




Mining the evolutionary process of knowledge through multiple relationships between keywords

Xinyuan Zhang¹ · Qing Xie² · Chaemin Song³ · Min Song³ 

Received: 9 September 2021 / Accepted: 11 January 2022 / Published online: 21 February 2022
© Akadémiai Kiadó, Budapest, Hungary 2022

Abstract

Knowledge evolution offers a road map for understanding knowledge creation, knowledge transfer, and performance in everyday work. Understanding the knowledge evolution of a research field is crucial for researchers, policymakers, and stakeholders. Further, paper keywords are considered efficient knowledge components to depict the knowledge structure of a research field by examining relationships between keywords. However, multiple relationships between keywords provided by papers are rarely used to explore knowledge evolution. Three relationships were applied: a direct co-occurrence relationship, indirect relationship by keyword pair citation, and same author trace, providing temporal and sequential knowledge evolution. The direct co-occurrence relationship is constructed by keyword co-occurrence pair and acts as the temporal structure of knowledge pairs. The indirect relationship is constructed by a keyword pair-based citation relationship, meaning the citation relationship between keyword co-occurrence pairs, acting as the sequential structure of knowledge pairs. Additionally, the same author trace represents an indirect relationship that a keyword pair provided by the same author in a different paper. Thus, knowledge evolution could be mined quantitatively from a different perspective. Therefore, we present an empirical study of the informetrics field with five evolution stages: knowledge generation, growth, obsolescence, transfer, and intergrowth. The results indicate that knowledge evolution is not a continuous trend but alternating growth and obsolescence. During evolution, knowledge pairs stimulate each other's growth, and some knowledge pairs transfer to others, demonstrating a small step toward knowledge change. According to the indirect keyword relationship paired with the same author trace, creators and followers of knowledge evolution are different.

Keywords Knowledge evolution · Multiple relationships of keywords · Same author trace · Keyword co-occurrence pair · Keyword pair-based citation relationships

✉ Min Song
min.song@yonsei.ac.kr

Extended author information available on the last page of the article

Introduction

Knowledge evolution offers a unique and powerful road map for understanding knowledge creation, learning, and performance in everyday work (Allee, 2012). For researchers and policymakers, an accurate understanding of the state-of-the-art of a research field is needed to maintain scientific creativity.

Further, academic entities, such as authors, papers, and journals, are also carriers of knowledge. The keywords given by authors are concentrated summaries of documents. Author-selected keywords refer to the lists of keywords selected by author(s) to represent the content of the underlying scientific article (Uddin & Khan, 2016). Many studies have used keywords and their relationships to map the knowledge structure in a series of articles (Peset et al., 2020; Su & Lee, 2010). Co-occurrence, citation, and co-citation relationships link keywords provided in the literature to shape the knowledge structure (Choudhury & Uddin, 2016; Zhu et al., 2015). Literature keywords comprise an efficient corpus for shaping the scientific knowledge structure because of the informative relationships between keywords and because keywords provide understandable content to represent knowledge. Keyword co-occurrence networks constitute the basic structure of knowledge map analysis (Uddin et al., 2015).

However, the keyword relationship approach to shaping the knowledge evolutionary process considers only one kind of relationship at a time. Both temporal and sequential knowledge structures are considered to describe the long-term evolutionary process of knowledge. Moreover, scientific innovation and advancement have been based on combining existing knowledge (Uzzi et al., 2013; Wang et al., 2017). Scholars have argued that papers and patents are a combination of knowledge elements rather than elements of knowledge themselves (Lee et al., 2010; Su & Lee, 2010). As a trace of knowledge evolution, traditional keyword network analysis uses the keyword as a network node, and the network edge represents the co-occurrence and citation relationships. The evolutionary process of network edges has not been as thoroughly discussed as that of network nodes. Additionally, as the keyword providers, authors decide which keywords are chosen in their research and whether to continue their use in future papers (Lu et al., 2021). Single relationships of keywords lack information to explore the role of the same author in the knowledge evolutionary process.

Therefore, we analyze the knowledge evolution process within a research field using multiple relationships, including keyword pair-based citation relationships, the same author trace, and keyword co-occurrence relationships. Keyword co-occurrence pairs are divided by year and represent the temporal knowledge structure of the research field of informetrics. The indirect citation relationship consists of keyword co-occurrence pairs and citations for these keyword pairs. The keyword pair citation relationship represents the sequential trace of knowledge pair evolution, and the same author trace indicates keywords provided by the same author in different papers. The same author plays a vital role in the diffusion of knowledge, which could be traced using the keywords employed in their series of papers. In particular, we address the following research questions:

RQ1: How do the characteristics of the direct keyword co-occurrence pair depict the knowledge evolutionary process?

RQ2: What are the unique characteristics of knowledge evolution revealed by indirect keyword relationships based on the same author trace?

RQ3: What the keyword pair-based citation relationship explains for the knowledge evolutionary process?

Multiple relationships between keywords, citations, and same author trace were assessed to better understand the knowledge evolutionary pattern within a scientific field to answer these research questions. To this end, we first extracted keywords from papers related to informetrics research. Second, we constructed the direct keyword relationship paired with the keyword co-occurrence relationship and indirect keyword relationship paired with the edge-based citation and same author trace. Third, we investigated the characteristics of knowledge evolution revealed by the direct and indirect keyword relationships, and we discovered knowledge evolutionary patterns in the informetrics field over time.

The contribution of this paper is primarily to quantitatively measure the evolutionary process of knowledge within a discipline using multiple relationships. Both keyword co-occurrence and citation relationships are considered to shape the knowledge evolution structure. Additionally, five evolution stages are recognized within the research field of informetrics: knowledge generation, growth, obsolescence, transfer, and intergrowth. This work explains these evolutionary stages in detail and presents examples in the informetrics field. Moreover, the keywords provided by the same authors in different papers were labeled to trace the same author's effect on promoting knowledge evolution.

Related work

Process stage of knowledge evolution

The previous scientific and technological knowledge trajectories imply that knowledge evolution is not random and follows a periodic life cycle (Dosi, 1982). Mina et al. (2007) mapped the medical knowledge evolutionary process and divided it into the emergence, growth, and transformation stages. In terms of the topic evolution model, the life-cycle events of a topic usually evolve by emerging, disappearing, merging, and splitting (Qian et al., 2020). With patent citation data, technological knowledge depreciation was identified (Liu, Grubler, et al., 2021; Liu, Yang, et al., 2021). Three general processes and two special processes are summarized in this research.

Knowledge generation

As is shown in Fig. 1. The edge $\langle A, E \rangle$ appears for the first time in a certain year in the evolutionary process. The new emerging edge $\langle A, E \rangle$ is used continuously for at least two years.

The generation of new scientific knowledge often occurs through the coming together, re-working, and re-formulating of previously distinct pieces of older knowledge and techniques into a new scientific synthesis (Hoch, 1985). The creation of new scientific knowledge builds on combining existing pieces of knowledge (Wang et al., 2017). Scholars have proved that novelty can be defined as unprecedentedly recombining pre-existing knowledge components (Arthur, 2009; Burt, 2004; Fleming, 2001). Therefore, the new knowledge generation reflected in the knowledge evolution within a discipline could be recognized with the new knowledge co-occurrence pair.

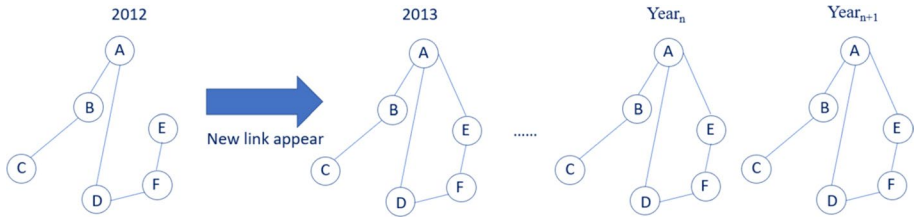


Fig. 1 Schematic diagram of knowledge generation

Knowledge growth

As is shown in Fig. 2. The weight of edge <D, F> keeps increasing in a certain period in the node-perspective evolutionary process.

Scientific revolution describes the next stage after knowledge generation as growth, which includes rapid and stable growth periods (Kuhn, 1962a, 1962b). Kuhn explains that the rapid growth period is the formation of a new scientific paradigm. When a new scientific paradigm encounters a new development bottleneck, the growth rate slows and enters a stable growth period. Earlier studies have attempted to mathematically assess the growth of knowledge as the consequence of knowledge diffusion (Modis, 2007). The growth of scientific knowledge is primarily due to the diffusion process of transmitting new ideas from person to person between scientific communities. The exponential increase in the literature is a good indicator of the knowledge growth processes (Chen & Hicks, 2004).

Knowledge obsolescence

As is shown in Fig. 3. The number of co-occurrence edges <D, F> becomes increasingly smaller in the evolutionary process.

The term *obsolescence* was first proposed by Gosnell (1944) and meant becoming out of date or something that is increasingly less used. He pointed out that the obsolescence of information conceived by the literature is much slower than the literature itself. The obsolescence of literature was measured in two ways. The synchronic approach uses the publication timeline of the newer half of the total literature currently in use in a discipline

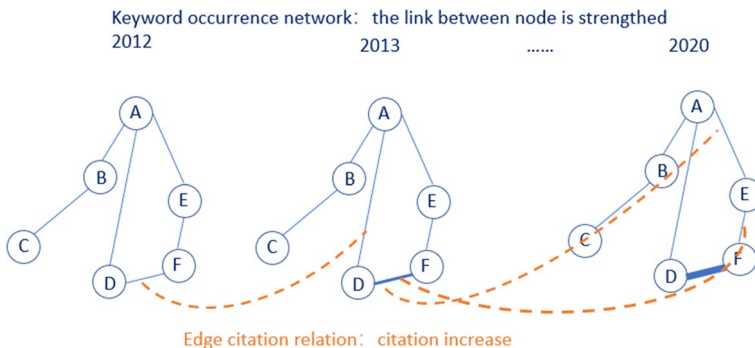


Fig. 2 Schematic diagram of knowledge growth

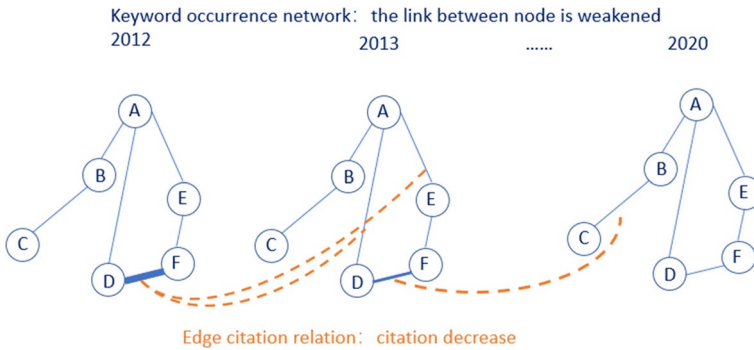


Fig. 3 Schematic diagram of knowledge obsolescence

(Burton & Kebler, 1960). The diachronic approach is to collect the annual citation count of the literature (Wang et al., 2019).

The underlying reason for obsolescence in literature is that the knowledge involved in the papers follows the knowledge evolutionary life cycle. When the knowledge in the research field has become dated or even wrong, the literature related to the dated knowledge is no longer cited. When the knowledge becomes part of the consensus of the scientific community, it continues to be in the knowledge graph every year but rarely receives many citations by future work. When knowledge is widespread in the scientific community, it is mentioned and cited with high frequency. However, many papers contribute to widespread knowledge, accelerating the obsolescence of earlier similar literature, corresponding to rapid scientific research activities in a discipline (Swanson, 1993). Obsolescence is proved to be reflected in the tendency of the citation rate of articles or patents to decay over time because of their reduced relevance for ongoing knowledge advancement (Higham et al., 2017). Knowledge obsolescence can be recognized when the knowledge no longer exists (is wrong) or is much less cited in successive years.

Knowledge transfer

As is shown in Fig. 4. The co-occurrence <D, F> in 2012 is cited by pair <A, E>, and the pair <A, E> continues to be cited in future years.

Citations are the cornerstone of knowledge transfer in science. In addition, knowledge transfer refers to transfer in the process of both paper and patent citations (Abramo

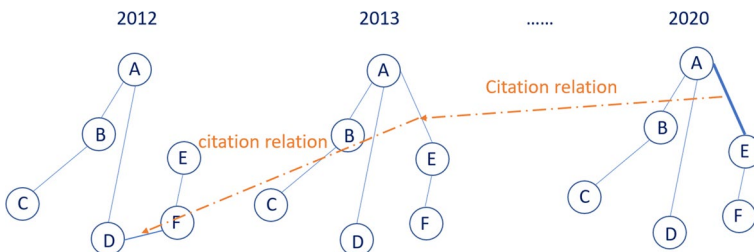


Fig. 4 Schematic diagram of knowledge transfer

& D'Angelo, 2020; Hassan et al., 2018; Silvello, 2018). The transfer of knowledge is a condition of progress in human society (Loasby, 2002; Pu et al., 2021) because it offers a new use-value for aging knowledge. Qian et al. (2020) defined topic transfer as when the similarity between the two periods is greater than the predefined threshold value, and topic transfer is the underlying process behind knowledge transition.

Knowledge intergrowth

As is shown in Fig. 5. The edge <D, F> was strengthened with the growth of edge <A, E> in the node-perspective evolutionary process. In addition, there is a citation link between edge <D, F> and edge <A, E> in the edge-perspective evolutionary process.

The intergrowth status of knowledge comes from the definition of the symbiosis theory in biology. Biology describes symbiosis as an organism participating in a symbiotic relationship in which two species derives a benefit from each other. However, the co-existence condition of knowledge in scientific activities is not as rigid as biological symbiosis; thus, the symbiosis relationship is not a unique relationship between knowledge to support continued knowledge growth. Scholars have called it *knowledge co-existence* (Réale et al., 2020; Urbano & Ardanuy, 2020) or knowledge intergrowth (Shi & Wang, 2009). This paper employs the term *intergrowth*, considering that scientific behavior should represent the benefit attachment relationship between the co-existing knowledge. Citations between literature provide a good way to depict the process in which one knowledge unit derives a benefit from another. The synchronized growth of these two co-existing knowledge units depicts the process in which the other knowledge unit is still active in the current scientific domain.

Content-based approaches to shape the knowledge evolutionary process

Scientific entities provide many approaches to trace knowledge evolution. Journals, articles, and authors act as knowledge carriers. The knowledge diffusion path could be identified through a network with the above entities as nodes (Börner et al., 2004). As with natural language processing technology, topics and keywords can be extracted efficiently and provide a more accessible knowledge representation than knowledge carriers. Topics and keywords, clustering algorithms, and network analyses have recently garnered considerable attention in identifying the knowledge structure and tracking academic knowledge evolution.

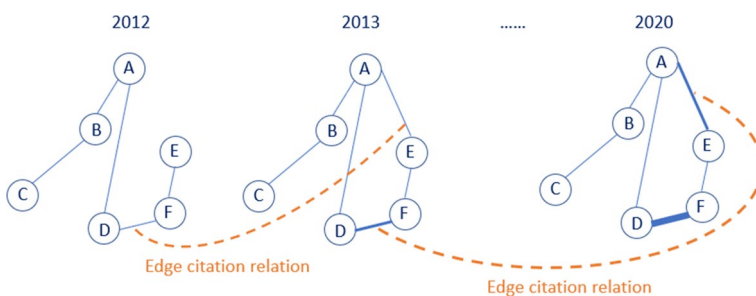


Fig. 5 Schematic diagram of knowledge transfer

Topic modeling

Topic modeling methods aim to extract semantic topics from unstructured documents. Topic evolution is one branch that seeks to analyze how temporal topics in a set of documents evolve and has successfully identified content transitions (Jung & Yoon, 2020). Named entities or distinctive phrases could act as topic terms (b; Xu, Ding, et al., 2019). The last 20 years have witnessed a rise in the number of studies using topic modeling in numerous articles (Figueroa et al., 2017). The latent Dirichlet allocation (Blei et al., 2003) has been the most discussed among these. The knowledge evolution reflected by topic evolution requires field knowledge interpretation because topic modeling generates the probability distribution of a group of words that is not interpretable without human knowledge (Chang et al., 2009). Further, the topic evolution traced with citation path and co-occurrence relation tends to provide persistent terms, which could be regarded as the scientific meme or gene (Kuhn et al., 2014).

Additionally, the Dirichlet multinomial regression (DMR) model (Mimno & McCallum, 2012), where the publication date is set as the observed feature, was applied to analyze the distribution of the top 10 topics from 1978 to 2020 in informetrics. The evolutionary topic process reflected by topic modeling is at a more macro level than the evolutionary knowledge process because the topic association defined in the topic evolutionary process is always based on the similarity and dissimilarity between topics in several different time windows (Xu, Ding, et al., 2019; Xu, Zhang, et al., 2019). Even topic modeling could be combined with citation relationships to identify the topic evolution path. Knowledge depicted in the topic evolutionary process is in the form of conceptual clustering. When and what kind of knowledge innovation is generated in the evolutionary process is ambiguous. Topic modeling is good at tracking the flow of innovation and knowledge rather than describing knowledge structure (Kim et al., 2018).

Keywords analysis

The descriptor keywords for an article are usually abstract definitions of the research context focus of the article (Su & Lee, 2010). In bibliometric research, keyword analysis of publications provides an effective way to investigate the knowledge structure of research domains and explore the developing trends within domains (Hu et al., 2018). Scientific literature is often fragmented, implying that certain scientific questions can only be answered by combining information from various articles. In this way, knowledge in a given paper is represented by series of keyword pairs (van der Eijk et al., 2004). Keyword networks formed from published academic articles were analyzed to examine how keywords are associated and identify important keywords and their changes over time.

Keyword co-occurrence pairs are employed to discern relationships among various scientific concepts in scientific papers to reveal the temporal structure of scientific knowledge (Choudhury & Uddin, 2016; Khasseh et al., 2017; Sedighi, 2016). Keyword co-occurrence pairs have a direct relationship extracted from citing papers. The co-occurrence edge could be combined with keyword frequency weights, occurrence time, and semantic distance to emphasize the critical relationships between keywords. The keyword co-occurrence network is best suited to understanding the current topic knowledge (Lee et al., 2017).

Citation analysis symbolizes the transfer of scientific knowledge (Choi, 1988) and is prominent in scientific knowledge discovery (Lee et al., 2015). Keyword citation pairs have

an indirect relationship, requiring at least two papers to provide a citation link from the citing and cited papers. Cheng et al. (2020) demonstrated that the keyword citation network could detect indirect connections between keywords, understand critical knowledge units, and find significant topics. New keyword connections that have not appeared in the existing co-keyword analysis can be captured. Garfield (1965) claimed that citing between literature indicates a topical association between two papers. The relationship between keywords in the cited and citing papers is also similar (Hu & Zhang, 2015; Khasseh et al., 2017). A high semantic similarity exists between the keywords of the two papers in the citation relationship.

The indirect keyword relationship could be associated with a co-citation paper and the same author. The keyword co-citation network is proposed to measure the keyword similarity in terms of the topological network structure. The indirect keyword relationship paired by the same author in a different paper is less discussed in previous research. The research on author name disambiguation has confirmed that two articles with the same author published in a short time interval have high topic similarity and knowledge similarity (D'Angelo & van Eck, 2020). Therefore, keywords within these two same author papers are more likely to be reusable. The keywords of previous and subsequent papers written by the same author include reused keywords and new emerging keywords. The reused keyword represents the follow-up study of a certain topic. The new emerging keywords form an indirect pair of keywords, indicating the evolution of the author (Lu et al., 2020). The small processes of individual knowledge evolution combined constitute the entire process of knowledge evolution in scientific communities.

Keyword pair-based citation relationships

A paper generally refers to and cites multiple references and may be cited by others, thus reflecting the inheritance and variability of scientific development (b; Liu, Grubler, et al., 2021). Previous studies have used papers, authors, and keywords as the nodes for the citation network to predict paper citations, assess scholar influence, identify the research front, trace the knowledge diffusion path, and so on (Abrishami & Aliakbary, 2019; Kim et al., 2018; Muñoz-Écija et al., 2019). The knowledge evolutionary process has emphasized that novel ideas are spurred by the original combination of different existing knowledge (Mukherjee et al., 2017; Uzzi et al., 2013). The advancement of knowledge depends on the novelty and quality of an idea and whether the idea attracts researchers' attention (i.e., diffusion) and their research builds on it (i.e., adoption as a citation) (Furman & Stern, 2011; Liang et al., 2019; Parolo et al., 2015).

The process ignores the characteristic of the keyword combination pair using single keyword phrases as the node of a citation network. In a keyword-based knowledge graph, nodes are noun phrases of keywords that could represent knowledge concepts, whereas the edges between knowledge concepts illustrate a triple form of knowledge unit. Edge features contain essential information about graphs. In the traditional node-perspective network, the weight of the edge is calculated with the strength of two nodes, and the weight belongs to the first-order proximity between two nodes. The edge-perspective network involves the link and the related nodes and enables capturing higher-order structures in a graph (Zhao et al., 2018).

A study exists on a patent citation network and technology convergence of the patent pairs, where edge outliers are assumed to be the opportunity for innovation emerging (Zhang et al., 2017). In biomedicine, gene expression, drug treatment, and virus infection

evolve in a stochastic and temporal fashion (Wu & Wu, 2013). The edge-network analysis, where a node represents a pair of connecting nodes (i.e., an edge in the traditional node network), has been proved effective to predict the predisease state (the state of an individual before the appearance of clinical symptom) and thus achieves early disease diagnosis (Yu et al., 2014). The edge network transforms the ‘node expression’ data in the node space into ‘edge expression’ data, making it possible to explore the edge space (i.e., correlation space) to classify a single sample (Zeng et al., 2014).

Proposed approach

Research overview

The study was designed to mine the knowledge evolutionary process through a direct keyword relationship paired within a citing paper, an indirect keyword relationship paired by the same author trace in a different paper, and an indirect keyword citation relationship formed by the keyword co-occurrence edge, as illustrated in Fig. 6. In the results, we divided the above knowledge pairs by year and analyzed the evolutionary process, including knowledge generation, growth, obsolescence, migration, and intergrowth.

Data collection

The field of informetrics is a broad term comprising all metrics studies related to information science, including bibliometrics (bibliographies, libraries, etc.), scientometrics (science policy, citation analysis, research evaluation, etc.), webometrics (metrics of the web, Internet, or other social networks, such as citation or collaboration networks), and so on (Egghe, 2005). Informetrics was selected for this study as the discipline case because it is fast-developing (Bar-Ilan, 2008) and highly interdisciplinary and is affected by the incessant evolution of information technology (Prebor, 2010). The authors working on this study have focused on informetrics research for many years and are familiar with the knowledge of informetrics.

Papers and the bibliographic information were collected from Scopus because it includes a more expanded spectrum of journals than PubMed and Web of Science (Klavans & Boyack, 2009; Leydesdorff et al., 2016) to collect papers on informetrics research with as much detail as possible. First, all papers published in the *Journal of Informetrics* or *Scientometrics* were selected because they are well-known international journals focusing on quantitative features and characteristics of science and scientific research. Afterward, we refer to Bar-Ilan’s (2008) retrieval strategy when she conducted a detailed review of the status quo of informetrics in the twenty-first century. High-frequency keywords from informetrics or scientometrics were extracted and used to improve the retrieval strategy. In the same way, papers related to informetrics that were published in other journals were collected from Scopus.

Only English journal articles were considered. After cleaning the irrelevant records, 8732 papers were related to the field of informetrics. Additionally, the bibliographic information of the references of these 8732 papers was also collected to provide a direct citation relationship between keywords. The references consist of papers published in non-library and information science journals. Collected reference papers with at least three citation relationships were considered to build the direct citation relationship to capture the stable and convincing

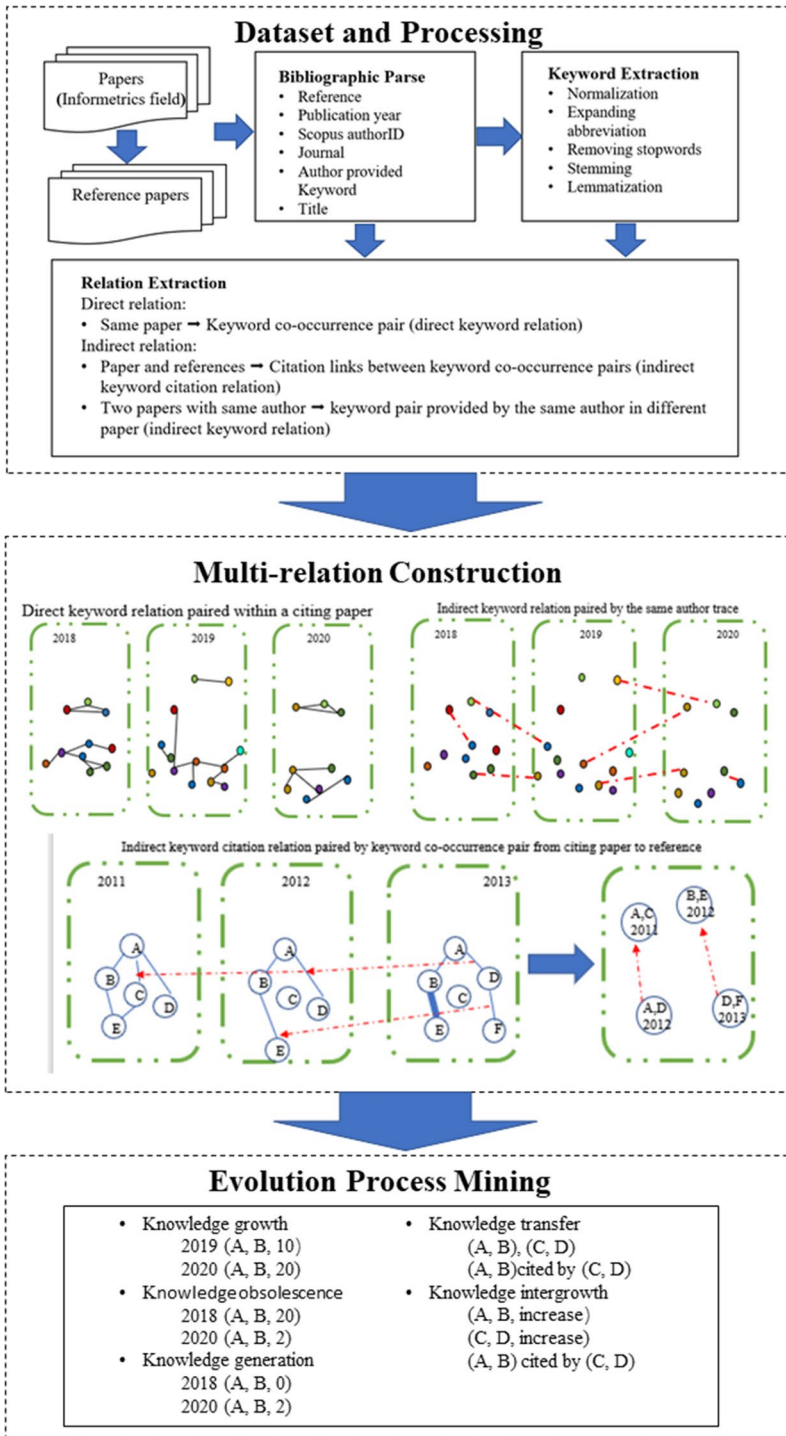


Fig. 6 Study structure

knowledge evolutionary process of informetrics (Waltman, 2016). The retrieval strategy and dataset description are presented in Table 1.

The missing keyword rate is high in the collected dataset; thus, we extracted the keyword from the paper title (Xie et al., 2021). The counts of co-occurrence pairs, citation pairs, and the same author pairs are listed in Table 2.

Multirelationship construction of the knowledge evolution process

Figure 7 shows the data preprocessing process for multirelationship construction between keywords. The raw data collected from Scopus are parsed into bibliographic information and then keyword are extracted. Based on the bibliographic information and keywords, three multirelationships are constructed. The multirelationship of the knowledge evolutionary process involves direct and indirect keyword relationships. The former refers to the keyword co-occurrence relationship, and the latter refers to the keyword co-occurrence edge's citation relationship and the keyword provided by the same author but not within the same paper.

Direct keyword relation construction

As Fig. 8 illustrates, $\langle A, C \rangle$, $\langle B, C \rangle$, $\langle B, E \rangle$, $\langle A, D \rangle$, and $\langle D, F \rangle$ are co-occurrence keyword pairs. The co-occurrence relationship has no direction, meaning the pairs $\langle A, B \rangle$ and $\langle B, A \rangle$ are the same. These pairs are divided into different time slices according to the publication date of the respective papers. The links between two keywords are the annual co-occurrence frequency of the keyword pairs. The bold blue line indicates an increase in the frequency of co-occurrence in a given year. The blue line of disappearance indicates the disappearance of the co-occurrence in a given year.***

Indirect keyword citation relationship construction

As Fig. 9 notes, $\langle A, C \rangle$, $\langle B, E \rangle$, $\langle A, D \rangle$, and $\langle D, F \rangle$ are co-occurrence keyword pairs. The paper citation path from the citing to cited paper indicates the citation relationship from $\langle A, C \rangle$ to $\langle A, D \rangle$, $\langle A, D \rangle$ to $\langle A, D \rangle$, and $\langle B, E \rangle$ to $\langle D, F \rangle$. Then, the indirect keyword citation relationship involves the link between these pairs above, as marked with the dashed red line. The pair-to-pair relationship represents the knowledge evolution path.

Indirect keyword relationship construction

The same author publishing two papers provides two groups of keywords, and the indirect keyword relationship is from the above academic process. As Fig. 10 presents, the edge of two keywords is the number of the same authors to whom those keywords belong, and we did not count the keywords listed in the same paper.

Results

Knowledge evolutionary process based on direct keyword co-occurrence pairs

The keyword co-occurrence pairs are grouped by year. Hence, the year of the first occurrence is identified. The evolutionary increment of the direct keyword co-occurrence

Table 1 Retrieval strategy and description of the dataset

Paper collection	Paper counts	Time distribution	Missing keyword rate
Papers	8732	1978–2021	0.68
References	36,152	1978–2021	0.51
Retrieval strategy	TITLE-ABS-KEY (Informetric* OR Scientometrics * OR webometric* OR scientometric* OR altmetric* OR "citation analysis" OR "citation analyses" OR "co citation analysis" OR "co-citation analyses" OR "co-citation analysis" OR "co-citation analyses" OR "link analysis" OR "link analyses" OR "link structure" OR "self citation" OR "self citations" OR "self-citation" OR self-citations OR "S&T indicator" OR "S&T indicators" OR citation map* OR citation visualization* OR "science policy" OR "research policy" OR "impact factor" OR "impact factors" OR h-index OR "h index" OR Hirsch index OR "patent analysis" OR "patent analyses" OR Zipf OR Bradford OR Lotka OR "collaboration network" OR "collaboration networks" OR "coauthorship network" OR "coauthorship networks" OR "co-authorship network" OR "co-authorship networks" OR "scholar performance") AND (LIMIT-TO (PUBSTAGE, "final")) AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English")) AND (LIMIT-TO (SRCTYPE, "j")) AND (EXCLUDE (EXACTSRCTITLE, "Scientometrics")) OR EXCLUDE (EXACTSRCTITLE, "Journal of Informetrics"))		

Table 2 Pair counts of the datasets of the paper collection

	Co-occurrence pair counts	Citation pair counts	Same author pair counts
Papers	111,740	543,329	16,532
References	413,460		16,490

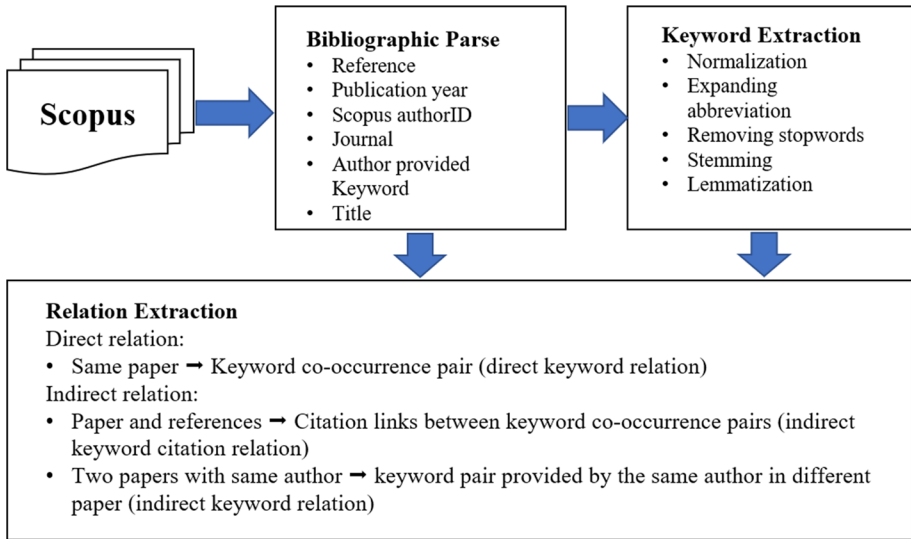


Fig. 7 Data preprocessing process

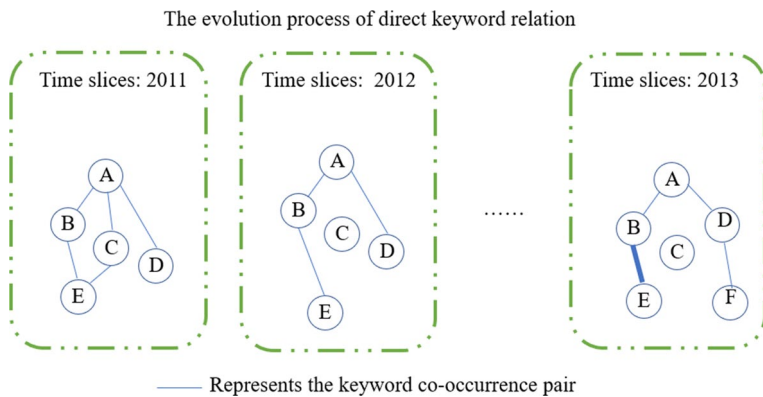


Fig. 8 Direct keyword relationships

pairs is calculated using the keyword co-occurrence frequency of one pair at the current year minus the keyword co-occurrence frequency of the same pair during the last year. The evolutionary increment represents the annual change in one keyword co-occurrence pair. When the increment is greater than zero, the keyword co-occurrence pair indicates

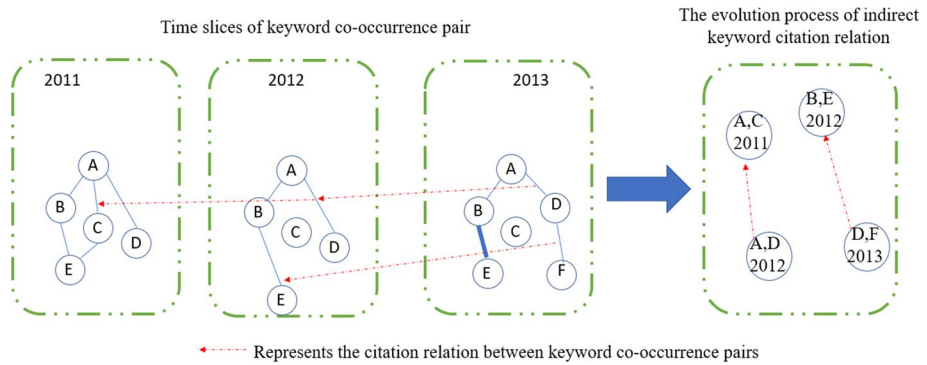


Fig. 9 Indirect keyword citation relationships

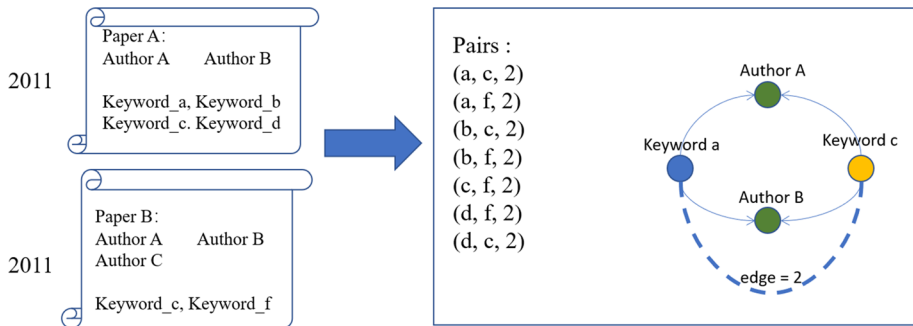


Fig. 10 Indirect keyword relationships (the same author trace)

the temporary knowledge growth stage in the evolutionary process. In contrast, when the increment is negative, the corresponding pair indicates the temporary knowledge obsolescence stage in the evolutionary process. Thus, the evolutionary process has had rises and falls since its first emergence in the research field.

Table 3 presents the pairs of keyword co-occurrences that fluctuated the most during the evolutionary process. Columns 2 and 6 list the generation year of the keyword pairs. The growth and obsolescence years represent the year combination list in the incremental calculation of the evolutionary process. The year combination is written as the two years joined by an underscore. These keyword pairs in Table 3 might contain more than one set of year combinations.

Figure 11 presents the visualized evolutionary flow of these keyword pairs from Table 3. The starting point for each keyword pair evolutionary diagram is on the far right, corresponding to the generation year in Table 3. From right to left, the path in the Sankey diagram represents a fragment of the evolutionary process between two years. Then, the number beside the year is the current frequency of co-occurrence. The node height depends on the frequency of the keyword co-occurrence. If the node on the right is taller than the node on the left, the keyword pair experiences an obsolescence stage between two years. In addition, the reverse exhibits a growth stage between two years for this keyword pair.

Table 3 Knowledge evolutionary process based on direct keyword co-occurrence relationships

Keyword pairs	Generation years	Growth years	Obsolescence years	Keyword pairs	Generation years	Growth years	Obsolescence years
Bibliometric and citation analysis	1997	2010_2009	2014_2013	Journal rank and journal impact factor	1984	1985_1984	2015_2014
		2011_2010	2018_2017			2008_2007	2019_2018
		2015_2014	2019_2018			2010_2009	2020_2019
		2016_2015				2014_2013	
		2017_2016				2017_2016	
Scientific product and bibliometric	1997	2020_2019		Citation and bibliometric	1996	1997_1996	2010_2009
		2021_2020	2013_2012			2011_2010	2014_2013
		1998_1997				2013_2012	2016_2015
		2016_2015				2015_2014	2018_2017
						2017_2016	2021_2020
Research evaluation and h-index	2006	2007_2006	2011_2010	Citation analysis and h-index	2002	2019_2018	2015_2014
						2016_2015	
						2020_2019	
Research trend and bibliometric analysis	2007	2008_2007	2021_2020	University and bibliometric	2010	2021_2020	2014_2013
		2016_2015				2020_2019	
		2018_2017					
H-index and bibliometric	2007	2008_2007	2011_2010	Scientometric and gender	2014	2015_2014	2017_2016
		2013_2012				2020_2019	
		2021_2020					

Table 3 (continued)

Keyword pairs	Generation years	Growth years	Obsolence years	Keyword pairs	Generation years	Growth years	Obsolence years
Research evaluation and bibliometric	2009	2010_2009 2013_2012 2015_2014 2017_2016 2021_2020	2012_2011/2016_2015/2020_2019	Altmetric and bibliometric	2013	2014_2013	2016_2015 2019_2018

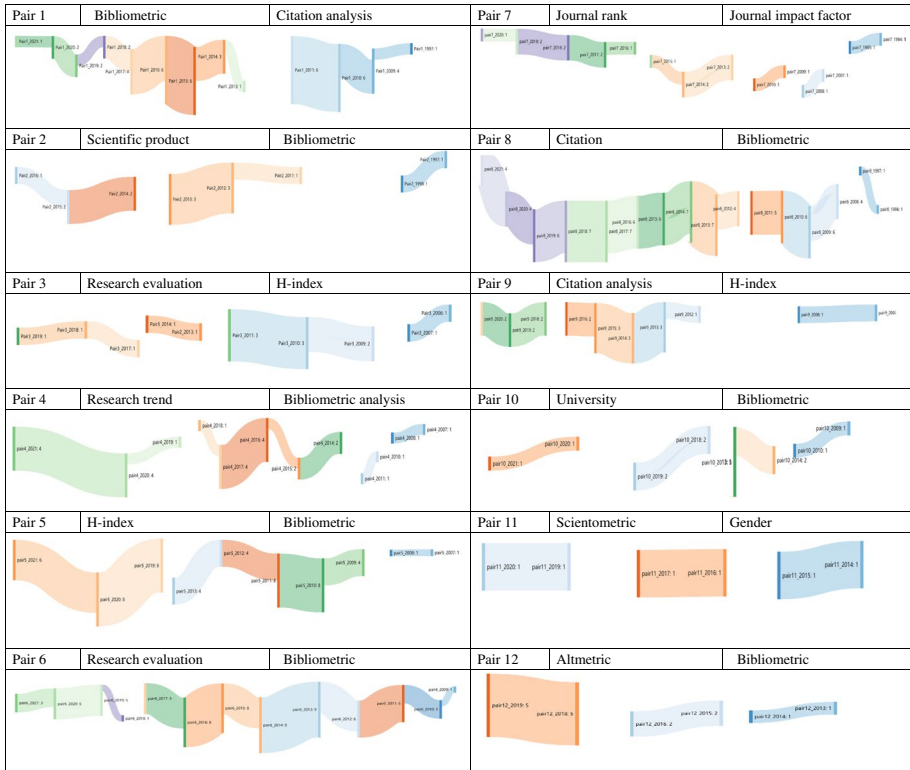


Fig. 11 Knowledge evolution flow based on direct keyword co-occurrence relationships

The pairs of keyword co-occurrences that fluctuate the most during the evolutionary process have a long history in the research field of informetrics. The keyword pair of *journal rank* and *journal impact factor* emerged in 1984 and experienced three growth stages after its generation (Pair 7). However, it has experienced two transitory obsolescence stages in recent years. As the basic knowledge elements for *bibliometric research*, *citation* and *citation analysis* have two pairs (1 and 8). The results reveal that Pairs 1 and 8 have the most growth stage fragments, indicating that the bibliometrics- and citation-related knowledge pairs are permanently influential. Additionally, Pair 11 indicates research on scientometrics and gender issues, and Pair 12 indicates research on bibliometrics and altmetrics issues. Pair 11 and 12 are recent hot topics in the research field of informetrics, and the number of their growth stage fragments is less than for traditional informetrics knowledge, like Pair 9 (*citation analysis* and *h-index*) and Pair 5 (*h-index* and *bibliometric*). Regarding research evaluation, as a well-known application area of scientometrics, the evolutionary process of Pair 6 indicates the continuous and prosperous growth of the knowledge combination in the field of informetrics.

Knowledge evolutionary process based on the indirect keyword relationship

The indirect keyword relationship is based on the same author trace whose relationships were constructed using the keyword as the knowledge pair element, and the two selected

keywords were provided by the same author but in a different paper. Previous work has proved that articles by the same author share similar knowledge or ideas (Koppel & Winter, 2014; Mihaljević & Santamaría, 2021). In this way, the keyword pair represents that the author contributes to a short fragment of knowledge evolution.

Table 4 lists 10 representative indirect keyword pairs with the most same authors, including five indirect keyword pairs with the most same authors for the generation stage and five for the growth stage. Columns 3, 5, and 7 list the year of generation, growth stage, and obsolescence, respectively. Column 4 lists the same authors' ScopusID for generation status, and Column 6 is for the growth stage. The number in parentheses is the number of the same authors. The year of obsolescence stage is listed without the same author trace because the obsolescence stage of the indirect keyword pair is not necessarily the result of the same author trace. Then, Fig. 12 presents a detailed evolutionary flow of each indirect keyword relationship paired by the same author trace. The direction of the evolutionary flow is from right to left, which is the same as in the Sankey diagram principle in Fig. 11.

Table 4 and Fig. 12 indicate that most pairs share different same author traces at different stages of evolution, suggesting that some authors have derived new combinations of knowledge elements, and some other groups of authors have followed these pioneers. Pairs 1, 3, 4, 5, and 7 have experienced more growth in recent years than obsolescence, and fewer researchers have contributed to the indirect keyword pairs than at the beginning. Pairs 2, 6, 8, 9, and 10 show different evolutionary patterns with Pair 1, where Pair 2 experienced more obsolescence than growth. However, more researchers have contributed to the indirect keyword pairs than at the beginning. Additionally, the evolutionary flow in Fig. 12 is sparser than in Fig. 11, which indicates that the scale of knowledge evolution based on the indirect keyword relationship with the same author trace is smaller than that identified by direct keyword co-occurrence relationships.

Knowledge evolution based on indirect keyword pair-based citation relationships

In relationships based on keyword pairs and citations, keyword pairs in the same papers were treated as nodes, and the edge denotes the paper citation relationship. Tables 5 and 6 reveal citing keyword pairs, cited keyword pairs, and citation years, demonstrating a citation relationship from the cited keyword pair to the citing keyword pair in the year.

Table 5 displays the keyword pairs eligible for the knowledge intergrowth process, indicating that these cited and citing keyword pairs experience a temporal growth stage as long as the citation relationship appears. The knowledge pairs represented by *crown indicator* and *normal* and by *citation* and *normal* is a case of intergrowth process, indicating to the researchers that the normalization of citation-related indicators is becoming more discussed in informetrics. Moreover, the intergrowth process between the knowledge pair represented by *bibliometric* and *scientific collaboration* and by *coauthorship network* and *scientific collaboration* indicates that research about scientific collaboration is becoming more popular. Further, web of knowledge turns out to be an important tool for informetrics research.

Table 6 presents the keyword pairs eligible for the knowledge transfer process. The cited keyword pairs experience a temporal obsolescence stage as long as the citation relationship appears. The knowledge pairs represented by *bibliometric* and *h-index* and by *p-index* and *bibliometric* are a case of the transfer process, demonstrating a slight temporal reduction in the research focus on the h-index replaced with the p-index.

Table 4 Knowledge evolutionary process based on indirect keyword relationships: paired by the same author trace

No	Keyword pair	Generation year	Same author ID for pair generation (number of same authors)	Growth year	Same author ID for pair growth stage (number of same authors)	Obsolescence year
1	Citation analysis and normal	2010	(10): 7004058552; 7,102,107,028; 7102107028; 14632830700; 14632830700; 14632651000; 14632651000; 35611695700; 35611695700	2011_2010 2013_2012 2016_2015	(3): 8850037200; 7003954276; 7005502691	2017_2016
2	Bibliometric and h-index	2008	(7): 7404414337; 7404414337; 7103385531; 7202433367; 7202433367; 7402094045; 7402094045	2009_2008	(14): 7102347190; 14031883700; 14031883700; 36806096600; 7005436626; 7005436626; 24765951700; 57192167279; 7006458179; 6507407139; 6507407139; 7005500864; 7005500864; 6602256988	2011_2010 2012_2011 2016_2015 2017_2016 2020_2019 2021_2020
3	Impact factor and Scopus	2013	(7): 8396445300; 8396445300; 16302963700; 16302963700; 57221391750; 57221391750; 6603451410	2014_2013 2018_2017	(3): 14031883700; 24822361200; 7005500864	2015_2014
4	Co-citation analysis and scientometric	2017	(7): 57193352174; 56648784400; 7501950297; 55729771100; 57201024901; 57189064272; 57189064272	2018_2017 2021_2020	(1): 55665562600	2019_2018

Table 4 (continued)

No	Keyword pair	Generation year	Same author ID for pair generation (number of same authors)	Growth year	Same author ID for pair growth stage (number of same authors)	Obsolescence year
5	Network analysis and scientometric	2009	(6): 14060977700; 57194647093; 6507380175; 6507380175; 10043514200	2010_2009 2012_2011 2015_2014 2018_2017	(1): 55665562600	2019_2018
6	Research collaboration and bibliometric	2011	(5): 56962739400; 22833445200; 22833445200; 26654558800	2018_2017	(5): 57194647093; 56962739400; 56962739400; 22833445200; 22833445200	2012_2011 2014_2013 2015_2014 2016_2015 2020_2019
7	Main path analysis and bibliometric	2017	(4): 56959839200; 57085448800; 57192996020; 57193001656	—	(2): 24765951700; 57192167279	2018_2017
8	Web of Science and bibliometric	2008	(2): 9746425500; 8850037200	2009_2008 2011_2010 2013_2012	(13): 8396445300; 8396445300; 16302963700; 16302963700; 57221391750; 57221391750; 7402554934; 7402554934; 8396445100; 8396445100; 8850037200; 8850037200; 7005958879	2016_2015 2018_2017 2019_2018 2020_2019 2021_2020

Table 4 (continued)

No	Keyword pair	Generation year	Same author ID for pair generation (number of same authors)	Growth year	Same author ID for pair growth stage (number of same authors)	Obsolescence year
9	Research evaluation and bibliometric	2009	(2): 8704727600; 7404896227	2010_2009	(8): 56823197200; 56823197200; 56962739400; 56962739400; 56717393900; 56717393900; 22833445200; 22833445200	2013_2012 2015_2014 2017_2016 2018_2017 2020_2019 2021_2020
10	Bibliometric and Scopus	2007	(1): 57202052068	2008_2007 2011_2010 2013_2012	(10): 8396445300; 8396445300; 16302963700; 16302963700; 57194647093; 57194647093; 57221391750; 57221391750; 57203026911; 57203026911	2015_2014 2016_2015 2018_2017 2019_2018

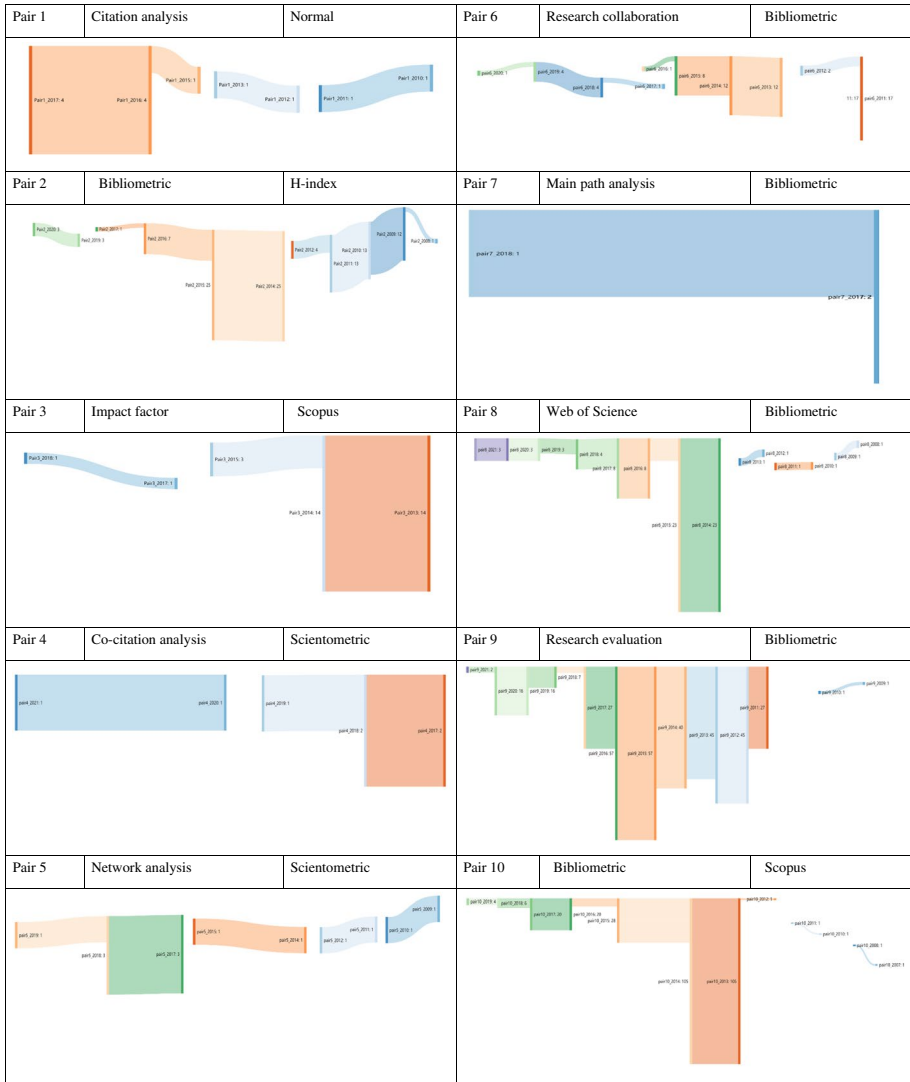


Fig. 12 Knowledge evolution flow based on indirect keyword relationships paired by the same author

Dirichlet multinomial regression topic trend

In this section, we set the publication date as the observed feature of the documents of the DMR topic model, which has a log-linear prior on the document-topic distribution function of the observed features of the document. Further, the DMR was applied by analyzing the distribution of the top ten topics from 1978 to 2020. The topic distribution for DMR topic modeling is presented. Figure 13 depicts the relative proportion of these 10 topic clusters. Table 7 lists the word in each topic cluster. The results of the DMR

Table 5 Keyword pairs conforming to the knowledge intergrowth process

No	Citing keyword pair	Cited keyword pair	Citation year	No	Citing keyword pair	Cited keyword pair	Citation year
1	Journal rank & Google Scholar	H-index & rank	2010	13	Impact & citation	Energy & quality	2012
2	Normal & citation	Normal & crown indicator	2011	14	Impact & citation	Quality & quantity	2012
3	Crown indicator & normal	Impact factor & bibliometric	2011	15	G-index & bibliometric	Energy & quality	2012
4	Crown indicator & normal	Citation & normal	2011	16	Quality & quantity	Energy & quality	2012
5	Impact factor & bibliometric	Normal & crown indicator	2011	17	G-index & bibliometric	Quality & quantity	2012
6	Impact factor & bibliometric	Citation & normal	2011	18	Bibliometric & scientific collaboration	Coauthorship network & scientific collation	2012
7	Web of knowledge & research evaluation	Scientometric & web of knowledge	2012	19	Citation & product	Citation & evaluation	2012
8	Web of knowledge & research product	Scientometric & web of knowledge	2012	20	Citation & evaluation	Citation & product	2012
9	Web of knowledge & scientometric	Research evaluation & web of knowledge	2012	21	Impact factor & bibliometric	Journal evaluation & impact factor	2012
10	Web of knowledge & research product	Research evaluation & web of knowledge	2012	22	Research trend & scientometric	Social network analysis & bibliometric	2015
11	Web of knowledge & research evaluation	Web of knowledge & research product	2012	23	Gender differ & scientific product	Gender & bibliometric	2015
12	Web of knowledge & scientometric	Web of knowledge & research product	2012	24	Impact & scientific product	Gender & bibliometric	2015

Table 6 Keyword pairs conforming to the knowledge transfer process

No	Citing keyword pair	Cited keyword pair	Citation year
1	H-index & rank	H-index & research evaluation	2010
2	P-index & bibliometric	Bibliometric & h-index	2011
3	Citation analysis & bibliometric	Research evaluation & h-index	2011
4	H-index & citation	Research evaluation & bibliometric	2012

only describe the growth or obsolescence trend. The trends of the top 10 topics have increasing, consistent, or decreasing patterns.

Table 8 shows the change of topic terms of five topics since 2016–2020. The five topic clusters are generated from each year’s keyword information. It can be seen that topic terms in each cluster change over time. The term that scientific collaboration appears in the topic cluster #0 with terms “bibliometrics, citations, interdisciplinarity, and economics” in 2016 and then moved to the topic cluster #1 with terms “bibliometrics, altmetrics, scientometrics, and web of science” in 2017. After that, the term scientific collaboration transferred into topic cluster#1 in 2018, topic cluster #1 in 2019, and topic cluster #4 in 2020. However, how the relationship changes between the term scientific collaboration and others remain unknown.

Compared with the topic evolutionary process of DMR topic modeling, the knowledge evolutionary process shaped using keyword pairs and multiple relationships could describe a more fine-grained evolutionary process. The knowledge components are represented by keyword pairs, which is comprehensible. The changes in the number of keyword pairs could be measured by the smallest time interval because a small number of new papers is enough to cause a change in the number of keyword pairs. Topic words involved in each topic cluster are numerous and highly overlapped. The overlap is high because the scope of informetrics is not large, and research paradigms have not undergone many dramatic changes. Due to a large and highly overlapping distribution of topic words, changes in the topics require the participation of more papers and takes more time to accumulate.

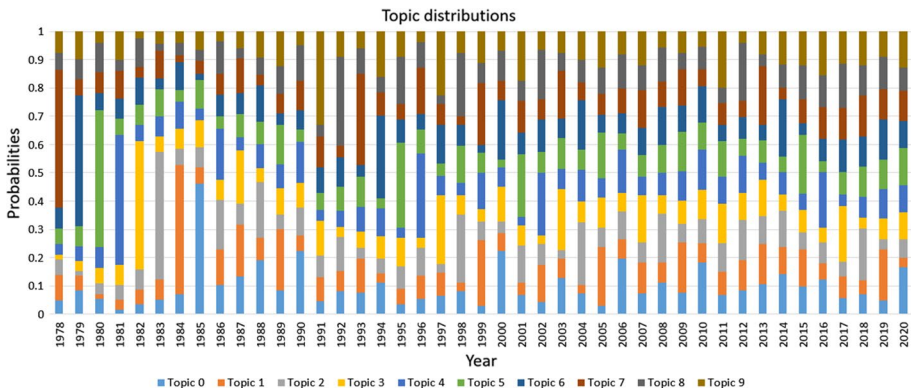


Fig. 13 Topic distribution (Dirichlet multinomial regression topic model)

Table 7 Top ten words for each topic

Topic 0	Topic 1	Topic 2	Topic 3	Topic 4
Impact	Research	Research	Research	H-index
Number	Impact	Scientific	Academic	Access
Patent	Citation	Science	Science	Media
Citation	Journal	Scientists	H-index	Research
Scientific	Information	Indicators	Articles	Women
Paper	Factor	Literature	Data	Cited
Index	Collaboration	Distribution	Network	Springer
Research	Publications	Data	Scopus	Peer-review
Countries	Quality	Paper	Global	Journal
Science	Performance	National	Metrics	Publications
Topic 5	Topic 6	Topic 7	Topic 8	Topic 9
Journal	Academic	International	Number	Science
Citation	Research	Impact	Publication	Journals
Web	Faculty	Journal	Study	Analysis
Assessment	Model	Indicators	Ranking	Research
Country	Gender	Articles	Article	Citation
Bibliometric	Authors	Author	Citations	Scientific
Significant	Education	Data	Scientific	Articles
Model	Field	Citation	Index	Factors
Evaluation	Publication	Field	Field	Index
Influence	Trends	Paper	Network	Activity

Discussion

In this research, three relationships represent knowledge evolution: direct keyword relationships paired with keyword co-occurrence pairs, indirect keyword relationships paired with the same author trace, and indirect keyword citation relationships paired with the citation relationship. The former two offer insight into the knowledge evolutionary process with the annual change in the quantity of the same keyword pairs. The latter provides more information about how different keyword pairs interact during the knowledge evolutionary process.

First, a direct keyword relationship was constructed with keyword co-occurrence pairs to depict the knowledge components. The knowledge evolutionary process can be measured by the annual changes in the number of keyword co-occurrence pairs (Kim et al., 2021). The results demonstrate that annual increments in the quantity of keyword co-occurrence pairs can demonstrate three evolutionary stages: new generation, growth, and obsolescence. At least one growth stage is required for effective generation identification to ensure the meaningful new generation of knowledge pairs. A nonnegative increment represents a fragment of the growth stage, and a positive increment represents a fragment of the obsolescence stage. In this way, the fragmentary knowledge evolutionary process was measured in the pair view in informetrics. Moreover, the year is

Table 8 Top five topic clusters from 2016 to 2020

Year	Top five topic clusters (top five terms of each topic)
2016	Topic #0: bibliometrics. scientific collaboration. citations. interdisciplinarity. economics
	Topic #1: bibliometrics. web of science. scopus. research productivity. citations
	Topic #2: network analysis. scientometrics. bibliometric indicators. research performance. universities
	Topic #3: social network analysis. co-authorship. h-index. citation. indicators
	Topic #4: scientometrics. scopus. research output. efficiency. bibliometrics
2017	Topic #0: Citation analysis. open access. library and information science. co-word analysis. universities
	Topic #1: Bibliometrics. altmetrics. scientometrics. scientific collaboration. web of science
	Topic #2: citations. bibliometric analysis. research collaboration. scholarly communication
	Topic #3: network analysis. bibliometric. productivity. impact. research evaluation
	Topic #4: bibliometrics. citation. h-index. journal impact factor. gender
2018	Topic #0: interdisciplinarity. google scholar. scientific collaboration. bibliometrics. collaboration
	Topic #1: scopus. citations. bibliometric analysis. research performance. citation analysis
	Topic #2: bibliometrics. scientometrics. citation analysis. peer review. journal impact factor
	Topic #3: co-authorship. machine learning. china. impact factor. psychology
	Topic #4: bibliometrics. research productivity. citation impact. altmetrics. web of science
2019	Topic #0: scientometrics. impact factor. scopus. library and information science. bibliometrics
	Topic #1: h-index. bibliometric analysis. citation analysis. scopus. google scholar. scientific collaboration
	Topic #2: bibliometrics. citations. research evaluation. text mining. information retrieval. field normalization
	Topic #3: bibliometrics. research productivity. scientometrics. ranking. self-citation
	Topic #4: bibliographic coupling. evaluation. india. science policy. g-index
2020	Topic #0: scientometric. citation. scientific impact. gender disparity. bibliometric indicators
	Topic #1: bibliometrics. open access. scholarly communication. web of science. fractional counting
	Topic #2: bibliometrics. h-index. china. scopus. citation analysis
	Topic #3: social network analysis. web of science. citation. h-index. gender
	Topic #4: scientometrics. bibliometric analysis. research evaluation. bibliometrics. scientific collaboration

an effective evolutionary interval. Such an evolutionary fragment can shorten the interval to months or weeks for rapidly changing knowledge domains.

Second, papers completed by the same author in a similar period have higher knowledge similarity. The indirect keywords relationship based on the same author trace determines keyword pairs provided by the same author in a different paper. Changes in researchers can be identified in the same knowledge evolutionary process. The results reveal that in the evolutionary process of most knowledge pairs, the authors who promote them are not those who first proposed them.

Finally, the keyword pair-based citation relationship provides a citation-based trace for knowledge evolution because the citation trace has been proved to be available for knowledge diffusion measurement (Yu et al., 2010). Keyword pair-based citation relationships reveal how keyword co-occurrence pairs interact with each other during the evolutionary process. According to the keyword pair-based citation relationship, two evolutionary

statuses exist: knowledge pair transfer and intergrowth. Knowledge pair transfer represents a knowledge flow from a cited to citing keyword pair, and the cited pair experiences segmental obsolescence as the citation relationship disappears. Tracing the continuous knowledge transfer process is beneficial to identify the frontier knowledge and ensure research innovation. Knowledge pair intergrowth indicates that cited and citing keyword pairs all experienced segmental growth stage along with the citation relationship appears. Intergrowth borrows from biology and refers to a mutually beneficial evolutionary connection between two sets of knowledge over a period. Identifying knowledge intergrowth pairs helps determine potentially influential combinations of scientific knowledge. References can be considered as knowledge providers (Wu et al., 2017), the citing papers are knowledge recipients. The citation network between citing papers and references has been applied to trace the diffusion process of scientific knowledge, which contains the knowledge meme. A knowledge meme is a text unit in scientific publications, which is distributed in many citing publications without being broken apart or altered (Kuhn et al., 2014; Mao et al., 2020). In this way, the citation relationship from the cited keyword pair to the citing keyword pair indicate a hereditary of combination of knowledge gene. Kuhn's the structure of scientific revolution (1962a, 1962b) explained that science is an episodic model where periods of conceptual continuity are interrupted by periods of revolutionary science. Therefore, the keyword combination pair is dynamic and ever-changing over time. The continuous knowledge combination has become an important meme supporting the development of the present knowledge domain. For sociological perspective, citing relations may infer networks of diffusion and influence (Gomez-Rodriguez et al., 2012). Link the social survey keyword pair from different time series, can clarify the relationship between things, and propose the possible social processes plan to achieve the most beneficial social outcomes. In addition, tracking the events provided a coherent representation of the news cycle (Dou et al., 2012). For political sides, tracking the information epidemics in the social media identified the political discussions community (Guerrero-Solé, 2017), and the political polarization on Twitter (Conover et al., 2011).

Conclusions

A single keyword using the frequency of a continuous period has difficulty exhibiting the structure of knowledge evolution. Thus, in this work, the direct co-occurrence relationship, indirect relationship by keyword pair citation relationship, and same author trace were analyzed, which could mine the evolutionary process quantitatively from a different perspective.

In the results, an empirical study of the informetric field was analyzed. The five evolutionary process stages include knowledge generation, growth, obsolescence, transfer, and intergrowth. Further, a DMR-based topic trend was compared with those results. First, knowledge evolution is a process of rises and falls, not a continuously smooth state. The life cycle of knowledge is longer than the research topic, and old knowledge can be renewed in recent years. The keyword co-occurrence pair (Table 3) of *journal rank* and *journal impact factor* emerged in 1984 and experienced three growth stages after its generation (Pair 7).

Second, prolocutors and promoters of knowledge are not the same scholars during the evolutionary process shaped by keyword pairs. Table 5 and Fig. 12 demonstrate that most pairs share different same author traces at different stages of knowledge evolution,

suggesting that some authors generated new combinations of knowledge elements and that other authors followed these pioneers.

Third, citations trace the knowledge diffusion path and provide insight into how different knowledge pairs interact. In relationships based on keyword pairs and citations, the pairs (*crown indicator* and *normal*) and (*citation* and *normal*) and the pairs (*bibliometric* and *scientific collaboration*) and (*coauthorship network* and *scientific collaboration*) are two typical sets of intergrowth keyword pairs. Compared with relationships based on keyword pairs and citations, the DMR topic trend has difficulty mining the knowledge transform term at the micro-level, especially in the migration and intergrowth processes.

The main limitation of this study is that we only used informetrics papers and related reference papers. Future research should focus on verifying the applicability of the datasets for multiple disciplines.

Acknowledgements This study was partially funded by the National Natural Science Foundation of China (NSFC) Grant No. 72104220. This work was also supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2020S1A5B1104865).

References


- Abramo, G., & D'Angelo, C. A. (2020). The domestic localization of knowledge flows as evidenced by publication citation: The case of Italy. *Scientometrics*, *125*(2), 1305–1329.
- Abrishami, A., & Aliakbary, S. (2019). Predicting citation counts based on deep neural network learning techniques. *Journal of Informetrics*, *13*(2), 485–499.
- Allee, V. (2012). *The knowledge evolution*. Routledge.
- Arthur, W. B. (2009). *The nature of technology: What it is and how it evolves*. Simon and Schuster.
- Bar-Ilan, J. (2008). Informetrics at the beginning of the 21st century—A review. *Journal of Informetrics*, *2*(1), 1–52.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *The Journal of Machine Learning Research*, *3*, 993–1022.
- Börner, K., Maru, J. T., & Goldstone, R. L. (2004). The simultaneous evolution of author and paper networks. *Proceedings of the National Academy of Sciences*, *101*(suppl 1), 5266–5273.
- Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology*, *110*(2), 349–399.
- Burton, R. E., & Kebler, R. W. (1960). The “half-life” of some scientific and technical literatures. *American Documentation*, *11*(1), 18–22.
- Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Proceedings of the 22nd international conference on neural information processing systems (NIPS)* (pp. 288–296).
- Chen, C., & Hicks, D. (2004). Tracing knowledge diffusion. *Scientometrics*, *59*(2), 199–211.
- Cheng, Q., Wang, J., Lu, W., Huang, Y., & Bu, Y. (2020). Keyword-citation-keyword network: A new perspective of discipline knowledge structure analysis. *Scientometrics*, *124*(3), 1923–1943.
- Choi, J. M. (1988). Citation analysis of intra-and interdisciplinary communication patterns of anthropology in the USA. *Behavioral & Social Sciences Librarian*, *6*(3–4), 65–84.
- Choudhury, N., & Uddin, S. (2016). Time-aware link prediction to explore network effects on temporal knowledge evolution. *Scientometrics*, *108*(2), 745–776.
- Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on twitter. In *5th international AAAI conference on weblogs and social media*.
- D'Angelo, C. A., & van Eck, N. J. (2020). Collecting large-scale publication data at the level of individual researchers: A practical proposal for author name disambiguation. *Scientometrics*, *123*(2), 883–907.
- Dosi, G. (1982). Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. *Research Policy*, *11*(3), 147–162.
- Dou, W., Wang, X., Ribarsky, W., & Zhou, M. (2012). Event detection in social media data. In *IEEE VisWeek workshop on interactive visual text analytics-task driven analytics of social media content* (pp. 971–980).
- Egghe, L. (2005). Expansion of the field of informetrics: Origins and consequences. *Information Processing and Management*, *41*(6), 1311–1316.

- Figuerola, C. G., Marco, F. J. G., & Pinto, M. (2017). Mapping the evolution of library and information science (1978–2014) using topic modeling on LISA. *Scientometrics*, *112*(3), 1507–1535.
- Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science*, *47*(1), 117–132.
- Furman, J. L., & Stern, S. (2011). Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *American Economic Review*, *101*(5), 1933–1963.
- Garfield, E. (1965). Can citation indexing be automated? In *Statistical association methods for mechanized documentation, symposium proceedings* (Vol. 269, pp. 189–192).
- Gomez-Rodriguez, M., Leskovec, J., & Krause, A. (2012). Inferring networks of diffusion and influence. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *5*(4), 1–37.
- Gosnell, C. F. (1944). Obsolescence of Books in College Libraries. *College & Research Libraries*, *5*(2), 115–125.
- Guerrero-Solé, F. (2017). Community detection in political discussions on Twitter: An application of the retweet overlap network method to the Catalan process toward independence. *Social Science Computer Review*, *35*(2), 244–261.
- Hassan, S. U., Safder, I., Akram, A., & Kamiran, F. (2018). A novel machine-learning approach to measuring scientific knowledge flows using citation context analysis. *Scientometrics*, *116*(2), 973–996.
- Higham, K. W., Governale, M., Jaffe, A. B., & Zülicke, U. (2017). Unraveling the dynamics of growth, aging and inflation for citations to scientific articles from specific research fields. *Journal of Informetrics*, *11*(4), 1190–1200.
- Hoch, P. K. (1985). Migration and the generation of new scientific ideas. *Minerva*, *25*, 209–237. <https://doi.org/10.1007/bf01097783>
- Hu, J., & Zhang, Y. (2015). Research patterns and trends of recommendation system in China using co-word analysis. *Information Processing & Management*, *51*(4), 329–339.
- Hu, K., Wu, H., Qi, K., Yu, J., Yang, S., Yu, T., Zheng, J., & Liu, B. (2018). A domain keyword analysis approach extending term frequency-keyword active index with Google word2vec model. *Scientometrics*, *114*(3), 1031–1068.
- Jung, S., & Yoon, W. C. (2020). An alternative topic model based on common interest authors for topic evolution analysis. *Journal of Informetrics*, *14*(3), 101040.
- Khasseh, A. A., Soheili, F., Moghaddam, H. S., & Chelak, A. M. (2017). Intellectual structure of knowledge in iMetrics: A co-word analysis. *Information Processing & Management*, *53*(3), 705–720.
- Kim, M., Baek, I., & Song, M. (2018). Topic diffusion analysis of a weighted citation network in biomedical literature. *Journal of the Association for Information Science and Technology*, *69*(2), 329–342.
- Kim, M. C., Feng, Y., & Zhu, Y. (2021). Mapping scientific profile and knowledge diffusion of Library Hi Tech. *Library Hi Tech*, *39*(2), 549–573. <https://doi.org/10.1108/LHT-08-2019-0164>
- Klavans, R., & Boyack, K. W. (2009). Toward a consensus map of science. *Journal of the American Society for Information Science and Technology*, *60*, 455–476.
- Koppel, M., & Winter, Y. (2014). Determining if two documents are written by the same author. *Journal of the Association for Information Science and Technology*, *65*(1), 178–187.
- Kuhn, T. (1962a). The nature and necessity of scientific revolutions, from the structure of scientific revolutions. In *The philosophy of science* (pp. 148–157). MIT Press.
- Kuhn, T. S. (1962b). *The structure of scientific revolutions* (1st ed., p. 3). University of Chicago Press. ISBN:0-226-45807-5.
- Kuhn, T., Perc, M., & Helbing, D. (2014). Inheritance patterns in citation networks reveal scientific memes. *Physical Review X*, *4*(4), 041036.
- Lee, D., Kim, W. C., Charidimou, A., & Song, M. (2015). A bird’s-eye view of Alzheimer’s disease research: Reflecting different perspectives of indexers, authors, or citers in mapping the field. *Journal of Alzheimer’s Disease*, *45*(4), 1207–1222.
- Lee, K., Kim, S., Kim, E. H. J., & Song, M. (2017). Comparative evaluation of bibliometric content networks by tomographic content analysis: An application to Parkinson’s disease. *Journal of the Association for Information Science and Technology*, *68*(5), 1295–1307.
- Lee, P.-C., Su, H.-N., & Chan, T.-Y. (2010). Assessment of ontology-based knowledge network formation by Vector-Space Model. *Scientometrics*, *85*(3), 689–703.
- Leydesdorff, L., de Moya-Anegón, F., & de Nooy, W. (2016). Aggregated journal–journal citation relations in scopus and web of science matched and compared in terms of networks, maps, and interactive overlays. *Journal of the Association for Information Science and Technology*, *67*(9), 2194–2211.
- Liang, G., Hou, H., Lou, X., & Hu, Z. (2019). Qualifying threshold of “take-off” stage for successfully disseminated creative ideas. *Scientometrics*, *120*(3), 1193–1208.

- Liu, J., Grubler, A., Ma, T., & Kogler, D. F. (2021). Identifying the technological knowledge depreciation rate using patent citation data: A case study of the solar photovoltaic industry. *Scientometrics*, *126*(1), 93–115.
- Liu, Y., Yang, L., & Chen, M. (2021b). A new citation concept: Triangular citation in the literature. *Journal of Informetrics*, *15*(2), 101141.
- Loasby, B. J. (2002). The evolution of knowledge: Beyond the biological model. *Research Policy*, *31*(8–9), 1227–1239.
- Lu, W., Huang, S., Yang, J., Bu, Y., Cheng, Q., & Huang, Y. (2021). Detecting research topic trends by author-defined keyword frequency. *Information Processing & Management*, *58*(4), 102594.
- Lu, W., Liu, Z., Huang, Y., Bu, Y., Li, X., & Cheng, Q. (2020). How do authors select keywords? A preliminary study of author keyword selection behavior. *Journal of Informetrics*, *14*(4), 101066.
- Mao, J., Liang, Z., Cao, Y., & Li, G. (2020). Quantifying cross-disciplinary knowledge flow from the perspective of content: Introducing an approach based on knowledge memes. *Journal of Informetrics*, *14*(4), 101092.
- Mihaljević, H., & Santamaría, L. (2021). Disambiguation of author entities in ADS using supervised learning and graph theory methods. *Scientometrics*, *126*(5), 3893–3917.
- Mimno, D., & McCallum, A. (2012). Topic models conditioned on arbitrary features with Dirichlet-multinomial regression. arXiv preprint [arXiv:1206.3278](https://arxiv.org/abs/1206.3278).
- Mina, A., Ramlogan, R., Tampubolon, G., & Metcalfe, J. S. (2007). Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge. *Research Policy*, *36*(5), 789–806.
- Modis, T. (2007). Strengths and weaknesses of S-curves. *Technological Forecasting and Social Change*, *74*, 866–872.
- Mukherjee, S., Uzzi, B., Jones, B. F., & Stringer, M. (2017). How atypical combinations of scientific ideas are related to impact: The general case and the case of the field of geography. In *Knowledge and Networks* (pp. 243–267). Springer.
- Muñoz-Écija, T., Vargas-Quesada, B., & Rodríguez, Z. C. (2019). Coping with methods for delineating emerging fields: Nanoscience and nanotechnology as a case study. *Journal of Informetrics*, *13*(4), 100976.
- Parolo, P. D. B., Pan, R. K., Ghosh, R., Huberman, B. A., Kaski, K., & Fortunato, S. (2015). Attention decay in science. *Journal of Informetrics*, *9*(4), 734–745.
- Peset, F., Garzón-Farínos, F., González, L. M., García-Massó, X., Ferrer-Sapena, A., Toca-Herrera, J. L., & Sánchez-Pérez, E. A. (2020). Survival analysis of author keywords: An application to the library and information sciences area. *Journal of the Association for Information Science and Technology*, *71*(4), 462–473.
- Prebor, G. (2010). Analysis of the interdisciplinary nature of library and information science. *Journal of Librarianship and Information Science*, *42*(4), 256–267.
- Pu, T., Huang, M., & Yang, J. (2021). Migration knowledge graph framework and its application. *Journal of Physics: Conference Series*, *1955*(1), 012071.
- Qian, Y., Liu, Y., & Sheng, Q. Z. (2020). Understanding hierarchical structural evolution in a scientific discipline: A case study of artificial intelligence. *Journal of Informetrics*, *14*(3), 101047.
- Réale, D., Khelifaoui, M., Montiglio, P. O., et al. (2020). Mapping the dynamics of research networks in ecology and evolution using co-citation analysis (1975–2014). *Scientometrics*, *122*, 1361–1385. <https://doi.org/10.1007/s11192-019-03340-4>
- Sedighi, M. (2016). Application of word co-occurrence analysis method in mapping of the scientific fields (case study: The field of informetrics). *Library Review*, *65*(1/2), 52–64.
- Shi, H. W., & Wang, F. N. (2009). Knowledge intergrowth and tech-information utility optimization in the process of endogenous growth. *Studies in Science of Science*, *27*(11), 1700–1711. (in Chinese).
- Silvello, G. (2018). Theory and practice of data citation. *Journal of the Association for Information Science and Technology*, *69*(1), 6–20.
- Su, H. N., & Lee, P. C. (2010). Mapping knowledge structure by keyword co-occurrence: A first look at journal papers in technology foresight. *Scientometrics*, *85*(1), 65–79.
- Swanson, D. R. (1993). Intervening in the life cycles of scientific knowledge. *Library Trends*, *41*(4), 606–632.
- Uddin, S., & Khan, A. (2016). The impact of author-selected keywords on citation counts. *Journal of Informetrics*, *10*(4), 1166–1177.
- Uddin, S., Khan, A., & Baur, L. A. (2015). A framework to explore the knowledge structure of multidisciplinary research fields. *PLoS ONE*, *10*(4), e0123537.

- Urbano, C., & Ardanuy, J. (2020). Cross-disciplinary collaboration versus coexistence in LIS serials: Analysis of authorship affiliations in four European countries. *Scientometrics*, *124*, 575–602. <https://doi.org/10.1007/s11192-020-03471-z>
- Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, *342*(6157), 468–472.
- van der Eijk, C. C., van Mulligen, E. M., Kors, J. A., Mons, B., & van den Berg, J. (2004). Constructing an associative concept space for literature-based discovery. *Journal of the American Society for Information Science and Technology*, *55*(5), 436–444.
- Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, *10*(2), 365–391.
- Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, *46*(8), 1416–1436.
- Wang, M., Zhang, J., Chen, G., & Chai, K. H. (2019). Examining the influence of open access on journals' citation obsolescence by modeling the actual citation process. *Scientometrics*, *119*(3), 1621–1641.
- Wu, S., & Wu, H. (2013). More powerful significant testing for time course gene expression data using functional principal component analysis approaches. *BMC Bioinformatics*, *14*(1), 1–13.
- Wu, C., Hill, C., & Yan, E. (2017). Disciplinary knowledge diffusion in business research. *Journal of Informetrics*, *11*(2), 655–668.
- Xie, Q., Zhang, X., & Song, M. (2021). A network embedding-based scholar assessment indicator considering four facets: Research topic, author credit allocation, field-normalized journal impact, and published time. *Journal of Informetrics*, *15*(4), 101201.
- Xu, J., Ding, Y., Bu, Y., Deng, S., Yu, C., Zou, Y., & Madden, A. (2019). Interdisciplinary scholarly communication: An exploratory study for the field of joint attention. *Scientometrics*, *119*(3), 1597–1619.
- Xu, Y., Zhang, S., Zhang, W., Yang, S., & Shen, Y. (2019). Research front detection and topic evolution based on topological structure and the PageRank algorithm. *Symmetry*, *11*(3), 310.
- Yu, G., Wang, M. Y., & Yu, D. R. (2010). Characterizing knowledge diffusion of Nanoscience & Nanotechnology by citation analysis. *Scientometrics*, *84*(1), 81–97.
- Yu, X., Li, G., & Chen, L. (2014). Prediction and early diagnosis of complex diseases by edge-network. *Bioinformatics*, *30*(6), 852–859.
- Zeng, T., Zhang, W., Yu, X., Liu, X., Li, M., Liu, R., & Chen, L. (2014). Edge biomarkers for classification and prediction of phenotypes. *Science China Life Sciences*, *57*(11), 1103–1114.
- Zhang, H., Kiranyaz, S., & Gabbouj, M. (2017). Outlier edge detection using random graph generation models and applications. *Journal of Big Data*, *4*(1), 1–25.
- Zhao, H., Xu, X., Song, Y., Lee, D. L., Chen, Z., & Gao, H. (2018). Ranking users in social networks with higher-order structures. In *32nd AAAI conference on artificial intelligence*.
- Zhu, L., Liu, X., He, S., Shi, J., & Pang, M. (2015). Keywords co-occurrence mapping knowledge domain research base on the theory of big data in oil and gas industry. *Scientometrics*, *105*(1), 249–260.

Authors and Affiliations

Xinyuan Zhang¹ · Qing Xie² · Chaemin Song³ · Min Song³ 

Xinyuan Zhang
zhangxinyuan@zzu.edu.cn

Qing Xie
xieqing@yonsei.ac.kr

Chaemin Song
chaemin11@yonsei.ac.kr

¹ Henan Academy of Big Data, Zhengzhou University, No. 75, Daxuebei-ro, Zhengzhou 450000, Henan, China

² School of Management, Shenzhen Polytechnic, Shenzhen 518055, Guangdong, China

³ Department of Library and Information Science, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea