information

MDPI

Article

MFT-PTM: A Multisource-Fused and Temporally-Aware Framework for Evolutionary Analysis of Rare Earth Patent Topics Model

Haofei Zhang †, Jingyu Wang †, Jinling Yu *,† and Lixin Liu †

School of Digital and Intelligent Industry, Inner Mongolia University of Science and Technology, Arding Street No.7, Baotou 014010, China
* Correspondence: yjl1977@imust.edu.cn
† These authors contributed equally to this work.

Abstract

Rare-earth elements are critical to a wide range of high-technology applications, and analyzing patents involving rare-earth elements is essential for understanding technological progress and innovation trends. Traditional topic models cannot fully exploit patent network structures and temporal information, limiting their ability to capture the dynamic evolution of technology topics. To overcome these limitations, we propose a novel multisource-fused framework (MFT-PTM), which integrates three types of multisource features: textual, network, and temporal features via the time-aware TemporalK-Means algorithm. Specifically, we use SciBERT to extract text embeddings, TransR to generate network embeddings, and derive temporal scalars from patent data. After fusing and reducing these features with Uniform Manifold Approximation and Projection (UMAP), we apply TemporalK-Means clustering with a time-decay mechanism to capture evolutionary trends. Experiments on 43,322 rare-earth-related patents indicate that the proposed framework achieves improved performance compared with traditional methods such as LDA and BERTopic in terms of topic coherence, cluster quality, and cluster separation. Furthermore, the analysis suggests a noticeable technological transition in rare-earth applications, gradually shifting from environmental catalysis toward advanced energy and biomedical domains. Overall, the framework provides a quantitative approach for integrating multisource patent information and exploring technological evolution patterns.

Keywords: multisource fusion; data mining; patent topic modeling; patent evolution; BERTopic

Check for updates

Academic Editor: Arkaitz Zubiaga

Received: 23 January 2026
Revised: 10 March 2026
Accepted: 27 March 2026
Published: 2 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Rare-earth elements (REEs) are indispensable to the modern high-tech and green energy sectors, acting as critical components in permanent magnets, solid oxide fuel cells, and advanced catalysts. As global supply chain security and environmental sustainability become paramount, the rare-earth industry is undergoing rapid, cross-disciplinary technological transformations. Tracking the evolutionary trajectory of rare-earth-related patents provides a quintessential case study for understanding how foundational materials science drives broader technological breakthroughs. The technological lineage of rare earths traces back to the late 19th century, notably with Carl Auer von Welsbach's pioneering patents on incandescent gas mantles utilizing a mixture of rare-earth oxides (e.g., cerium and thorium) [1]. Since then, the patent landscape has fundamentally evolved

from basic illumination and extraction techniques to highly complex, multi-functional material applications. Patent literature is a vital resource for assessing technological innovation and competitiveness and serves as a comprehensive record of technological evolution. Accurately identifying technological themes from patent texts has become a crucial foundation in informetrics, underpinning tasks such as disruptive-technology detection, emerging-technology identification, evolution-path analysis, and opportunity discovery. Early bibliometric methods—like citation and keyword analysis—provided only basic insights into technological structures, often missing semantic depth and dynamic changes. Topic modeling techniques such as LDA [2] enabled latent-theme extraction across large corpora. More recently, transformer-based models like BERTopic [3] have been adopted to capture richer semantic relationships within patent texts.

Although traditional topic models (e.g., LDA, BERTopic) can extract semantic topics, they often overlook the network structures embedded in patents—such as IPC co-occurrence [4], patent family relations [5], and citation chains [6]—as well as strong temporal dependencies. Patent documents are primary carriers of technological innovation. They contain structured text, like claims and abstracts, and unstructured network data, such as IPC classifications and citation links, forming a goldmine for studying technological evolution. According to the World Intellectual Property Organization (WIPO, 2025), global PCT filings reached 273,900 in 2024, marking a return to growth with a 0.5% increase compared to 2023. The clean energy sector maintained strong momentum, while advanced materials continued to be a key driver of innovation in international patent filings [7]. Altuntas and Gök demonstrated in a renewable-energy patent-family network analysis that family-network metrics correlate with technology diffusion rates [8]. However, conventional bibliometric approaches—such as LDA and bibliographic coupling analysis [9]—face two significant limitations: (1) text modeling ignores complex inter-patent network relations; and (2) static topic assignments fail to capture dynamic technological transitions [10]. Consequently, technology forecasting frequently trails behind actual innovation cycles; for example, patents on key breakthroughs in solid-state batteries emerged 3–5 years before commercialization, yet existing models failed to detect these advancements in a timely manner [11]. Multimodal patent representation has been applied to technology fusion prediction by integrating textual semantics, IPC co-occurrence networks, and knowledge graph structures into a unified learning framework. Joint semantic–structural feature learning via deep neural networks significantly improves prediction accuracy compared with single-modality approaches [12]. Similarly, the fusion of text embedding and heterogeneous network embedding enables patents to be mapped into a shared representation space, enhancing the consistency and reliability of technology evolution identification beyond text-only modeling strategies [13]. Nevertheless, most existing methods rely on simple feature concatenation and neglect temporal decay effects on technology associations. Liu et al. observed that the influence of a cited patent on a citing patent diminishes over time; applying a time-decay weighting function to citation links can more accurately assess inter-technology relatedness [14].

Despite its advances, BERTopic-based methods for technology-theme identification exhibit several limitations. First, most implementations rely on general-domain pre-trained language models for text embedding, which often underperform when applied to specialized patent corpora [15]. Second, while BERTopic allows replacing HDBSCAN with standard K-Means, such static clustering approaches do not incorporate temporal weighting during optimization [16]. So, this study proposes a multisource, time-aware patent topic modeling framework (MFT-PTM) designed to capture technological evolution patterns from heterogeneous patent data. The rare-earth domain is employed as an empirical validation context to demonstrate the framework's analytical capability.

2. Related Work

While BERTopic has advanced patent text topic modeling, its off-the-shelf implementation still faces key challenges. First, extracting high-quality semantic embeddings from patent texts remains difficult, as general domain pretrained language models (PLMs) often underperform on specialized jargon. To address this, we employ SciBERT [17] in place of traditional BERT. SciBERT is a pretrained language model trained on scientific text, which provides more accurate, domain-adapted embeddings. Second, the default HDBSCAN clustering in BERTopic can yield unstable topic assignments and offers limited control over topic granularity, which hampers interpretability. To address these gaps, this paper investigates two research questions:

(1) Leveraging multisource patent data alongside SciBERT-based embeddings to extract deeper semantic features and improve theme discrimination.
(2) Enhancing clustering within the topic-modeling pipeline by replacing HDBSCAN with a time-aware K-Means variant to boost the coherence and interpretability of extracted technology topics.

2.1. Bibliometric Applications of Multisource Fusion

Fusing patent text, citation networks, and graph embeddings captures complementary signals missed by single-mode methods, revealing hidden relationships and innovation trajectories that underpin precise topic extraction.

Graph-based patent knowledge representation has also been extensively investigated. The PatNet patent knowledge graph has been leveraged to train various graph-embedding models, quantifying entity relatedness through cosine similarity between embeddings [18]. Hybrid frameworks combining NLP and KGE techniques have been developed for patent retrieval tasks [19]. Network-embedding approaches fused with citation analysis have been adopted to model technological ecosystems and measure knowledge-exploration distances [20]. Semantic-based similarity methods integrating multiple textual fields have also demonstrated improved performance over traditional approaches [21].

2.2. Patent Topic Modeling and Embedding Methods

BERTopic is a modular topic modeling framework combining pretrained Transformer embeddings, UMAP dimensionality reduction, clustering algorithms, and class-based TF-IDF for topic representation. It has been applied in patent analysis. To effectively capture both the semantic meaning of patent texts and the structural relationships embedded in patent networks, recent studies have increasingly integrated deep semantic embeddings with graph embedding techniques.

Deep semantic embeddings: Pretrained language models (PLMs) have revolutionized patent-text analysis. Chikkamath et al. systematically evaluated BERT and PatentBERT on USPTO IPC-classification tasks and found that domain-adaptive pre-training yields a 2% micro-F1 improvement [22]. Christofidellis et al. improved patent classification accuracy by 4% using a domain-adaptive BERT variant, BERT-LIKE [23]. However, pure text embeddings cannot capture structural relationships among technologies [24]. For example, graphene synthesis and energy-storage applications have low textual similarity but are strongly linked in the IPC co-occurrence network, underscoring the importance of structural information. SciBERT improves the performance of downstream scientific natural language processing tasks by performing unsupervised pre-training on a large-scale, multidisciplinary scientific publication corpus. For example, in the biomedical field, the best performance can be achieved in medical entity extraction tasks using PubMedBERT (F1 value = 88.8%) [25].

Graph embedding techniques: To complement deep semantic embeddings, knowledge-graph embeddings (KGEs) have been applied to patent-network analysis [26–28]. Node2Vec excels at capturing local structures, such as citation chains [29], whereas GraphSAGE is better suited for global community discovery. Dynamic network-embedding methods jointly model temporal and structural information, improving the accuracy of patent-community evolution analysis. Heterogeneous graph-attention networks (HGATs) map multi-dimensional patent networks into a unified embedding space for the first time, but their high computational cost limits large-scale application [30]. Wang et al. combined patent KGEs with Transformer-based semantic models to enhance citation-recommendation performance [31].

2.3. Time-Aware Topic Models and Dynamic Evolution Analysis

Time-aware topic mining aims to track topic changes over time, but traditional methods have limitations. The Dynamic Topic Model (DTM) is widely used, yet Müller-Hansen et al. reported that DTM responds to technological shifts with an average lag of two years [32]. One strategy is to introduce time-decay mechanisms. Gao et al. proposed a continuous time-decay function to capture the temporal proximity of text-publication timestamps for cross-network user identification [33]. By incorporating an exponential-decay distance metric, these methods reduce the influence of older data, enabling faster adaptation to new topic changes [34].

Unlike Dynamic Topic Models (DTM), which model temporal evolution through generative priors across discrete time slices, and embedding-based frameworks such as BERTopic that typically apply temporal analysis after static clustering (e.g., post-hoc temporal slicing), most existing approaches treat time either as a generative dependency or as an external analytical dimension.

Another strategy optimizes time-window selection. Li et al. compared fixed windows, sliding windows, and event-driven windows, finding that event-driven windows achieve the highest recall in shift detection [35]. Ryu & Lee applied a time-series topic model for early warning in patent-hotspot detection, more accurately capturing the timing of technological shifts [36]. Early work by Chen et al. laid the groundwork for subsequent time-aware topic-mining approaches [37]. Holder C. introduced a partition-clustering algorithm with elastic-distance measures in K-Means to assess temporal impacts on clustering [38]. Miyamoto, Nozomu et al. modeled topic dependencies via self-attention mechanisms to capture branching and merging processes, outperforming prior dynamic embedded-topic models in perplexity and Coherence [39].

3. Methodology

We propose an end-to-end multisource fusion framework (MFT-PTM) for patent topic evolution analysis, as illustrated in Figure 1. The framework comprises four sequential modules: Data Source Preparation, Data Embedding and Dimensionality Reduction, Temporal K-Means Clustering, and Topic Representation and Evolution Analysis.

1. Data Source Preparation: We gather three modalities of raw patent inputs: textual content (titles/abstracts/claims), normalized temporal information (application dates), and structured KG networks (IPC, Citations, DWPI families).
2. Data Embedding and Dimensionality Reduction: We utilize the domain-specific allenai/scibert_scivocab_uncased model to generate 768-dimensional embeddings of patent textual content, thereby capturing specialized terminology and long-tail technical phrases more effectively. We then apply UMAP to reduce these embeddings to 15 dimensions, striking a balance between structural fidelity and computational efficiency.

3. Temporal KMeans Clustering: To ensure controllable and reproducible topic discovery, we fuse an exponential timedecay factor into reduced embeddings and perform KMeans clustering with an explicit topic number K, determined via elbow and silhouette analysis, enhancing the reproducibility and comparability of the clustering results.
4. Topic Representation and Evolution Analysis: We use topics_over_time (nrbins = 20, globaltuning = False) to obtain $p_k^{(w)}$, build a transition matrix T, and compute entropy $H^{(w)}$, then link centroids to trace topic evolution paths.

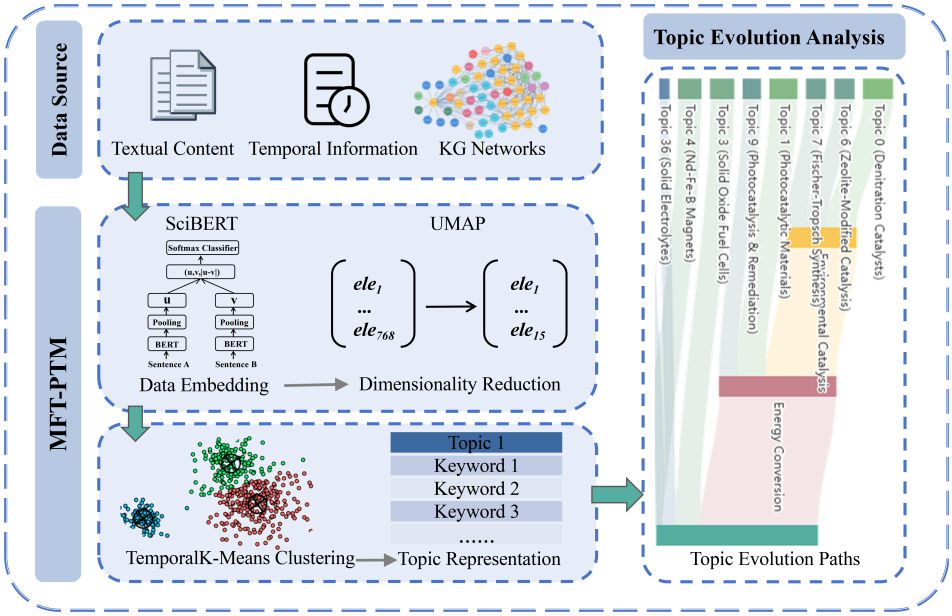


Figure 1. Model framework diagram.

3.1. Data Collection and Preprocessing

We base our study on 43,322 global patents involving the application, extraction, processing, or material integration of rare-earth elements (covering all 17 rare-earth elements, with the period 2004–2024), retrieved from the Derwent World Patents Index (https://www.derwentinnovation.com) [40]. Patent retrieval was performed using a keyword-based Boolean query targeting rare-earth elements and their common variants (e.g., "rare earth", lanthanide, neodymium, cerium, praseodymium, samarium, and yttrium), combined with date filtering to ensure coverage of the study period. The full search query is provided in Appendix A. The search was restricted to application dates between 1 January 2004, and 31 December 2024. From each patent node in the knowledge graph, we extract: 1. Semantic text fields: titles, abstracts, and claims. 2. Network features: IPC codes, citation links, and DWPI patent-family relations.

For text semantics, we encode the cleaned patent text (stopwords, section numbers, and punctuation removed) using the fine-tuned SciBERT model (allenai/scibert_scivocab_uncased) from our enhanced BERTopic pipeline, producing 768-dimensional embeddings.

Using the retrieved patent metadata, we constructed a heterogeneous patent knowledge graph integrating three distinct relation types: citation relations, DWPI family relations, and IPC-based technological similarity relations. Specifically, directed edges (Patent_A, cites, Patent_B) capture patent citations; undirected edges (Patent_A, same_family, Patent_B) connect patents within the same DWPI family; and undirected edges (Patent_A, shares_IPC, Patent_B) link patents sharing an IPC subclass code (e.g., C22B). These (head, relation, tail) triples form a multi-relational graph.

For each of the three network subgraphs (IPC, Citations, DWPI families), we compare multiple KGE methods—TransE, TransR, ComplEx, and DistMult—on a held-out validation set using MRR and Hits@K metrics. Given the heterogeneous schema, we adopt the TransR model to learn entity embeddings: TransR projects entities into relation-specific spaces, allowing each edge type to be modeled independently, yielding a 128-dimensional vector for each patent node $\boldsymbol{e}^j$, where $j \in \{\text{IPC, Cite, DWPI}\}$.

Simultaneously, we linearly normalize each patent's application year $t \in [2004,2024]$ to a scalar $\varphi(t_i) \in [0,1]$. This module outputs four embedding vectors: $\boldsymbol{e}_i^{(T)}$, $\boldsymbol{e}_i^{(\text{IPC})}$, $\boldsymbol{e}_i^{(\text{Cite})}$, $\boldsymbol{e}_i^{(\text{DWPI})}$, plus the normalized time scalar $\varphi(t_i)$, forming the basis for subsequent fusion and clustering.

3.2. Multisource Feature Fusion

To enhance the integration of textual, network, and temporal information for clustering, we implement a feature-level fusion strategy that combines these modalities into a unified representation. Each feature component is scaled by a tunable weight to balance its contribution, followed by horizontal concatenation into a single vector: a_T, a_{IPC}, a_{Cite}, a_{DWPI}, a_t to balance modality contributions. Next, we horizontally concatenate the weighted features into a single vector:

$$f_i = \left[\alpha_T e_i^{(T)}, \quad \alpha_{IPC} e_i^{(IPC)}, \quad \alpha_{Cite} e_i^{(Cite)}, \quad \alpha_{DWPI} e_i^{(DWPI)}, \quad \alpha_t t_i\right] \in \mathbb{R}^{768+3\times128+1} \tag{1}$$

Optimal weights are determined through grid search on a validation set, maximizing topic coherence (C_v) and silhouette score. This yields $\alpha_T = 0.6$, $\alpha_{IPC} = 0.3$, $\alpha_{Cite} = 0.3$, $\alpha_{DWPI} = 0.3$, and $\alpha_t = 0.1$. Before weighted concatenation, each feature embedding is L2-normalized to prevent scale differences (e.g., text embeddings often having larger norms) from causing any single modality to dominate the fusion. The resulting high-dimensional vectors f_i are reduced to 15 dimensions using UMAP (cosine metric, n_neighbors = 30), preserving local and global structures while minimizing computational complexity.

Textual content, comprising titles, abstracts, and claims, is encoded into 768-dimensional embeddings using SciBERT, which effectively captures domain-specific semantics involving rare-earth element patent texts. Knowledge graph (KG) networks—encompassing IPC co-occurrence, citation links, and DWPI family relations—are transformed into 128-dimensional embeddings via TransR, modeling structural relationships among patents. Temporal features, derived from application years (2004–2024), are normalized to a scalar $\varphi(t_i) = t_i - 2004$ reflecting patent recency. Through weighted concatenation, these modalities are integrated to form a comprehensive feature space, with UMAP ensuring efficient representation for clustering. This fused representation, particularly the embedded temporal scalar t_i, facilitates the TemporalK-Means algorithm's ability to emphasize recent patents, enabling dynamic tracking of technological evolution.

3.3. TemporalK-Means: Time-Aware Clustering

In our implementation, rather than using the default BERTopic pipeline, we developed a custom TemporalK-Means class that overrides the standard clustering estimator—this algorithm, built on classical K-Means, is designed to introduce temporal sensitivity and improve clustering controllability by converting the time feature into an exponential-decay weight embedded within the distance metric, enabling the active injection of our time-decay distance array directly into the model's fit process and making cluster assignments more responsive to the temporal evolution of patents. The design is as follows:

3.3.1. Fundamentals of the K-Means Algorithm

The classical K-Means algorithm is based on Lloyd's method, iteratively performing the "assignment–update" steps to minimize the sum of squared distances between samples and their cluster centroids:

$$\min_{\{C_k\}} \sum_{k=1}^{K} \sum_{i \in C_k} \|u_i - u_k\|^2, \quad u_k = \frac{1}{|C_k|} \sum_{i \in C_k} u_i \tag{2}$$

where u_i is the UMAP-reduced fused vector, the number of clusters K is jointly determined by the elbow method and the silhouette score. Compared to density-based clustering methods, K-Means provides deterministic outputs, explicit control over the number of clusters, and a computational complexity of O(NK), making it particularly suitable for processing datasets on the order of millions of patents.

3.3.2. Innovation in Time-Decay Mechanism

Although text and graph embeddings capture semantic and structural aspects of patents, they neglect temporal evolution signals. To address this, we transform the normalized time feature $t_i \in [0, 1]$ into a time weight via an exponential-decay function:

$$t_i' = \exp[-\lambda(1 - t_i)], \quad \lambda > 0 \tag{3}$$

where $t_i = 1$ denotes the most recent patent and $t_i = 0$ the earliest. We replace the original time component in the fused vector with $t_i^{'}$. Hence: When $t_i \approx 1$ (newest patents), $t_i^{'} = 1$. When $t_i \approx 0$ (earliest patents), $t_i^{'}$ decays to $e^{-\lambda}$.

Grid search on the validation set indicates $\lambda = 0.10$ achieves the best trade-off between semantic coherence and temporal sensitivity. Operating over normalized time values, this temporal decay parameter serves as a regularization mechanism rather than a dominant clustering force, and sensitivity analysis further confirms it balances temporal responsiveness and semantic stability—avoiding excessive recency bias while preserving chronological smoothness.

3.3.3. Updated Fusion Vector and Clustering Workflow

We substitute $t_i^{'}$ for the original time scalar in each fused vector to obtain the updated representation:

$$f_i' = \left[\alpha_T e_i^{(T)}, \quad \alpha_{IPC} e_i^{(IPC)}, \quad \alpha_{Cite} e_i^{(Cite)}, \quad \alpha_{DWPI} e_i^{(DWPI)}, \quad \alpha_t t_i'\right] \tag{4}$$

In each Lloyd iteration, every fused vector $f_i^{'}$ is assigned to its nearest centroid based on Euclidean distance:

$$z_i = \text{argmin}_k \parallel f_i' - \bar{k} \parallel^2 \tag{5}$$

Centroids are then recomputed as the mean of their assigned points:

$$\bar{k} = \frac{1}{|C_k|} \sum_{i \in C_k} f_i' \tag{6}$$

This weighting mechanism elevates the influence of recent patents during clustering, effectively enhancing the detection of emerging technology themes. Advantages of the Time-Decay Mechanism:

Nonlinear decay: Compared to linear scaling, the exponential form assigns higher weight to recent patents. For example, with $\lambda = 0.1$, a 2024 patent ($t_i = 1$) has weight 1.0, whereas a 2004 patent ($t_i = 0$) decays to $e^{-0.1} \approx 0.905$, reflecting technology-lifecycle

decline. Distance sensitivity: Temporal differences are nonlinearly amplified in the distance metric, encouraging recent patents to form tighter clusters. For instance, two patents from 2010 differ by 0.05 in normalized time, which after decay shrinks to $e^{-0.1\cdot(1-0.05)} - e^{-0.1} \approx 0.005$; whereas a 2020 vs. 2024 difference from 0.20 reduces only slightly to 0.02. Parameter interpretability: A grid search identified the optimal $\lambda = 0.1$, balancing semantic coherence with temporal proximity.

3.3.4. Time-Window Segmentation and Evolutionary Metrics

We leverage BERTopic's topics_over_time interface with nr_bins = 20, dividing 2004–2024 into 20 equal annual segments, and turn off smoothing (global_tuning = False) to obtain precise topic distributions per time window. For each window w, we compute the relative frequency of topic k:

$$p_k^{(w)} = \frac{1}{|W_w|} \sum_{i \in W_w} \mathbb{I}(z_i = k) \tag{7}$$

Build an inter-window transition matrix T,

$$T_{k\to l}^{(w)} = \frac{\sum_{i \in W_w} \mathbb{I}\left(z_i^{(w)} = k\right) \mathbb{I}\left(z_i^{(w+1)} = l\right)}{\sum_{i \in W_w} \mathbb{I}\left(z_i^{(w)} = k\right)}, \tag{8}$$

To quantify technological diversity and innovation intensity, we calculate the Shannon entropy for each window:

$$H^{(w)} = -\sum_k p_k^{(w)} \ln p_k^{(w)} \tag{9}$$

Higher entropy indicates greater topic diversity and serves as an indicator for identifying key transition points, which informs subsequent qualitative analysis.

4. Experiments and Results

This section presents experimental findings across four interconnected dimensions—intrinsic evaluation (assessing the quality of topics generated, using metrics like coherence and silhouette score), clustering quality (optimizing model parameters such as the number of clusters K and time-decay coefficient λ), topic distributions (revealing the landscape of identified topics and their sizes), and evolutionary analysis (tracking how topics change over time to identify technological trends)—while comparing our proposed multi-source fusion approach against BERTopic and LDA baselines.

4.1. Intrinsic Evaluation

For each method, we computed two intrinsic metrics—topic Coherence and topic diversity. All baseline models (LDA and BERTopic) were tuned via a grid-search strategy over their number-of-topics parameter (K ranged from 20 to 100, matching our model's search space) and other relevant hyperparameters to optimize coherence. We then report the best-achieved metrics. Experimental results indicate that the standard BERTopic model on the rare-earth patent corpus achieves an average coherence score of approximately 0.42 and a silhouette score of approximately 0.32; the corresponding values for LDA are 0.30 and 0.25, respectively. Our proposed multi-source fusion with the TemporalK-Means method jointly models textual, graph, and temporal features, raising the Coherence to 0.58 and the silhouette score to 0.50. As illustrated in Figure 2, a bar chart visually compares these three methods across the two metrics. Table 1 presents the exact scores and relative improvement percentages, confirming our approach's substantial gains in topic cohesion and differentiation.

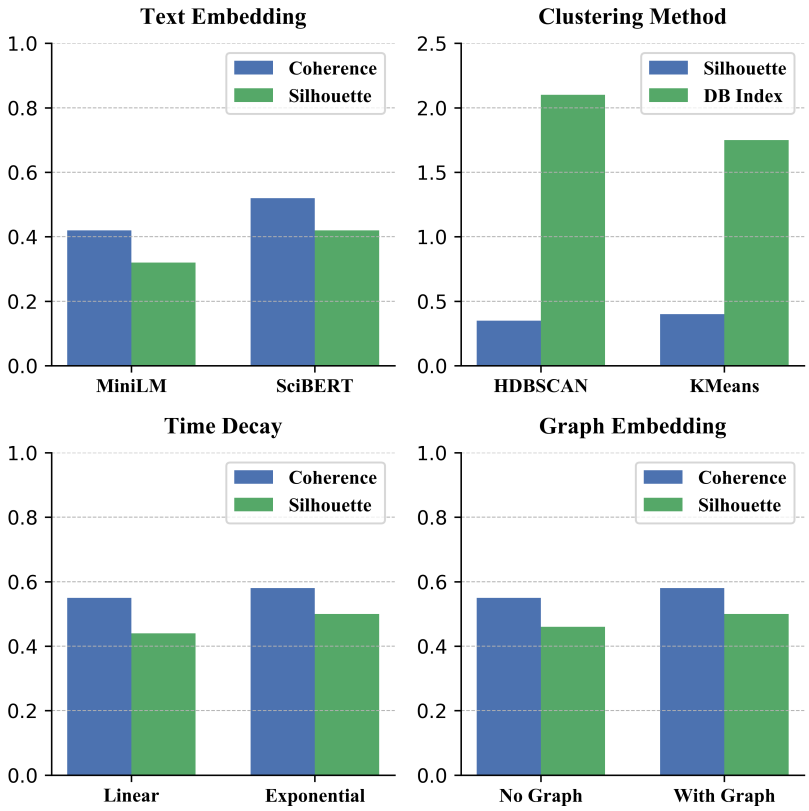


Figure 2. Presents module-level efficacy experiments.

Table 1. Comparison of Result Indicators for Technical Topic Identification Using Different Methods.

Matchs	*C_v*	Silhouette
LDA	0.30	0.25
BERTopic (MiniLM+HDBSCAN)	0.42	0.32
SciBERT+HDBSCAN	0.48	0.35
SciBERT+K-Means	0.50	0.40
SciBERT+TransR+K-Means	0.52	0.42
SciBERT+TransR+TemporalK-Means	0.58	0.50

4.2. Baseline Comparison Experiments

To elucidate the factors contributing to the efficacy of our proposed method, we conducted a comprehensive set of experiments evaluating its core components against established benchmarks. Since the core objective of this study is to evaluate multisource feature fusion and temporally aware clustering, we adopt LDA and vanilla BERTopic as controlled and methodologically aligned baselines. These experiments systematically address key challenges in patent topic modeling, as discussed earlier in this paper. Specifically, we investigate the handling of domain-specific terminology (Section 4.2.1), the assurance of reproducibility and control over the number of clusters (Section 4.2.2), the integration of temporal dynamics to capture technological evolution (Section 4.2.3), and the incorporation of structural insights derived from citation networks (Section 4.2.4). Furthermore, an ablation study (Section 4.2.5) demonstrates the synergistic impact of these components, underscoring their collective contribution to the enhanced performance of our method.

4.2.1. Text-Embedding Enhancement

Because patent-domain terminology is rich and long-tailed, general-purpose pretrained models struggle to capture technical phrases. In isolation, we compare domain-

adaptive SciBERT against generic MiniLM, measuring improvements in topic coherence (C_v) and cluster compactness (Silhouette). As shown in Table 2.

Table 2. Comparison of text embedding models on topic coherence and silhouette scores.

Embedding Model	C_v	Silhouette
MiniLM	0.42	0.32
SciBERT	0.48	0.35

Coherence increases from 0.42 → 0.48 (+0.06) and Silhouette from 0.32 → 0.35 (+0.03), indicating SciBERT better aggregates patent terminology and long-tail phrases into coherent topics and yields tighter, more separated clusters in the reduced space.

4.2.2. Clustering-Method Improvement

The original BERTopic uses HDBSCAN, whose cluster count is hard to control and whose results are less reproducible.

Replacing HDBSCAN with controllable K-Means (fixed K determined via the elbow method) markedly improves cluster quality. Silhouette rises 0.35 → 0.40 (+0.05), and Davies–Bouldin falls 2.10 → 1.75 (–0.35). Thus, K-Means's explicit K control preserves reproducibility while enhancing cluster separation and cohesion.

4.2.3. Time-Decay Mechanism Evaluation

Patents evolve over time; weighting recent patents more heavily can better reveal their potential. We isolate linear time normalization vs. exponential-decay encoding ($\lambda = 0.10$) to assess their impact on intrinsic metrics.

Coherence improves 0.55 → 0.58 (+0.03), and Silhouette 0.44 → 0.50 (+0.06). The decay mechanism boosts clustering of recent patents without sacrificing overall semantic Coherence, confirming its necessity and effectiveness for capturing evolving technology hotspots. As shown by our ablation results (Tables 3 and 4), In contrast, our proposed TemporalK-Means directly integrates a normalized time-decay factor into the clustering optimization process. This embedded time-weighted clustering significantly improves topic coherence and cluster quality over static baselines.

Table 3. Performance comparison of different clustering algorithms.

Clustering Algorithm	Silhouette	Davies-Bouldin
SciBERT+HDBSCAN	0.35	2.10
SciBERT+K-Means (K = 50)	0.40	1.75

Table 4. Impact of time-decay mechanisms on clustering performance.

Time Encoding	C_v	Silhouette
Linear normalization (t)	0.55	0.44
Exponential decay ($\lambda = 0.10$)	0.58	0.50

4.2.4. Graph-Embedding Impact

Although textual and temporal features capture patents' semantic and evolutionary dimensions, citation links and patent-family networks embed rich structural relationships that pure text alone cannot reveal. To assess the effectiveness of network-structure information, as shown in Table 5. we augment our enhanced model (SciBERT + Time Decay + TemporalK-Means) with knowledge-graph embeddings and evaluate their indepen-

dent contribution to topic-model intrinsic quality (Coherence) and clustering compactness (Silhouette).

Table 5. Ablationstudy on the effect of graph embeddings.

Model Configuration	*C_v*	Silhouette
(no graph embedding)	0.55	0.46
(with graph embedding)	0.58	0.50

Coherence rises 0.55 → 0.58 (+0.03) and Silhouette 0.46 → 0.50 (+0.04), showing TransR embeddings effectively capture text semantics and produce tighter, better-separated clusters. This isolated evaluation confirms the critical value of graph embeddings within the multisource fusion framework.

4.2.5. Ablation Study

From these results, we observe that Figure 2 presents module-level efficacy experiments. From these results, we observe that: Text-Embedding Enhancement: SciBERT improves topic coherence and cluster compactness; Clustering-Method Improvement: K-Means enables explicit control of K and enhances both separation and cohesion; Time-Decay Mechanism: Exponential decay balances weights between recent and older patents, further boosting model coherence and sensitivity; Graph-Embedding Integration: Knowledge-graph embeddings fill semantic blind spots left by text alone, enhancing both topic quality and cluster clarity.

The independent validation of each module and the synergistic gains exhibited by the full model (SciBERT + Graph Embedding + ExponentialTime + UMAP + TemporalK-Means) in intrinsic evaluations fully demonstrate the innovative value and practical effectiveness of our end-to-end multisource-fusion, time-aware framework.

4.3. Clustering Quality

We investigated the impact of the number of clusters K and the time-decay coefficient λ on clustering metrics (Silhouette Score and Davies–Bouldin Index). Elbow Method (Figure 3a): The total within-cluster sum of squares exhibits a clear "elbow" at $K = 50$. Silhouette Analysis (Figure 3b): The Silhouette Score peaks at 0.44 when $K = 50$. Time-Decay Coefficient Tuning (Figure 3c): With K fixed at 50, we evaluated $\lambda \in \{0.05, 0.10, 0.20, 0.50\}$. The optimal $\lambda = 0.10$ yields the highest Silhouette Score (0.47) and the lowest Davies–Bouldin Index (1.39), indicating that this setting best balances intra-cluster compactness with effective temporal integration.

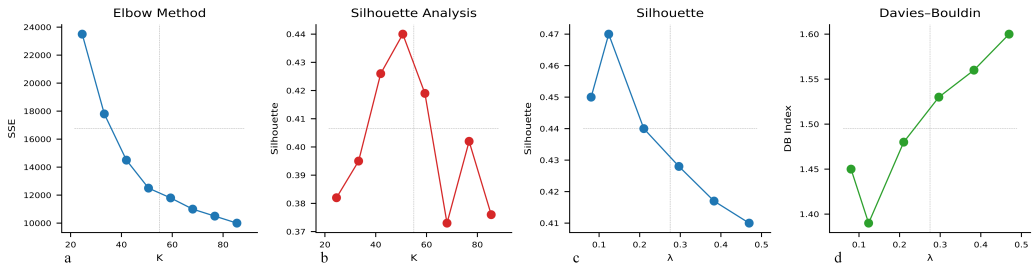


Figure 3. Clustering quality experiment diagram (**a**–**d**).

Figure 3 shows that the within-cluster sum of squares (SSE) continuously decreases as K increases from 20 to 90, but a clear "elbow" appears at $K = 50$. Beyond this point, further reductions in SSE diminish, and the Silhouette Score rises with K peaks at 0.44 when $K = 50$ before declining or fluctuating at larger K. Thus, $K = 50$ optimally balances topic

granularity and cluster compactness: it ensures high intra-cluster similarity without over-fragmentation. With K fixed at 50, we evaluated $\lambda \in \{0.05, 0.10, 0.20, 0.50\}$. The Silhouette Score reaches its maximum (0.47) at $\lambda = 0.10$, dropping at lower and higher values. Similarly, the Davies–Bouldin Index attains its minimum (1.39) at $\lambda = 0.10$, indicating the best trade-off between intra-cluster tightness and inter-cluster separation.

A decay rate of $\lambda = 0.10$ not only weights newer and older patents appropriately to reflect temporal evolution but also maximizes clustering compactness and distinguishability.

4.4. Topic Distribution

Using K = 50 clusters, we tabulated each topic's document count, label name, and top keywords. From 43,322 rare-earth patents (2004–2024), we derived 50 topics.As shown in Table 6)

Table 6. Summary of top involving rare-earth elements patent topics and document counts.

Topic ID	Count	Name	English Category
0	1922	0_denitration_cerium zirconium …	Flue Gas Denitration Catalysis
1	1797	1_roasting_photocatalyst	Sintered Photocatalytic
2	1579	2_denitration_cerium zirconium…	Oxide-Based Denitration Catalysts
3	1548	3_solid fuel_solid fuel cell…	Solid Oxide Fuel Cells
4	1541	8_neodymium boron_iron boron…	Neodymium–Iron–Boron Permanent
…	…	…	…
49	10	49_thick film_tellurium_thick…	Thick-Film Tellurium-Based Materials

The five most significant clusters (Topics 0–4) focus on flue-gas denitrification catalysis (1922 documents), sintered photocatalytic materials (1797 documents), oxide denitrification catalysis (1579 documents), solid-oxide fuel cells (1548 documents), and NdFeB permanent-magnet materials (1514 documents), together accounting for approximately 19.3% of the dataset. See Figure 4.

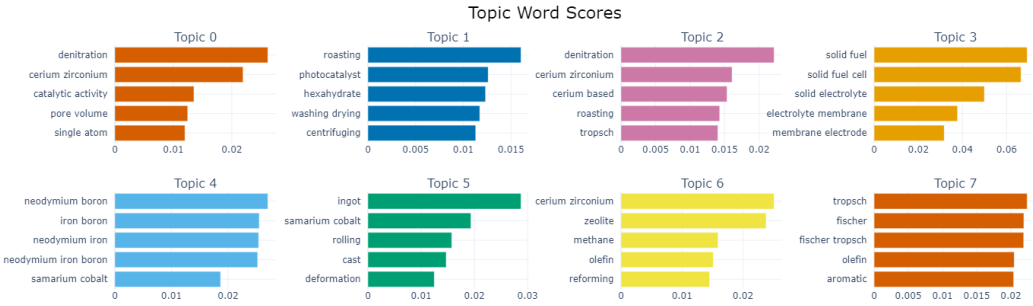


Figure 4. Presents the characteristic terms and weights for the top eight topics by associated document count.

A similarity matrix reveals these topics exhibit high pairwise similarity in the embedding space (average ≈ 0.85), indicating strong technological overlap and methodological cross-fertilization in catalysis and energy-materials research, core industrial hotspots, see Figure 5. Medium-sized clusters (Topics 5–20; 850–1437 patents) cover applications such as Sm–Co alloys, Fischer–Tropsch synthesis, and membrane separation.

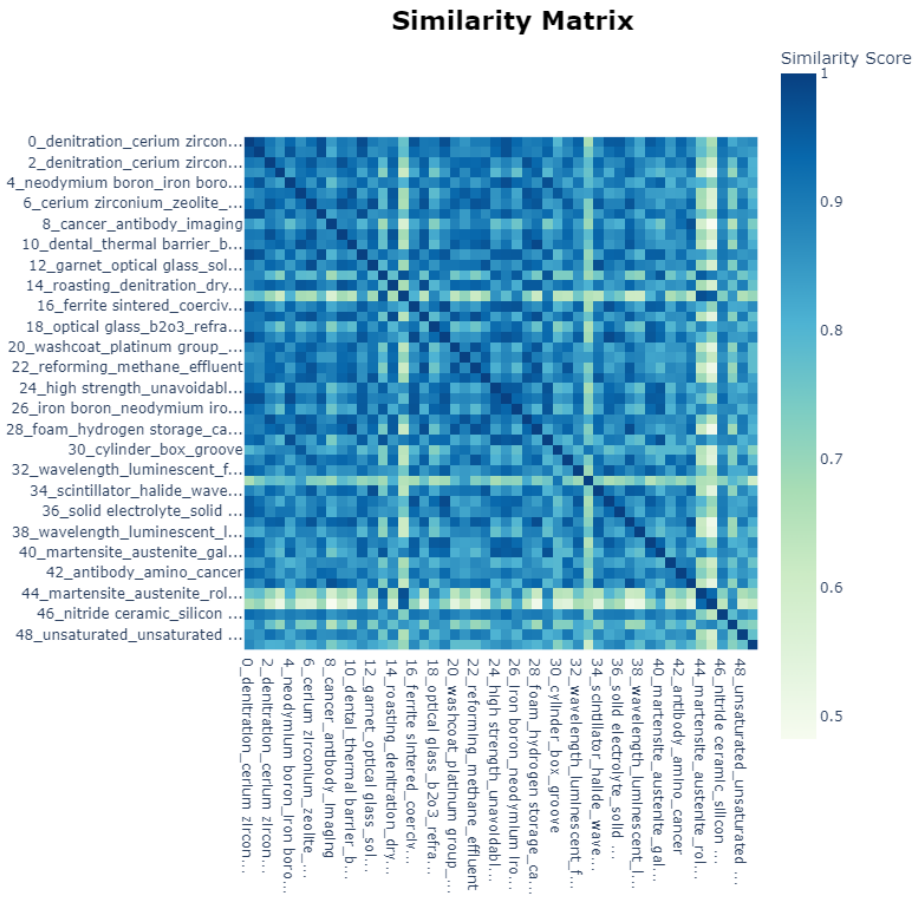


Figure 5. Cosine-Similarity Heatmap.

Smaller, long-tail topics (Topics 21–49; 10–902 patents) represent emerging frontiers like high-entropy alloys, nanoparticle drug carriers, and thin-film materials. This "few hotspots + diverse long tail" pattern validates that rare-earth technology develops through concentrated breakthroughs and diverse innovation. The heat-map color intensity reflects keyword representativeness, visually confirming the consistency and interpretability of topic labels and keywords.

These topics include both prominent "functional hubs" (e.g., Topic 0 focusing on environmental catalysis) and smaller "long-tail" innovation areas (e.g., high-entropy alloys and thick-film tellurium) with fewer patents. The substantial increase in Silhouette score reflects that our model effectively separates these diverse themes into coherent clusters. Thus, the quantitative improvements align with high topic interpretability and thematic diversity observed in our qualitative analysis.

4.5. Evolutionary Analysis

Using an equal-width 20-year window (nr_bins = 20), we traced each topic's occurrence frequency $p_k^{(w)}$ over time.

Hot topics, such as Topic 0 "denitration catalysis," exhibited slow beginnings from 2004 to 2008, accelerated growth from 2009 to 2015, and stabilization from 2016 to 2024; "photocatalysis" (Topics 1 and 9) experienced explosive growth after 2010, reflecting industry drivers of environmental demand and breakthroughs in new materials; "permanent-magnet materials" (Topics 4, 21, and 26) climbed steadily from 2012 to 2018 before converging, indicating that this field has entered a maturity and optimization phase. Long-tail topics, such as Topic 28 "hydrogen storage foam," have only intermittently emerged in the past five years, representing potential technological breakout points, see Figure 6.

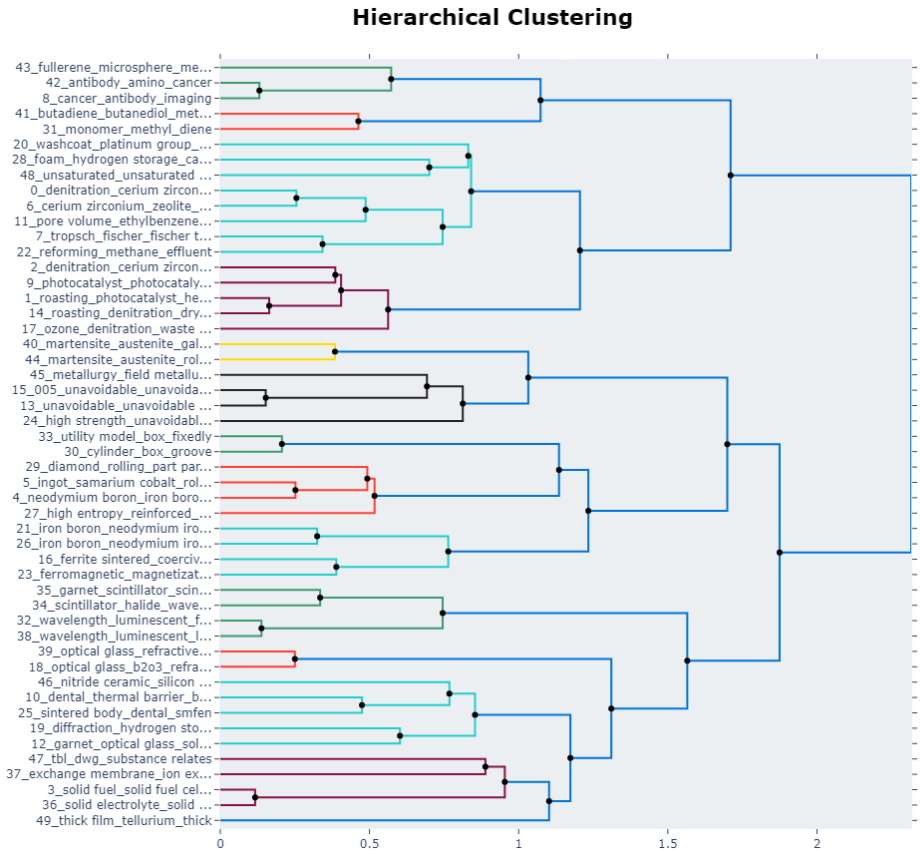


Figure 6. Hierarchical Clustering Dendrogram.

Using the end-to-end, multi-source-fusion MFT-PTM model, we identified 50 topic clusters that capture the technological evolution of global rare-earth patents from 2004 to 2024. The "Topics over Time" analysis shows that early research (2004–2008) was dominated by rare-earth–based denitration catalysis (Topic 0), indicating a relatively concentrated thematic structure. This concentration is reflected in the low Shannon entropy value of 1.2 in 2004, which increased to 2.4 by 2010 during the incubation stage.

Between 2009 and 2015, research expanded into Fischer–Tropsch synthesis and zeolite-modified catalysis (Topics 6 and 7), marking rapid diversification. Correspondingly, entropy rose sharply and peaked at 3.8 during 2010–2018, confirming the explosive multi-domain growth observed in the topic trajectories. After 2018, entropy slightly declined and stabilized around 3.2, suggesting increasing structural maturity and integration rather than further fragmentation.

Beginning in 2010, studies on photocatalytic materials (Topics 1 and 9)—notably rare-earth-doped TiO_2 catalysts—exploded, driven by emerging environmental standards and renewable-energy demands. From 2012 to 2018, research on high-performance permanent-magnet materials—exemplified by the Nd–Fe–B system (Topics 4, 21, 26)—intensified and gradually expanded into magneto-optical functional materials (Topic 35) and high-entropy alloy ceramics (Topic 27), demonstrating trends in functional integration and microstructure optimization. After 2015, solid-oxide fuel cells and proton-exchange membranes (Topics 3, 36, and 37) rapidly rose, revealing a clear trajectory from rare-earth catalysis chemistry to the electrochemical-storage and hydrogen-energy value chains. Notably, after 2020, the large-scale emergence of luminescent and display materials (Topics 32, 38) and antibody imaging and tumor targeting (Topics 8, 42) signaled deep coupling between rare-earth technologies, biomedicine, and optoelectronic integration. The dendrogram and similarity matrix jointly confirm these thematic linkages in semantic space: from denitration catalysis

to photocatalysis, then to membrane separation and fuel cells, illustrating systematic evolution in rare-earth functional materials research; while relatively isolated long-tail topics—such as hydrogen-storage foam (Topic 28) and thick-film tungsten-based coatings (Topic 49)—represent potential future innovation breakthroughs. Overall, see Figure 7. The rare-earth patent landscape is progressing along a gradient of "environmental catalysis → energy conversion → electrochemical storage → multifunctional integration." As multi-field material coupling, hydrogen energy, and biomedical markets accelerate, we anticipate the emergence of more multifunctional rare-earth materials integrating magnetic, optical, electrochemical, and biosensing properties, providing critical support for green manufacturing and intelligent health applications.

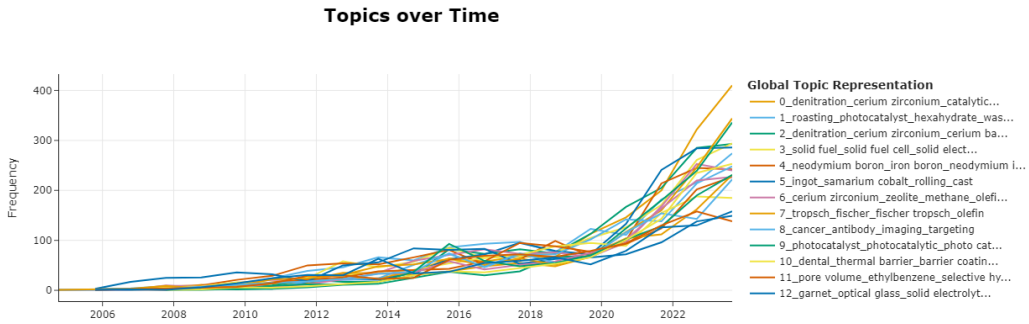


Figure 7. Trends of themes changing over the years.

5. Conclusions

Traditional topic models often overlook the rich network structures and temporal dynamics inherent in patent data, resulting in delayed or incomplete detection of emerging technological trends. To address this, we present MFT-PTM, an end-to-end multisource-fused framework that integrates textual embeddings from SciBERT, network embeddings from TransR, and normalized temporal scalars. After feature fusion and UMAP reduction, an exponential time-decay TemporalK-Means algorithm uncovers coherent, high-quality clusters. Experiments on 43,322 involving rare-earth element patents demonstrate that MFT-PTM achieves substantially higher topic coherence and cluster separation than LDA and vanilla BERTopic, successfully tracking the significant evolution of rare-earth technologies from environmental catalysis to advanced energy and biomedical applications.

Through 20 equal-width annual windows (2004–2024) and BERTopic's topics_over_time, we compute Shannon entropy per window to delineate three key phases: 2004–2010 (entropy increase indicating growing diversity), 2010–2018 (entropy peak during explosive growth in photocatalysis and permanent magnet materials), and 2018–2024 (entropy decline marking maturation and a shift toward hydrogen energy and biomedical domains). These entropy-based observations provide quantitative indicators of changes in technological diversity, which may help inform patent analysis and strategic planning. Future work will explore event-driven temporal windows and end-to-end learning of fusion weights to enable adaptive, real-time topic monitoring across evolving patent landscapes.

Finally, this study has several limitations. The dataset relies on the English-translated patent texts provided by the Derwent Innovation database. Although this enables the use of the SciBERT model, some subtle technical nuances from the original languages may be lost during translation. In addition, the proposed multisource representation combines textual, knowledge-graph, and temporal embeddings from heterogeneous spaces. While normalization and weighted fusion improve clustering performance empirically, they do not guarantee a fully unified geometric representation. Future work may explore more principled alignment strategies, such as shared latent-space learning or graph neural network–based joint embedding.

Author Contributions: Methodology, H.Z.; validation, H.Z. and L.L.; formal analysis, J.W. and L.L.; investigation, J.Y.; resources, J.Y.; writing—original draft preparation, H.Z.; supervision, J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Basic Scientific Research Operating Expenses Projects of Universities Directly Affiliated to the Inner Mongolia Autonomous Region, grant numbers 2024XKJX028 and 2024RCTD003, and the National Natural Science Foundation of China, grant number 62466045.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request.

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

Appendix A. Patent Retrieval Query

To ensure transparency and reproducibility, the patent retrieval strategy used in this study is provided in this Appendix A. All patent data were collected from the Derwent Innovation database. The dataset covers patent applications related to rare-earth technologies filed between 2004 and 2024.

The search strategy combines keyword-based queries and classification-based filtering to capture patents involving rare-earth elements and their technological applications.

Appendix A.1. Keyword-Based Search

```
CTB=(("rare earth" OR lanthanide* OR lanthanum* OR cerium*
OR praseodymium* OR neodymium* OR promethium* OR samarium*
OR europium* OR gadolinium* OR terbium* OR dysprosium*
OR holmium* OR erbium* OR thulium* OR ytterbium*
OR lutetium* OR yttrium* OR scandium*))
AND AD >= (20040101) AND AD <= (20241231)
```

Appendix A.2. Classification-Based Search

```
IC=(A61K010330 OR A61K0006822 OR A61K0006842 OR A61K0033244
OR A61K0051* OR A61K0033*)
OR IC=(B01J002310 OR B01J002383 OR B01J0027125 OR B01J0023*)
OR IC=(C08F000412 OR C08F000444 OR C08G006384)
OR IC=(C01F000100 OR C01F001700)
OR IC=(C03C0003068 OR C03C0003095 OR C03C000315)
OR IC=(C04B003540 OR C04B003550)
OR IC=(C09K001177)
OR IC=(H01J0001148 OR H01J0019068)
OR IC=(C22C002306 OR C22C0038* OR C22C0033*
OR B22F0001* OR B22F0003* OR B22F0009* OR C22C1)
OR IC=(C30B002924 OR C30B002926 OR C30B002928)
OR IC=(H01F0001053 OR H01F0041* OR H01F0001*)
OR IC=(H01M0008126 OR H01M0008127 OR H01M0008*)
OR IC=(F23Q000204 OR F23Q000212 OR F23Q000214
OR F23Q000220 OR F23Q000222 OR F23Q0002*)
```

The combination of keyword-based retrieval and IPC classification filtering ensures comprehensive coverage of patents related to rare-earth elements and their technological applications.

References

1. Greinacher, H. History of Rare Earth Applications. In *Rare Earths: Science, Technology, Production and Use*; ACS Symposium Series; American Chemical Society: Washington, DC, USA, 1981; pp. 1–22. Available online: https://pubs.acs.org/doi/pdf/10.1021/bk-1981-0164.ch001 (accessed on 23 February 2026).
2. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
3. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* **2022**, arXiv:2203.05794.
4. Watanabe, I.; Takagi, S. Technological trajectory analysis of patent citation networks: examining the technological evolution of computer graphic processing systems. *Rev. Socionetw. Strateg.* **2021**, *15*, 1–25. [CrossRef]
5. Kabore, F.P.; Park, W.G. Can patent family size and composition signal patent value? *Appl. Econ.* **2019**, *51*, 6476–6496. [CrossRef]
6. Rodriguez-Prieto, O.; Araujo, L.; Martinez-Romo, J. Discovering related scientific literature beyond semantic similarity: A new co-citation approach. *Scientometrics* **2019**, *120*, 105–127. [CrossRef]
7. WIPO. *Reseña Anual del Tratado de Cooperación en Materia de Patentes 2025—Resumen: El Sistema Internacional de Patentes*; Reseña Anual del PCT; World Intellectual Property Organization: Geneva, Switzerland, 2025. Available online: http://tind.wipo.int/record/58378 (accessed on 15 January 2025).
8. Altuntas, F.; Gök, M.Ş. Technological evolution of wind energy with social network analysis. *Kybernetes* **2021**, *50*, 1180–1211. [CrossRef]
9. Song, Y.; Lei, L.; Wu, L.; Chen, S. Studying domain structure: A comparative analysis of bibliographic coupling analysis and co-citation analysis considering all authors. *Online Inf. Rev.* **2023**, *47*, 123–137. [CrossRef]
10. Vayansky, I.; Kumar, S.A. A review of topic modeling methods. *Inf. Syst.* **2020**, *94*, 101582. [CrossRef]
11. Thomas, F.; Mahdi, L.; Lemaire, J.; Santos, D.M. Technological advances and market developments of solid-state batteries: A review. *Materials* **2024**, *17*, 239. [CrossRef]
12. Zhang, H.; Chen, Y. Technology feature fusion prediction method based on multidimensional information extraction and deep neural networks. *Knowl. Inf. Syst.* **2025**, *67*, 9051–9084. [CrossRef]
13. Yan, Z.; Xiang, B.; Shi, J. A Hybrid Embedding Method for Identifying Technology Evolution Paths of Patents: The Case on Battery Electric Vehicle Industry. *IEEE Trans. Eng. Manag.* **2026**, *73*, 52–64. [CrossRef]
14. Liu, Z.; Xiang, X.; Feng, J. Tracing evolutionary trajectory of charging technologies in electric vehicles: Patent citation network analysis. *Environ. Dev. Sustain.* **2024**, *26*, 12789–12813. [CrossRef]
15. Chen, B.; Zhang, S. Technology Topic Extraction and Analysis Based on BERTopic ChatGPT4o. In *Proceedings of the 2025 8th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), Nanjing, China, 9–11 May 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 83–86.
16. Chen, X.; Güttel, S. Fast and explainable clustering based on sorting. *Pattern Recognit.* **2024**, *150*, 110298. [CrossRef]
17. Beltagy, I.; Lo, K.; Cohan, A. SciBERT: A pretrained language model for scientific text. *arXiv* **2019**, arXiv:1903.10676. [CrossRef]
18. Li, G.; Siddharth, L.; Luo, J. Embedding knowledge graph of patent metadata to measure knowledge proximity. *J. Assoc. Inf. Sci. Technol.* **2023**, *74*, 476–490. [CrossRef]
19. Siddharth, L.; Li, G.; Luo, J. Enhancing patent retrieval using text and knowledge graph embeddings: a technical note. *J. Eng. Des.* **2022**, *33*, 670–683. [CrossRef]
20. Choi, J.; Yoon, J. Measuring knowledge exploration distance at the patent level: Application of network embedding and citation analysis. *J. Inf.* **2022**, *16*, 101286. [CrossRef]
21. Wang, Z.; Liu, Y. SEA-PS: Semantic embedding with attention to measuring patent similarity by leveraging various text fields. *J. Inf. Sci.* **2024**, *50*, 831–850. [CrossRef]
22. Endres, M.; Chikkamath, R.; Parmar, V.R.; Otiefy, Y. Patent classification using BERT-for-patents on USPTO. In Proceedings of the MLNLP '22: Proceedings of the 2022 5th International Conference on Machine Learning and Natural Language Processing, Sanya, China, 23–25 December 2022.
23. Christofidellis, D.; Lehmann, M.M.; Luksch, T.; Stenta, M.; Manica, M. Automated patent classification for crop protection via domain adaptation. *Appl. AI Lett.* **2023**, *4*, e80. [CrossRef]
24. Zhang, J.; Shi, J.; Zhang, P. An approach for identifying complementary patents based on deep learning. *J. Inf.* **2024**, *18*, 101561. [CrossRef]
25. Wang, X.; Liu, G.; Zhu, B.; He, J.; Zheng, H.; Zhang, H. Pre-trained Language Models and Few-shot Learning for Medical Entity Extraction. In *Proceedings of the 2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA), Xi'an, China, 28–30 March 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 1243–1247.
26. Xiao, L.; Yin, C. An Integrating Semantic and Structural Features Method for Knowledge Graph Embedding. In *Proceedings of the 2023 8th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), Okinawa, Japan, 23–25 November 2023*; IEEE: Piscataway, NJ, USA, 2023; Volume 8, pp. 160–165.
27. Li, Y.; Zhang, X.; Wang, F.; Zhang, B.; Huang, F. Fusing visual and textual content for knowledge graph embedding via dual-track model. *Appl. Soft Comput.* **2022**, *128*, 109524. [CrossRef]

28. Zamini, M.; Reza, H.; Rabiei, M. A review of knowledge graph completion. *Information* **2022**, *13*, 396. [CrossRef]
29. Wang, Y.; Richard, R.; McDonald, D. Competitive analysis with graph embedding on patent networks. In *Proceedings of the 2020 IEEE 22nd Conference on Business Informatics (CBI), Antwerp, Belgium, 22–24 June 2020*; IEEE: Piscataway, NJ, USA, 2020; Volume 1, pp. 10–19.
30. Vo, T. An integrated topic modelling and graph neural network for improving cross-lingual text classification. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2022**, *22*, 1–18. [CrossRef]
31. Wang, F.; Shi, X.; Wang, D.; Lou, Y. P-BERT: Towards Long Sequence Modeling by Enabling Language Representation with Prefix Sequence Compression, Soft Position Embedding, and Data Augmentation for Patent Relevance Assessment. *IEEE Access* **2025**, *13*, 41928–41942. [CrossRef]
32. Müller-Hansen, F.; Callaghan, M.W.; Lee, Y.T.; Leipprand, A.; Flachsland, C.; Minx, J.C. Who cares about coal? Analyzing 70 years of German parliamentary debates on coal with dynamic topic modeling. *Energy Res. Soc. Sci.* **2021**, *72*, 101869. [CrossRef]
33. Gao, H.; Wang, Y.; Shao, J.; Shen, H.; Cheng, X. User identity linkage across social networks with the enhancement of knowledge graph and time decay function. *Entropy* **2022**, *24*, 1603. [CrossRef]
34. Gao, W.; Zhu, C.; Si, B.; Zhou, L.; Zhou, K. Precision-dependent modulation of social attention. *NeuroImage* **2025**, *310*, 121166. [CrossRef]
35. Li, W.; Saigo, H.; Tong, B.; Suzuki, E. Topic modeling for sequential documents based on hybrid inter-document topic dependency. *J. Intell. Inf. Syst.* **2021**, *56*, 435–458. [CrossRef]
36. Ryu, S.; Lee, S. Development of a technology tree using patent information. *Adv. Eng. Inform.* **2024**, *59*, 102277. [CrossRef]
37. Chen, H.; Wu, Y.; Tang, H.; Lei, J.; Wang, G.; Zhao, W.; Liao, J.; Wang, F.; Wang, Z. Visual Analysis Method for Traffic Trajectory with Dynamic Topic Movement Patterns Based on the Improved Markov Decision Process. *Electronics* **2024**, *13*, 467. [CrossRef]
38. Holder, C.; Middlehurst, M.; Bagnall, A. A review and evaluation of elastic distance functions for time series clustering. *Knowl. Inf. Syst.* **2024**, *66*, 765–809. [CrossRef]
39. Miyamoto, N.; Isonuma, M.; Takase, S.; Mori, J.; Sakata, I. Dynamic structured neural topic model with self-attention mechanism. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2023, Toronto, ON, Canada, 9–14 July 2023; pp. 5916–5930.
40. Clarivate Analytics. Derwent World Patents Index (DWPI). Online Database 2024. Available online: https://www.derwentinnovation.com (accessed on 15 February 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.