南京理工大学
NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

Technology Convergence Prediction From a Timeliness Perspective: An Improved Contribution Index in a Dynamic Network

Jinzhu Zhang, Bing Yan

Department of Information Management, School of Economics and Management, Nanjing University of Science and Technology, Nanjing China

The 5th Extraction and Evaluation ofKnowledge Entities from Scientific Documents (EEKE2024)

2024/04/23

1 Research Perspective

- Current methods for technological convergence using patent data include approaches based on **patent co-classification**, **patent cross-referencing**, and **text mining methods**. In these methods, technology categories are typically identified by patent classification numbers and technical topics.
- Now, scholars have expanded the research on technology convergence to a broader perspective, such as the construction of **market characteristics**, **social impacts** and **time characteristics**, etc., to further improve the prediction index system of technology convergence.

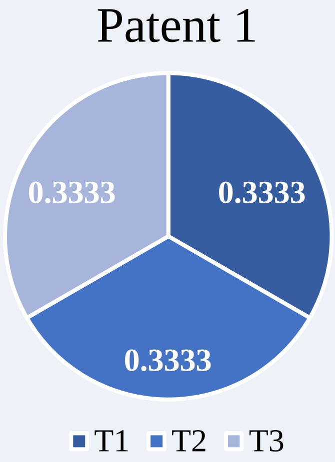
2 Research Question

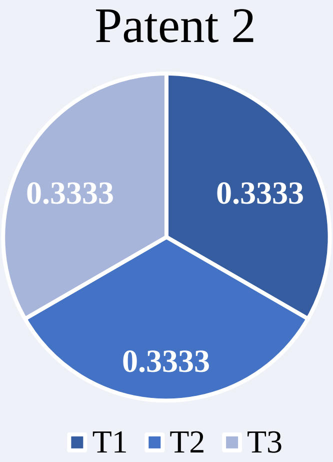
- **The differences in the importance of technologies in each convergence.**
- **Changes in timeliness.**

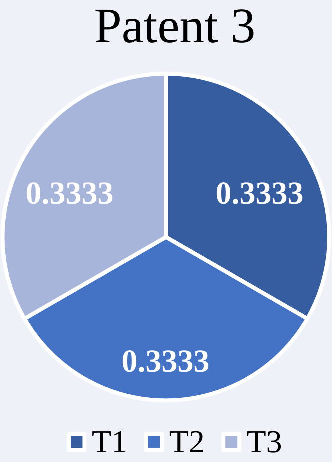
- They assume that the importance of each technology **is the same in each case of technological convergence**. In addition, the technology timeliness is often distinguished by technology lifecycle segmentation and linear weight assignment, which are too broad and difficult to capture small differences between different technologies.

2 Research Question

Figure 1: Changes in the contribution of technology within patents

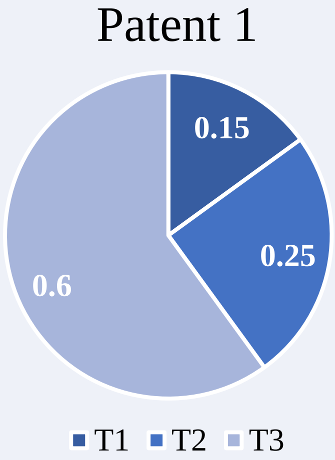


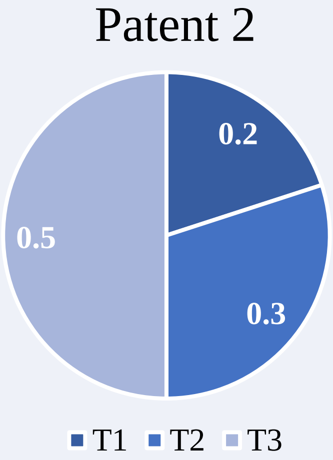


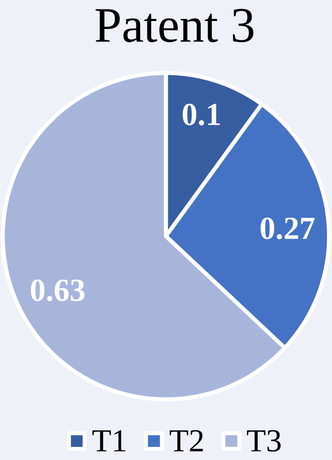


Total technical importance

T1	1.0
T2	1.0
T3	1.0







Total technical importance

T1	0.45
T2	0.82
T3	1.73

T_1 T_2 T_3

2 Research Question

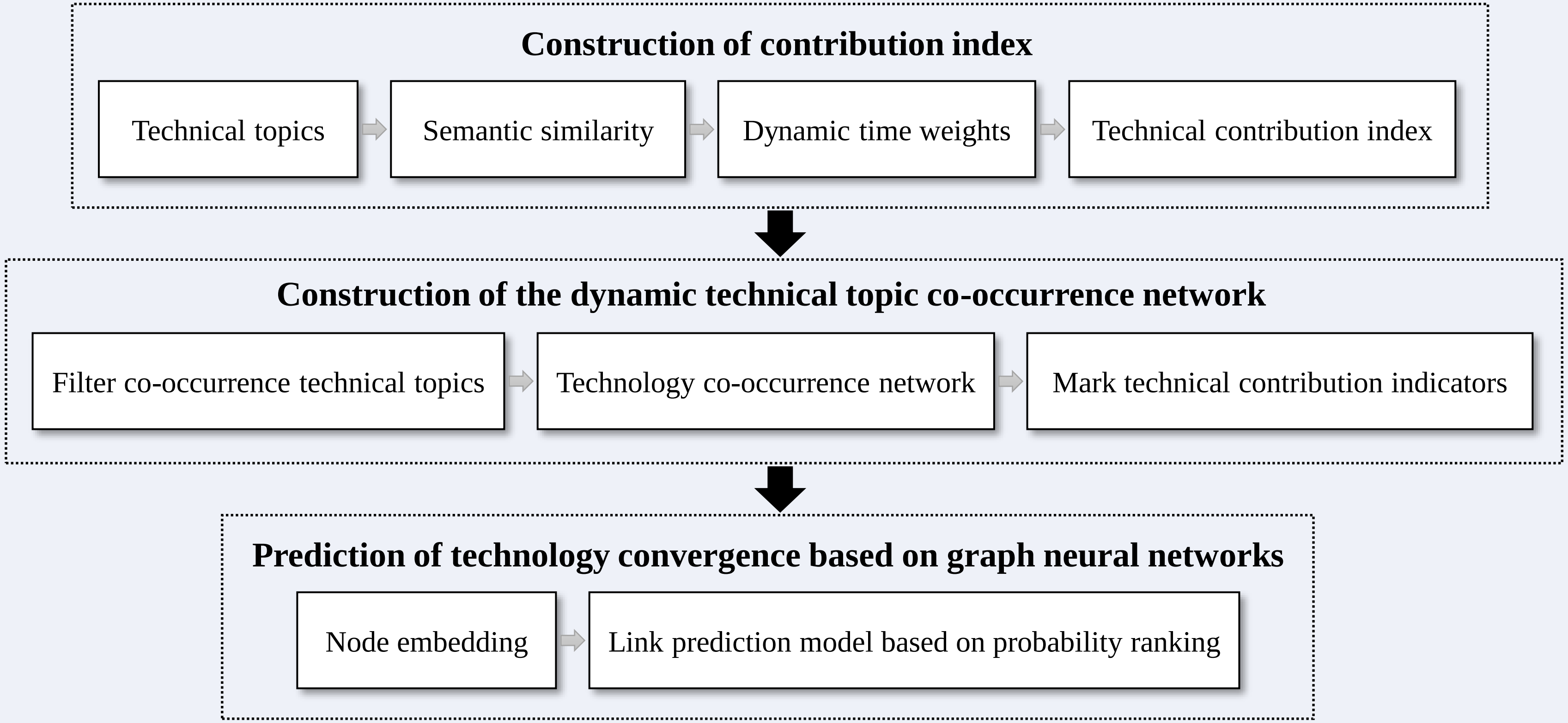
- Actually, different technologies **contribute differently** to the overall technological combination and have different timeliness with each co-occurrence.
- As a result, their impact within the technological network **differs in scope and extent**.
- For example, as shown in Figure 1(b), although technology T_1 is present in each convergence, its contribution declines over time, indicating declining importance and possibly gradual obsolescence. On the other hand, technology T_2 maintains a stable contribution, suggesting that it may be a foundational technology or in a phase of steady development. Meanwhile, technology T_3 shows a higher contribution, indicating a greater impact or greater timeliness within the technology combination, making it more likely to combine with other technologies.

3 Research Methods

- We proposes an index designed from a timeliness perspective to measure changes in the importance of technology.
- 1. The contribution of the technology in each co-occurrence by calculating the semantic similarity between the technical topic and the patent.
- 2. Combining the dynamic time weights to obtain the final value, constructing a dynamic technological co-occurrence network.
- 3. Using link prediction to explore the prediction of technology convergence, aiming to better evaluate the timeliness of technology and its impact on convergence.

3 Research Methods

Figure 2: Framework of the method



3.1 Construction of Contribution Index and Dynamic Network

- **3.1.1 Extraction of technical topics**
- **Technical topics** offer a more flexible and comprehensive expression of technical content, making them more explainable. Therefore, we choose to use technical topics to represent different technical categories.
- This paper determines the optimal number of topics based on the topic coherence score. And each technical topic has **20 representative keywords** to reduce overlap between topics.
- The **TF-IDF weighting** is applied to improve the LDA model's process of generating feature words for technical topic extraction, with the aim of improving the representativeness of the topic words.

3.1 Construction of Contribution Index and Dynamic Network

- **3.1.2 Calculation of technical contribution index**
- We use the semantic similarity between technical topics and patent texts to represent the contribution of technology in each co-occurrence. The study uses Doc2vec to obtain semantic representations of technical topics and patent texts respectively.
- Then, it applies cosine similarity to calculate the semantic similarity between them, obtaining the contribution values of different technologies in each co-occurrence, as shown in Formula (1).

$$con_{T_{i_n}} = cos(\theta) = \frac{\sum_{i=1}^{m}(P_i \times T_i)}{\sqrt{\sum_{i=1}^{m}(P_i)^2} \times \sqrt{\sum_{i=1}^{m}(T_i)^2}}\ , \qquad (1)$$

3.1 Construction of Contribution Index and Dynamic Network

南京理工大学
NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

- **3.1.2 Calculation of technical contribution index**
- It is important to consider the timeliness of technology. To address this, dynamic time weights are introduced, based on the retention function of memory capacity. This assigns weighted sums to the contribution of technical topics in each convergence, resulting in the final contribution index score for the given technical topic, as shown in Formula (2).

$$con_{T_i} = \sum_{1}^{n} con_{T_{i_n}} \times Time_{weight}\,, \tag{2}$$

$$Time_{weight} = \frac{e^{0.42}}{(t_0 + t_i)^{0.0225}}\,, \tag{3}$$

3.1 Construction of Contribution Index and Dynamic Network

- **3.1.3 Construction of the dynamic technical topic co-occurrence network**
- First step is to identify the technical topics present in the patent, we set the probability distribution threshold to **0.2**. Technical topics exceeding this threshold are considered to be present in the patent, resulting in the generation of a technology co-occurrence matrix.
- Then we extract co-occurrence relationships using the **networkx** package, forming node pairs that represent technical topics.
- Finally, we mark the obtained technical topic contributions from Section 3.1.2 on the matching nodes, establishing a dynamic co-occurrence network of technical topics.

3.2 Prediction of technology convergence based on graph neural networks

- **3.2.1 Node embedding based on graph neural network model**
- We first use the co-occurrence relationships in the training set as the graph structure. The contribution index of corresponding nodes is input as node attribute information into the graph neural network for training, thereby obtaining the embedding vectors of known technical topics.
- Secondly, using a link prediction model, we calculate the probability of fusion between technical nodes, obtaining fusion scores between nodes.

3.2 Prediction of technology convergence based on graph neural networks

南京理工大学
NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

- **3.2.2 Link prediction model based on probability ranking**
- The link prediction method proposed in this paper relies on a co-occurrence graph of technical topics, where the relationships between technical topics serve as edges. The technology contribution is trained as node features on the co-occurrence relationships of technical topics.
- This **probability score** can be regarded as the link prediction score. The higher the score, the greater the possibility of a future link between the two nodes, indicating a higher probability of convergence between these two technical topics.
- Finally, we choose **AUC** as the evaluation metric to assess the performance of the prediction model based on graph neural networks.

4 Empirical Analysis

4.1 Data collection

- In this paper, the full text data of patent applications were batch downloaded from the USPTO (United States Patent and Trademark Office) patent search platform in December 2023, parsed and stored in a PostgreSQL database. We use SQL queries to search for relevant patents in the field of new energy vehicles, as shown in Figure.
- A total of 23,792 relevant patents were retrieved and the titles, abstracts and application time of the patents were extracted as the data source for the study.

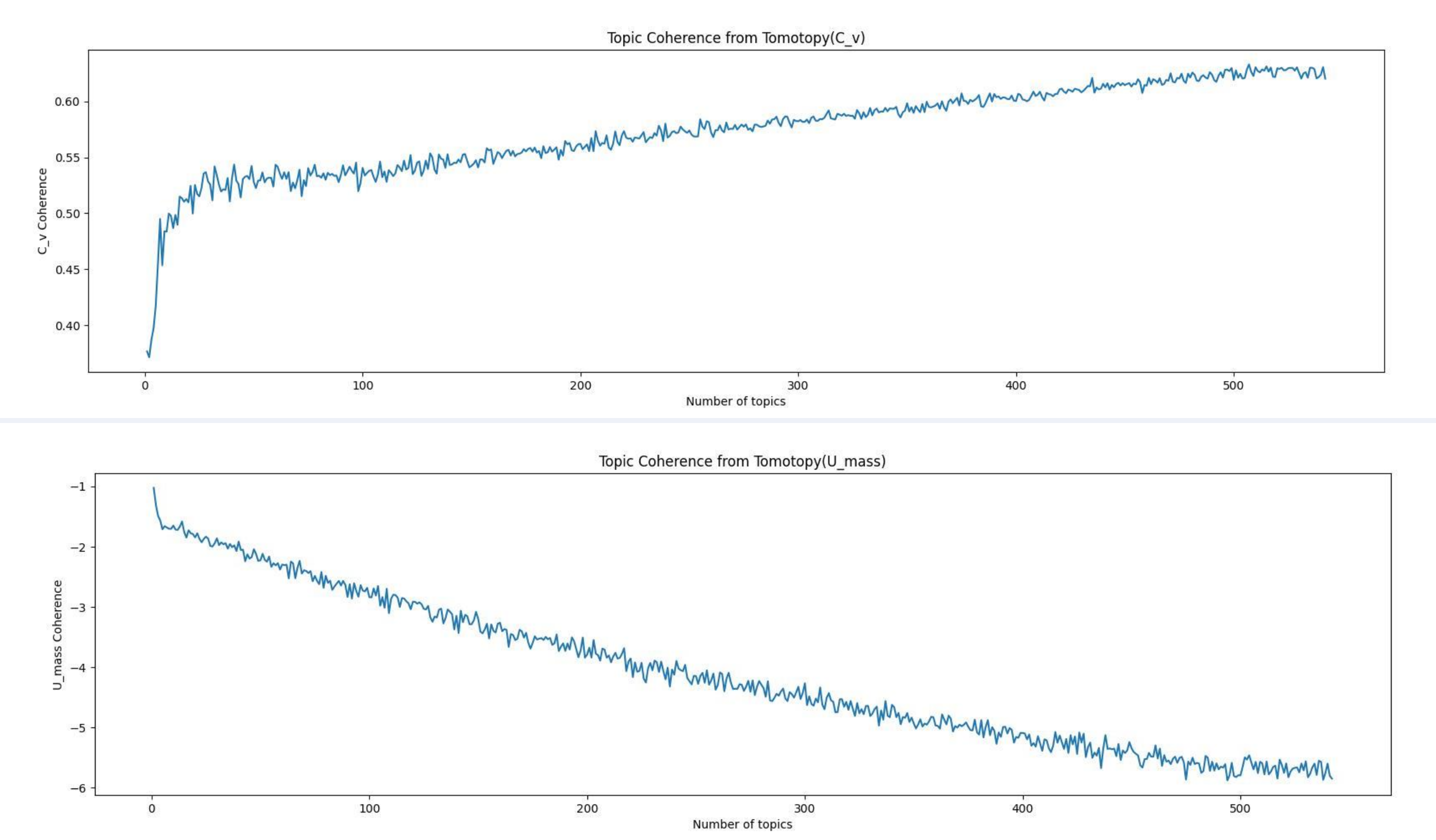
```
SELECT *
FROM uspto_patents
WHERE
    (publication_date >= '2012-01-01' AND publication_date <= '2023-12-31')
    AND (
        LOWER(publication_title) LIKE '%new energy vehicles%'
        OR LOWER(publication_title) LIKE '%nevs%'
        OR LOWER(publication_title) LIKE '%new energy automobile%'
        OR LOWER(abstract) LIKE '%new energy vehicles%'
        OR LOWER(abstract) LIKE '%nevs%'
        OR LOWER(abstract) LIKE '%new energy automobile%'
        OR LOWER(section_class_subclass_groups) LIKE '%B60L%'
    );
```

Figure 3: SQL statement for querying patents related to new energy vehicles

Training data	2012-2021	16,975
Test set	2022-2023	6,817

4.2 Construction of Contribution Index and Dynamic Network

Figure 4: C_v and U_mass variation curves



- As shown in Figure 4, ***u_mass*** and ***c_v*** gradually converged when the number of topics was around 500. After comparing extreme values, <u>**507**</u> was identified as the optimal number of topics for this paper.

4.3 Construction of the dynamic technical topic co-occurrence network

- This paper generated three co-occurrence networks with different features, to compare and validate the effectiveness of the proposed method.
- The first network, **T-Co1**, only considers the frequency of co-occurrence of technical topics.
- The second network, **T-Co2**, includes centrality indices as features for technical topic nodes.
- The third network, **T-Co3**, integrates technical contribution as features for technical topic nodes. The centrality measure chosen here is degree centrality, which reflects the number of connections a node has. A higher degree centrality indicates a stronger node centrality, signifying greater importance.

4.4 Prediction of technology convergence based on graph neural networks

- For the three types of co-occurrence networks, we use three graph neural network models, namely **GCN, GNN** and **GAT**, to learn node representations, using link prediction for quantitative evaluation.
- The main difference between **GCN** and traditional **GNN** lies in the use of convolutional operators for information aggregation, while GAT uses self-attention mechanisms for node weight allocation.

4.5 Results

AUC of Different Methods

	T-Co1	T-Co2	T-Co3
GCN	0.7120	0.7106	**0.7998**
GNN	0.7106	0.7125	**0.7236**
GAT	0.6379	0.6411	**0.6731**

- The results show that the performance of T-Co3 is generally superior to T-Co1 and T-Co2 across different model representations, with GCN performing best on T-Co3. In the GCN model, the AUC value of T-Co3 has increased by **8.78%** compared to T-Co1 and **8.92%** compared to T-Co2. In the GNN and GAT models, the AUC value of T-Co3 has also increased by **1.3%** and **3.52%**, respectively.

4.5 Results

AUC of Different Methods

	T-Co1	T-Co2	T-Co3
GCN	0.7120	0.7106	**0.7998**
GNN	0.7106	0.7125	**0.7236**
GAT	0.6379	0.6411	**0.6731**

- Compared to other indicators of importance, the contribution index reflecting technological timeliness provides better, more comprehensive, and accurate clues for predicting technological convergence. And in this experiment, the GCN model performed better and showed better discriminative capabilities for different features. It is more suitable for the technology convergence prediction task in this paper.

5 Conclusion and Prospect

- This paper refines the assessment of technological importance from a timeliness perspective, shifting from traditional distinctions based on lifecycle and dates to a more precise measurement within each convergence event.
- As a next step, we aim to improve the technological timeliness index by incorporating additional temporal cues. In addition, the exploration of more efficient embedding models is expected to improve predictive performance.

南京理工大学
NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

Thanks for Watching

Jinzhu Zhang, Bing Yan

Department of Information Management, School of Economics and Management, Nanjing University of Science and Technology, Nanjing China

The 5th Extraction and Evaluation ofKnowledge Entities from Scientific Documents (EEKE2024)

2024/04/23