Scientometrics (2025) 130:2137–2166
https://doi.org/10.1007/s11192-025-05279-1

Evolution of journal preference based on topic focus: A case study in the technology innovation management field

Xinhang Zhao[1] · **Xuefeng Wang**[1] · **Yuqin Liu**[2] · **Hongshu Chen**[1] · **Rui Guo**[3]

Received: 14 May 2024 / Accepted: 2 March 2025 / Published online: 24 March 2025

Abstract
Topic evolution is essential for exploring a field; however, the journal's contribution has not been explored in topic evolution research. In this work, we interpret a journal's contribution as a journal preference and investigate the concept based on topic focus, as shown in the journal-topic distribution. To analyse the topic focus, we first processed the data into documents consisting of only fine-grained topic words. Document vectors were generated using Sci-BERT and clustered using the k-means algorithm after dimensionality reduction. By matching journals with topic clusters, we calculated the journal preference score based on topic focus and then added a time factor to represent the evolution of journal preference. Simultaneously, we used the Zipfian distribution to classify fine-grained topic words into core and rare topic words, which were then used to establish topic relations in the evolutionary analysis and calculate the novelty scores of journal topic words. We use the technology innovation management (TIM) field to conduct a case study. There were 8 typical and 16 derivative topics, totalling 24 different topics. We focused on four important topics: R&D activity, technology management, innovation activity, and climate change, and found that they all have a relatively innovative evolution in a given year. The study indicates that within a given topic, while the composition and ranking of top journal preferences fluctuate over time, a subset of journals consistently exhibits dominance, appearing in the top ranks across most years. Although no clear relationship exists between journal preferences and ratings, A- and B-rated journals often dominate preferences for specific topics. Additionally, A- and B-rated journals with high or long preferences showed limited novelty. Most journals that preferred to interact with novel issues were C-rated.

Keywords Journal preference · Topic focus · Sci-BERT · K-means · Zipfian · Technology innovation management

✉ Xuefeng Wang
wxf5122@bit.edu.cn

[1] School of Management, Beijing Institute of Technology, Beijing 100081, PR China

[2] School of Publication, Beijing Institute of Graphic Communication, Beijing 102600, PR China

[3] School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing 100049, PR China

Introduction

Topic evolution is a classical and popular method of investigating the development and trends of academic research topics. This is typically conducted through co-citation analysis, co-word analysis, or topic clustering to track topic changes using temporal documents (Gao et al., 2022; Zhu et al., 2022). With the continuous development of society, the cross-integration of disciplines, and an increasing number of publications, topic evolution analysis has become useful for exploring how a topic accumulates, develops, and evolves (Bai et al., 2021). In contrast to analysing research topics within a specific field, recent research has considered incorporating other academic entities such as authors (Rosen-Zvi et al., 2012; Schäfermeier et al., 2023), organisations (Jin et al., 2022), and journals (Song et al., 2017) in topic evolution analysis.

Journals have always provided an important and useful foundation for studying topic evolution. Some studies use journals to delimit the research field, using journal papers as research data. Papers published in 40 marketing-related journals were confined and analysed in the top marketing journals to select the most important keywords for qualitative methods (Murgado-Armenteros et al., 2015). Bai et al. (2021) identified 20 journals from Q1-level journals in the SCImago Journal Rank in the e-learning field, excluded journals unrelated to e-learning, and finally retained 10 journals. The use of journals as qualifications in a research field ensures consistency in the meaning of topic words and avoids noise from unrelated fields. Simultaneously, it is easy to filter meaningful data from journals with significant influence. Other studies set journals as analysis elements. For example, the top 11 most productive Web of Science journals were obtained by conducting a bibliometric analysis of life-cycle assessment-related publications (Chen et al., 2014).

However, current research seldom focuses on revealing the interactions between topics and journals when analysing topic changes. It is beneficial to explore this interaction because: (1) the relationship between journals and topics can be clarified from a topic evolution perspective. Furthermore, (2) the role of journals in topic evolution can be clarified. Song et al. (2017) proposed a novel journal–topic–time model that calculates journal rankings separately for different topics and explores journal contributions to topics. They interpreted this contribution as the degree of attention that the journal paid to the topic. However, their research problem concerned journal ranking; therefore, they only used topic distribution under journals to indicate the journal's contribution to the topics as a small part of the analysis. They analysed attention but did not explicitly explore the types of relationships between topics and journals. In this study, we defined the concepts of topic focus and journal preference. These two behavioural concepts are commonly used in recommendation systems: user preference is used earlier, and user focus is often ignored (Chen et al., 2019). We draw on these two concepts to describe the relationship between journals and topics and further explain this relationship through topic distribution.

Therefore, this study aims to develop a method for analysing journal preferences based on topic focus and exploring temporal changes along with topic evolution. First, we defined the concepts of topic focus and journal preference. Using the topic cluster method, we analysed the relationships among them. The data were processed into fine-grained topic words and document vectors, which were then clustered using the k-means algorithm. We conducted an in-depth exploration of journal preferences using temporal and topical analyses. Our research is confined to fields where journals serve as the primary medium for publication. Using journal papers in the field of technology innovation management (TIM) published between 2013 and 2022 as the research data, we defined two types of topics:

typical and derivative. We also analysed the evolution of journal preferences for the four typical topic clusters and their novelty. The contributions of this study can be summarised as follows: First, we defined the concept of journal preference by regarding topic focus as the representation used to analyse which journal has a predisposition in favour of a certain topic. Second, considering the dynamic characteristics of journal preferences, we did not simply consider the relationship between journal preferences and topics under the topic of a certain year, but explored the change in journal preferences in the evolution of topics with comprehensive analysis.

Literature review

Topic evolution

Topic evolution delineates dynamic changes in a topic over time, encompassing its maturity, integration of knowledge from other domains, amalgamation or division into new topics, and the ascendancy or decline of specific topics (Chen et al., 2017). It incorporates time-based topic modelling, which can show changes in topics in the field more dynamically with multi-perspective analysis (Gao et al., 2022; Zhu et al., 2022). Some studies consider time when modelling topics, and the most prominent method is dynamic topic modelling (DTM). The DTM incorporates a time factor into latent dirichlet allocation (LDA) (Miao et al., 2020) by introducing dynamic changes in topic distributions at each time point (Blei & Lafferty, 2006). The partitioning of documents into temporal slices is highly influenced by the choice of time granularity. To address such issues, a continuous-time DTM has been proposed, which uses Brownian motion to model topics through continuous document collection (Wang et al., 2012). “Topics over time” has also been proposed based on DTM, which jointly models words and timestamps within a probabilistic graphical model (Wang & McCallum, 2006). “Topics over time” incorporates the time factor more directly, explicitly modelling the change in topics over time. The recurrent Chinese restaurant process is another typical topic evolution model. It is a stochastic process used in Bayesian nonparametric models, particularly in the context of topic modelling, and captures the birth, growth, and death of topics (Ahmed & Xing, 2008). Its main advantage lies in providing a nonparametric method to deal with dynamic topic changes and the ability to adapt automatically to data.

Other studies establish topic relations after obtaining topics from different time slices. Topic evolution relations can be established by computing topic similarities between different time slices. Some studies use words as research topics to directly link the same topic words in different time slices (Amiri et al., 2021; Burmaoglu et al., 2017) or use co-word networks and adjacency matrices to link them (Wang et al., 2022a). Semantic distance, or Kullback–Leibler divergence, is also used to calculate the coherence of a topic across different time slices (Mei & Zhai, 2005; Zhu et al., 2022). In addition, an environment-adapted relation identification function was constructed to identify topic relations, considering the different types of relations between topics: evolution, fusion, and death (Zhang et al., 2017). Topic evolution pathway tracking has also been applied to establish interactions between topics (Zhang et al., 2021).

Topic evolution research shows a diversified trend, providing more abundant methodology and analytical aspects for in-depth analysis. Huang et al. (2024) analysed the patterns of topic evolution in science, especially from the perspective of the semantic coherence of topics in the semantic vector space, and explored the possible reasons for this. The k-means clustering algorithm was used to identify four general semantic consistency evolution patterns. Considering that the temporal distribution of research keywords in the literature may reflect the evolutionary stage of topics over time, the time distribution of research topics in different development stages based on research heat curves was mined (Zhang et al., 2024b). Topic evolutionary paths are also explored based on embeddings (Jin et al., 2024). They proposed a method that combined word embedding, document embedding, clustering, and network analysis to extract topics, measure their semantic similarity, and quantitatively distinguish their evolving states. Additionally, academia values the evolution of technological topics. Zhang et al. (2024a) presented a method for analysing technology development from an entity-centric perspective, which is more accurate than the traditional coarse-grained topic analysis. Liu et al. (2024) proposed an integrated method to map the technological evolution paths of scientific papers and patents. Recently, there has been a trend toward using popular topic modelling methods for topic evolution, demonstrating the superiority of these methods. A framework for interdisciplinary topic identification and evolution analysis has been proposed based on BERTopic (Wang et al., 2024). Using BERTopic, the model extracts topics, identifies interdisciplinary topics based on topic diversity and cohesion, and analyses their evolution. Invernici et al. (2024) introduced the CORD-19 Topic Visualizer (CORToViz), a method and a related visualisation tool for examining the CORD-19 scientific abstract text corpus. BERTopic was implemented in this tool for topic modelling.

In the early years, most studies used software to derive the evolution of a topic, which involved only topics. Recently, the contributions of other academic entities such as authors, organisations, and journals to topics have been considered. The most frequently used academic entity is the author, and a graph structure helps analyse topic flows between authors with different research interests (Schäfermeier et al., 2023). The author–topic model was proposed by extending the LDA to include authorship information (Rosen-Zvi et al., 2012), and this extension adds a time factor to investigate the interactions between authors and topics (Xu et al., 2014). As for organisations, the associations between funding agencies and the topics they fund are considered; they analyse funding patterns at both organisational and topic levels (Jin et al., 2022). Song et al. (2017) extended the Dirichlet multinomial regression to a journal–time–topic model to rank journals and then analysed the journals' contributions. They interpreted this contribution as the degree of attention the journal paid to the topic.

Topic modelling

Topic modelling is a core component of topic evolution, and the accuracy of the topic identification results has a considerable impact on topic evolution. Topic modelling methods can be divided into four main types: co-citation analysis, co-word analysis, LDA-related models, and semantic cluster approach. Co-citation analysis uses the relations between citing and cited publications and has been used to identify research topics in various disciplines (Chang et al., 2015; Hou et al., 2018; Réale et al., 2020). However, the connection between citations and the originality, significance, or quality of work, differences among source materials in technical and applied fields, and biases in databases have been extensively criticised (Zhang et al., 2017). With the development of natural language processing tools, the content analysis of topic evolution has become increasingly convenient and popular. Co-word analysis is an easy and useful method for scholars, often used in research on topic evolution (Bai et al., 2021; Murgado-Armenteros et al., 2015). LDA is a classical model of topic detection. Researchers have conducted several meaningful topic evolution analyses using LDA and its related models. The popularity of the semantic cluster approach stems from the successful use of Word2Vec (Mikolov et al., 2013), which contributes significantly to representation learning. Studies are more likely to use Word2Vec to obtain semantic relations (Huang et al., 2022). Bidirectional Encoder Representations from Transformers (BERT), a new language model, has achieved better results in most natural language processing tasks and is trained on massive amounts of text; thus, it has been proven to have excellent semantic representation (Devlin et al., 2019). This makes it useful for improving the accuracy of topic detection. Regarding document vectors, Sentence-BERT achieved superior results in many tasks (Reimers & Gurevych, 2019). SciBERT accurately represents academic terms (Beltagy et al., 2019). After generating semantic embeddings, researchers tend to use cluster algorithms (e.g. k-means and DBSCAN (Density-Based Spatial Clustering of Applications with Noise)) to cluster vectors into topics.

BERTopic and Top2Vec are two popular algorithms used for topic modelling. BERTopic is an integrated method with the following components: Sentence-BERT, UMAP (Uniform Manifold Approximation and Projection), HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise), and c-TF-IDF (Grootendorst, 2022). This addresses the inconsistency between density-based clustering and centroid-based sampling. It first uses Sentence-BERT to generate document vectors. To reduce the complexity of high-dimensional vectors, UMAP is used to reduce the vectors' dimensionality (McInnes et al., 2018). Subsequently, a hierarchical density-based algorithm, HDBSCAN, is employed to cluster the vectors by identifying the dense and stable regions in the data. Finally, c-TF-IDF is used to generate topic-representative words. Top2Vec is an algorithm that automatically detects topics and generates jointly embedded topics, documents, and word vectors (Angelov, 2020). Top2Vec's processes are similar to BERTopic in that they involve common steps in topic modelling.

Unlike previous studies, we extended the meaning of topic focus and regarded it as a representation of journal preferences by conducting topic evolution and constructing relations between journals and topics. We used SciBERT to represent document embeddings and then used k-means to cluster topics, which served as the foundation for calculating the journal preference score. Moreover, we did not incorporate time into the topic model because (1) we wanted to explore the journal's contribution to the topic clearly and accurately without other factors. Further, (2) the classical methods do not integrate time or use semantics for topic clustering. However, it is necessary to consider the dynamic characteristics of journal preferences, and it is also meaningful to study topic evolution based on journals rather than simply treating topics as research objects. Therefore, we added a time factor after detecting topics in different time slices.

Methodology

In this section, we present the three-stage method shown in Fig. 1. Journal articles were used as inputs, with each article treated as a document and analysed using this three-stage method. The first stage provided a conceptual framework for topic focus and journal preferences. In the second stage, we generated a topic vocabulary for fine-grained topic words through data preprocessing. Subsequently, we processed the data into documents consisting of only fine-grained topic words and generated document vectors. After applying the

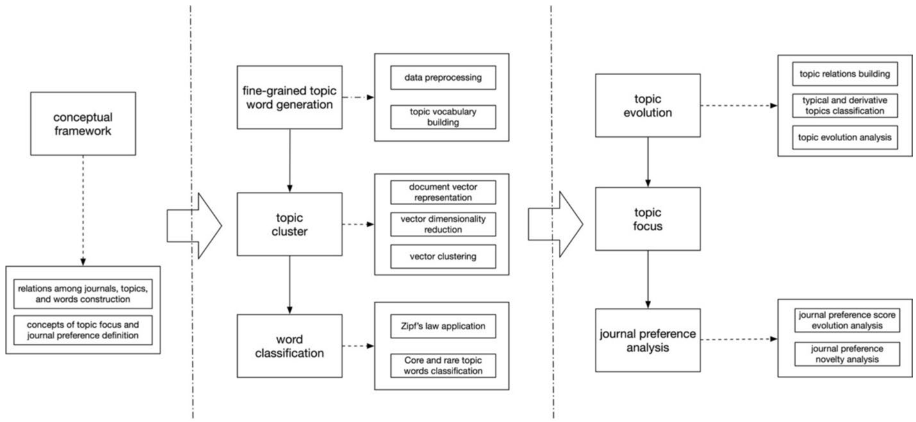

Fig. 1 The framework of the methodology

UMAP algorithm to reduce vector dimensionality and using the k-means algorithm to cluster the vectors, we classified core and rare topic words for each topic. In the third stage, we matched the same topic across different years to establish topic relations. Then, we obtained the results of topic evolution by matching journals and topics and analysed journal preferences based on topic focus.

Topic focus and journal preference

In this section, we define the concepts of topic focus and journal preference, and clarify their relationships, approaching them from the perspective of topic evolution. As mentioned by Song et al. (2017), a journal's contribution is measured by the attention it devotes to the topic. We extend this statement and claim that topic focus is a journal's act or state of applying the mind to a topic. Furthermore, topic focus was used to represent journal preferences. Journal preference is a journal's predisposition to favour something; therefore, it is the behavioural tendency of the journal. When based on topic focus, it is the journal's predisposition to favour the topic. Topic focus analysis describes how topics capture user focus, whereas journal preferences describe how topics suit user preferences. In this situation, we have the opportunity to investigate journal preferences by analysing the interaction between journals and topics.

First, we clarified the basic definition of a topic cluster. Let W be the set of words $W = \{w_1, w_2, \ldots, w_N\}$, where N is the total number of words. Words can refer to individual words or n-grams. Let D be the set of documents$D = \{d_1, d_2, \ldots, d_M\}$, where M is the total number of documents. Let J be the set of journals$J = \{j_1, j_2, \ldots, j_K\}$, where each journal is represented by an element of$J$. A mapping function $f : D \rightarrow J$ is defined to represent the assignment of each document in the document set D to its corresponding journal within the journal set J. For every document$d_i \in D$, the mapping function f maps it to a journal$j \in J$, such that$f\left(d_i\right) = j$. For any two distinct documents d_i and d_j ($d_i \neq d_j$), if they belong to the same journal, then$f\left(d_i\right) = f\left(d_j\right)$; if they belong to different journals, then$f\left(d_i\right) \neq f\left(d_j\right)$.

Based on the topic cluster, a topic set Z is defined as the probability distribution of W, and the word distribution of topic z is a mapping $\beta_z : W \rightarrow [0,1]$ (topic-word distribution). Document-topic distribution is defined as a probability distribution over a set of topics for a given document. Specifically, for document$d$, the document-topic distribution is a mapping$\theta_d : Z \rightarrow [0,1]$, where Z is the set of topics, and each element $\theta_{d,z}$ of the vector θ_d denotes the proportion of topic z in document$d$:$\theta_d = (\theta_{d,1}, \theta_{d,2}, \ldots, \theta_{d,K})$, where θ_d follows a dirichlet distribution, and the sum of the proportions for all topics in a document equals 1:$\sum_{z=1}^{K} \theta_{d,z} = 1$. If it is a topic cluster for documents, then the document belongs to one topic. For each document, there is exactly one topic z for which$\theta_{d,z} = 1$, and for all other topics$z\prime \neq z$,$\theta_{d,z\prime} = 0$.

Therefore, topic focus is defined as the probabilistic mapping of the distribution of journal J for each topic z (journal-topic distribution). Let $\beta_j : Z \rightarrow [0,1]$ be the topic distribution for journal$j$, representing the journal's degree of focus on various topics. This mapping can be specifically expressed as$\beta_j(z) = P(z|j)$, where $P(z|j)$ denotes the probability of selecting (i.e. focusing on) topic z given journal$j$. The sum of the probabilities for all topics of the same journal j equals 1:$\sum_{z \in Z} \beta_j(z) = 1$. Additionally, journal-topic-word distribution can be interpreted as follows: for each journal j from$J$, a mapping from words to [0,1] is yielded by the sum over $\beta_j(z) * \beta_z(w)$ for all pairs (z, w) in$Z \times W$.

Journal preferences based on topic focus refer to the interpretation of journal preferences through the lens of topic focus. We use the ordering relation to define journal preference, which indicates which topic is better chosen by the journal between topics z_1 and z_2. Let Z be a selection collection, and let the preference relation be denoted by the symbol $\succsim$. For any two topics $z_1, z_2 \in Z$ that are available for selection, $z_1 \succsim z_2$ indicates that selecting z_1 is considered at least as good as (or better than) selecting z_2. If the journal defines a strict preference relation such that selecting topic z_1 is considered better than selecting topic z_2, this can be represented as $z_1 \succ z_2$. Figure 2 presents a schematic of the relations among topic clusters, topic focus, and journal preferences, illustrating the three layers: journals, topics, and words. Entities corresponding to each layer are depicted, with various relationships among these entities clearly marked.

Based on journal preference, we define the degree of journal preference based on topic focus, which is called the journal preference score. Let $u : Z \rightarrow \mathbb{R}^+$ be a real-valued function. Define u as a utility function that represents the utility value of each option z in the set Z. For any $z_1, z_2 \in Z$, if $u(x) > u(y)$, it indicates that the journal has a higher preference for topic z_1 over topic z_2. The strict preference relation $z_1 \succ z_2$ is equivalent to $u\left(z_1\right) > u(z_2)$.

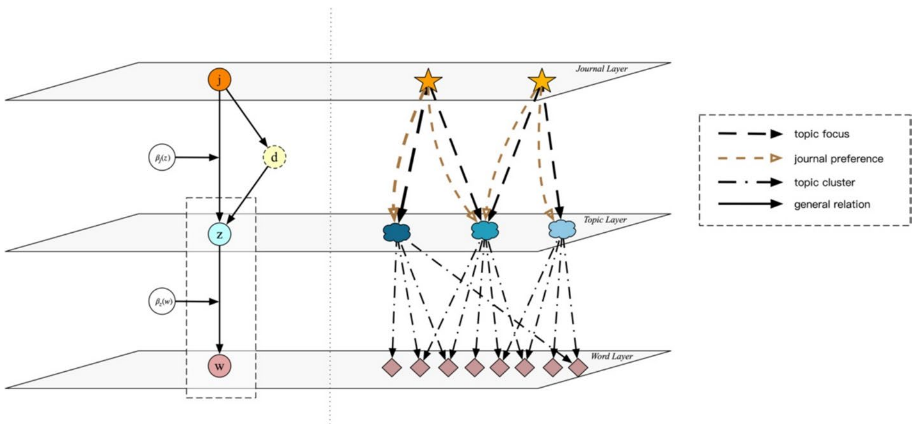

Fig. 2 The schematic of relations among topic clusters, topic focus, and journal preferences

We used a preference score to calculate the degree of journal preference. By analysing the degree of topic focus, comparable journal preference scores were calculated for the same topic. It is worth noting that this type of preference is different from the journal's publication preference, which is reflected in the acceptance rate of papers—opaque data for academics. The journal preference score calculation considers that when a document is near the cluster centre, the document topic is closely related to the cluster's topic. The contribution of documents to the same topic is different; therefore, the contribution of documents to a topic should be used in the journal preference score calculation. The journal preference score of Journal A for Topic B can be calculated as:

$$JP_{A-B} = \sum_i^n \left(\frac{1}{index_i} \times \frac{1}{1+\ln\left(1+distance_i\right)} \right). \tag{1}$$

where n is the number of documents in Journal A clustered in Topic B, $distance_i$ is the Euclidean distance between the document vector i and the cluster centre vector, and $index_i$ is the index of the current document in all documents in this cluster, ranked by distance. The preference score of Journal A in Topic B was calculated by summarising the preference scores of all documents belonging to Journal A. The accumulation reflects the large number of papers themselves and also reflects the journal's attention to the topic. It is worth noting that only the preferences computed for the two journals on the same topic are comparable. However, this method is still sensitive to the number of papers published by the journals. This score eliminates the impact of the number of papers published in a single journal. Based on the distance, the index acts as a balancing factor, and together with the *ln*() function, both work jointly to ensure a more even contribution of different journals to the topics, thereby reducing the excessive influence of large-scale data on the results. However, if a journal publishes a large number of papers on that topic, those papers are also likely to be located near the cluster centre, so the journal preference score will be large. The analysis of journal preferences used these scores.

Data processing

The data processing included fine-grained topic word generation, topic clustering, and word classification. First, research data were obtained and processed into fine-grained topic words. To explore the contributions of the journals, we selected a specific journal list to retrieve research data. Using these journal lists, papers were obtained from a specific database. To better represent semantic meanings, fine-grained topic words were used, as research has shown that a single word cannot reflect the complete meaning of a research topic (Li et al., 2018; Zhang & Yu, 2020). We mainly used ITGInsight software (Wang et al., 2022b) to extract fine-grained topic words. The first step was to generate n-grams excluding unigrams. Data were imported into ITGInsight, where the function 'Data Cleaning' was used. The minimum n-grams length is set to 2 and then lengths are set based on the range of n-gram lengths, which is determined by the distribution of more widely occurring n-grams. Based on the process, n-grams were extracted as the candidate set of fine-grained topic words. The second step involved selecting meaningful n-grams as fine-grained topic words and constructing a topic vocabulary. After stemming and synonym merging, we used software to calculate the TF-IDF values of n-grams and sorted the fine-grained topic words according to their TF-IDF values from highest to lowest. When the TF-IDF values of the n-grams fell below a certain threshold, they became difficult to interpret. This value was selected as the threshold, and n-grams with TF-IDF values below it were filtered out. The selected n-grams were then queried to check whether they are domain-specific terms. The ones identified as relevant were retained as fine-grained topic words and collected into a topic vocabulary. Experts manually reviewed the final topic vocabulary. The final step involves processing the raw document data $D = [w_1, w_2, \dots, w_n]$, where D is an ordered sequence of words. The fine-grained topic words are treated as unified units in our study, without delving into the internal structure of the individual words within them. After preprocessing, the document $D\prime$ consists of only fine-grained topic words $D\prime = [p_1, p_2, \dots, p_m]$, where each fine-grained topic word p_i is a contiguous subsequence of words from D. The original order of words in D is preserved in $D\prime$, meaning that the fine-grained topic words $p_1, p_2, \dots, p_m$ appear in the same sequence as their constituent words in D. Non-fine-grained topic words are removed during this process. By ensuring that $D\prime$ retains the sequential structure of D, it remains an ordered list of fine-grained topic words rather than a set.

Second, the topics were clustered. All documents $D\prime$ consisting of only fine-grained topic words were represented using Sci-BERT. By default, Sci-BERT maps each document onto a 768-dimensional dense vector space. We employed the UMAP algorithm for vector dimensionality reduction The k-means clustering algorithm was then used to form topic clusters, and the silhouette coefficient was used to determine the optimal number of clusters. Considering the volume of data and the anticipated extent of topic analysis, we determined the range of topic numbers to be set for the experiment. Topic clusters were generated according to the best K for each year.

Third, fine-grained topic words were classified to extract those that represented the topics. Based on the word distribution of the topic, fine-grained topic words—hereinafter referred to as 'topic words'—are categorised into two distinct types: core and rare topic words. Core topic words are the foundation of a topic and reflect its intrinsic meaning. Rare topic words appear less frequently in a topic and change substantially each year, reflecting the evolution of the topic owing to its interactions with other issues. Zipf's law was used to generate representative topic words, followed by word classification (Wang et al., 2023). This process includes the following steps: First, we considered the classical Zipfian distribution, which requires the word size and rank list. Words were ranked according to their frequencies, and words with the same frequency were ranked in any order but assigned different ranks. The word list was treated as a Zipf word list, and the ranks were called Zipf rank lists. Next, we plotted the Zipfian curves, as shown in Fig. 3a. Point r_0 distinguishes between the head and tail. Words in the head are defined as common words, whereas those in the tail are defined as rare words. r_0 is the inflection point of the curve. In our study, because common words contributed more to the topic, we termed them 'core topic words'. If topic words with the same frequency were encountered while classifying topic words, all

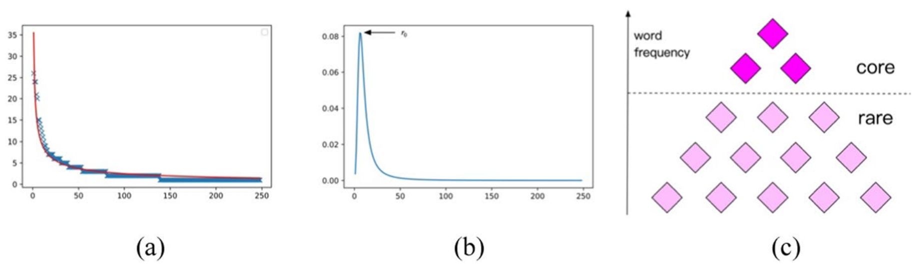

Fig. 3 Word classification through Zipfian distribution

topic words were selected for inclusion in this study. Any topic word with the same term frequency can be considered the first rank of the term frequency. To determine r_0, we plotted a curvature curve in Fig. 3b. Equation (2) is derived by combining the Zipf distribution with a curvature formula, where C and α are constants specific to the document type and can be estimated through linear regression. r_0 was determined by calculating the maximum value of the curvature curve in (3). We then used r_0 to classify topic words as core or rare, as shown in the schematic figure in Fig. 3c.

$$k(r) = \frac{C\alpha(\alpha+1)r^{-\alpha-2}}{\left[1+\left(C\alpha\cdot r^{-\alpha-1}\right)^2\right]^{\frac{3}{2}}} \tag{2}$$

$$r_0 = argmaxk(r) \tag{3}$$

Analysis of topic evolution, topic focus, and journal preferences

In this section, we describe topic evolution, topic focus, and journal preferences. First, we established the relations between topics in different years and analysed the topic evolution results. An annual time-slice was used. After obtaining the topics for each year, we matched them using the similarity of topics, which was computed based on the ratio of shared core topic words. Based on this similarity, we were able to construct topic relations between years. Thus, the topic evolution results were obtained. Based on the topic evolution results, topic clusters were divided into typical and derivative clusters. Typical topics are those that have maintained their importance throughout their evolution. At certain stages, they may receive less explicit attention owing to the emergence of derivative topics but continue to serve as a focal point for journal publications. Derivative topics are new topics that have evolved from the typical ones. After their emergence, they may reintegrate into typical topics or continue to develop independently in subsequent stages. Within a specific analysis period, these topics may exhibit various evolutionary characteristics: some emerge and gradually integrate into typical topics, some persist continuously or intermittently after their emergence, and others initially exist independently before eventually merging into typical topics.

After obtaining the journal preference scores, an evolutionary analysis of journal preferences was conducted. We used a journal preference evolution graph to show which journals exhibited a preference for the current topic and to what extent they concentrated on it. For each year, the order of the journals represents their level of preference for the topic in that year. The journal with a higher preference was placed closer to the top. The width of the journal box represents the number of articles published in the journal on the current topic during the year. Thus, the overall cumulative width per year is the number of papers published on the topic. Our analysis consisted of three parts. First, journals close to the top position were analysed. They contribute more to the current topic; therefore, it is meaningful to emphasise them. Second, we analysed journals that had a constant preference for all years. It is important to analyse persistent efforts as they represent journals that are more likely to pay attention to the topic in the future. Finally, we incorporated the analyses from top-rated journals into the above two sections because most scholars are willing to publish their papers in higher-quality journals; thus, top-rated journals will receive more attention in academia. Through the above analysis, our method can provide a clear understanding of topics in specific fields and their corresponding journal preferences, which can contribute to the selection of journals by scholars.

Based on the novelty analysis of fine-grained topic words, we explored the innovative evolution of topics from the perspective of journal preferences and their interactions with other hot issues. By calculating the novelty of core and rare topic words, we can determine whether a topic has undergone substantive innovative evolution. Changes in the novelty of core topic words reflect the evolution of the essence of topics preferred by journals, whereas changes in the novelty of rare topic words reveal the interactions between topics preferred by current journals and other trending issues. We obtained all data from these journals in the database and defined the age of the topic word as the time interval between the first appearance of the topic word in the full-year data and the current year. Thus, we obtained the novelty of the journal topic words for a specific topic within a specific year. For the topic of a specific journal in a certain year, we calculated two novelty scores of the topic by selecting the lower quartile of the rank corresponding to the core topic word and the rare topic word and used a scatter graph to show their novelty. This allowed us to determine the novelty of a journal topic word for a specific topic and explore the relationship between topic word novelty and journal preference. The above two analysis patterns and graph representations will be embedded in ITGInsight software in the future.

Case study: evolution of journal preference in the TIM field

We considered the TIM field as a case study. Technology Innovation Management (TIM) integrates multiple disciplines, including management, economics, and engineering, making it an essential interdisciplinary research area. The diversity of journal topics in the TIM field provides a robust foundation for analysing journal preferences and research trends. In the field of TIM, journal articles are the primary vehicles for disseminating academic findings. These articles typically exhibit high quality, substantial theoretical depth, and methodological rigour, ensuring that the data derived from them are reliable and stable. Leading journals in the TIM field, such as *Research Policy* and the *Journal of Product Innovation Management*, are known for their rigorous quality standards. The research they publish is not only of high academic value but also possesses broad applicability and generalisability, effectively capturing the overall development and emerging trends in the TIM field. Moreover, the topics explored in the TIM field are closely aligned with contemporary technological advancements, ensuring high relevance, considerable academic interest, and practical applicability. This alignment provides a wealth of research material and analytical opportunities. The TIM field is important and popular, particularly among scholars, policymakers, and business managers. This study aims to provide a better understanding of topic evolution and journal preference in the TIM field, which will aid scholars in effectively selecting journals for submitting papers and assist policymakers and business managers in formulating policies and enterprise development strategies.

Data preparation

The target domain was the TIM field. To reflect this field more accurately, we employed the science, technology, and innovation management fields of the Federation of Management Societies of China (FMS Journal Rating Guide). This rating list was developed using rigorous scientific methods and drew upon several prominent and authoritative journal lists, including *UTD 24*, *FT50*, *ABS*, *ABDC*, *CNRS*, and *VHB*. High-quality journals were selected and ranked from a comprehensive perspective to ensure the reliability and importance of journal ratings. Consequently, our choice to study TIM journals based on this ranking was both reasonable and representative, yielding meaningful results. There are four journal ratings (A, B, C, and D) in the FMS list. According to experts' opinions, the quality of journals decreases from A to D. The field of science and technology innovation management in China is often considered synonymous with TIM because it originally emerged from technology developed in more advanced countries, shaping the TIM field. However, its development in China occurred later, coinciding with an increasing emphasis on science in the TIM field. Hence, the term 'science and technology innovation management' is used in China. This field includes 22 journals at levels A, B, and C (D-rated journals are typically absent from prominent journal ranking lists), suggesting that these journals do not adequately represent research in the TIM field. Consequently, they were excluded from the selection process. Table 1 presents the journal classification results.

All articles used in this research were retrieved from the Web of Science (WoS) due to its extensive coverage of diverse journals and comprehensive data. We used data spanning 10 years, from 2013 to 2022. The search date was 1 February 2023. Although *Economics of Innovation and New Technology* and *European Journal of Innovation Management* were included only in the WoS core index in 2015 and 2016, they had a minimal influence on our research. We collected 13,730 papers and retained their titles and abstracts as the research data. ITGInsight was then used to generate the topic vocabulary. The initial length of the fine-grained topic words was set to 2–6, with their TF-IDF values set to greater than or equal to 20. After applying TF-IDF filtering, all resulting topic words fell within the length range of 2–4. Considering that the data for each year were mostly around 1000–2000, we selected a number between 10 and 20 as the candidates for K. The K values for different years are shown in Fig. 4.

Overall topic evolution in 2013–2022

The result of topic evolution is shown in Fig. 5, which is a flow graph drawn using the Highcharts website and then manually labelled based on it; the topic intensity is calculated using the document numbers. It can be easily observed that there has been notable growth

Table 1 FMS ABC-rated journals

Journal rank	Journals
A	Journal of Product Innovation Management (JPIM); Research Policy (RP)
B	Industrial and Corporate Change (ICC); Technological Forecasting and Social Change (TFSC); Technovation (Tech); R&D Management (RDM); Science, Technology & Human Values (STH); Journal of Technology Transfer (JTT)
C	Futures(Futu); Industry and Innovation (IAI); Research Technology Management (RTM); Science Technology and Society (STS); Evaluation (Eval); Economics of Innovation and New Technology (EINT); International Journal of Technology Management (IJTM); Journal of Engineering and Technology Management (JET-M); Technology Analysis & Strategic Management (TASM); Research Evaluation (RE); Creativity and Innovation Management (CIM); Science and Public Policy (SPP); Innovation-Organization & Management (IOM); European Journal of Innovation Management (EJIM)

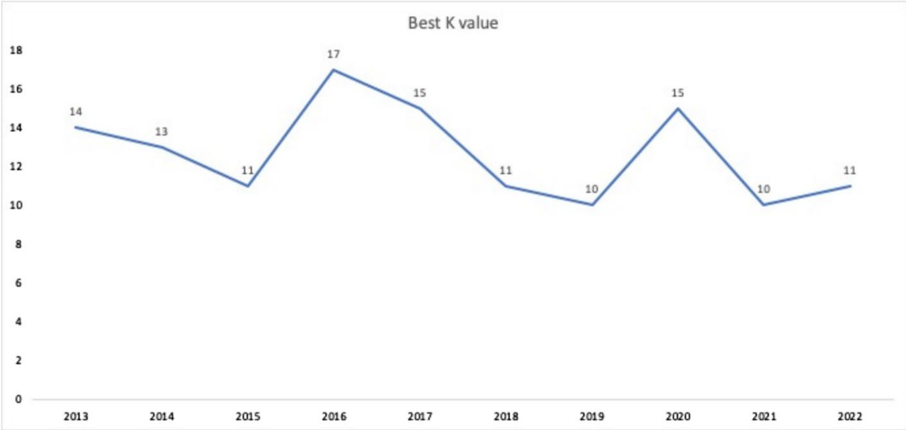

Fig. 4 The best K value in each year

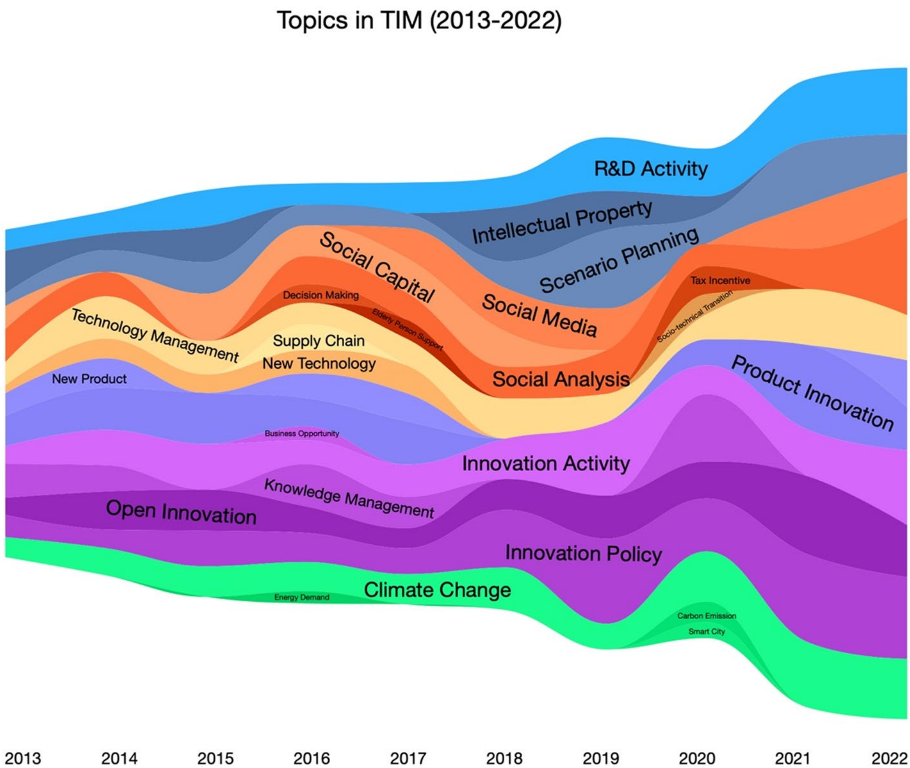

Fig. 5 Overall topic evolution of the TIM field

within the TIM field. The number of papers increased from 966 to 1922, nearly doubling over the period. This phenomenon is caused not only by the emergence of new topics but also by the expansion of some existing topics. Table 2 lists the specific topics and their

Table 2 Topics in the TIM field

Type	Topic
Typical topic clusters	R&D activity, scenario planning, social analysis, technology management, innovation activity, open innovation, innovation policy, climate change
Derivative topic clusters	Intellectual property, social media, social capital, decision making, tax incentive, elderly person support, socio-technical transition, supply chain, new technology, new product, product innovation, business opportunity, knowledge management, carbon emission, energy demand, smart city

corresponding classifications. There were 8 typical and 16 derivative topics, totalling 24 different topics. A single topic cluster is referred to as a 'topic' in the following analysis. Many derivative topics, including 'tax incentive', 'elderly person support', 'sociotechnical transition', 'business opportunity', 'supply chain', 'carbon emission', 'energy demand', and 'smart city', appear only once. These topics are covered from either 2016 or 2020. 'Decision making', 'tax incentive', and 'elderly person support' are derived from 'social analysis', while 'socio-technical transition' represents the integration of two distinct typical topics: 'social analysis' and 'technology management'. 'Business opportunity' is highly correlated with 'innovation activity'. Business opportunities drive innovation activities by creating demand for novel solutions, whereas innovation activities enable businesses to seize opportunities through the development of new products, services, and processes. 'Supply chain' is derived from 'technology management'. As a complex, multistage system, the supply chain relies on technology management to achieve the efficient coordination of information, logistics, and financial flows. 'Carbon emission', 'energy demand', and 'smart city' were derived from 'climate change'. Climate change has driven the need to reduce carbon emissions because the accumulation of greenhouse gases in the atmosphere is a primary cause of global warming. In response to climate change, rising energy demands necessitate a shift toward sustainable solutions, and smart cities have emerged as innovative frameworks to mitigate these impacts through efficient resource management and reduced environmental footprints. Some derivative topics also appear frequently. 'Intellectual property', 'product innovation', and 'knowledge management' all appear intermittently in the domain, with 'intellectual property' being highly related to 'R&D activity', while 'product innovation' and 'knowledge management' are derived from 'innovation activity'. Similarly, 'new product' is highly related to 'product innovation', so it is also a derivative topic of 'innovation activity'. 'Social capital' and 'social media' were derived from 'social analysis' in different time slices. 'New technology' existed independently from 2013 to 2017 and was integrated into 'technology management' in 2018.

Specific topic analysis

Considering the abundance of topics, we sought advice from experts and selected four typical topics and their respective preferred journals for analysis: R&D activity, technology management, innovation activity, and climate change. These are mainstream topics that have garnered substantial scholarly attention and remain central to the field because of their enduring relevance and contribution.

Evolution of journal preference in 'R&D activity'

R&D activity is the foundation of technological innovation, driving business development and fuelling economic growth. A journal preference evolution graph for this topic is shown in Fig. 6. The topic has been active for the past ten years and has become increasingly attractive to journals, with the number increasing from a minimum of 11 journals to all journals that publish on the topic by 2022. Since 2019, the data size of this topic has grown rapidly, with *Technological Forecasting and Social Change* making significant contributions to its development. We consider common top journals to be those that have ranked in the top 5 for more than 5 out of 10 years. The ranking of preferences for the topic was consistently diverse, and the common top journals included *Research Evaluation*, *Science and Public Policy*, *Research Policy*, *Journal of Technology Transfer*, and *Technological Forecasting and Social Change*, which appeared 10, 10, 8, 7, and 6 times, respectively. One A-rated, two B-rated, and two C-rated journals were included. These five journals had many papers focusing on the topic, showing that they were deeply devoted

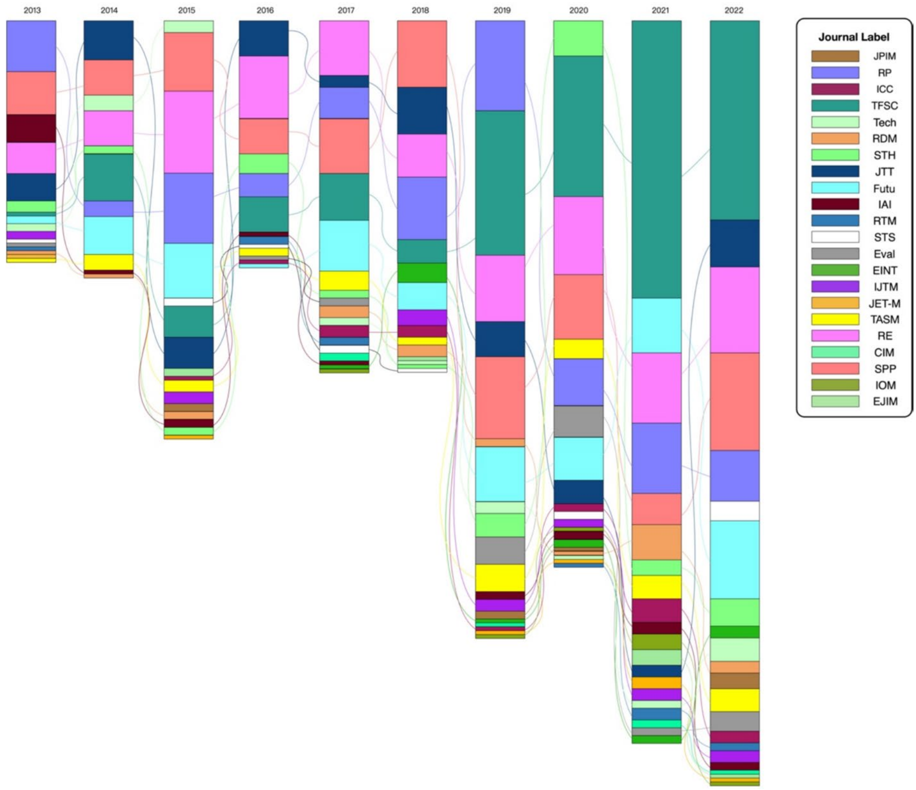

Fig. 6 Journal preference evolution graph of R&D activity

to it. *Technovation* and *Science, Technology & Human Values* ranked first in 2015 and 2020, with a relatively small number of papers. They also ranked high, with a relatively small number of papers published in some years. Many journals have long preferred this as the basic topic in the TIM field. Eight journals had a ten-year preference for the topic, including *Research Policy*, *Journal of Technology Transfer*, *Technological Forecasting and Social Change*, *Science, Technology & Human Values*, *Futures*, *Science and Public Policy*, *Technology Analysis & Strategic Management*, and *Research Evaluation*. The long-preference journals include one A-rated journal, three B-rated journals, and four C-rated journals. Another A-rated journal, *Journal of Product Innovation Management*, showed little interest in this topic, publishing totally nine papers in 2015, 2019, 2020, and 2022, and ranked low in terms of preferences in these four years.

We then analyse the novelty of journals' topic words in 'R&D activity', which is an auxiliary analysis of journal preference. A novelty scatter plot is shown in Fig. 7. When the colour of the circle is darker, the novelty of the corresponding journal topic words is higher in that year. It is worth emphasising that the data used to calculate novelty used the lower quartile, which will not be mentioned in the following section to avoid redundancy. Figure 7 shows core topic word novelty in 2021 because of the global spread of COVID-19, and therefore 'COVID-19 pandemic' becomes core topic words in the current year. *R&D Management* and *Research Technology Management* are the most dominant journals on this topic. This represents a substantial innovative evolution of the topic, whereas the preference for *R&D Management* and *Research Technology Management* shows high novelty. Of the nine papers published by *R&D Management* in 2021, seven used the term 'COVID-19 pandemic', which appeared 12 times. *Research Technology Management* has published three papers on the topic, two of which refer to the term 'COVID-19 pandemic', appearing twice. For rare topic words, *Technovation* is an example. In its preference for 'R&D activity' in 2021, its rare topic words were very novel. Although *Technovation* did not contribute to the core topic words in 2021, its rare topic words such as 'digital transformation' (appearing 4 times), 'entrepreneurial ecosystem' (appearing 9 times), and 'business ecosystem' are quite innovative. These terms represent *Technovation*'s interactions with hot issues within the 'R&D activity', and these interactions have not impacted the substantive content of the topic. At the same time, we also found that among the rare topic words, the relatively novel journals were *Industrial and Corporate Change, Technovation,*

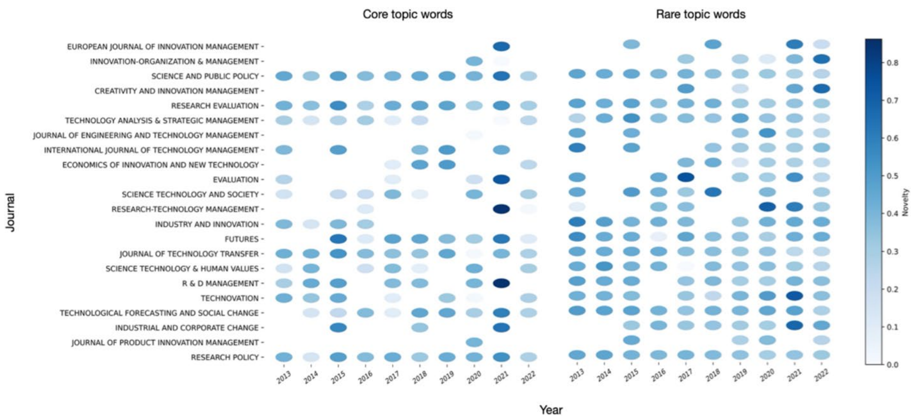

Fig. 7 Novelty scatter in R&D activity

Futures, Industry and Innovation, Research Technology Management, Science Technology and Society, Evaluation, International Journal of Technology Management, Creativity and Innovation Management, Innovation-Organization & Management, and *European Journal of Innovation Management*. Most journals were C-rated journals.

We then considered the topic word novelty of the high-preference journals. *Technovation* ranked first in preference in 2015, and *Science, Technology & Human Values* ranked first in 2020, and both had average novelty. *Research Policy*, *Research Evaluation*, and *Science and Public Policy* both had high and long preferences. Their core topic words also had average novelty, and their rare words were novel in the early years but old in recent years. We also explored topic word novelty in long-preference journals. We found that long-preferred journals do not usually show high novelty. Except for *Futures,* neither the core nor rare topic words of *Research Policy*; *Journal of Technology Transfer*; *Technological Forecasting and Social Change*; *Science, Technology & Human Values*; *Science and Public Policy*; *Technology Analysis & Strategic Management*; or *Research Evaluation* showed strong novelty.

Evolution of journal preference for 'technology management'

Technology management is a classic research topic in the TIM field, which not only promotes the innovative development of enterprises but also drives the innovation of management ideas and methods. A journal preference evolution graph for this topic is shown in Fig. 8. As shown in Fig. 8, numerous journals have focused on this topic, ranging from a minimum of 16 journals to all journals that published on the topic by 2021. Common

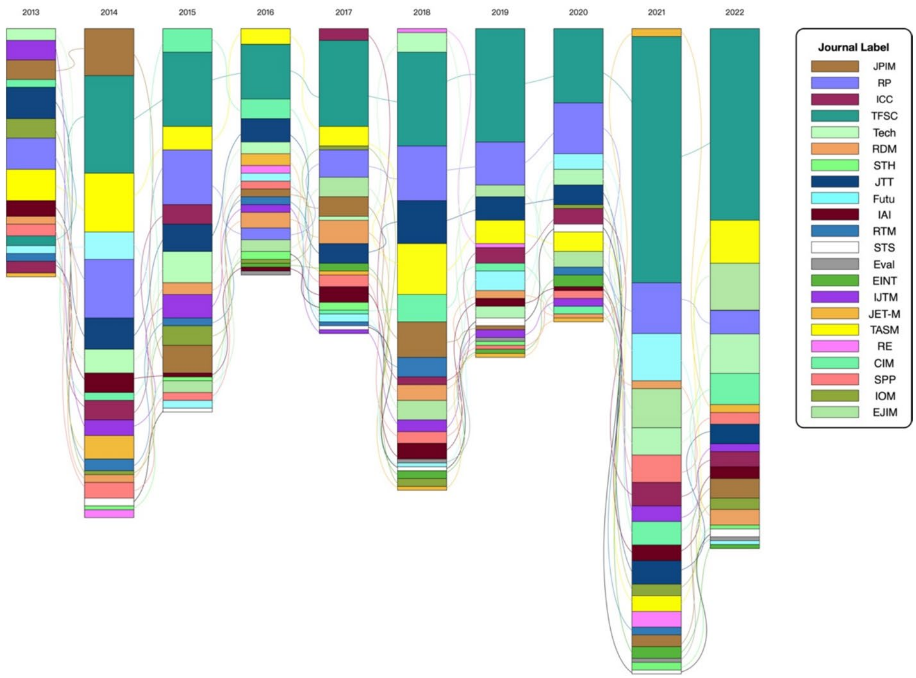

Fig. 8 Journal preference evolution graph of technology management

top journals included *Technological Forecasting and Social Change*, *Research Policy*, and *Technology Analysis & Strategic Management*, which appeared 9, 8, and 6 times, respectively. There were two B-rated journals and one C-rated journal. In 2021 and 2022, the topic increased in size, largely owing to contributions from *Technological Forecasting and Social Change*. The journals with the highest preference changed over time until *Technological Forecasting and Social Change* had the highest preference in 2019. *Journal of Product Innovation Management* had a strong focus on this topic for several years. Although it was not as prominent as *Technological Forecasting and Social Change*, it ranked first in 2014, using less data than *Technological Forecasting and Social Change*. Another A-rated journal, *Research Policy*, has also paid considerable attention to this topic. Its ranking was generally moderate but high in 2019, 2020, and 2021. *Technovation*, *Journal of Product Innovation Management*, *Creativity and Innovation Management*, *Technology Analysis & Strategic Management*, *Industrial and Corporate Change*, *Research Evaluation*, and *Journal of Engineering and Technology Management* ranked first in different years, showing ranking uncertainty in the evolution of this topic. Eleven journals have had a long-term preference for this topic: *Research Policy*, *Technological Forecasting and Social Change*, *Technovation*, *R&D Management*, *Journal of Technology Transfer*, *International Journal of Technology Management*, *Futures*, *Technology Analysis & Strategic Management*, *Industry and Innovation*, *Creativity and Innovation Management*, and *Science and Public Policy*.

A scatter plot of the novelty of the topic words in 'technology management' is shown in Fig. 9. Compared to other topics, both core and rare topic words were relatively novel for this topic. Most journals that contributed core topic words demonstrated several years of core topic word novelty. Although the novelty of *Futures* was very low in the core topic words, it was very high in the rare topic words in some years. *Evaluation* and *Research Evaluation* were similar. Many journals were innovative in core topic words, indicating that most made positive contributions to the evolution of the topic. The core topic words of this topic also showed obvious novel performance in 2021, which was influenced by 'digital technology', 'artificial intelligence' and 'big data'. These advancements in digital technology, artificial intelligence, and big data have collectively revolutionised technology management by enabling data-driven decision-making, intelligent automation, and

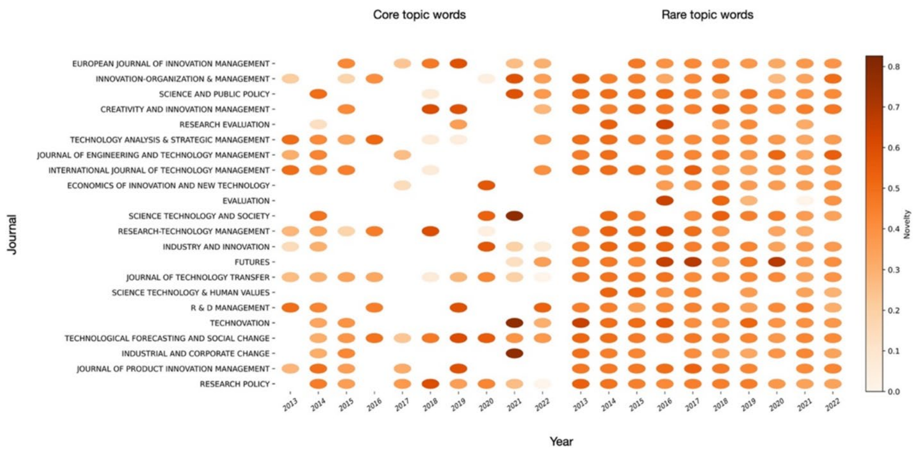

Fig. 9 Novelty scatter in technology management

innovation at an unprecedented speed and scale. *Industrial and Corporate Change*, *Technovation,* and *Science, Technology and Society* contributed greatly to the substantive innovation of this topic because of their preferences. Journals that exhibited significant novelty in rare topic words included *Futures*, *Evaluation*, and *Research Evaluation*. All of these were C-rated journals.

As representative journals with high preferences, *Technological Forecasting and Social Change* had novel core words in 2019 and 2020. *Technovation* had a strong topic preference in 2013, and its rare words showed strong novelty in that year compared to other years. *Journal of Product Innovation Management*, *Creativity and Innovation Management,* and *Technology Analysis & Strategic Management* all showed a relatively general level of novelty in both core and rare topic words when they showed a strong preference for this topic. *Research Evaluation* and *Journal of Engineering and Technology Management* did not contribute any core topic words, and their rare topic words had very low novelty in their high-preference years. We then analysed the topic word novelty of journals with long preferences. *Technological Forecasting and Social Change*, *Research Policy*, *R&D Management*, *Industry and Innovation*, *Science and Public Policy*, and *Technology Analysis & Strategic Management* had relatively general-level core and rare topic words. *Journal of Technology Transfer*, *Futures*, and *International Journal of Technology Management* were not novel in core topic words but were average in rare topic words. Conversely, *Creativity and Innovation Management* was novel in core topic words, with mediocre performance in rare topic words. Only *Technovation* exhibited novelty in both core and rare topic words in certain years.

Evolution of journal preference for 'innovation activity'

Innovation activity is a general but important topic in the field of TIM, which can drive the long-term development and performance improvement of enterprises. We plotted a journal preference evolution graph for this topic, as shown in Fig. 10. At its highest, this topic received attention from 21 journals in 2019. The paper sizes on the topic were very large in both 2019 and 2022 owing to contributions from *Technological Forecasting and Social Change.* Common top journals included *Technological Forecasting and Social Change*, *Research Policy*, and *Journal of Product Innovation Management*, which appeared 9, 8, and 6 times, respectively. All of these were A- or B-rated journals. *Technological Forecasting and Social Change* accounted for a large proportion of the preferences for innovation activities. It had a high preference for this topic, ranking first and second most of the time. Other journals, such as *Research Policy* and *Technology Analysis & Strategic Management*, often had strong preferences, but their rankings fluctuated more widely than *Technological Forecasting and Social Change*. *European Journal of Innovation Management* has exhibited a high preference for this topic over the past 5 years. By 2022, it reached top position. Note that *R&D Management* had a high journal preference score in 2016 despite the small number of papers related to this topic. *Technological Forecasting and Social Change, Research Policy, Journal of Product Innovation Management, R&D Management, Journal of Technology Transfer, Technology Analysis & Strategic Management, Creativity and Innovation Management, International Journal of Technology Management,* and *Technovation* have all persisted on this topic for 10 years. Most journals are A- or B-rated.

Figure 11 shows the novelty of journal topic words in 'innovation activity'. Many journals were innovative in core topic words, meaning that most made significant contributions to the evolution of the topic. Although the novelty of this topic did not show clear

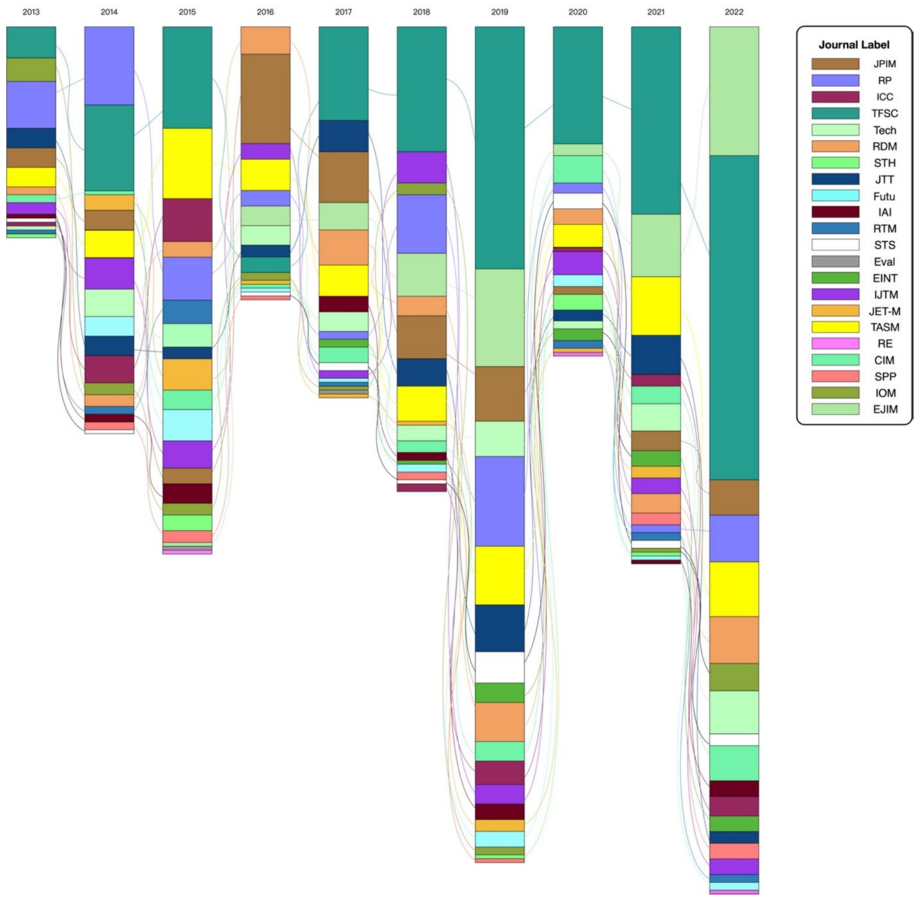

Fig. 10 Journal preference evolution graph of innovation activity

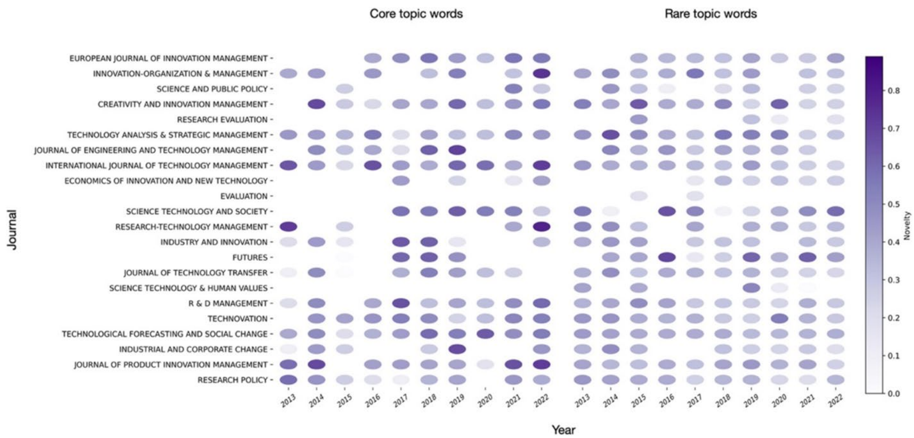

Fig. 11 Novelty scatter in innovation activity

year-based clustering, 2022 had relatively more journals with notable novelty than other years. By analysing the topic words of novel journals, we found that this topic was also influenced by digital innovation. Digital innovation accelerates processes, enables new capabilities, and fosters collaboration across boundaries, thereby leading to increased efficiency and novel outcomes. This leads to a more innovative evolution of the topic. Journals that demonstrated novelty in rare topic words in certain years included *Futures*, *Science, Technology and Society*, *Technology Analysis & Strategic Management*, *Creativity and Innovation Management*, and *Innovation-Organization & Management*. All of these were C-rated journals. We also found that journals that were highly novel in their core topic words did not show strong novelty in rare topic words.

Although *Technological Forecasting and Social Change* has paid considerable attention to this topic, its core topic words and rare topic words are not novel, meaning that this journal tends to publish typical papers with high and long preferences for this topic. The topic words of *Research Policy*, *R&D Management*, and *European Journal of Innovation Management* were not novel in their high-preference years. Regarding long-preferred journals, *Technovation*, *Research Policy*, and *Journal of Technology Transfer* had novel core or rare topic words. The core topic words of *R&D Management*, *Journal of Product Innovation Management*, and *International Journal of Technology Management* were novel for some years, but their rare topic words were not. In contrast, the core topic words of *Technology Analysis & Strategic Management* were not novel, and their rare topic words showed novelty in 2014. The core and rare topic words of *Creativity and Innovation Management* have shown novelty in certain years.

Evolution of journal preference for 'climate change'

Climate change has recently emerged as a prominent topic in various fields, including TIM. We plotted a journal preference evolution graph for this topic, as shown in Fig. 12. This topic has attracted journals, ranging from a minimum of 15 journals to all journals publishing on the topic by 2021. Although some journals focused on this topic, they had a low preference for it. The year 2021 had the highest number of publications, owing to the journal *Technological Forecasting and Social Change*'s strong attention to the topic. *Technological Forecasting and Social Change* and *Research Policy* were the most common top journals, appearing 10 and 9 times, respectively. From 2015 to 2018, *Creativity and Innovation Management*, *Industrial and Corporate Change*, *R&D Management*, *Technology Analysis & Strategic Management*, and *Journal of Product Innovation Management* had a higher preference for this topic than *Technological Forecasting and Social Change*. Moreover, nine journals had long preferences for this topic on a year-on-year basis, including *Technological Forecasting and Social Change*, *Science and Public Policy*, *Technovation*, *Research Policy*, *Journal of Engineering and Technology Management*, *Futures*, *Industrial and Corporate Change*, *Creativity and Innovation Management*, and *Journal of Technology Transfer*.

Figure 13 shows the novelty of the journal topic words in 'climate change'. Most journals have a weak sense of contribution to the core topic words, but most contributing journals have relatively novel performances. *Technovation* and *Journal of Engineering and Technology Management* are the most novel performers in this topic in 2022. A common feature of both journals was the emergence of research on autonomous cars. The emergence of autonomous cars in the context of climate change reflects the potential for technological innovation to address global environmental issues. This connection promotes the

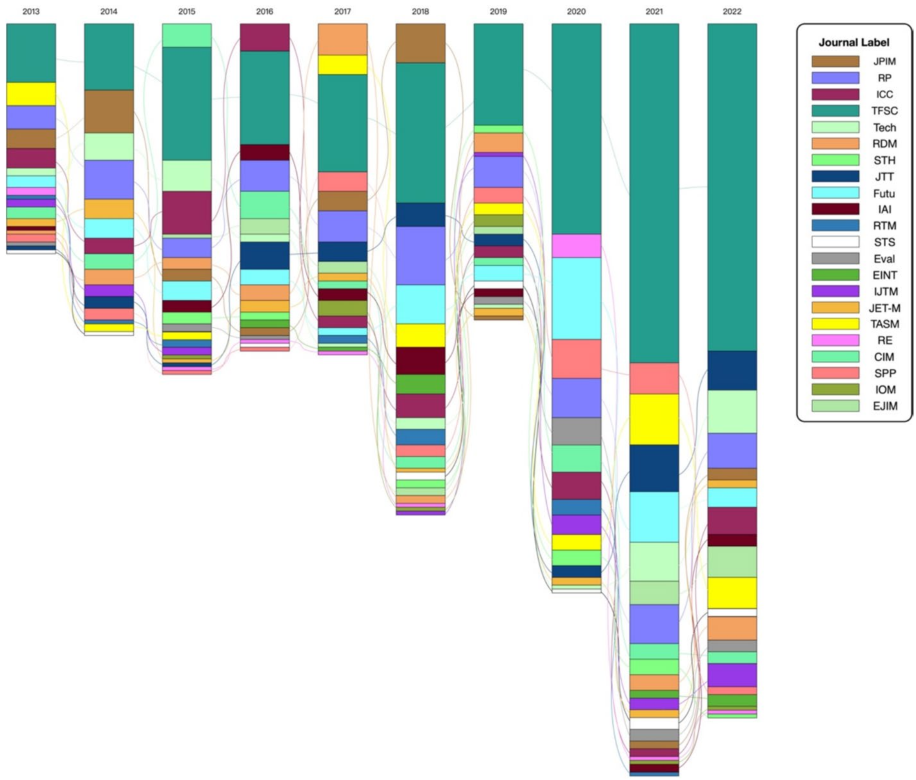

Fig. 12 Journal preference evolution graph of climate change

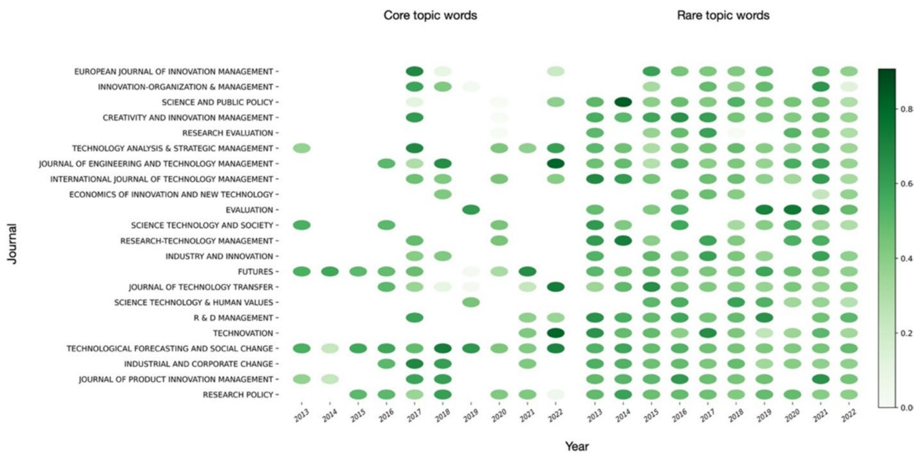

Fig. 13 Novelty scatter in climate change

novelty of the research content and interdisciplinary integration, making the topic more innovative. This has led to a substantial evolution of the topic. Journals that were novel in the rare topic words in certain years include: *Technovation*, *R&D Management*, *Journal of*

Technology Transfer, *Research Technology Management*, *Science, Technology and Society*, *Evaluation*, *International Journal of Technology Management*, *Creativity and Innovation Management*, *Science and Public Policy*, and *Innovation-Organization & Management*. Most of these were C-rated journals.

As for the core topic words, *Technovation* and *Journal of Engineering and Technology Management* performed excellently in novelty in 2022. Regarding rare topic words, *Research Technology Management* and *Science and Public Policy* showed strong novelty in 2014. *Evaluation* showed strong novelty in 2019, 2020, and 2021. Although the high-preference journal *Technological Forecasting and Social Change* did not have the most novel core or rare topic words, it performed relatively well in the novelty analysis. *Creativity and Innovation Management* showed the strongest topic preference in 2015 but did not contribute to core topic words, and its rare topic words showed only moderate novelty. *Industrial and Corporate Change*, *R&D Management*, and *Journal of Product Innovation Management* showed the strongest topic preferences in three different years, but their core and rare topic words performed equally well. *Technological Forecasting and Social Change*, *Technovation*, *Journal of Engineering and Technology Management*, and *Journal of Technology Transfer* had novel core topic words and average novelty in rare topic words over certain years. Unlike the above four journals, the core topic words of *Science and Public Policy* were not novel; however, their rare topic words were novel for some years. Neither the core topic words nor the rare topic words for *Industrial and Corporate Change* were novel. *Research Policy*, *Futures*, *and Creativity and Innovation Management* were moderate in core and rare topic words.

Summary of journal preference evolution

The study indicates that within a given topic, while the composition and ranking of top journal preferences fluctuate over time, a subset of journals consistently exhibits dominance, appearing in the top ranks across most years. For 'R&D activity', common top journals include *Research Evaluation*, *Science and Public Policy*, *Research Policy*, *Journal of Technology Transfer*, and *Technological Forecasting and Social Change*, which appeared 10, 10, 8, 7, and 6 times, respectively. As for 'technology management', common top journals include *Technological Forecasting and Social Change*, *Research Policy*, and *Technology Analysis & Strategic Management*, which appeared 9, 8, and 6 times, respectively. For 'innovation activity', common top journals included *Technological Forecasting and Social Change*, *Research Policy*, *and Journal of Product Innovation Management*, which appeared 9, 8, and 6 times, respectively. *Technological Forecasting and Social Change* and *Research Policy* are the common top journals, appearing 10 and 9 times, respectively, in 'climate change'. For a single topic, the preference for a small number of journals is advantageous. Among all four topics, *Technological Forecasting and Social Change* and *Research Policy* showed advantages. There is no clear relationship between ranking and journal preference, but A- and B-rated journals both had dominant topics in specific years. The two A-rated journals exhibited the highest preference for two topics in a given year. *Research Policy* preferred 'R&D activity' and 'innovation activity', while *Journal of Product Innovation Management* preferred 'technology management' and 'climate change'. For B-rated journals, *Technological Forecasting and Social Change* showed a preference advantage in all four topics; *Industrial and Corporate Change* showed a preference advantage in two topics, 'technology management' and 'climate change'; *Technovation* showed an advantage in 'R&D activity' and 'technology management'; *R&D Management* showed an advantage in 'innovation activity' and 'climate change'; and *Journal of Technology Transfer* and *Technology & Human Values* showed an advantage in 'R&D activity'.

All four topics evolved relatively innovatively in the given years. In 2020, 'R&D activity' was affected by the COVID-19 pandemic, and topics evolved more innovatively. In 2021, both 'technology management' and 'innovation management' were influenced by digital technology and digital innovation. Climate change exhibited a unique pattern of interdisciplinary evolution. We also analysed the relationship between A- and B-rated journals with long and high preferences for topic novelty and found that these journals do not show high novelty. This may be because they have always focused on the same topic, reflecting more conservative preferences. We also found that C-rated journals are better at interacting with other issues. Specifically, in 'R&D activity', nine out of the eleven journals demonstrated novelty through the use of rare topic words are C-rated. In 'technology management' and 'innovation activity', all the journals that showed novelty in rare topic words were C-rated journals. As for 'climate change', seven out of the ten C-rated journals showed novelty in their use of rare topic words.

Method comparison and parameter evaluation

In this section, we first employ a comparative analysis to evaluate the different document vector representation methods and the optimal topic numbers for different years. Next, we explain the selection of length and TF-IDF threshold for fine-grained topic words.

We used three types of SentenceBERT (SBERT): MiniLM-L6, MiniLM-L12, and MPNet-based. MiniLM-L6 and MiniLM-L12 have 6-layer and 12-layer transformer structures, respectively, and encode text into 384-dimensional vectors. MPNet-based has a 12-layer transformer structure and is a larger model, encoding text into 768-dimensional vectors. Additionally, we used Sci-BERT because our texts are scientific documents. We also decided whether to apply UMAP for dimensionality reduction prior to topic clustering based on the set topic range. The results are summarised in Table 3, bold values highlight the highest NPMI values. We used the NPMI to evaluate the performance of the topic clusters. Its value ranges from − 1 to 1; 1 indicates that the topic is completely consistent, and − 1 indicates that the topic is completely inconsistent. It is worth noting that, because we only used fine-grained topic words for clustering—specifically high-frequency co-occurrence term pairs—the NPMI score is expected to be large. It has been demonstrated that SciBERT, combined with UMAP, achieved the best performance across all ten years, which is why we selected this combination for our research.

Next, we determined the optimal number of topic divisions, which is the most important parameter for k-means. Through a manual analysis, we identified that the number of topics corresponding to the topic granularity and scale that this study aimed to explore ranged between 10 and 20. Although NPMI is highly advanced, when used to determine the optimal cluster number K, the best K is usually distributed between 100 and 150. Such a cluster number is too large for this study because the research granularity is too fine, making it unsuitable for adoption. Therefore, we used the classical silhouette coefficient method to determine the optimal number of topics for k-means clustering on the current data. The value range of the silhouette coefficient is [− 1,1], where 1 indicates perfect clustering, 0 indicates that the clustering effect is not obvious, and a negative value indicates incorrect clustering. The stable annual optimal clustering number K after multiple experiments is shown in Table 4, bold values identify the best cluster numbers with peak silhouette coefficients. The silhouette coefficient values did not appear to be very high because of the

Table 3 Method comparison

Method	Year				
	2013	2014	2015	2016	2017
SBERT(L6)	0.6916	0.6900	0.6600	0.6608	0.6632
SBERT(L6) + UMAP	0.6851	0.6893	0.6602	0.6440	0.6418
SBERT(L12)	0.6900	0.6937	0.6638	0.6459	0.6640
SBERT(L12) + UMAP	0.6643	0.6938	0.6520	0.6652	0.6434
SBERT(MPNet)	0.6850	0.6862	0.6610	0.6636	0.6698
SBERT(MPNet) + UMAP	0.6835	0.7013	0.6677	0.6619	0.6493
SciBERT	0.7900	0.7481	0.7611	0.7578	0.7570
SciBERT + UMAP	**0.8126**	**0.7927**	**0.7983**	**0.7873**	**0.7923**
	2018	2019	2020	2021	2022
SBERT(L6)	0.6760	0.6487	0.6493	0.6212	0.6108
SBERT(L6) + UMAP	0.6660	0.6376	0.6296	0.5900	0.6149
SBERT(L12)	0.6763	0.6424	0.6431	0.6211	0.6152
SBERT(L12) + UMAP	0.6533	0.6320	0.6451	0.6156	0.6098
SBERT(MPNet)	0.6716	0.6580	0.6595	0.6239	0.6198
SBERT(MPNet) + UMAP	0.6709	0.6387	0.6410	0.6180	0.6095
SciBERT	0.7510	0.7628	0.7520	0.7033	0.7029
SciBERT + UMAP	**0.7983**	**0.7668**	**0.7668**	**0.7550**	**0.7500**

information density of the fine-grained topic words. In our primary analysis, we chose the value with the largest silhouette coefficient within the range as the optimal number of topics for each year.

We then discuss the parameter selection for topic words, which includes the length of fine-grained topic words and the TF-IDF threshold for fine-grained topic word selection. As mentioned in the data processing section, we manually filtered and constructed a topic vocabulary so that each document retained only the fine-grained topic words within the topic vocabulary and carried out a topic clustering experiment. This is explained in detail in this section. As mentioned above, phrases express semantic information better than a single word and are more suitable for topic modelling tasks. Therefore, single words were not considered. By using ITGInsight to generate n-grams, it was shown in Table 5 that the n-gram length was primarily concentrated within the range of two to six. After performing stemming and synonym merging on the n-grams, we calculated their TF-IDF values. When the TF-IDF value falls below a certain threshold, the majority of the n-grams fail to convey specific meaning. We chose this threshold value to filter out n-grams with TF-IDF scores below it. Consequently, a TF-IDF value of 20 was selected as the threshold.

Discussion

Research implications

This study has both theoretical and practical implications. First, we define the concept of journal preference. By using topic focus as a representation, we can explore a journal's predisposition in favour of certain topics. Second, considering the dynamic characteristics

Table 4 Best cluster number selection

Year	Best K	Max score	Silhouette coefficient(10–20)
2013	14	0.4477	[0.4236, 0.4354, 0.4476, 0.44496, **0.4477**, 0.4464, 0.4431, 0.4268, 0.4295, 0.4293, 0.4350]
2014	13	0.4533	[0.4250, 0.4359, 0.4378, **0.4533**, 0.4438, 0.4340, 0.4232, 0.4255, 0.4230, 0.4259, 0.4405]
2015	11	0.4552	[0.4429, **0.4552**, 0.4483, 0.4490, 0.4506, 0.4469, 0.4360, 0.4321, 0.4260, 0.4196, 0.4170]
2016	17	0.4541	[0.4327, 0.4314, 0.4454, 0.4293, 0.4333, 0.4402, 0.4497, **0.4541**, 0.4480, 0.4422, 0.4284]
2017	15	0.4460	[0.4124, 0.4182, 0.4126, 0.4348, 0.4434, **0.4460**, 0.4385, 0.4372, 0.4329, 0.4249, 0.4329]
2018	11	0.4635	[0.4578, **0.4635**, 0.4534, 0.4584, 0.4429, 0.4301, 0.4101, 0.4001, 0.4289, 0.4259, 0.4172]
2019	10	0.4333	[**0.4333**, 0.4181, 0.4172, 0.4032, 0.4065, 0.4195, 0.4082, 0.4138, 0.4177, 0.4098, 0.4116]
2020	15	0.4281	[0.4062, 0.4069, 0.3984, 0.3969, 0.4095, **0.4281**, 0.4211, 0.4252, 0.4254, 0.4251, 0.4203]
2021	10	0.4258	[**0.4258**, 0.4187, 0.4088, 0.4214, 0.4143, 0.4126, 0.3988, 0.4105, 0.4069, 0.4147, 0.4061]
2022	11	0.4208	[0.3934, **0.4208**, 0.4180, 0.4135, 0.4117, 0.4073, 0.4011, 0.4103, 0.4098, 0.4078, 0.3953]

Table 5 The document number of N-grams

N-gram's N	2	3	4	5	6	7	8+
Document numbers	18,795	13,962	5171	2430	1187	399	201

of journal preferences, we did not simply consider the relationship between journal preferences and topics for a given year but explored how journal preferences evolve over time. Evolution analysis can help understand the rising or falling trends of journal preferences, which helps in exploring the development of the field and changes in a journal's future preference. It is important to emphasise that the aforementioned theoretical contributions should be understood within the context of our research, which focuses on academic fields in which journals serve as the primary medium of publication.

By analysing journal preferences, an appropriate contribution strategy for journal selection can be identified. Choosing a suitable journal is crucial for the publication of a paper, as it not only increases the visibility and impact of the paper but also brings more academic opportunities and resources to the researcher. Understanding journal preferences helps authors more accurately locate journals that are suitable for them, thereby improving the success rate of submissions. Additionally, this analysis can promote the alignment of academic research with the needs of both the country and society. By analysing journal preferences, we can monitor their impact on the academic ecosystem and prevent the neglect of strategic and fundamental research.

Research limitations

Our study has some limitations. First, the removal of topic words with TF–IDF values of less than 20 may result in the omission of some emerging words. However, these less frequent words were not used in other common topic clustering methods. Second, we acknowledge that the high rankings of *Technological Forecasting and Social Change* and *Research Policy* are affected by their significantly higher volume of scientific publications. Although it has been verified that many papers in the core position also come from these two journals, their high ranking in relation to the attention of a certain topic reflects the objective fact that these journals have high publication volumes and numerous core papers. In future studies, we aim to consider journal preferences more comprehensively and expand our method to other fields, including those where journals are not the primary medium of publication.

Conclusion

This study explored the evolution of journal preferences based on topic focus. First, we extended the concept of topic focus, defined the concept of journal preference based on topic focus, and clarified the relationship between topic focus and journal preference. Second, we used specific clustering methods to model the proposed journal preference problem. Fine-grained topic words were processed as the smallest unit of analysis, and SciBERT was used to represent the semantics of fine-grained topic words, facilitating more precise clustering results. After dimensionality reduction and application of k-means clustering to topics, we mapped journal-topic distribution to topic focus and proposed a method to calculate the journal preference score, which comprehensively considers the distance of the paper from the topic centre and the number of papers, thereby producing precise preference ranking results. Third, we considered temporal information in our research. Evolution analysis is used for understanding the status and fluctuations of journal preferences over time. Fourth, we classified fine-grained topic words into core and rare topic words using Zipfian distribution, which contributed to establishing topic relations and novelty analysis of journal topic words. Finally, our method provides a clear understanding of topics in specific fields and their corresponding journal preferences, which is helpful for scholars, policymakers, and business managers when selecting journals, formulating policies, and developing enterprise strategies.

In this study, we used the TIM field to conduct a case study that included eight typical and sixteen derivative topics, totalling 24 different topics in the TIM field. We focused on four important topics: R&D activity, technology management, innovation activity, and climate change, and found that they all exhibit relatively innovative evolution in a given year. We explored journals with high and long preferences through a temporal analysis of journal preferences and observed the performance of A- and B-rated journals. Within a given topic, while the composition and ranking of top journal preferences fluctuate over time, a subset of journals consistently exhibits dominance, appearing in the top ranks across most years. Despite the absence of a direct correlation between preferred journals and their rankings, A- and B-rated journals often dominate preferences for specific topics. R&D activity, technology management, and innovation management have undergone innovative evolution under the influence of the COVID-19 pandemic, digital technology, and digital innovation, respectively, whereas climate change has demonstrated an interdisciplinary innovation evolution. Moreover, A- and B-rated journals with high or long preferences typically do not introduce considerable novelty, and C-rated journals are more inclined to interact with newer issues.

Acknowledgements This work was supported by the General Program of National Natural Science Foundation of China under (Grant No. 72074020), the Young Scientists Fund of National Natural Science Foundation of China under (Grant Nos. 72004009 and 72204245), and the Beijing Natural Science Foundation (Grant No. 9252016). The findings and observations in this paper are those of the authors and do not necessarily reflect the views of the supporters.

Funding National Natural Science Foundation of China,72074020,Xuefeng Wang,72004009,Hongshu Chen,72204245,Rui Guo,Natural Science Foundation of Beijing Municipality,9252016,Hongshu Chen

Declarations

Conflict of interest Authors have no conflict of interest to declare.

References

Ahmed, A., & Xing, E. (2008). Dynamic non-parametric mixture models and the recurrent Chinese restaurant process: with applications to evolutionary clustering. *Proceedings of the 2008 SIAM International Conference on Data Mining*, 219–230. https://doi.org/10.1137/1.9781611972788.20

Amiri, B., Karimianghadim, R., Yazdanjue, N., & Hossain, L. (2021). Research topics and trends of the hashtag recommendation domain. *Scientometrics, 126*(4), 2689–2735. https://doi.org/10.1007/s11192-021-03874-6

Angelov, D. (2020). Top2vec: Distributed representations of topics. Preprint retrieved from https://arxiv.org/abs/quant-ph/2008.09470

Bai, Y., Li, H., & Liu, Y. (2021). Visualizing research trends and research theme evolution in E-learning field: 1999–2018. *Scientometrics, 126*(2), 1389–1414. https://doi.org/10.1007/s11192-020-03760-7

Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A pretrained language model for scientific text. Preprint retrieved from https://arxiv.org/abs/quant-ph/1903.10676

Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. *Proceedings of the 23rd International Conference on Machine Learning*, 113–120. https://doi.org/10.1145/1143844.1143859

Burmaoglu, S., Saritas, O., Kıdak, L. B., & Berber, İC. (2017). Evolution of connected health: A network perspective. *Scientometrics, 112*(3), 1419–1438. https://doi.org/10.1007/s11192-017-2431-x

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. https://doi.org/10.1007/s11192-015-1762-8

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. https://doi.org/10.1016/j.joi.2017.10.003

Chen, H., Yang, Y., Yang, Y., Jiang, W., & Zhou, J. (2014). A bibliometric investigation of life cycle assessment research in the web of science databases. *The International Journal of Life Cycle Assessment, 19*(10), 1674–1685. https://doi.org/10.1007/s11367-014-0777-3

Chen, J., Wang, C., Shi, Q., Feng, Y., & Chen, C. (2019). Social recommendation based on users' attention and preference. *Neurocomputing, 341*, 1–9. https://doi.org/10.1016/j.neucom.2019.02.045

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Preprint retrieved from https://arxiv.org/abs/quant-ph/1810.04805

Gao, Q., Huang, X., Dong, K., Liang, Z., & Wu, J. (2022). Semantic-enhanced topic evolution analysis: A combination of the dynamic topic model and word2vec. *Scientometrics, 127*(3), 1543–1563. https://doi.org/10.1007/s11192-022-04275-z

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. Preprint retrieved from. https://arxiv.org/abs/quant-ph/2203.05794

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), 869–892. https://doi.org/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, 127*(9), 5353–5383. https://doi.org/10.1007/s11192-022-04273-1

Huang, S., Lu, W., Cheng, Q., Luo, Z., & Huang, Y. (2024). Evolutions of semantic consistency in research topic via contextualized word embedding. *Information Processing & Management, 61*(6), 103859. https://doi.org/10.1016/j.ipm.2024.103859

Invernici, F., Bernasconi, A., & Ceri, S. (2024). Exploring the evolution of research topics during the COVID-19 pandemic. *Expert Systems with Applications, 252*, 124028. https://doi.org/10.1016/j.eswa.2024.124028

Jin, Q., Chen, H., Wang, X., Ma, T., & Xiong, F. (2022). Exploring funding patterns with word embedding-enhanced organization–topic networks: A case study on big data. *Scientometrics, 127*(9), 5415–5440. https://doi.org/10.1007/s11192-021-04253-x

Jin, Q., Chen, H., Zhang, Y., Wang, X., & Zhu, D. (2024). Unraveling scientific evolutionary paths: An embedding-based topic analysis. *IEEE Transactions on Engineering Management, 71*, 8964–8978. https://doi.org/10.1109/TEM.2023.3312923

Li, B., Yang, X., Zhou, R., Wang, B., Liu, C., & Zhang, Y. (2018). An efficient method for high quality and cohesive topical phrase mining. *IEEE Transactions on Knowledge and Data Engineering, 31*(1), 120–137. https://doi.org/10.1109/TKDE.2018.2823758

Liu, P., Zhou, W., Feng, L., Wang, J., Lin, K. Y., Wu, X., & Zhang, D. (2024). Mapping and comparing the technology evolution paths of scientific papers and patents: An integrated approach for forecasting technology trends. *Scientometrics, 129*(4), 1975–2005. https://doi.org/10.1007/s11192-024-04961-0

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. Preprint retrieved from https://arxiv.org/abs/quant-ph/1802.03426

Mei, Q., & Zhai, C. (2005). Discovering evolutionary theme patterns from text—An exploration of temporal text mining. *In Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, 198–207. https://doi.org/10.1145/1081870.1081895

Miao, Z., Du, J., Dong, F., Liu, Y., & Wang, X. (2020). Identifying technology evolution pathways using topic variation detection based on patent data: A case study of 3D printing. *Futures, 118*, 102530. https://doi.org/10.1016/j.futures.2020.102530

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Proceedings of the 26th International Conference on Neural Information Processing Systems,* 2, 3111–3119. https://dl.acm.org/doi/10.5555/2999792.2999959

Murgado-Armenteros, E. M., Gutiérrez-Salcedo, M., Torres-Ruiz, F. J., & Cobo, M. J. (2015). Analysing the conceptual evolution of qualitative marketing research through science mapping analysis. *Scientometrics, 102*(1), 519–557. https://doi.org/10.1007/s11192-014-1443-z

Réale, D., Khelfaoui, M., Montiglio, P. O., & Gingras, Y. (2020). Mapping the dynamics of research networks in ecology and evolution using co-citation analysis (1975–2014). *Scientometrics, 122*(3), 1361–1385. https://doi.org/10.1007/s11192-019-03340-4

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. Preprint retrieved from https://arxiv.org/abs/quant-ph/1908.10084

Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. (2012). The author-topic model for authors and documents. Preprint retrieved from https://arxiv.org/abs/quant-ph/1207.4169

Schäfermeier, B., Hirth, J., & Hanika, T. (2023). Research topic flows in co-authorship networks. *Scientometrics, 128*(9), 5051–5078. https://doi.org/10.1007/s11192-022-04529-w

Song, M., Kim, S., & Lee, K. (2017). Ensemble analysis of topical journal ranking in bioinformatics. *Journal of the Association for Information Science and Technology, 68*(6), 1564–1583. https://doi.org/10.1002/asi.23840

Wang, X., & McCallum, A. (2006). Topics over time: A non-Markov continuous-time model of topical trends. *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 424–433. https://doi.org/10.1145/1150402.1150450

Wang, C., Blei, D., & Heckerman, D. (2012). Continuous time dynamic topic models. Preprint retrieved from. https://arxiv.org/abs/quant-ph/1206.3298

Wang, X., He, J., Huang, H., & Wang, H. (2022a). MatrixSim: A new method for detecting the evolution paths of research topics. *Journal of Informetrics, 16*(4), 101343. https://doi.org/10.1016/j.joi.2022.101343

Wang, X., Zhang, S., & Liu, Y. (2022b). ITGInsight–discovering and visualizing research fronts in the scientific literature. *Scientometrics, 127*(11), 6509–6531. https://doi.org/10.1007/s11192-021-04190-9

Wang, Z., Ren, M., Gao, D., & Li, Z. (2023). A Zipf's law-based text generation approach for addressing imbalance in entity extraction. *Journal of Informetrics, 17*(4), 101453. https://doi.org/10.1016/j.joi.2023.101453

Wang, Z., Chen, J., Chen, J., & Chen, H. (2024). Identifying interdisciplinary topics and their evolution based on BERTopic. *Scientometrics, 129*(11), 7359–7384. https://doi.org/10.1007/s11192-023-04776-5

Xu, S., Shi, Q., Qiao, X., Zhu, L., Jung, H., Lee, S., & Choi, S.-P. (2014). Author-Topic over Time (AToT): A Dynamic Users' Interest Model. *In Mobile, Ubiquitous, and Intelligent Computing: MUSIC 2013* (pp. 239–245). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40675-1_37

Zhang, H., Zhang, C., & Wang, Y. (2024a). Revealing the technology development of natural language processing: A Scientific entity-centric perspective. *Information Processing & Management, 61*(1), 103574. https://doi.org/10.1016/j.ipm.2023.103574

Zhang, J., Sun, X., & Liu, Z. (2024b). Measuring the evolving stage of temporal distribution of research topic keyword in scientific literature by research heat curve. *Scientometrics, 129*(11), 7287–7328. https://doi.org/10.1007/s11192-024-04937-0

Zhang, J., & Yu, W. (2020). Early detection of technology opportunity based on analogy design and phrase semantic representation. *Scientometrics, 125*(1), 551–576. https://doi.org/10.1007/s11192-020-03641-z

Zhang, Y., Cai, X., Fry, C. V., Wu, M., & Wagner, C. S. (2021). Topic evolution, disruption and resilience in early COVID-19 research. *Scientometrics, 126*(5), 4225–4253. https://doi.org/10.1007/s11192-021-03946-7

Zhang, Y., Zhang, G., Zhu, D., & Lu, J. (2017). Scientific evolutionary pathways: Identifying and visualizing relationships for scientific topics. *Journal of the Association for Information Science and Technology, 68*(8), 1925–1939. https://doi.org/10.1002/asi.23814

Zhu, H., Qian, L., Qin, W., Wei, J., & Shen, C. (2022). Evolution analysis of online topics based on 'word-topic' coupling network. *Scientometrics, 127*(7), 3767–3792. https://doi.org/10.1007/s11192-022-04439-x

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.