Journal of Informetrics 16 (2022) 101343

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

MatrixSim: A new method for detecting the evolution paths of research topics

Xiaoguang Wang [a,b], Jing He [a], Han Huang [a,*], Hongyu Wang [c]

[a] School of Information Management, Wuhan University, Wuhan 430072, China
[b] Big Data Institute, Wuhan University, Wuhan 430072, China
[c] School of Management, Wuhan University of Technology, Wuhan 430070, China

ARTICLE INFO

Keywords:
Research topic evolution
Matrix similarity
Evolution path detection
MatrixSim
Co-word network

ABSTRACT

In this study, MatrixSim, a new method for detecting the evolution paths of research topics based on matrix similarity, was proposed. In the analysis of research topic evolution with the help of co-word networks, in contrast to traditional methods of topic evolution path detection, such as cosine similarity and edge similarity, MatrixSim is based on the local community structure of topic communities in co-word networks and considers the similarity of research topics in both nodes and edges, that is, words and inter-word relations. Using the library and information science field as an example, two sets of experiments were designed for topic similarity detection and subject-specific research topic evolution analysis to evaluate and verify the performance of MatrixSim in detecting the evolution paths of research topics and its validity and feasibility in research topic evolution analysis. The results confirm that MatrixSim performs well in detecting the evolution paths of research topics. It can correlate important research topics, help describe the research development process in scientific fields, reveal the internal evolutionary features of research topics, and thus discover and track the research frontiers in scientific fields. This study provides significant methodological support for researchers conducting prospective research activities.

1. Introduction

Research topic evolution aims to reflect the dynamics of research topic relevance, including content shifts in individual research topics and the relevance between research topics. Specifically, it profiles how topics change over time, such as whether they are developed maturely, import knowledge from other topics, merge or split into others, and which topics are gaining importance or dying out (Chen et al., 2017; Liu et al., 2020). Such knowledge evolution characteristics are of great significance for researchers to automatically, timely, and accurately clarify scientific domains, track the development process of topics, recognize relationships between topics, and grasp scientific frontiers and research trends (Huang et al., 2022; Katsurai & Ono, 2019; Zhang et al., 2022). The exploration and analysis of research topic evolution have always been a research hotspot in the field of informetrics (Ding & Stirling, 2016; Jung & Yoon, 2020). Several researchers use content analysis (Chu, 2015), model-based methods (Han, 2020; Xie et al., 2020), and bibliometric methods such as co-words (Callon et al., 1983), citations, or co-citations (Hou et al., 2018; Yang et al., 2016) to explore the distribution dynamics of research topics, and their research covers various aspects such as major research topic identification (Lu & Liu, 2016), inter-industry technology linkage (Yang et al., 2021), and disciplinary classification (Zhang et al., 2016).

* Corresponding author.
E-mail address: huanghan@whu.edu.cn (H. Huang).

https://doi.org/10.1016/j.joi.2022.101343
Received 15 April 2022; Received in revised form 11 October 2022; Accepted 13 October 2022
Available online 31 October 2022

In previous studies, our research team constructed a four-step framework for analyzing the evolution of research topics based on co-word networks, including topic representation, topic identification, evolution path detection, and visualization (Wang et al., 2014). In detail, the target literature set is divided into time periods in chronological order, and the research topics are identified from the keywords of each time period by using topic models, cluster analysis, or community identification based on co-word networks. The similarity of research topics between adjacent periods is analyzed to establish the association of research topics in different time periods, and finally, the association between research topics in each time period (i.e., the topic evolution relationship) is visualized. In this framework, evolution path detection is transformed into the calculation of the similarity between two research topics in adjacent periods, and an evolutionary relationship is considered to exist when the similarity is greater than a certain threshold. The analytical approach has been widely used in existing topic evolution analysis (Huang et al., 2022; Jung & Yoon, 2020), and similar methods have been used in the longitudinal analysis of research topics in scientific mapping tools, such as CiteSpace (Chen, 2006), VOSviewer (Van Eck & Waltman, 2009), Bibliometrix (Aria & Cuccurullo, 2017), and SciMAT (Cobo et al., 2012). When measuring the similarity between topics, it is mainly calculated from the perspective of node overlap or edge overlap of topic communities using methods such as cosine similarity, Jaccard index, and related deformation methods (Wang et al., 2014). Thus, such methods tend to consider only the same keywords or the same inter-word relations contained in research topics in adjacent periods without a comprehensive consideration of the keywords and their related relations. However, in co-word networks, keywords and their inter-word relations together constitute the representation of topics; analyzing only from the perspective of node overlap without considering the emergence, extinction, or weight changes of inter-word relations may overlook the shift in the research focus of specific research topics and the emergence of new research objects or methodological paradigms. Analyzing from the perspective of relationship overlap without considering the size and weight changes of nodes may also lead to the neglect of the core nodes, that is, the research focus (Behrouzi et al., 2020; Choudhury & Uddin, 2016). Based on such single-dimensional methods of calculating the similarity of research topics, it is difficult to detect the correlation between topics, and thus reflect the true overall picture of the evolution of research topics.

To address these limitations, a novel method for calculating the similarity of research topics, MatrixSim, was proposed for evolution path detection in topic evolution analysis based on previous research. The method is based on the local network structure of research topics in co-word networks, that is, the community structure, and closely combines the two-dimensional factors of nodes and relations through adjacency matrices instead of calculating them with simple weight assignment and addition, subtraction, multiplication, and division, avoiding the errors and limitations caused by manual assignment. To verify the effectiveness of MatrixSim and its application to the process of domain topic evolution analysis, an empirical analysis was conducted using the library and information science (LIS) field as an example. Under the guidance of domain experts, first, the set of research topic associations in the LIS field in adjacent periods was constructed, and the manually constructed standard set was used as the benchmark model. Then, the similarity of the topic communities in adjacent periods was calculated using MatrixSim and traditional similarity methods to identify the associations between research topics. The results were then compared with the manually constructed set to verify the effectiveness of MatrixSim. Furthermore, MatrixSim was used to conduct domain-oriented research topic evolution analysis on a broader dataset in the LIS field to further reveal the scientific validity and applicability of MatrixSim in research topic evolution analysis.

Concretely, this study aims to answer the following questions:

(1) How does MatrixSim calculate the correlation between research topics in adjacent periods? How does it differ from the traditional similarity methods?
(2) How does MatrixSim detect the evolution paths of research topics? How does it work?
(3) When portraying the evolution paths of research topics, can MatrixSim reflect the actual situation of field development?

The remainder of this paper is organized as follows. First, relevant work is reviewed, and then the methodology is presented. Next, the proposed method is verified. The experiments and results are introduced and the discussion and implications of this study are described. Finally, conclusions and directions for future work are presented.

2. Related work

2.1. Research topic evolution path detection

The purpose of research topic evolution path detection is to reveal the development context and evolution law of subject areas by tracking and analyzing the development and changing trends of research topics characterized by words in a time series and the interaction between different research topics (Park & Magee, 2017; Xu, 2020), which is an important step in the evolution analysis of research topics. Current studies mainly detect the evolution paths by judging the similarity between research topics in adjacent periods, and an evolutionary relationship is considered to exist when the similarity is greater than a threshold value given in advance, that is, an evolution path is detected (White & Jose, 2004). Existing measures of research topic similarity can be broadly divided into two basic approaches: the keyword-based approach and word-to-word relation-based approach. The keyword-based approach, also known as the node-based approach in co-word network analysis, is based on the consistency of keywords between research topics by calculating the cosine similarity (Chen et al., 2017; Jeong & Song, 2014; Koylu, 2019), Jaccard index (Jian et al., 2018; Wang et al., 2015), Kullback-Leibler (KL) scatter (Mei & Zhai, 2005), Jensen-Shannon (JS) scatter (Lee, 2001; Wu et al., 2021), Hellinger distance (Beykikhoshk et al., 2018), core node overlap (Cobo et al., 2011), and other methods to evaluate the similarity of topics. Bibliometrix (Aria & Cuccurullo, 2017) and SciMAT (Cobo et al., 2012) mainly adopt keyword-based methods in the longitudinal analysis of research topics, and SciMAT provides a variety of implementations, including the Jaccard index, Salton's

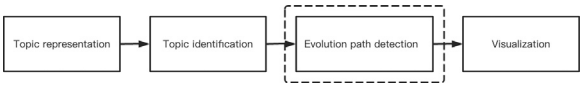

Fig. 1. General steps of research topic evolution analysis.

cosine measure, and equivalence index. The relation-based approach, that is, the edge-based approach in co-word network analysis, refers to the method of measuring the similarity between research topics by the consistency of the relationships between the keywords contained in the topics, which is usually implemented specifically with the help of Jaccard index. Wang et al., 2014 and Berger-Wolf and Saia (2006) have used the relation-based method in their studies.

In the analysis of research topic evolution based on co-word networks, the network community formed by nodes and their connected edges represents the research topic. However, existing studies only consider the unidimensional factors of nodes or edges when detecting the evolution paths and lack comprehensive consideration and analysis of keywords and their related relationships. Therefore, a similarity measure based on the network structure of topic communities that considers both nodes and edges between topics may be useful for improving the efficiency of evolution path detection toward better disciplinary knowledge discovery.

2.2. Matrix similarity

In graph theory, the topology of a network is usually represented by a matrix, and there have been several explorations and applications to judge the similarity between matrices. Similar to the vector inner product, the Frobenius inner product is also an important concept in algebraic operations. Chehab and Raydan (2008) pointed out that the Frobenius inner product allows the definition of the angle cosine between two given symmetric matrices. Thus, some researchers have proposed methods to measure the similarity between two matrices based on the Frobenius inner product and have applied them to different fields. Riaz et al. (2020) created a new similarity measure (SM) based on cosine similarity and the Frobenius inner product of matrices and then validated the method in the case of patient psychological disorder detection. Cristianini et al. (2006) introduced the concept of kernel alignment based on the Frobenius inner product to measure the similarity between two kernel functions or between a kernel and a target function, which was later developed as a more efficient optimization setup by (Lankriet et al. 2004). Mijangos et al. (2016, 2017) used a matrix similarity measure based on the Frobenius inner product in two studies of sentence similarity measure and document spectrum clustering, and both obtained better results than traditional similarity methods. In audio processing, Wang et al. (2010) used the spatial correlation captured by the covariance matrix of the mean super-vector for speaker verification and measured the similarity between speech utterances in terms of the spatial correlation through two kernel metrics, namely, the log-Euclidean inner product and Frobenius angle. Similarity matrices have also been used to identify repetitive elements in audio signals in order to achieve music/sound separation (Rafii and Pardo, 2012). In image processing, Demirc (2007) detected the color edges of color images based on similarity matrices. Wang et al. (2018) compared six graph similarity methods, including matrix similarity, and proposed two machine-learning-based graph similarity metrics. Yan et al. (2019) further proposed two similarity matrix construction methods, namely, the cosine-Euclidean similarity matrix and cosine-Euclidean dynamic weighted similarity matrix, in hyperspectral image clustering.

Applications of the Frobenius inner product and matrix similarity measures based on the Frobenius inner product in several fields show that the similarity between matrices can be measured based on the Frobenius inner product, and it is more effective than traditional similarity methods in certain specific tasks. Considering that the adjacency matrix can be used to characterize the network structure of topic communities in the analysis of research topic evolution based on co-word networks, in this study, the Frobenius inner product was applied to the similarity measure of research topics, and a new method for detecting the evolution paths of research topics, MatrixSim, was proposed. The feasibility and applicability of MatrixSim were verified by empirical cases.

3. Methodology

Fig. 1 illustrates the general steps of research topic evolution analysis. Focusing on the steps of evolution path detection, MatrixSim is proposed and applied to the evolution path detection of research topics, and its performance is explored in depth.

The methodological framework of the research topic evolution analysis adopted in this study is shown in Fig. 2. First, the scientific literature data in a specific field are collected, and key information is extracted, including keywords, titles, abstracts, and years. Subsequently, the extracted data are preprocessed, which involves word form conversion, acronym matching, and noise filtering, and the data are sliced by time. Second, the co-word network of each time period is constructed based on the author's keywords, that is, different authors' keywords in the literature of each time period are used as nodes. Then, the edges between two keywords appearing in the same paper are constructed, and the number of occurrences of the two keywords together is used as the edge weights. The constructed network is the co-word network of each time period, and the temporal co-word network sequence is formed according to the time order. Then, a community discovery algorithm is used to identify topics in each period. The next step is to detect the evolution paths, which is divided into two steps. The first step is to label the research topics in each time period, and the second step is to use MatrixSim to analyze the correlation between research topics in adjacent periods to obtain the research topic evolution paths, which is the focus of this study. Finally, the evolution paths of the research topics are visualized.

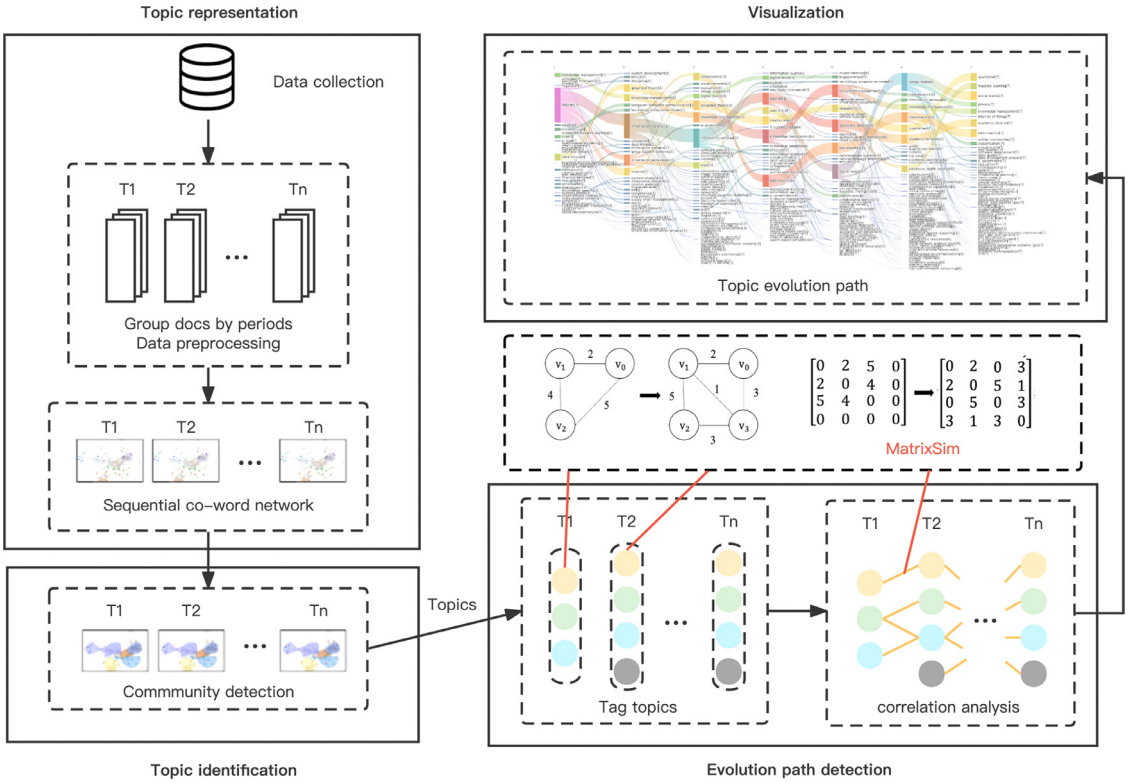

Fig. 2. Methodological framework of the research topic evolution analysis.

3.1. Definition of MatrixSim

The inner product of the vectors reflects the magnitude of the angle θ between vectors a, b, that is, $\langle a, b\rangle = \|a\| \cdot \|b\| \cdot \cos\theta$. The magnitude of the angle is a numerical index reflecting the correlation or similarity degree of two vectors, and its cosine value is commonly expressed as cosine similarity. Similarly, the matrix similarity, that is, the Frobenius inner product, reflects the angle between two matrices, and it is with the help of the Frobenius inner product that MatrixSim characterizes the similarity of two matrices.

The specific definition of MatrixSim is as follows:

For two matrices A and B with m rows and n columns, the Frobenius inner product $\langle \cdot, \cdot \rangle$ is defined as follows:

$$\langle A, B\rangle = tr\left(B^T A\right) \tag{1}$$

where $tr(\cdot)$ denotes the sum of the main diagonal elements of matrix. Similar to the inner product of the vectors, the Frobenius inner product denotes the sum of the products of the corresponding positional elements of the two matrices. From the Frobenius inner product, its parametric $\| \cdot \|$ can be derived as

$$\| A \| = \sqrt{\langle A, A\rangle} \tag{2}$$

In this way, MatrixSim r is defined as

$$r = cos\theta = \frac{\langle A, B\rangle}{\| A \| \cdot \| B \|} \tag{3}$$

where θ is the angle between the two matrices, and the range of r is $[-1, 1]$. When r tends to be closer to 1, the two matrices become more similar. When r tends to 0, it means the two matrices are less similar. When r is negative, it can be said to be negatively correlated in the physical sense, but in the similarity calculation, a negative correlation is often meaningless; thus, it is set to 0, that is, not similar.

3.2. MatrixSim in evolution path detection

From the perspective of research topic evolution, the evolution path detection of research topics can be transformed into the calculation of the similarity between two research topics in adjacent periods. After topic identification using the community discovery algorithm, each network community in the co-word network corresponds to a research topic. MatrixSim can be used to calculate the similarity of research topics after transforming topic communities into matrices with the same dimensions.

The adjacency matrix is the most common representation of the network structure, and in this study, it is used to characterize the research topic structure. To construct matrices of the same dimensions, the node union is first constructed based on the set of two topic community nodes in adjacent periods. Let two topic communities in adjacent periods be $P = \{V_p, E_p, W_p\}$ and $Q = \{V_q, E_q, W_q\}$, where V and E are the set of nodes and edges of the topic communities, respectively, and W is the edge weight, which is the number of co-occurrences among keywords in the topic communities. For the research topic communities P, Q, their node union V_s can be expressed as Eq. (4):

$$V_s = V_p \cup V_q = \{v_i | v_i \in (V_p \cup V_q)\} \tag{4}$$

Let V_s contain a total of N elements, that is, topics P and Q contain a total of N different keywords. Then, according to the network structure of the topic communities, the N-order adjacency matrices of topics P and Q are obtained by zero-padding. Let the adjacency matrix of the topic community P be A_p. For topic P, only when node $v_i^p, v_j^p \in V_p$ and there is a connected edge between the two nodes, that is, $(v_i^p, v_j^p) \in E_p$, the value of element a_{ij}^p in row i and column j of A_p is the connected edge weight between nodes v_i^p and v_j^p, and the rest of the cases are zero, as shown in Eq. (5).

$$a_{ij}^p = \begin{cases} w_{ij}^p, & if\ v_i^p, v_j^p \in V_p,\ \left(v_i^p, v_j^p\right) \in E_p \\ 0, & else \end{cases} \tag{5}$$

where $i, j \in [1, N]$, and w_{ij}^p represents the weight of the connected edge between node v_i^p and node v_j^p, which is the number of co-occurrences of the keyword corresponding to node v_i^p and the keyword corresponding to node v_j^p in the co-word network. Similarly, the elements a_{ij}^q in the N-order adjacency matrix A_q of topic Q can be expressed as

$$a_{ij}^q = \begin{cases} w_{ij}^q, & if\ v_i^q, v_j^q \in V_q,\ \left(v_i^q, v_j^q\right) \in E_q \\ 0, & else \end{cases} \tag{6}$$

The characteristics of the adjacency matrix determine that it can preserve the network structure of the communities to the maximum extent. It not only directly reflects the association (co-occurrence) between nodes but also fully accounts for the weights of the nodes themselves. In an adjacency matrix, the degree value of node i can be obtained by adding all elements of the ith row or column.

Then, the adjacency matrices A_p and A_q of topic communities P and Q, respectively, can be expressed as

$$A_p = \begin{bmatrix} a_{11}^p & a_{12}^p & \cdots & a_{1N}^p \\ a_{21}^p & a_{22}^p & \cdots & a_{2N}^p \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1}^p & a_{N2}^p & \cdots & a_{NN}^p \end{bmatrix} \tag{7}$$

$$A_q = \begin{bmatrix} a_{11}^q & a_{12}^q & \cdots & a_{1N}^q \\ a_{21}^q & a_{22}^q & \cdots & a_{2N}^q \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1}^q & a_{N2}^q & \cdots & a_{NN}^q \end{bmatrix} \tag{8}$$

Finally, MatrixSim r between two research topics in adjacent periods can be calculated as follows:

$$r = cos\theta = \frac{\langle A_p, A_q \rangle}{\| A_p \| \cdot \| A_q \|} \tag{9}$$

Because the number of co-occurrences of keywords cannot be negative, the range of r is $[0, 1]$, and the similarity is maximum when $r = 1$, and minimum when $r = 0$.

A simple example is provided to explain how MatrixSim detects the evolution paths of research topics. The network structures of the two communities in adjacent periods are shown in Fig. 3.

According to MatrixSim, the set of all nonrepeating nodes contained in the communities of two adjacent periods is obtained as $V_s = \{v_0, v_1, v_2, v_3\}$. The adjacency matrix of the co-word network at T_0 is $\begin{bmatrix} 0 & 2 & 5 & 0 \\ 2 & 0 & 4 & 0 \\ 5 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, and the adjacency matrix of the co-word network at T_0' is $\begin{bmatrix} 0 & 2 & 0 & 3 \\ 2 & 0 & 5 & 1 \\ 0 & 5 & 0 & 3 \\ 3 & 1 & 3 & 0 \end{bmatrix}$. MatrixSim can be obtained by substituting them into Eq. (8): $r = 48/(\sqrt{90} \times \sqrt{96}) = 2\sqrt{15}/15 \approx 0.5164$.

The traditional basic metrics of node overlap are dot product, cosine similarity, weighted cosine coefficient, Jaccard index, and generalized Jaccard index. Considering cosine similarity as an example, the community nodes (keywords) are mapped as vectors, and

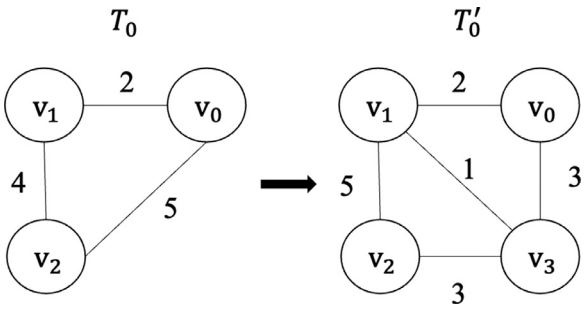

Fig. 3. Example of two communities in adjacent periods.

Table 1
Overview of test dataset distribution.

No.	Time period	Keywords	Papers
T1	2016	18636	3465
T2	2017	18571	3298
T3	2018	19455	3455
T4	2019	20493	3567
T5	2020	23292	3975

the similarity of two research topics is determined by calculating the cosine of the vector angle between these two research topics in adjacent periods, regardless of the node weights. The cosine similarity is

$$r = \cos\theta = \frac{A \cdot B}{\| A \| \cdot \| B \|} = (1,1,1,0)\cdot(1,1,1,1)/\sqrt{[(1,1,1,0)]^2} \times \sqrt{[(1,1,1,1)]^2} = \sqrt{3}/2 \approx 0.8660.$$

According to the formula used to measure the similarity of communities by edge relations, we obtain $r = \frac{E(x)\cap E(y)}{E(x)\cup E(y)} = 1/3 = 0.3333$.

3.3. Method verification

The case in Fig. 3 demonstrates the difference between the calculation process of MatrixSim and the common node overlap and relational similarity methods. To verify the applicability and effectiveness of MatrixSim in topic evolution path detection, a model-verification experiment was conducted. The LIS field was chosen owing to the disciplinary background of our research team members (Wang et al., 2021). Due to the authority and comprehensiveness of the Web of Science (WoS) core collection, the scientific literature in the field of LIS included in the WoS core collection was selected as the data source. The time span was limited to 2016-2020, the paper type was limited to research articles, and the language was limited to English. After downloading the corresponding bibliographic data and filtering out the repeated data and the data without the authors' keywords, 17,760 papers were obtained. Considering the existence of different forms of the same words and mixed abbreviations in the author keywords field, the keywords were preprocessed according to the process shown in Fig. 4, and a test dataset of 41,653 valid keywords was obtained. It should be noted that in the data preprocessing process, relevant keywords were not merged semantically to respect the authors' choice of keywords and avoid getting into cumbersome processing. However, the original keywords chosen by the authors were retained as much as possible. Previous studies have also shown that these semantically related keywords are basically divided into the same topic community; therefore, this processing method still yields reasonable analysis results (Wang et al., 2021).

To explore the sensitivity of MatrixSim to the differences between topics in adjacent periods during evolution path detection, each year was considered as a time window, and the test dataset was divided into five time periods with the data distribution shown in Table 1.

After constructing five keyword co-word networks based on the keywords in each time period, the fast unfolding algorithm (Blondel et al., 2008) was used to prune and identify the communities in the network and assign topic labels to the communities based on the z-score values (Wang et al., 2014). A total of 214 core topic communities were identified (topic communities with the number of nodes greater than 10 are called core topic communities), and the number of core topic communities in each time period between T1 and T5 (2016-2020) was 43, 41, 41, 41, 41, and 48, respectively.

After identifying the core research topics in each time period, an evolution path standard set of "research topic -> research topic" was constructed manually; that is, the researchers determined the association between two topics based on the keywords and their co-occurrence in adjacent periods (focusing on the core keywords, i.e., the top 20 keywords and their co-occurrence), and thus determined whether there was an evolution path between the two topics. The standard set was constructed with the participation of two doctoral candidates, one master's candidate, and one professor with more than 20 years of research experience in the LIS field. First, the association relationship of research topics in adjacent periods in the test dataset was judged by three students in pairs. After completing the construction of their respective artificial path sets, the results of the three were compared, and the inconsistencies

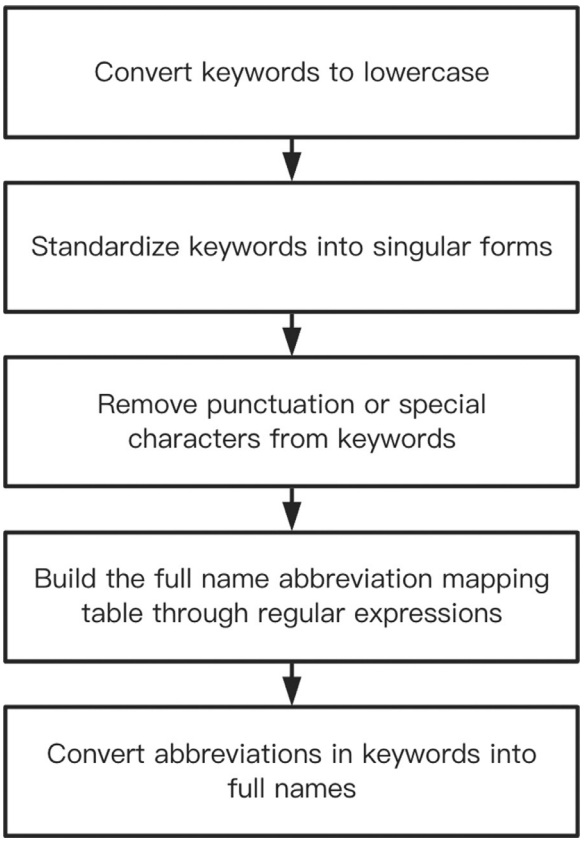

Fig. 4. Keyword preprocessing process.

Table 2
Distribution of evolution path standard set.

Time period	T1-T2	T2-T3	T3-T4	T4- T5	Sum
Path num	44	55	54	53	206
Same label path num	8	6	10	8	32

Table 3
Evolution path standard set (partial results).

Time	Path	Time	Path
T1-T2	bibliometrics-> bibliometrics	T3-T4	classification-> classification
T1-T2	sociomateriality-> monitoring	T3-T4	metadata-> classification
T1-T2	open data-> data sharing	T3-T4	uncertainty-> uncertainty
T1-T2	social media-> social media	T3-T4	spatial analysis-> uncertainty
T1-T2	information literacy-> data sharing	T3-T4	knowledge management-> technostress
T2-T3	qualitative-> qualitative	T4-T5	internet of thing-> human capital
T2-T3	trust-> social network	T4-T5	social media-> simulation
T2-T3	action research-> social network analysis	T4-T5	machine learning-> machine learning
T2-T3	social media-> social network analysis	T4-T5	academic library-> open innovation
T2-T3	electronic health record> machine learning	T4-T5	e-government-> blockchain

were submitted to the professor for adjudication and evaluation. Finally, an evolution path standard set, including 206 research topic pairs, was obtained, among which there were 32 topic pairs with the same topic labels in adjacent periods, as shown in Table 2. Due to space constraints, only some manually detected research topic evolution paths are listed in Table 3.

Subsequently, MatrixSim was used for evolution path detection of the research topics. It is noteworthy that although topic evolution methods based on machine learning and deep learning have been proposed by scholars with the wide application of information technology, methods such as cosine similarity are still commonly used for evolution path detection in research topic evolution, that is, to determine the relationship between topics in adjacent periods (Deligiannis et al., 2021; Huang et al., 2022; Jiang et al., 2022).

Table 4
Comparison of the paths of similarity methods.

Method/threshold	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Standard set	206	206	206	206	206	206	206	206
MatrixSim	259	225	202	182	167	155	146	138
Edge	274	237	208	188	173	161	150	138
Cos	2160	1590	1144	809	617	454	344	265
Cos-weighted	697	491	380	320	275	241	227	197
Jaccard	2055	1514	1069	763	581	447	336	271

Table 5
Comparison of evaluation metrics of similarity methods.

Method	Metric/threshold	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	Average
MatrixSim	P	0.7670	0.7476	0.7233	0.6893	0.6408	0.6068	0.5777	0.5583	0.6639
	R	0.6100	0.6844	0.7376	0.7802	0.7904	0.8065	0.8151	0.8333	0.7572
	F1	0.6796	0.7146	0.7304	0.7320	0.7078	0.6925	0.6761	0.6686	0.7002
Edge	P	0.7816	0.7427	0.6845	0.6845	0.6456	0.6214	0.5777	0.5437	0.6602
	R	0.5876	0.6456	0.6779	0.7500	0.7688	0.7950	0.7933	0.8116	0.7287
	F1	0.6708	0.6907	0.6812	0.7157	0.7018	0.6975	0.6685	0.6512	0.6847
Cos	P	0.9903	0.9854	0.9369	0.8835	0.8301	0.7476	0.6796	0.6214	0.8344
	R	0.0944	0.1277	0.1687	0.2250	0.2771	0.3392	0.4070	0.4830	0.2653
	F1	0.1724	0.2261	0.2859	0.3586	0.4156	0.4667	0.5091	0.5435	0.3722
Cos-weighted	P	0.9223	0.8544	0.7864	0.7136	0.6602	0.6068	0.5825	0.5291	0.7069
	R	0.2726	0.3585	0.4263	0.4594	0.4945	0.5187	0.5286	0.5533	0.4515
	F1	0.4208	0.5050	0.5529	0.5589	0.5655	0.5593	0.5543	0.5409	0.5322
Jaccard	P	0.9903	0.9806	0.9175	0.8641	0.7961	0.7136	0.6019	0.5680	0.8040
	R	0.0993	0.1334	0.1768	0.2333	0.2823	0.3289	0.3690	0.4317	0.2568
	F1	0.1805	0.2349	0.2965	0.3674	0.4168	0.4502	0.4576	0.4906	0.3618

Therefore, four representative traditional similarity methods, namely, edge overlap (edge) (Wu et al., 2011), cosine similarity (cos) (Chen et al., 2017), weighted cosine coefficient (cos-weighted) (Pribadi et al., 2017), and Jaccard index (Jaccard) (Niwattanakul et al., 2013), were selected as baselines to evaluate the detection effectiveness of MatrixSim.

Similarity thresholds must be set in advance when using similarity methods for topic evolution path detection. In this study, the thresholds were set in the range of [0.1,0.8], and the evolution paths of research topics under different threshold settings by several methods were analyzed. The detected research topic evolution paths are shown in Table 4.

Next, the topic evolution paths detected by the different methods were compared with the standard set established manually, and the comparison results of *precision, recall, and F-measure* values are shown in Table 5.

It can be seen that MatrixSim has the highest *F-measure* value in most cases, and its average *F-measure* value under different thresholds is also higher than the other methods, which indicates that MatrixSim is effective in performing the evolution path detection of research topics to certain extent, and outperforms edge, cos, cos-weighted, and Jaccard, which only considers node overlap or relational similarity from a single perspective.

Furthermore, the effectiveness of different methods in detecting the evolution paths of specific topics was analyzed. Fig. 5 shows the result of evolution path detection with MatrixSim for a single research topic "social media" (the similarity threshold δ is 0.3). In Fig. 5, rectangles of different colors represent specific topic communities, and the height of the rectangles reflects the size of the keywords within the topic. The curves between the topics in adjacent periods represent the evolutionary process of research topics, that is, as long as the similarity between two research topics is greater than 0.3, an evolutionary relationship can be considered to exist between them; namely, there is an edge on the Sankey diagram. Therefore, the sum of the similarities between the research topics may be greater than one, and the width of the curve is adjusted based on the relative magnitude of these similarities; that is, the width of the curve reflects the degree of similarity between research topics. It is easy to find from Fig. 5, that the research topic of "social media" has been splitting from T1 to T5, and with the development of social media and related research, the research in the field of LIS has gradually formed a system for specific issues such as social networks (Badri et al., 2017), open data (Kassen, 2018), information sharing (Wang et al., 2019), information privacy (Youn and Shin, 2019), and google trends (Turki et al., 2020), and become independent from the topic of "social media," which is consistent with previous studies. The standard set of evolution paths of the research topic "social media" is shown in Table 6. Fig. A1–A4 gives the evolution paths of "social media" obtained using the edge, cos, cos-weighted, and Jaccard similarity methods, respectively.

The results of the evolution path detection of "social media" for different methods with a similarity threshold δ of 0.3 were compared with the standard set, and the resulting *precision, recall, and F-measure* values are shown in Table 7. It can be intuitively observed that the *F-measure* value corresponding to MatrixSim is far greater than that of the other methods, which further reflects the superiority of MatrixSim.

In summary, MatrixSim is effective in the process of research topic evolution path detection, and it performs better in terms of both the overall evolutionary trend of the domain research topics and the evolution path detection of specific topics.

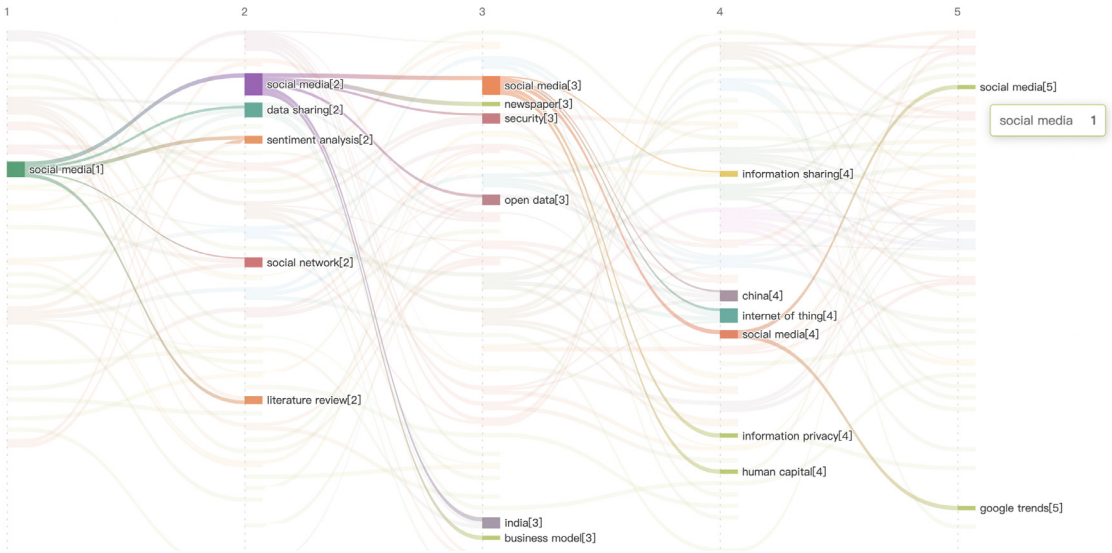

Fig. 5. Evolution paths of a single research topic “social media” (MatrixSim).

Table 6
Standard set of evolution paths of a single research topic "social media".

Time	Path	Time	Path
T1-T2	social media-> social media	T3-T4	social media-> machine learning
T1-T2	social media-> data sharing	T3-T4	social media-> china
T1-T2	social media-> sentiment analysis	T3-T4	social media-> internet of thing
T1-T2	social media-> literature review	T3-T4	social media-> e-government
T2-T3	social media-> social media	T3-T4	social media-> regulation
T2-T3	social media-> social network analysis	T3-T4	social media-> information privacy
T2-T3	social media-> newspaper	T3-T4	social media-> human capital
T2-T3	social media-> open data	T4-T5	social media-> social media
T2-T3	social media-> india	T4-T5	social media-> simulation
T2-T3	social media-> business model	T4-T5	social media->e-government
T3-T4	social media-> social media	T4-T5	social media-> google trends
T3-T4	social media-> information sharing		

Table 7
Comparison of evaluation metrics for evolution paths of a single research topic "social media".

Method	Precision	Recall	F-Measure
MatrixSim	0.7391	0.8947	0.8095
Edge	0.5652	0.8667	0.6842
Cos	0.9565	0.2683	0.4190
Cos-weighted	0.4348	0.7143	0.5405
Jaccard	0.8696	0.3077	0.4545

4. MatrixSim-based domain-oriented research topic evolution case

4.1. Data selection and preprocessing

To verify the effectiveness of MatrixSim in detecting evolution paths in a relatively broad regional dataset and the rationality and reliability of topic evolution analysis, based on method verification, a topic evolution analysis oriented to a specific subject field was executed, revealing the development lineage of the LIS field over the past 20 years.

Similar to the aforementioned data collection and preprocessing steps, scientific literature in the field of LIS from 1996 to 2020 was collected using the WoS core collection as the data source, which contains 41,888 papers, and a case dataset of 76,190 valid keywords was obtained through the data preprocessing in Fig. 4. Referring to Han (2020), every five years was chosen as a time unit, and the case data were divided into five time periods of data, and the data distribution is shown in Table 8.

Table 8
Overview of the case dataset distribution.

No.	Time period	Keywords	Papers
t1	1996-2000	6136	1239
t2	2001-2005	14524	2913
t3	2006-2010	37281	7335
t4	2011-2015	68830	12641
t5	2016-2020	100447	17760

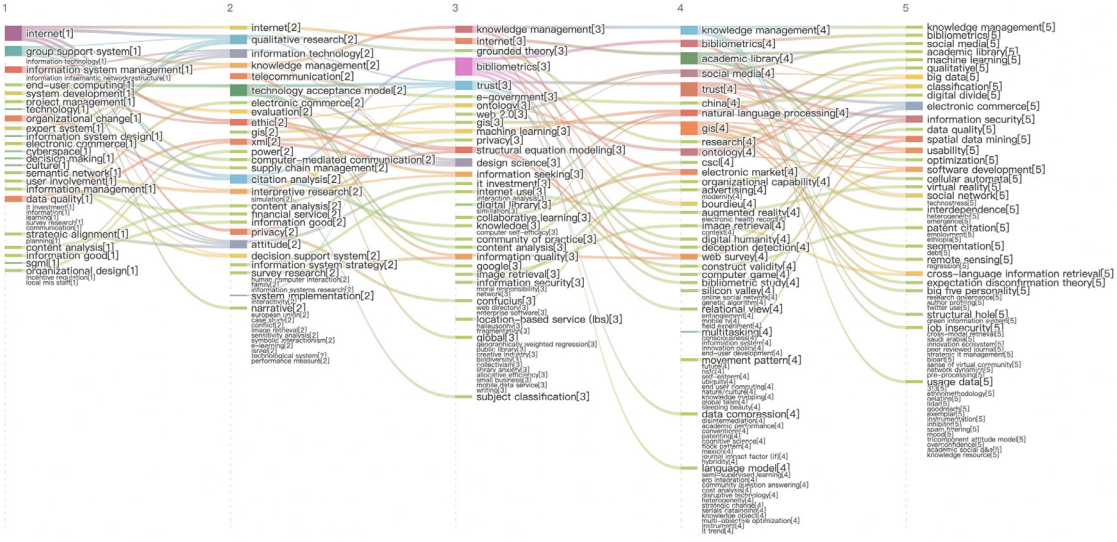

Fig. 6. Evolution of research topics in the field of LIS from 1996 to 2020.

4.2. Analysis of results

Similar to the analysis steps of research topic evolution in the method verification section, the similarity threshold δ of MatrixSim was set to 0.3, and the overall trend of the evolution of research topics in the LIS field for more than 20 years is shown in Fig. 6.

As can be observed, there is an evident evolution of research topics in the LIS field from t1 to t5 (1996–2020). (a) There has been an overall increase in the number and scope of research topics. (b) There exists merging and splitting of research topics. For example, the research topics of "internet" and "organizational change" in t1 (1996–2000) were merged into "citation analysis" in t2 (2001-2005), which reflected the merging of research topics. And "citation analysis" in t2 (2001–2005) was then split into "bibliometrics," "subject classification," and "knowledge" in t3 (2006–2010), reflecting the splitting of research topics. In addition, research topics such as "knowledge management" and "bibliometrics" were split into several sub-topics in the later period, reflecting that the splitting phenomenon is more frequent than the merging phenomenon and indicating the gradual improvement of the research system in this field. (c) The contraction and growth of research topics also appear in LIS field studies. For example, the research topic "internet" was presented throughout t1-t3 (1996–2010), with the size of the corresponding community changing continuously. (d) A new research topic "gis" emerged in t2 (2001-2005), and "data compression" in t4 (2011–2015) disappeared in t5 (2016–2020), indicating the birth and death of research topics.

The critical paths in Fig. 6 were selected to further clarify the evolutionary development of the important research topics, as shown in Fig. 7. Research topics in the LIS field from 1996 to 2020 can be classified into three dimensions: measurement (blue), management (green), and technology (orange).

(1) The measurement dimension includes quantitative research on bibliometrics, information metrics, scientometrics and webometrics. With the development of "citation analysis" in t2 (2001–2005), the evolving research topic of "bibliometrics" has become highly stable across time periods (2006–2020) as well as the evolution of "content analysis," "subject classification," "patent citation," and other research topics, which are commonly used in bibliometrics and are dedicated to quantitative assessment of the impact of journals, scholars, and scientific research. It can be seen that quantitative research is an important research direction and an inevitable trend in the LIS field.
(2) The management dimension refers to further deepening and expansion of management in the field of LIS. For example, the research topic of "knowledge management," which evolved from the research topic of "group support system" in t1 (1996–2000), emerged steadily from t2 to t5 (2001–2020) and continuously differentiated into different sub-topics. From t1 (1996–2000) to

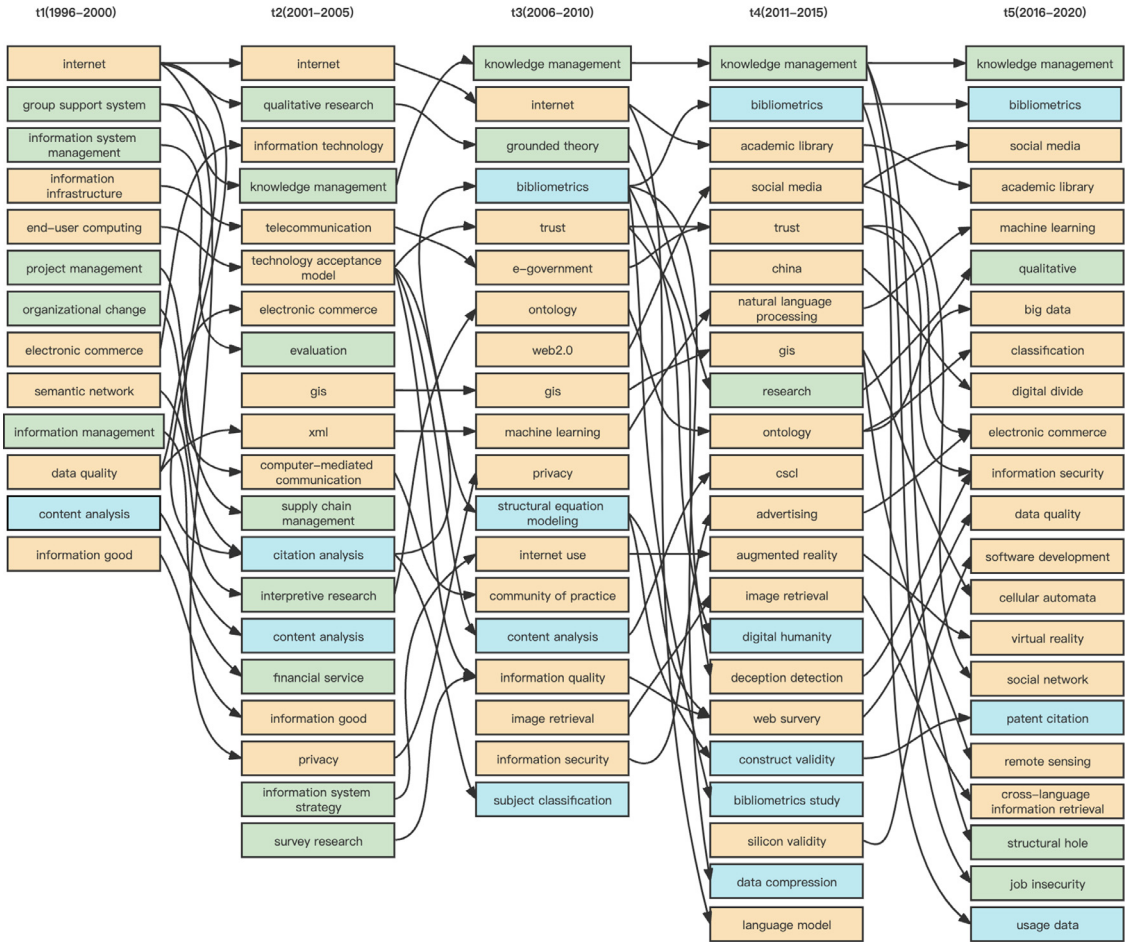

Fig. 7. Path map of important research topics.

t2 (2001–2002), "project management" evolved into "supply chain management" and "information management" evolved into "financial services."

(3) In the technology dimension, information technology has a profound impact on the field of LIS and promotes the development of LIS towards information science. The LIS field continues to introduce mathematical models, computer algorithms, and other tools for new research, including "gis," "natural language processing," "machine learning," "image retrieval," and other research topics. These research topics are important emerging trends in the field of LIS and deserve more attention. In addition, the prosperity of the Internet drives the LIS field towards diversification and intersectionality. For example, with the development of "internet" and "web2.0," research topics such as "social media," "social network," and "software development" have undergone a change of technological context from the networks and Internet to social media and mobile applications. Research topics such as "electronic commerce" and "trust" have been constantly discussed and closely linked to "information technology" and the emergence of thinking about issues such as "digital divide" and "information security," all of which reflect LIS scholars' fundamental interest in the interactions between information, people, and technology. Meanwhile, the flourishing development of "internet" also encourages the transformation of "academic library," which means that the emergence of digital libraries and smart libraries has injected new vitality into library science.

4.3. Comparison study-based validation

One empirical study based on a latent Dirichlet allocation model to explore the evolution of research topics in the field of LIS was published recently (Han, 2020), with the years 1996–2019 divided into five periods, and influential journals from each period were selected to generate dynamic journal lists for analysis. Despite the slight difference (i.e., time period and source journals), it is still interesting to compare our results with his work (we abbreviate this as Han's work).

In the field of LIS, these two studies have certain similarities in their overlapping periods. For example, the research topics of information management, information systems/design in t1 (1996–2000), the Internet, information systems, and citation analysis in t2 (2001–2005), e-government, modeling algorithms (structural equation modeling, gis, machine learning) in t3 (2006–2010),

research topics such as bibliometrics, online social networks in t4 (2011–2015), and research topics such as social media, knowledge management, and e-commerce in t5 (2016–2020) appear in both studies (topic labels and keywords in the topics are similar).

We consider that some of the results of our study reasonably match Han's work (Han, 2020), and the differences can be explained as follows.

(1) The research topics obtained in this study are more extensive. This may reflect the different data sources used in the two studies. Han selected influential journals for analysis in each time period, hence the scope was relatively narrow and focused on the core areas of LIS. Our study collected data directly from the Web of Science core collection, where journal articles are more flexible and sensitive to the external social environment and technological development.
(2) However, the experimental results yielded different conclusions. For example, Han argues that library science has become less popular over time because there were no clusters of popular topics related to library issues from 2000 to 2005. By contrast, library issues were revisited in this study from 2011 to 2020, which can be interpreted as the gradual digital transformation of library science with the development of the Internet. This may be because of the effects of different algorithm choices.

In addition, Chang et al. (2015) identified three broad subjects of LIS based on 580 highly cited journal articles from 1995 to 2014, namely "bibliometrics," "information systems (IS) and information retrieval (IR)," and "application of Internet technology (AIT)". Our previous study (Wang et al., 2021) also demonstrated that research in the field of LIS from 2014 to 2019 focused on academic libraries, information literacy, social media, knowledge management, bibliometrics, and qualitative research. These findings correspond to the dimensions we summarized, further justifying that MatrixSim is scientific and reliable in exploring the developmental lineage of the LIS field.

5. Discussion

In this study, the correlation between research topics was analyzed from the perspective of MatrixSim, and the traditional similarity methodology was improved by introducing the concept of the Frobenius inner product. The two factors, node overlap and relational similarity, are tightly combined in the form of a matrix instead of a simple weight assignment or basic mathematical calculation, avoiding the errors and limitations caused by manual weighting. Specifically, the topic communities in adjacent periods were transformed into adjacency matrices to calculate MatrixSim. The adjacency matrix can not only intuitively determine the connection (co-occurrence) of edges between nodes, but also fully consider the weights of the nodes themselves, thus maximizing the preservation of the network structure of the topic communities.

When detecting the evolution paths of research topics, MatrixSim works similarly to other similarity methods by calculating the similarity between research topics in two adjacent periods and determining the existence of certain evolutionary relationships for topic pairs with similarity greater than a preset threshold. The effectiveness of MatrixSim in detecting the evolution paths of research topics was verified through empirical experiments. Compared with baseline methods such as edge, cos, cos-weighted, and Jaccard, it is shown that the *F-measure* values corresponding to MatrixSim are greater than those corresponding to traditional similarity methods in both the overall evolution paths of research topics and the evolutionary trajectory of a single research topic, which, to certain extent, indicates the superiority of MatrixSim in topic evolution path detection.

MatrixSim reflects the actual situation of the field development and is a reliable method for exploring the development context of the field. This is demonstrated by the consistency of individual research topic evolution paths in the method verification section with the validation results of authoritative journals, which is further validated by a subject-specific topic evolution analysis based on the LIS field literature collection. In this study, hot topics in LIS over the past 20 years were classified into three dimensions: measurement, management, and technology, which are consistent with the development context of the LIS field summarized by existing studies. In particular, the differences between our study and Han's study (e.g., the "library science" research topic) reflect that MatrixSim can track frontier topics and recent progress in the field to certain degree, thus providing more possibilities for researchers to detect emerging trends (Han, 2020).

More importantly, the basic principle of MatrixSim is to judge the similarity between topics based on the network structure of research topics, which may become a cornerstone for researchers to conduct sci-tech information research such as research topic evolution and emerging trend detection. The idea of substituting MatrixSim for traditional similarity methods in detecting research topic associations can be applied to the topic-time longitudinal analysis of software tools such as CiteSpace, VOSviewer, Bibliometrix, and SciMAT to obtain richer longitudinal analysis results. Although MatrixSim is currently applied only to research topics based on co-word network discovery, its algorithmic principle can support various types of network structures, methods such as Word2vec and BERT can be used subsequently to add more semantic information to construct word vector networks, and MatrixSim can be used to explore the association between topics. In addition, MatrixSim can be extended to the analysis of topics in basic disciplines such as mathematics and physics, interdisciplinary disciplines such as artificial intelligence (AI) and computational sociology, and technical industrial fields such as graphene and 3D printing. In particular, the applicability and discrepancy of MatrixSim in different fields are also of concern and may need to be further explored in future studies.

6. Conclusion

In this study, a new method, MatrixSim, was proposed to detect the evolution paths of research topics. Compared with the common methods in topic evolution path detection, such as cosine similarity and Jaccard index, MatrixSim is similar to them in the implementation principle; however, MatrixSim preserves the network structure of topic communities to the maximum extent

with the help of the adjacency matrix and integrates both node similarity and relationship similarity. This provides the possibility for the detection of new topic evolution paths, that is, the discovery of new knowledge evolution and transmission paths, which may become a new perspective and cornerstone for topic evolution analysis based on co-word networks. In particular, the LIS field was selected to compare the differences between MatrixSim and baseline methods such as edge, cos, cos-weighted, and Jaccard in topic similarity detection, to conduct subject-specific domain-oriented topic evolution analysis in conjunction with co-word network analysis methods, and to discuss the performance of MatrixSim in exploring the developmental lineage of the field. The results confirmed that MatrixSim is superior in evolution path detection, which can facilitate the exploration of the internal mechanism of scientific knowledge generation and evolution, and provide methodological support for scientific prediction and knowledge services.

There are also certain limitations, including the fact that empirical research is only conducted in the LIS field, and the applicability of MatrixSim in other fields needs to be further verified. In the next step, as well as further analysis of the applicability of MatrixSim in other fields, we will focus on the improvement of MatrixSim, such as matrix deformation and its combination with other similarity methods. Additionally, as a network structure-based similarity algorithm, the application of MatrixSim can be more extensive. It is also a future direction to apply MatrixSim to the analysis of scientific knowledge networks, such as citation networks and co-authorship networks, and to combine it with word embedding techniques, such as word2vec and BERT, to further apply it to the construction and analysis of word vector networks in the field of scientific and technological intelligence.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Xiaoguang Wang: Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Jing He:** Conceptualization, Data curation, Investigation, Writing – original draft, Writing – review & editing. **Han Huang:** Methodology, Data curation, Writing – original draft, Writing – review & editing. **Hongyu Wang:** Writing – review & editing.

Acknowledgment

This work was supported by National Natural Science Foundation of China (No. 71874129), Innovative Research Group Project of the National Natural Science Foundation of China (No. 71921002) and Hubei Provincial Natural Science Fund for Creative Research Groups (No. 2019CFA025).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.joi.2022.101343.

Appendix A

Figs. A1–A4

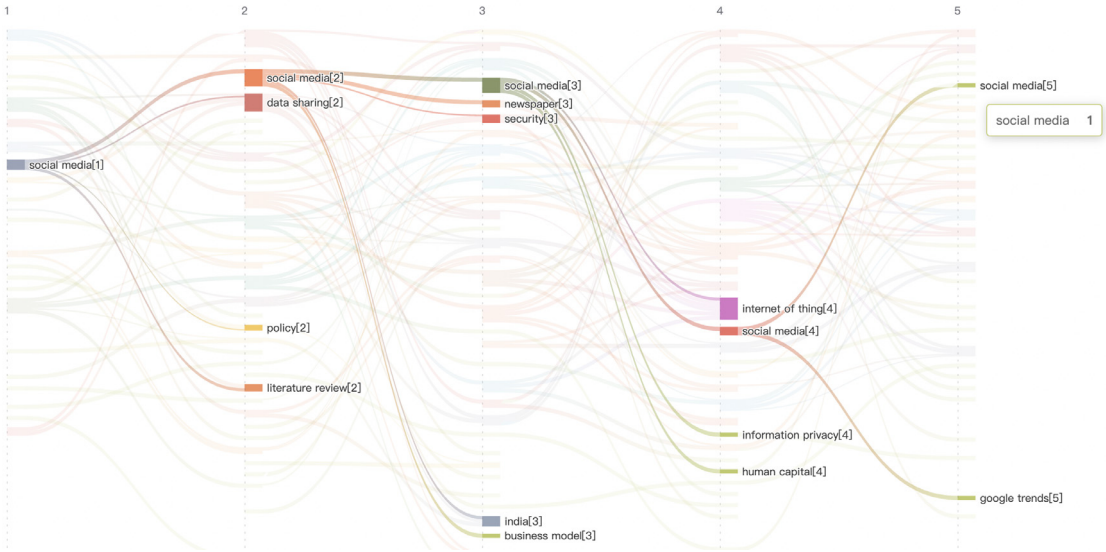

Fig. A1. Evolution paths of a single research topic “social media” (Edge).

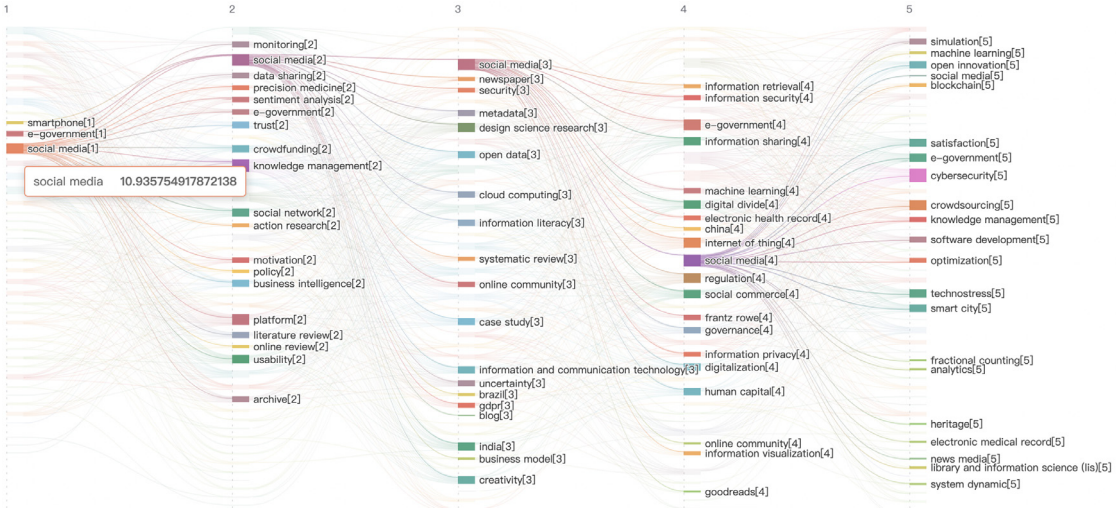

Fig. A2. Evolution paths of a single research topic “social media” (Cos).

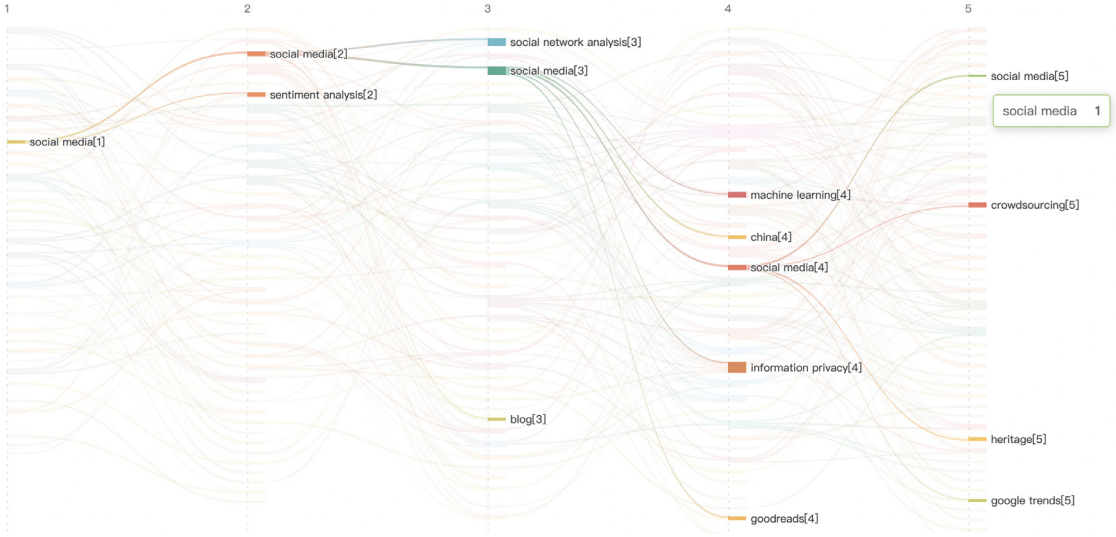

Fig. A3. Evolution paths of a single research topic “social media” (Cos-weighted).

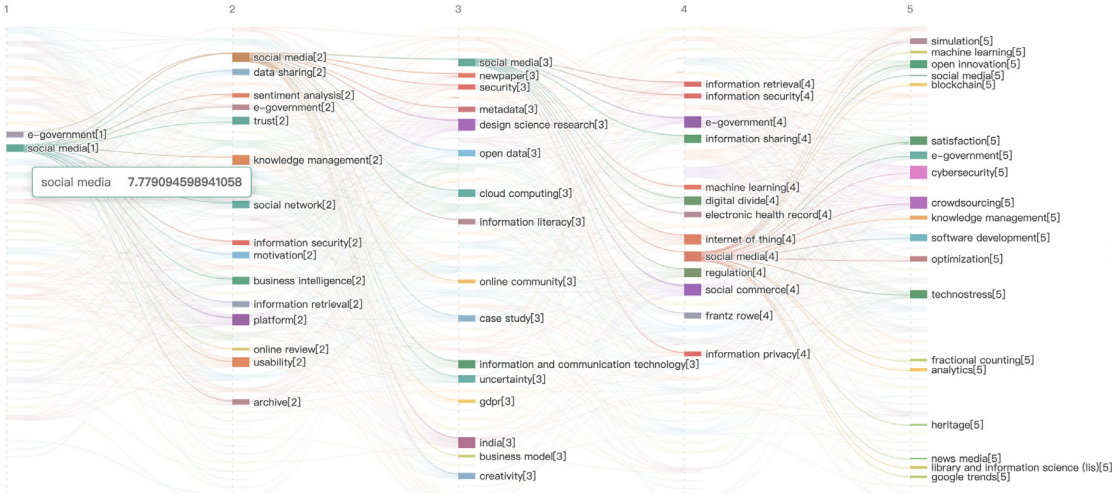

Fig. A4. Evolution paths of a single research topic “social media” (Jaccard).

References

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11*(4), 959–975. 10.1016/J.JOI.2017.08.007.

Badri, M., Nuaimi, A. al, Guang, Y., & Rashedi, A. al (2017). School performance, social networking effects, and learning of school children: Evidence of reciprocal relationships in Abu Dhabi. *Telematics and Informatics*. 10.1016/j.tele.2017.06.006.

Behrouzi, S., Shafaeipour Sarmoor, Z., Hajsadeghi, K., & Kavousi, K. (2020). Predicting scientific research trends based on link prediction in keyword networks. *Journal of Informetrics, 14*(4), Article 101079. 10.1016/J.JOI.2020.101079.

Berger-Wolf, T. Y., & Saia, J. (2006). A framework for analysis of dynamic social networks. In *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining, 2006*. 10.1145/1150402.1150462.

Beykikhoshk, A., Arandjelović, O., Phung, D., & Venkatesh, S. (2018). Discovering topic structures of a temporally evolving document corpus. *Knowledge and Information Systems, 55*(3). 10.1007/s10115-017-1095-4.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10). 10.1088/1742-5468/2008/10/P10008.

Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information, 22*(2). 10.1177/053901883022002003.

Chang, Y. W., Huang, M. H., & Lin, C. W. (2015). Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses. *Scientometrics, 105*(3), 2071–2087. 10.1007/S11192-015-1762-8/TABLES/9.

Chehab, J. P., & Raydan, M. (2008). Geometrical properties of the Frobenius condition number for positive definite matrices. *Linear Algebra and Its Applications, 429*(8–9). 10.1016/j.laa.2008.06.006.

Chen, B., Tsutsui, S., Ding, Y., & Ma, F. (2017). Understanding the topic evolution in a scientific domain: An exploratory study for the field of information retrieval. *Journal of Informetrics, 11*(4), 1175–1189. 10.1016/j.joi.2017.10.003.

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology, 57*(3), 359–377. 10.1002/ASI.20317.

Choudhury, N., & Uddin, S. (2016). Time-aware link prediction to explore network effects on temporal knowledge evolution. *Scientometrics, 108*(2), 745–776. 10.1007/S11192-016-2003-5/TABLES/10.

Chu, H. (2015). Research methods in library and information science: A content analysis. *Library and Information Science Research, 37*(1). 10.1016/j.lisr.2014.09.003.

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics, 5*(1). 10.1016/j.joi.2010.10.002.

Cobo, M. J., Lõpez-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2012). SciMAT: A new science mapping analysis software tool. *Journal of the American Society for Information Science and Technology, 63*(8), 1609–1630. 10.1002/ASI.22688.

Cristianini, N., Kandola, J., Elisseeff, A., & Shawe-Taylor, J. (2006). On kernel target alignment. *Studies in Fuzziness and Soft Computing, 194*. 10.1007/10985687_8.

Deligiannis, P., Vergoulis, T., Chatzopoulos, S., & Tryfonopoulos, C. (2021). Visualising scientific topic evolution. In *Proceedings of the web conference 2021 - companion of the world wide web conference, WWW: 5* (pp. 468–472). 10.1145/3442442.3451371.

Demirci, R. (2007). Similarity relation matrix-based color edge detection. *AEU - International Journal of Electronics and Communications, 61*(7). 10.1016/j.aeue.2006.08.004.

Ding, Y., & Stirling, K. (2016). Data-driven discovery: A new era of exploiting the literature and data. *Journal of Data and Information Science, 1*(4), 1–9. 10.20309/jdis.201622.

Han, X. (2020). Evolution of research topics in LIS between 1996 and 2019: An analysis based on latent Dirichlet allocation topic model. *Scientometrics, 125*(3). 10.1007/s11192-020-03721-0.

Hou, J., Yang, X., & Chen, C. (2018). Emerging trends and new developments in information science: A document co-citation analysis (2009–2016). *Scientometrics, 115*(2). 10.1007/s11192-018-2695-9.

Huang, L., Chen, X., Zhang, Y., Wang, C., Cao, X., & Liu, J. (2022). Identification of topic evolution: Network analytics with piecewise linear representation and word embedding. *Scientometrics*. 10.1007/s11192-022-04273-1.

Jeong, D. H., & Song, M. (2014). Time gap analysis by the topic model-based temporal technique. *Journal of Informetrics, 8*(3). 10.1016/j.joi.2014.07.005.

Jian, F., Yajiao, W., & Yuanyuan, D. (2018). Microblog topic evolution computing based on LDA algorithm. *Open Physics, 16*(1). 10.1515/phys-2018-0067.

Jiang, L., Zhang, T., & Huang, T. (2022). Empirical research of hot topic recognition and its evolution path method for scientific and technological literature. *Journal of Advanced Computational Intelligence and Intelligent Informatics, 26*(3), 299–308. 10.20965/JACIII.2022.P0299.

Jung, S., & Yoon, W. C. (2020). An alternative topic model based on common interest authors for topic evolution analysis. *Journal of Informetrics, 14*(3), Article 101040. 10.1016/j.joi.2020.101040.

Kassen, M. (2018). Adopting and managing open data: Stakeholder perspectives, challenges and policy recommendations. *Aslib Journal of Information Management, 70*(5). 10.1108/AJIM-11-2017-0250.

Katsurai, M., & Ono, S. (2019). TrendNets: Mapping emerging research trends from dynamic co-word networks via sparse representation. *Scientometrics, 121*(3), 1583–1598. 10.1007/s11192-019-03241-6.

Koylu, C. (2019). Modeling and visualizing semantic and spatio-temporal evolution of topics in interpersonal communication on Twitter. *International Journal of Geographical Information Science*, (4), 33. 10.1080/13658816.2018.1458987.

Lanckriet, G. R., Cristianini, N., Bartlett, P., Ghaoui, L. E., & Jordan, M. I. (2004). Learning the kernel matrix with semidefinite programming. *Journal of Machine Learning Research, 5*, 27–72.

Lee, L. (2001). On the effectiveness of the skew divergence for statistical language analysis. In *Proceedings of the international workshop on artificial intelligence and statistics* (pp. 176–183).

Liu, H., Chen, Z., Tang, J., Zhou, Y., & Liu, S. (2020). Mapping the technology evolution path: A novel model for dynamic topic detection and tracking. *Scientometrics, 125*(3), 2043–2090. 10.1007/s11192-020-03700-5.

Lu, L. Y. Y., & Liu, J. S. (2016). A novel approach to identify the major research themes and development trajectory: The case of patenting research. *Technological Forecasting and Social Change, 103*. 10.1016/j.techfore.2015.10.018.

Mei, Q., & Zhai, C. X. (2005). Discovering evolutionary theme patterns from text - An exploration of temporal text mining. In *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 198–207). 10.1145/1081870.1081895.

Mijangos, V., Sierra, G., & Herrera, A. (2016). A Word embeddings model for sentence similarity. *Research in Computing Science, 117*(1). 10.13053/rcs-117-1-5.

Mijangos, V., Sierra, G., & Montes, A. (2017). Sentence level matrix representation for document spectral clustering. *Pattern Recognition Letters, 85*. 10.1016/j.patrec.2016.11.008.

Niwattanakul, S., Singthongchai, J., Naenudorn, E., & Wanapu, S. (2013). Using of jaccard coefficient for keywords similarity. In *Proceedings of the international multiconference of engineers and computer scientists: 1* (pp. 380–384).

Park, H., & Magee, C. L. (2017). Tracing technological development trajectories: A genetic knowledge persistence-based main path approach. *PLoS ONE, 12*(1). 10.1371/journal.pone.0170895.

Pribadi, F. S., Adji, T. B., & Permanasari, A. E. (2017). Automated short answer scoring using weighted cosine coefficient. In Proceedings of the IEEE conference on e-learning, e-management and e-services, IC3e 2016, *70–74*. 10.1109/IC3E.2016.8009042.

Rafii, Z., & Pardo, B. (2012). Music/voice separation using the similarity matrix. In *Proceedings of the 13th international society for music information retrieval conference* (pp. 583–588).

Riaz, M., Naeem, K., & Afzal, D. (2020). A similarity measure under Pythagorean fuzzy soft environment with applications. *Computational and Applied Mathematics, 39*(4). 10.1007/s40314-020-01321-5.

Turki, H., Hadj Taieb, M. A., ben Aouicha, M., & Abraham, A. (2020). Nature Or science: What Google trends says. *Scientometrics, 124*(2). 10.1007/s11192-020-03511-8.
Van Eck, N. J., & Waltman, L. (2009). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538. 10.1007/S11192-009-0146-3.
Wang, C., Cui, L., Wu, X., & Yang, B. (2015). Semantic related feature analysis and dynamic evolution based on topic temporal chains under the social network. *Metallurgical and Mining Industry, 7*(8), 448–456.
Wang, E. Y., Guo, W., Dai, L. R., Lee, K. A., Ma, B., & Li, H. Z. (2010). Factor analysis based spatial correlation modeling for speaker verification. In *Proceedings of the 7th international symposium on Chinese spoken language processing, ISCSLP 2010*. 10.1109/ISCSLP.2010.5684490.
Wang, J. C., Wang, Y. Y., & Che, T. (2019). Information sharing and the impact of shutdown policy in a supply chain with market disruption risk in the social media era. *Information and Management, 56*(2). 10.1016/j.im.2018.09.005.
Wang, T., Lu, G., Liu, J., & Yan, P. (2018). Graph-based change detection for condition monitoring of rotating machines: Techniques for graph similarity. *IEEE Transactions on Reliability*. 10.1109/TR.2018.2866152.
Wang, X., Cheng, Q., & Lu, W. (2014). Analyzing evolution of research topics with NEViewer: A new method based on dynamic co-word networks. *Scientometrics, 101*(2), 1253–1271. 10.1007/S11192-014-1347-Y/FIGURES/6.
Wang, X., Wang, H., & Huang, H. (2021). Evolutionary exploration and comparative analysis of the research topic networks in information disciplines. *Scientometrics, 126*(6). 10.1007/s11192-021-03963-6.
White, R. W., & Jose, J. M. (2004). A study of topic similarity measures. In *Proceedings of the Sheffield SIGIR - twenty-seventh annual international ACM SIGIR conference on research and development in information retrieval*. 10.1145/1008992.1009100.
Wu, B., Wang, B., & Yang, S. Q. (2011). Framework for tracking the event-based evolution in social networks. *Ruan Jian Xue Bao/Journal of Software, 22*(7). 10.3724/SP.J.1001.2011.03841.
Wu, H., Yi, H., & Li, C. (2021). An integrated approach for detecting and quantifying the topic evolutions of patent technology: A case study on graphene field. *Scientometrics, 126*(8). 10.1007/s11192-021-04000-2.
Xie, Q., Zhang, X., Ding, Y., & Song, M. (2020). Monolingual and multilingual topic analysis using LDA and BERT embeddings. *Journal of Informetrics, 14*(3). 10.1016/j.joi.2020.101055.
Xu, H. (2020). Topic-linked innovation paths in science and technology. *Journal of Informetrics, 14*(2). 10.1016/j.joi.2020.101014.
Yan, Q., Ding, Y., Zhang, J. J., Xia, Y., & Zheng, C. H. (2019). A discriminated similarity matrix construction based on sparse subspace clustering algorithm for hyperspectral imagery. *Cognitive Systems Research, 53*. 10.1016/j.cogsys.2018.01.003.
Yang, S., Han, R., Wolfram, D., & Zhao, Y. (2016). Visualizing the intellectual structure of information science (2006–2015): Introducing author keyword coupling analysis. *Journal of Informetrics, 10*(1). 10.1016/j.joi.2015.12.003.
Yang, Z., Islam, N., Shi, Y., Venkatachalam, K., & Huang, L. (2021). The Evolution of Interindustry technology linkage topics and its analysis framework in three-dimensional printing technology. *IEEE Transactions on Engineering Management*. 10.1109/TEM.2021.3086760.
Youn, S., & Shin, W. (2019). Teens' responses to Facebook newsfeed advertising: The effects of cognitive appraisal and social influence on privacy concerns and coping strategies. *Telematics and Informatics, 38*. 10.1016/j.tele.2019.02.001.
Zhang, J., Liu, X., & Wu, L. (2016). The study of subject-classification based on journal coupling and expert subject-classification system. *Scientometrics, 107*(3). 10.1007/s11192-016-1890-9.
Zhang, X., Xie, Q., Song, C., & Song, M. (2022). Mining the evolutionary process of knowledge through multiple relationships between keywords. *Scientometrics, 127*(4), 2023–2053. 10.1007/s11192-022-04272-2.