Centrality rTBC(0–1)	Cluster		PathFinder Network
Bibliometric	4	0.60	c1	Open access	
Open access	22	0.57	c2	Electronic journal	
Academic libraries	6	0.56	c3	Digital library	
Repositories	5	0.53	c4	Information resource sharing	
World wide web	2	0.50	c5	Academic library	

during this period, Open Access (C1), Electronic Journal (C2), Digital Library (C3), Information Resource Sharing (C4), And Academic Library (C5) were formed. Till this time, traditional research topics such as the library's resource sharing and digital library could be seen as major clusters.

In the W3 stage (2011.1–2014.12), 323 papers were collected; during this period, the OA policy of publicly funded research achievements was specifically implemented. In the US, federal funding policy was implemented, and in UK, all articles funded from RCUK were mandated for open access within 6–12 months of publication. In addition, as the trend of big deal contracts by academic journals has grown and the number of library movements against increasing subscription prices has increased, reflexive illegal OA papers sites such as Sci-Hub have emerged. Meanwhile, the open data policy was implemented in various countries as discussions on public data openness and sharing continued throughout the world, triggered by the Open Data Charter (Department For Business, Innovation & Skills Prime Minister's Office 2013) agreed at the G8 Science Ministers' Meeting. Looking at the intellectual structure during this period, it can be seen that the concept of information resource sharing among libraries was greatly reduced, and many research papers were published based on a new set of topics such as institutional repositories and open data. As shown in Table 4, the centrality and k**eyword frequency** of *open access* was very high at 0.86/61, and the *journal* and *institutional repositories* showed high centralities of 0.61 and 0.45, respectively. As the increasing subscription prices in the Elsevier journal became an issue, the centrality of the keyword *big deal* (0.52) was also high. Table 7 shows that *open data* (0.17), which has not appeared before, began to show centrality in this period. Meanwhile, the clusters formed in segment 3 are Open Access (C1), Institutional Repository (C2), Academic Library (C3), and Open Data (C4). Institutional repositories that were previously absent formed clusters, and open data clusters were newly created. Meanwhile, in the OA cluster, the economic aspects of OA (Kellersohn et al. 2011), such as journal publishing markets and big deals, were the main topics of discussion.

In the W4 segment (2014.1–2016.12), projects such as SCOAP3 (2014) and OA2020 (2015) began, and the OA movement progressed further. As OA publishing volume increased, services such as OA button (openaccessbutton.org) have been launched, allowing users to easily browse the free version of research papers. The number of OA papers increased, and 626 cases were analyzed during this period. The frequency of keyword occurrence of *open access* has also increased to 135 cases. It can be observed from Table 5 that *open access* (0.69) *scholarly journals* (0.64) and *repositories* (0.62) showed high centrality. It can be seen Table 7 that new concepts such as *altmetrics* (0.19) also showed centrality. Also notable are the side effects of open access exemplified by *predatory journals* (0.42), and *research data* (0.51), whose centrality has risen due to the influence of open science. Looking at the cluster of keywords, diverse clusters were formed such as Open Access (C1), Information Resource Sharing (C2), Academic Communication (C3), Library (C4), Gold/Green OA (C5), Metadata (C6), Publishing (C7), Internet (C8) and LOD (C9). The scope of the discussion has expanded since the OA cluster now includes keywords such as *open data, APC, predatory journal, altmetrics,* and *research data* (Onyancha 2016, Al-Khatib 2016; González-Valiente, et al. 2016). Academic communication clusters include keywords such as *scholarly publishing, citation, google scholar, impact,* and *citation analysis,* indicating that the impact verification through citation analysis of OA papers (Pisoschi and Pisoschi 2016: Poplašen and Grgić 2016) was active.

In the W5 segment (2017.1–2019.1), 583 related papers were collected. Considering that it is a shorter period than that in W4, it can be observed that many papers had been published in the OA domain. During this period, as the volume of OA surged, services

Table 4 Main keyword centrality, cluster and network map of W3 (2011–2013.12)

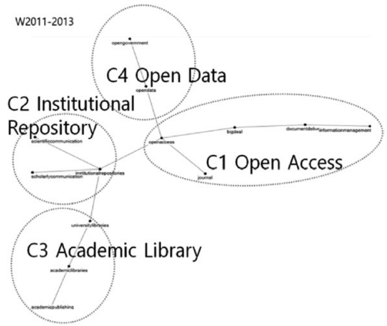
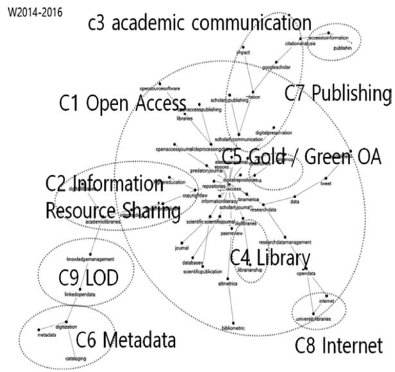
Node	Keyword frequency	Centrality rTBC(0–1)	Cluster		PathFinder Network
Open access	61	0.86	c1	Open access	
Journal	5	0.61	c2	Institutional repository	
Big deal	4	0.52	c3	Academic library	
Institutional repositories	16	0.45	c4	Open data	
Scholarly communication	7	0.39			

Table 5 Main keyword centrality; cluster and network map of W4(2014–2016.12)

Node	Keyword frequency	Centrality rTBC(0–1)	Cluster		PathFinder Network
Open access	135	0.68	c1	Open Access	
Scholarly journals	17	0.64	c2	Information resource sharing	
Repositories	8	0.62	c3	Academic communication	
Information literacy	4	0.59	c4	Library	
Scientific journal	6	0.52	c5	Gold/green OA	
			c6	Metadata	
			c7	Publishing	
			c8	Internet	
			c9	LOD	

and tools used for searching OA versions such as Unpaywall (https://unpaywall.org/) and Copernio (https://kopernio.com/) emerged. As a result, the library required a reasonable subscription agreement model from journal publishers. Therefore, the subscription models of journals such as offset and transformative agreement became prevalent. Looking at the centrality of keywords presented in Table 6, the centrality of unranked keywords such as *open science* (0.55) and *APC* (0.52) increased rapidly. Table 7 shows that the centralities of *semantic web* (0.24), *open data* (0.39), *LOD* (0.15) and *altmetrics* (0.44) were increased. When looking at the frequency of occurrence of words, *APC* and *open science* show high number of 10 and 8, respectively. Clusters are further divided into open Access (C1), altmetrics (C2), Open Data (C3), Journal (C4), LOD (C5), APC (C6), Library (C7), Wiki (C8), Metadata (C9), and Digital Library (C10). The information resource sharing clusters that existed until W4 disappeared, and instead, independent clusters of altmetrics, Open Data, LOD, Wiki, and APC emerged as subject clusters. Consequently, several new discussions (Leeuwen, et al. 2018; Pavan and Barbosa 2018; Segado-Boj et al. 2018) were activated. In addition, the OA cluster shows that new clusters discussing bibliometrics and altmetrics were derived separately.

Time series changes in centrality and cluster

Table 7 shows the centrality of time segments by selecting nodes with an average of more than 0.05. In terms of the average of all segments, keywords with the highest centralities were in the following order: *open access* (0.66), *institutional repository* (0.36), and *academic library* and *journal* (0.29). *Open access* maintains a high centrality of > 0.5 in all time segments. *Institutional repositories, academic libraries,* and *journals* do not show centrality in W1, but maintain the influence from W2 to the W5. The influence of keywords such as *document delivery* and w*orld wide web* is diminishing due to a decrease in centralities in the latest stage. Conversely, recently emerged nodes with high centralities are keywords such as *altmetrics, predatory journal,* and *LOD.* Through these observations, it can be explained that not only have a lot of new themes appeared, the corresponding centralities have also increased and the research front is expanding.

On the other hand, Table 8 shows how subject clusters change over time. It can be seen that the clusters are becoming more and more subdivided in recent years. During the initial stage of W1, discussions related to library information resource sharing such as document delivery were central, so OA at this time did not form an independent research area. The concept of OA has emerged in earnest since W2, but the institutional repository, which is a green OA infrastructure, did not form a major cluster. So far, traditional themes such as the library's resource sharing, digital libraries, academic libraries and electronic journals dominated. In W3, OA and Institutional Repositories formed major clusters, and new themed clusters such as Open Data also appeared. In the wake of the rapid increase in the volume of related papers in W4–W5, OA methodologies such as Gold/Green OA and APC formed major clusters, and new concept research topics such as LOD, altmetrics and Wiki emerged.

Correlation between time segments of centrality

Previously, the intellectual structure of research related to OA was captured over time, and the time series changes in centrality and cluster were observed. This study presents the results of the analysis of the correlation between time segments using the centrality

Table 6 Main keyword centrality, cluster and network map of W5 (2017–2019.1)

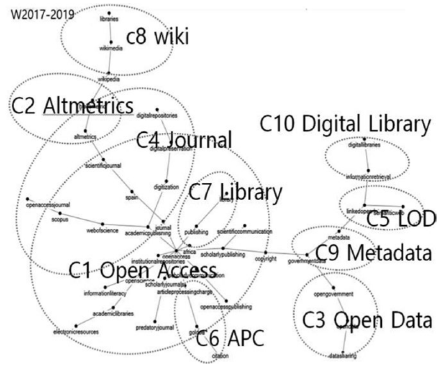
Node	Keyword frequency	Centrality rTBC(0–1)	Cluster		PathFinder Network
Open access	99	0.68	c1	Open access	
Institutional repositories	18	0.56	c2	Altmetrics	
Open science	8	0.55	c3	Open data	
Journal	4	0.54	c4	Journal	
Article processing charge (APC)	10	0.52	c5	LOD	
			c6	APC	
			c7	Library	
			c8	Wiki	
			c9	Metadata	
			c10	Digital library	

Table 7 Keyword centrality in each time segments

Node	W1	W2	W3	W4	W5	Average
Open access	0.50	0.57	0.86	0.69	0.68	0.66
Institutional repositories	0.00	0.35	0.45	0.44	0.56	0.36
Academic libraries	0.00	0.56	0.32	0.29	0.27	0.29
Journal	0.00	0.00	0.61	0.28	0.54	0.29
Bibliometric	0.00	0.60	0.00	0.43	0.36	0.28
Digital libraries	0.36	0.35	0.00	0.27	0.35	0.27
Copy right	0.36	0.22	0.00	0.27	0.46	0.26
Document delivery	0.47	0.37	0.21	0.24	0.00	0.26
World wide web	0.42	0.50	0.00	0.34	0.00	0.25
Scholarly communication	0.00	0.00	0.39	0.41	0.44	0.25
Repositories	0.00	0.53	0.00	0.62	0.00	0.23
Open access journal	0.00	0.13	0.00	0.43	0.43	0.20
Scholarly journals	0.00	0.00	0.00	0.64	0.33	0.19
Article processing charge	0.00	0.00	0.00	0.41	0.52	0.19
Digital storage	0.19	0.25	0.00	0.22	0.25	0.18
e-journal	0.00	0.36	0.00	0.36	0.17	0.18
Scientific communication	0.00	0.00	0.00	0.38	0.50	0.18
Open data	0.00	0.00	0.17	0.31	0.39	0.17
Predatory journal	0.00	0.00	0.00	0.42	0.43	0.17
Libraries	0.00	0.26	0.00	0.32	0.23	0.16
Scholarly publishing	0.00	0.00	0.00	0.30	0.51	0.16
Scientific journal	0.00	0.00	0.00	0.52	0.28	0.16
Information literacy	0.00	0.00	0.00	0.59	0.18	0.15
Electronic publishing	0.50	0.21	0.00	0.00	0.00	0.14
Open access publishing	0.00	0.00	0.00	0.29	0.40	0.14
Digital preservation	0.00	0.00	0.00	0.27	0.40	0.13
Publishing	0.00	0.00	0.00	0.25	0.39	0.13
Altmetrics	0.00	0.00	0.00	0.19	0.44	0.13
Citation	0.00	0.00	0.00	0.25	0.37	0.13
ebooks	0.00	0.18	0.00	0.43	0.00	0.12
Information management	0.00	0.44	0.17	0.00	0.00	0.12
Resource sharing	0.17	0.14	0.00	0.00	0.30	0.12
Open systems	0.00	0.33	0.00	0.25	0.00	0.12
Open science	0.00	0.00	0.00	0.00	0.55	0.11
Big deal	0.00	0.00	0.52	0.00	0.00	0.10
Research data	0.00	0.00	0.00	0.51	0.00	0.10
Africa	0.00	0.00	0.00	0.00	0.50	0.10
Metadata	0.00	0.00	0.00	0.24	0.25	0.10
Academic publishing	0.00	0.00	0.23	0.00	0.26	0.10
Peer review	0.00	0.00	0.00	0.45	0.00	0.09
Open government	0.00	0.00	0.00	0.00	0.43	0.09
Scientific publication	0.00	0.00	0.00	0.39	0.00	0.08
Databases	0.00	0.00	0.00	0.37	0.00	0.07
Digitization	0.00	0.00	0.00	0.22	0.13	0.07
Higher education	0.00	0.00	0.00	0.34	0.00	0.07
Public libraries	0.00	0.00	0.00	0.34	0.00	0.07

Table 7 (continued)

Node	W1	W2	W3	W4	W5	Average
Google scholar	0.00	0.00	0.00	0.33	0.00	0.07
Green OA	0.00	0.00	0.00	0.32	0.00	0.06
Information retrieval	0.00	0.20	0.00	0.00	0.11	0.06
Impact	0.00	0.00	0.00	0.28	0.00	0.06
LOD	0.00	0.00	0.00	0.10	0.15	0.05
Archives management	0.25	0.00	0.00	0.00	0.00	0.05
Semantic web	0.00	0.00	0.00	0.00	0.24	0.05
Web of science	0.00	0.00	0.00	0.00	0.24	0.05

Table 8 Gradual change of clusters

W1	W1	W2	W3	W4	W5
Archive	c1				
Information resource sharing	c2	c4		c2	
Open access		c1	c1	c1	c1
Electronic journal/journal		c2			c4
Digital library		c3			c10
Academic library/library		c5	c3	c4	c7
Institutional repository			c2		
Open data			c4		c3
Academic communication				c3	
Gold/green OA/Apc				c5	c6
Metadata				c6	c9
Publishing				c7	
Internet				c8	
LOD				c9	c5
Altmetrics					c2
Wiki					c8

of keywords. To summarize the research method once again, let's look at Table 7. Table 7 gives the centrality index for each segments of the keyword to be analyzed. Here, since the centrality index of individual keywords is assigned in five stages, correlation between segments can be analyzed using these variables. In the table, only keywords with a centrality index of 0.05 or higher in at least one segment were presented, but in the actual correlation analysis, all 69 keywords with a centrality of 0.01 or higher were included. Spearman correlation analysis was performed because the data did not satisfy the normality as a result of the Shapiro–Wilk test. Correlation analysis was performed by pairing each segment from W1 to W5, and a total of 25 results were obtained as shown in the following table.

As shown in Table 9, it can be seen that the index with the highest correlation between W1–W2 ($r = 0.49$) was shown. This is because in the W1 and W2 segments, information sharing issues such as *document delivery, resource sharing,* and the *world wide web* in addition to *digital libraries* and *copyrights* appeared as influential topics. At that time, research topics such as offline information resource sharing, e-library construction, and digital archiving were predominant, rather than full-scale implementation of open access. In this period, it can be assumed that the discussion of OA was passive in the boundary

Table 9 Relationship between time segments of intellectual structure

Time segments	W1	W2	W3	W4	W5
W1(2005.1–2007.12)	1.00	0.49	0.10	-0.06	0.00
W2(2008.1–2010.12)	0.49	1.00	0.27	0.24	0.08
W3(2011.1–2013.12)	0.10	0.27	1.00	0.08	0.24
W4(2014.1–2016.12)	– 0.06	0.24	0.08	1.00	0.19
W5(2017.1–2019.1)	0.00	0.08	0.24	0.19	1.00

of the mainstream research topic. On the other hand, let's take a look at the case where the value was low. The correlation ($r = -0.06$) of W1-W4 showed the low value. This is because in W4, topics such as *open access, journals, APC, open data, predatory journals, and altmetrics* with low centrality in W1 have emerged. Similarly, low values ($r = 0.00$) were found between W1–W5 with slightly larger time gap. This is because keywords such as *document delivery* that were high in the W1 segments disappeared from the W5. Also, as new themes such as *wikimedia, tweet, LOD, and predatory journal* emerged, there was a difference in intellectual structure. By observing the correlation index analyzed above, it can be seen that the intellectual structure of OA maintained some similarity until W3, 9 years after 2005, and then changed significantly.

In other words, the time when the research front expanded significantly was around 2014. New concepts emerged, showing high centrality, and the scope of research expanded. The W3 time segment is the time when the OA policy of research achievements that received public funds materialized. As countries announced their open data policies, public sector data utilization was promoted in earnest. In addition, in the journal market, there was a lot of debate on the price of journal subscriptions and the big deal contract became a major issue in the library world. Over this period, the scope of research topics broadened and diversified in the OA field.

Discussion

Public policy on open access and changes in the journal publishing market have changed the landscape of academic publishing and increased OA publication quantitatively. The increase in OA has led to changes in the library's journal subscription models. In addition, with increase in the citation of OA papers, various attempts to verify their impact have also been made. In addition, with the increase in new research impact evaluation systems such as altmetrics and the side effects of OA, many new research topics have attracted the attention of researchers in this field.

The intellectual structure related to OA analyzed over five periods was as follows. First, in the W1 time segment, OA was discussed with regard to the sharing or digital archiving of information resources. At this stage, it did not form an independent subject cluster. From the W2 stage onwards, OA had emerged as a cluster in the intellectual structure, but research subjects such as information resource sharing and electronic libraries still formed major clusters. By W3, discussions on sharing of information resources reduced. Instead, new topic clusters such as institutional repository and open data were created. In W4 and W5 stages a rapid increase in OA publication volume could be noticed, which also confirmed that the research topics were differentiated into more than 9 detailed clusters. New topics such as *altmetrics* and *APC* emerged. During this period, it was confirmed that new topics covering the economic aspects of the publishing market and OA were discussed and differentiated into detailed topics reflecting the progress of the OA phase. With the analysis of the maintenance and change timing of the research subject through correlation analysis of keyword centrality between time segments, the open access intellectual structure was interpreted to have changed significantly since 2014, 9 years from 2005. This was because the weak correlation ($r=0.10$) was maintained from the 2005 observations for a period of 9 years, but the correlation ($r=-0.06$) between the two sections with a gap of 9–12 years has no meaning. Topics such *as research data, APC, open data, predatory journals*, and *altmetrics*, which had no centrality in 2005, emerged from 2014, while main topics such as *document delivery* disappeared. In the meantime, as the nodes with no centrality in the past have been on the rise, the research area is expanding.

Conclusion

Although there are many studies that analyze intellectual structure, few studies have observed changes in intellectual structure. This study has theoretical value in that it infers how long the similarities between intellectual structures persist through correlations between intellectual structures captured by time. And this methodology has practical value in that it can be used in the analysis of intellectual structure in the future studies of various fields.

However, this study remained focused on observing the point of time for maintaining and changing the intellectual structure by using the time series index of keywords centrality. In future research, it will be possible to try to analyze the lead- lag relationship between social media reflecting the social environment and OA intellectual structure projected on the research results. Lead-lag analysis based on the similarity and linear relationship of time series data is mainly used in economics, but is also applied in the field of informatics. After revealing the intellectual structure through topic modeling or co-word analysis, research has been conducted to analyze lead-rag relationships between domains (Chen et al. 2016) using keywords time series index. Also, across countries with differences in open access maturity, there may be a lead-lag relationship and time difference in intellectual structure. In addition, in certain academic issues, there may be a lead-lag relationship between diverse academic disciplines. Analyzing the lead-lag relationship between intellectual structures of different fields could be another future research topic.

References

Alan, P., & Jack, M. (2008). The evolution of the intellectual structure of operations management—1980–2006: A citation/co-citation analysis. *Journal of Operations Management, 27*(3), 185–202.

Al-Khatib, A. (2016). Protecting Authors from Predatory Journals and Publishers. *Publishing Research Quarterly, 32*, 281–285.

Bae, Y. (2015). Analysis of input-performance structure for convergence research between technologies in government R&D projects [cited 2020. 7. 8]. https://scienceon.kisti.re.kr/srch/selectPORSrchReport.do?cn=TRKO201500014110.

Bailey, C. W. (2010). *Transforming scholarly publishing through open access: A bibliography*. Houston, TX: Digital Scholarship.

Boukacem-Zeghmouri, C., & Schöpfel, J. (2006). Document supply and open access: an international survey on grey literature. *Interlending & Document Supply, 34*(3), 96–104. https://doi.org/10.1108/02641610610686012.

Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences, 6*, 5303–5310. https://doi.org/10.1073/pnas.0307513100.
Chen, X., Chena, J., Wua, D., Xiea, Y., & Li, J. (2016). Mapping the research trends by co-word analysis based on keywords from funded project. *Procedia Computer Science, 91*, 547–555.
Chen, X., Li, J., Sun, X., & Wu, D. (2019). Early identification of intellectual structure based on co-word analysis from research grants. *Scientometrics, 121*(1), 349–369.
Cho, J. (2014). Intellectual structure of the institutional repository field: A co-word analysis. *Journal of Information Science, 40*, 1–12.
Cho, J. (2020). Analysis of open access status of domestic author's papers published in international journals : Based on highly cited paper. *Korean Society for Library and Information Science, 54*(1), 325–341.
Cho, J. (2018). The trends of media coverage about libraries in Korea: Using semantic network analysis of portal news. *Libri, 68*(4), 291–300.
Cho, J. (2019). Analysis of reading domain of men and women elderly using book lending data. *Korean Library and Information Science Society, 50*(1), 23–41.
Choi, H., Choi, Y., & Nam, S. (2018). Time series analysis of intellectual structure and research trend changes in the field of library and information science: 2003 to 2017. *Journal of the Korean society for Information Management, 35*(2), 89–114.
Department for Business, Innovation & Skills Prime Minister's Office. (2013). G8 Science Ministers Statement London UK. [cited 2020. 6. 15] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/206801/G8_Science_Meeting_Statement_12_June_2013.pdf.
Ding, Y., Chowdhury, G., & Foo, S. (1999). Mapping the intellectual structure of information retrieval studies: An author co-citation analysis, 1987–1997. *Journal of Information Science, 25*(1), 67–78.
González-Valiente, C. L., Pacheco-Mendoza, J., & Arencibia-Jorge, R. (2016). A review of altmetrics as an emerging discipline for research evaluation. *Learned Publishing, 29*(4), 229–238.
Heo, G., & Song, M. (2013). Examining the intellectual structure of a medical informatics journal with author co-citation Analysis and co-word analysis. *Journal of the Korean Society for Information Management, 30*(2), 207–225.
Kellersohn, A., Meyer, T., Mittermaier, B., & Schaeffler, H. (2011). Between pay-per-view and "big deal"—Licensing electronic information in Germany. *ZEITSCHRIFT FUR BIBLIOTHEKSWESEN UND BIBLIOGRAPHIE, 58*, 120–130.
Kim, K. (2014). Examining of knowledge structure of sports sociology through social network analysis of authors co-citation: Analysis on Journal of Sociology of Sport (KJSS). *Korean Society for the Sociology of Sport, 27*(4), 109–137.
Kwak, S., & Chung, E. (2012). Domain analysis on economics by utilizing co-citation analysis of multiple authorship. *Journal of the Korean Society for Information Management, 29*(1), 115–134.
Lee, J. (2006a). A study on the network generation methods for examining the intellectual structure of knowledge domains. *Journal of the Korean Society for Library and Information Science, 40*(2), 333–355.
Lee, J. (2006b). A novel clustering method for examining and analyzing the intellectual structure of a scholarly field. *Korea Society for Information management, 23*(4), 215–231.
Lee, J. (2006c). Centrality measures for bibliometric network analysis. *Journal of the Korean Society For Library And Information Science, 40*(3), 191–214.
Lee, J. (2013). A comparison study on the weighted network centrality measures of tnet and WNET. *Journal of the Korean society for information management, 30*(4), 241–264.
Leeuwen, T. N., Tatum, C., & Wouters, P. F. (2018). Exploring possibilities to use bibliometric data to monitor gold open access publishing at the national level. *JASIST, 69*(9), 1161–1173.
McGlashan, D., & Hadley, K. (2019). Adapting to a transformative future: Open Access and PlanS. [cited 2019. 12. 8] https://www.jstage.jst.go.jp/static/files/ja/pub_20190621_Seminar01.pdf.
Miguel, S., Oliveira, E. F. T., & Grácio, M. C. C. (2016). Scientific production on open access: A worldwide bibliometric analysis in the academic and scientific context. *Publications, 4*, 1–15.
National Institutes of Health. Request for Information: NIH Public Access Policy. [cited 2015. 8. 15] https://publicaccess.nih.gov/comments.htm.
OA2020. (2019). OA2020 progress report. [cited 2019. 12. 6.] https://oa2020.org/progress-report-nov2019/.
Onyancha, O. B. (2016). Open research data in sub-Saharan Africa: A bibliometric study using the data citation index. *Publishing Research Quarterly, 32*, 227–247.
Park, E. (2007). Perspectives on access to electronic journals for long-term preservation. *Serials Review, 33*(13), 22–25.

Park, J., & Jeong, D. (2013). A study on the intellectual structure of library and information science in Korea by author bibliographic coupling analysis. *Journal of the Korean Society for Information Management, 30*(4), 31–59.

Pavan, C., & Barbosa, M. C. (2018). Article processing charge (APC) for publishing open access articles: The Brazilian scenario. *Scientometrics, 117*(2), 805–823.

Pisoschi, A. M., & Pisoschi, C. G. (2016). Is open access the solution to increase the impact of scientific journals? *Scientometrics, 109*, 1075–1095.

Poplašen, L. M., & Grgić, I. H. (2016). Altmetric and bibliometric scores: Does open access matter? *Qualitative and Quantitative Methods in Libraries, 5*, 251–460.

RCUK Policy on Open Access and Supporting Guidance. [cited 2015. 8. 15]. https://www.ukri.org/files/legacy/documents/rcukopenaccesspolicy-pdf.

Rodrigues, R. S. R., Taga, V., & Passos, M. F. (2016). Research articles about open access indexed by scopus: A content analysis. *Publications., 4*(4), 31. https://doi.org/10.3390/publications4040031.

Rokaya, E., Atlam, M., Fuketa, T. C., & Dorji, J. I. (2008). Aoe Ranking of field association terms using co-word analysis. *Information Processing & Management, 44*(2), 738–755.

Ronda-Pupo, G. A., & Guerras-Martin, L. Á. (2012). Dynamics of the evolution of the strategy concept 1962–2008: A co-word analysis. *Strategic Management Journal, 33*(2), 162–188.

SCOAP3. https://scoap3.org/news102/.

Segado-Boj, F., Martín-Quevedo, J., & Prieto-Gutiérrez, J. J. (2018). Attitudes toward open access, open peer review, and altmetrics among contributors to spanish scholarly journals. *Journal of Scholarly Publishing, 50*(1), 48–70.

Suh, S., & Jeong, E. (2013). Domain analysis on the field of open access by co-word analysis. *Korean Biblia Society for Library and Information Science, 24*(1), 207–228.

Szell, M., Lambiotte, R., & Thurner, S. (2010). Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences of the United States of America, 107*(31), 13636–13641.

The office of Science and Technology Policy. (2013). Increasing access to the results of federally funded scientific research. [cited 2015. 8. 15]. https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

White, H. D. (2003). Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists. *Journal of the American Society for Information Science & Technology, 54*(5), 423–434.

Yiotis, K. (2005). The open access initiative: A New paradigm for scholarly communications. *Information Technology and Libraries, 24*(4), 157–162. https://doi.org/10.6017/ital.v24i4.3378.

Zhao, R., & Wu, S. (2014). Study on themes and authors' influence of open access in China. *Scientometrics, 101*, 1165–1177.