Knowledge and Information Systems (2022) 64:1723–1757
https://doi.org/10.1007/s10115-022-01697-2

REGULAR PAPER

A hierarchical interactive multi-channel graph neural network for technological knowledge flow forecasting

Huijie Liu[1,2] · Han Wu[1,2] · Le Zhang[3] · Runlong Yu[1,2] · Ye Liu[1,2] · Chunli Liu[4] · Minglei Li[5] · Qi Liu[1,2] · Enhong Chen[1,2]

Received: 27 January 2022 / Revised: 12 May 2022 / Accepted: 14 May 2022 /
Published online: 5 July 2022

Abstract
Technological advancement can provide new and more cost-effective solutions to challenges in critical areas. Therefore, as one of the important sources for technological progress, technological knowledge flow (TKF) forecasting, i.e., predicting the directional flows of knowledge from one technological field to another, has become a hot issue of widespread concern. However, existing researches either rely on labor-intensive empirical analysis or ignore the intrinsic characteristics inherent in TKF. To this end, we present a data-driven solution in this article, namely a hierarchical interactive multi-channel graph neural network (HIMTKF). Specifically, HIMTKF generates final predictions using two types of vector representations for each technology node (a diffusion vector and an absorption vector), which is realized by four components: high-order interaction module (HOI), co-occurrence module (CO), improved hierarchical delivery module (IHD) and technological knowledge flow tracing module (TFT). For one thing, HOI and CO are designed to represent high-order network relationships and co-occurrence relationships between technologies on the same hierarchy level. For another, IHD is aimed to model the hierarchical relationships between technologies while also taking their personalities into account. Then, TFT is intended for capturing the dynamic feature evolution of technologies with the above relations involved. Additionally, we develop a hybrid loss function and propose a new evaluation metric for better forecasting the unprecedented knowledge flows between technologies. Finally, we conduct extensive experiments on a large dataset of real-world patents. The results validate the effectiveness of our approach and shed light on several intriguing phenomena about technological knowledge flow trends.

Keywords Technological knowledge flow · Graph neural network · Patent mining · Cooperative patent classification · Link prediction

✉ Enhong Chen
cheneh@ustc.edu.cn

Extended author information available on the last page of the article

1 Introduction

In an era of fast-paced technology advancement, the importance of knowledge and technology is becoming increasingly apparent. The diffusion and transfer of technology and knowledge can bring technological advancement, which can provide new and more cost-effective solutions to critical challenges in essential sectors such as education, food and water, health, and industry, as well as opportunities for regions and firms that are less developed to close the technology gap and develop their own innovative capacity [1, 2]. Therefore, one effective way of catching up and mastering technology development trend is to capture the dynamics of technological knowledge flows, i.e., measuring the influence degree that a technological knowledge area transmits or receives and making predictions on future flows among different technological areas [3].

In general, technological knowledge flow (TKF) refers to the directional flow of knowledge from one technological field to another [4], which can be regarded as a kind of interaction between two different areas [5]. Many previous works have emphasized the importance of TKF analysis, and they are often carried out in large-scale scientific literature, among which patent mining plays an important role [6–8]. As a matter of fact, many researchers believe that patent citations can be used to reveal the linkage of technological knowledge [9, 10]. Furthermore, textual similarities between patents and their citations were employed to demonstrate that patent citations surely can reflect knowledge linkage [11].

In the field of patent, each patent often cites the issued ones, and they are all assigned to one or more technology classification codes at the beginning of the application [12], such as the well-known Cooperative Patent Classification (CPC),[1] International Patent Classification (IPC)[2] and so on. Taking the CPC system (shown in Fig. 1) as an example, we can see that each classification code (e.g., Subgroup codes) can be regarded as a kind of technology field [13, 14] and different technologies can be connected in three fundamental ways: (1) one is the hierarchical structure inherent in CPC system; (2) the second way is through technology co-occurrence, which occurs when two distinct CPC codes appear in the same patent; and (3) last but not least, the third way is through billions of patent citations. More importantly, the prior two ways condense the wisdom of patent experts, while the latter one represents the objective dynamics of knowledge flows between different patents and thus their associated technologies. Therefore, with the help of CPC system and patent citations, it is quite possible to construct technological knowledge flow networks each year and make a further study on TKF forecasting.

As far as we are concerned, numerous efforts on TKF forecasting have been made through the analysis of scientific and technological literature exemplified by patents. Existing methods can be broadly classified into two categories: One is the traditional expert experience-based methods [14], such as Delphi [15], theory of solving inventive problem (TRIZ) and indicator–experience combination methods [16] in which statistical analyses are widely used; and the other is machine learning-based methods, and the most commonly used is link prediction, which can identify potential links in a TKF network [13, 17]. However, current TKF forecasting solutions have some limitations for the following reasons. For one thing, methods based on expert experience are constrained by high labor costs and are confined to small areas; for another, existing machine learning methods are incapable of fully exploiting the characteristics of technology systems, for example, the hierarchical structure depicted in Fig. 1. In addition, nearly all of these methods focus exclusively on the final state of knowledge flows,

[1] https://www.cooperativepatentclassification.org/about.

[2] https://www.wipo.int/classifications/ipc/en/.

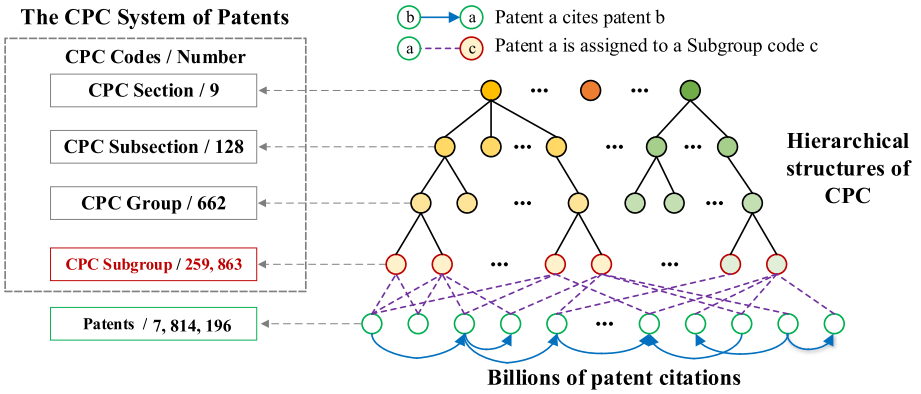


Fig. 1 An example of TKF in patent system

while ignoring their dynamic process. For these reasons, we attempt to find a solution by dynamically modeling the knowledge flow process in conjunction with characteristics of the technology system to provide some data-driven insights into future TKF.

Unfortunately, there are many technological and domain challenges inherent in designing effective solutions to the TKF forecasting problem. First, technological knowledge flow is a bidirectional process, implying that a flow must be generated from a source technology to a target technology. In the meantime, a technology field can serve as a dual identity, acting as both the source node flowing to others and the target node that other sources flow to. Therefore, how to represent a technology field's directional double-faced aspects is a worth considering problem. Second, although existing domain studies have generated numerous indicators for quantifying the degree of knowledge flow through technology nodes, it still remains a challenge to design metrics that adequately characterize their intrinsic principles, i.e., how to exploit high-order interactions between technologies, co-occurrence of technologies and hierarchical structures inherent in technologies (e.g., CPC), particularly in more complex situations involving thousands of technology nodes. Third, since TKF is a continuous and dynamic process [13] with consistent emergence of unprecedented knowledge flows, another challenge is determining how to dynamically model TKF process and capture potential flow trends.

To tackle these challenges, in our preliminary work [18], we proposed a novel framework for forecasting future TKF, named hierarchical interactive graph neural network for technological knowledge flow (HighTKF). Specifically, HighTKF is composed of three components: (1) high-order interaction module (HOI) is aimed to model the high-order relations among neighbors of the technology; (2) hierarchical delivery module (HD) is designed to better exploit the hierarchical relationships inherent in technology categories; and (3) technological knowledge flow tracing module (TFT) is introduced to model the dynamic interactions between technologies with the above relations involved. In addition, we design a hybrid loss function and propose a new evaluation metric for better predicting the unprecedented flows between technologies.

However, in our follow-up study, we discover that HighTKF is unable to exploit co-occurrence relationships between technologies and neglects the uniqueness of technologies on HD, which has a noticeable effect on its performance. As a result, in this article, we further develop an improved version of HighTKF and propose a hierarchical interactive multi-channel graph neural network (HIMTKF), in which we incorporate the co-occurrence

relationship and improve the HD of the original model HighTKF by taking into account personalized characteristics of technologies in the same category. To be more precise, in HIMTKF, we begin by exploring representations of technologies at the same level of the hierarchy. We employ high-order interaction module (HOI) and co-occurrence module (CO) to learn representations of technology nodes. Moreover, CO is a newly added module that is aimed to fully exploit co-occurrence information between technologies. Following that, we consider hierarchical relationships between technologies at various levels of the hierarchy. We replace HD with improved hierarchical delivery module (IHD), which finely considers personalized characteristics of technologies within the same category apart from taking advantage of hierarchical relationships inherent in technology categories. And after that, TFT is arranged to model the dynamic interactions of technologies. Finally, extensive experiments on real-world patent data demonstrate that the two models we proposed are capable of accurately forecasting future technological knowledge flows, and the performance of HIMTKF is also improved compared with HighTKF.

2 Related work

2.1 Technological knowledge flow forecasting

In this part, we will illustrate why patent is a good carrier for predicting technological knowledge flow, and will introduce common techniques for TKF forecasting. Generally, methods for TKF forecasting can be classified into two categories: expert experience-based methods and machine learning-based methods.

2.1.1 Patent as the footstone for studying TKF

The study of technological knowledge flow necessitates a substantial amount of reliable and accessible data. Patents are public documents that cover a broad range of subject areas and contain a wealth of detailed information, both structured and unstructured [7]. Moreover, each granted patent produces a highly structured document containing detailed information about the innovation itself, the technological areas to which it belongs and the references it cited [17]. Additionally, textual similarities between patents and their citations were used to demonstrate that patent citations most certainly can indicate knowledge linkage [11]. Since patents are widely regarded as up-to-date and reliable sources of knowledge that reflect rapidly evolving technological advancement, patent data have been extensively explored for technological knowledge flow forecasting [16, 19].

2.1.2 Expert experience-based methods

Expert experience-based technology trend forecasting is aimed to bridge the gap between observed trend patterns and technical indicators such as patents, scientific literature and R&D expenditure [20, 21]. Theory of solving inventive problem (TRIZ)-based method is one of the earliest expert experience-based methods that presents specific sequences of technological transitions or trends, which illustrate the evolution of a system or technology over time [22]. Methods for combining indicators and experience [16], such as combining bibliometric analysis and curve fitting-based approaches that heavily rely on statistical analysis, are the most widely accepted and adopted empirical technology forecasting techniques [14, 23].

Although these studies are undoubtedly helpful to draw directions of new technology development, their scope is restricted to a particular technology or industry. This type of method is primarily based on expert analysis in specific fields, which has the advantage of being easily validated but the disadvantages of being costly and time-consuming [24]. In addition, some researches [25] discover that these subjective strategies are not always precise and reliable.

2.1.3 Machine learning-based methods

Machine learning-based methods make use of a data-driven approach to discover complicated correlations between inputs and outputs. One of the most commonly used machine learning-based methods for TKF forecasting is link prediction which attempts to predict the emergence of future links in complex networks based on available information, such as the observed links and nodes' attributes [26]. At present, TKF prediction methods based on link prediction can be divided into following categories: (1) topology-based link metrics [17], including Resource Allocation index [27], Jaccard Coefficient, Adamic-Adar index [28, 29], Preferential Attachment and so on; and (2) social theory-based metrics, which employ classical social theories, such as triadic closure [30] and structural balance such as structural holes [31]. Additionally, text mining technology offers another possibility for TKF forecasting. Text mining is a data processing and information extraction technique. Since most parts of a patent document are expressed in text (natural language) format, the process of keyword extraction is applied to identify keywords and to measure similarity between patents [32]. However, with the help of patent citations and patent classification system such as CPC and IPC, we can quickly identify related technologies and keywords, as well as correlations between patents.

So far, there have been a large number of works on TKF prediction. Nevertheless, almost all of the above methods lose sight of relationships between technologies and the dynamic process of knowledge flows, thereby limiting their effectiveness.

2.2 Graph neural networks

Graph neural networks (GNNs) are a powerful machine learning technique for graphs. Given that our model is based on GNNs, we will discuss some related works below.

2.2.1 Early development of graph neural networks

The concept of GNN was first proposed by Gori et al. [33]. They put forward that in practical application, recurrent neural network can be extended to deal with all kinds of graphs, including directed graph, undirected graph and cyclic graph. In 2013, Bruna et al. proposed the spectral network on graph for the first time [34], and then, Defferrard et al. proposed the Chebyshev network [35], which reduced the complexity of the spectral network. At the same time, Kipf et al. simplified the Chebyshev network and produced the well-known graph convolutional neural network (GCN) [36]. Since then, various graph neural networks have emerged endlessly. Graph Attention Network (GAT) [37] uses the attention mechanism to define graph convolution, while GraphSage [38] extends the graph neural network from transductive learning to inductive learning.

2.2.2 Complex graph neural networks

With the rapid development of graph neural networks, numerous excellent works on modeling complex heterogeneous and dynamic graphs have been published. Nowadays, in order to integrate node features and topological structures in a complex graph with rich information, plenty of outstanding works have sprung up, including AM-GCN [39] which employs adaptive multi-channel graph convolutional networks for semi-supervised classification, NEC-DGT [40] which translates a multi-attributed input graph to a target graph. RGCN [41] is a node representation algorithm based on message-passing learning on heterogeneous graphs. The message on each edge of RGCN is obtained using the edge's unique linear transformation, which is the difference between it and GCN. Furthermore, an approach based on mixed hierarchical clustering and optimization for graph analysis in social media network [42] has been proposed for determining the hierarchical community structure in large networks.

In terms of dynamic graphs, DySAT [43] is a state-of-the-art network embedding method for link prediction on dynamic graphs that is capable of capturing both structural and temporal evolutionary patterns. Although GNNs have been around for a short period of time, there have yielded fruitful results. Nowadays, graph neural networks have been widely used in many fields such as natural language processing [44, 45], recommendation systems [46] and social network analysis [47].

In this paper, we propose a dynamic graph neural network approach based on the process of TKF, which combines not only technical indicators but also characteristics of the technology system.

3 Data description and statistics

In this section, we first describe the public patent data that we use, and then provide some supporting statistics.

3.1 Data description

Patent documents are important intellectual resources for protecting the interests of individuals, organizations and companies [7]. Since 1972, the United States Patent and Trademark Office (USPTO) has issued and granted more than 7 million patent documents, with 119 million references among them. As an important explicit knowledge carrier, patent literature has the strongest availability and measurability [48]. Furthermore, the level of redundancy in edges of patent citation network is much smaller than in that of academic paper citation network suggesting that patent citations are better representations of technological knowledge flow than other scientific publications [49, 50]. Therefore, it is quite suitable to explore TKF through patent citation analysis.

As far as we are concerned, the Cooperative Patent Classification (CPC) system [51, 52] is recognized as one of the most efficacious patent classification systems, whose structure is depicted in Fig. 2. In detail, it contains four levels, including 9 sections, 128 subsections, 662 groups and 259,863 subgroups, which are gradually refined from top to bottom and closely related to certain technology fields. For example, Section "F" stands for "mechanical engineering, lighting, heating, satellites, blastin," Subsection "F02" represents "combustion engines, hot-gas or combustion-product engine plants," Group "F02D" means "controlling combustion engines," and Subgroup "F02D2700/07" denotes "Mechanical control of speed

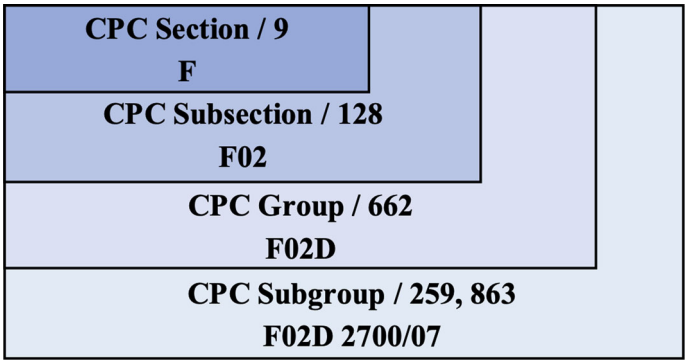


Fig. 2 Visualization of Cooperative Patent Classification (CPC)

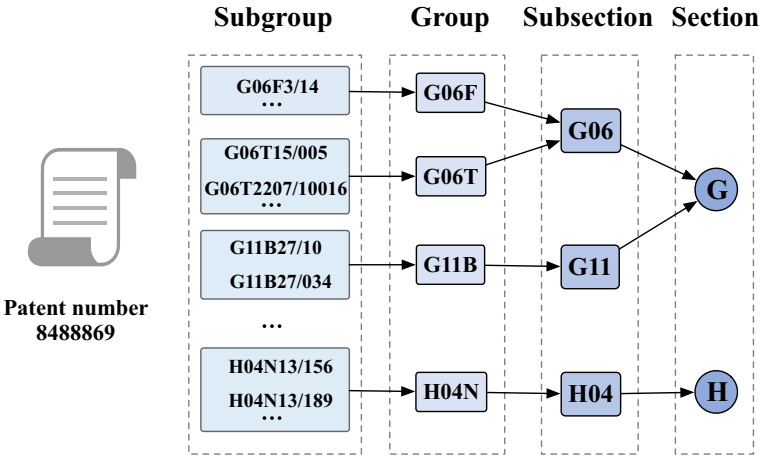


Fig. 3 Co-occurrence of technologies in Cooperative Patent Classification (CPC)

or power of a single cylinder piston engine-Automatic control systems according to one of the preceding groups in combination with control of the mechanism receiving the engine power." It is notable that in this paper we select the most detailed level (CPC Subgroup) as the resource of technology nodes, which is then exploited to construct TKF graphs with the combination of patent citations.

Additionally, as mentioned in Introduction section, each patent will be assigned to one or more classification codes at the beginning of the application which reflects the co-occurrence of technologies, as illustrated in Fig. 3, where technologies with co-occurrence relationships have been marked with gray dashed boxes. More specifically, the subgroup category codes of patent No.8496678 in CPC are "G06F3/14," "G06T15/005," "G06T2207/10016," "H04N13/156" and so on. These CPC codes appear in the same patent, which reflects the co-occurrence relationships among these technologies. Similarly, there exist co-occurrence connections for technology codes in upper levels such as Group, Subsection, and Section.

3.2 Technological knowledge flow

In order to achieve the final aim of TKF forecasting, the first step is to get the specific TKF graphs. As mentioned before, we can construct a TKF graph with the help of patent citations and CPC classification in any given time period, where technology nodes come from CPC Subgroup and flow edges are acquired by the aggregation of patent citations.

Specifically, as shown in Fig. 1, each technology node contains a certain number of patents, and there exists a large number of mutual references among patents. Therefore, we can naturally get the number of citations between any two technologies by aggregating the references of patents they contain. For instance, if patents in technology A cite patents in technology B n times, we can assume that there is a knowledge flow from technology B to technology A, with n denoting the strength of the flow. Obviously, a larger n indicates a more robust flow edge. Notably, we must constrain the scale of TKF network and eliminate the influence of noisy edges, which is why we define a N threshold as follows: Only when $n > N$, the corresponding edge can be reserved for TKF graph construction.

3.3 Statistics on technological knowledge flow graphs

In this section, we present some supportive statistics that illustrate the beneficial discoveries made about technological knowledge flow.

3.3.1 Double-faced aspects of technology nodes

Technological knowledge flow is a bidirectional process, which requires the production of a flow from a source technology to a target technology. Meanwhile, each technology node can act as both a source and a target node. To more thoroughly validate the duality of technology nodes, we conduct data analysis on the differential distribution between outflows and inflows of several technological areas [53]. The result is depicted in Fig. 4, where $\Delta = |\ \#outflows - \#inflows\ |$ denotes the absolute value of the difference between the outflows and inflows of each technology. Here we can see that Δ of more and more technologies changes across time, illustrating the diffusion and absorption abilities of technologies have been continuously changing. Therefore, it is necessary to learn two representations for the nodes' double-faced aspects by considering the diffusion and absorption features of technologies separately.

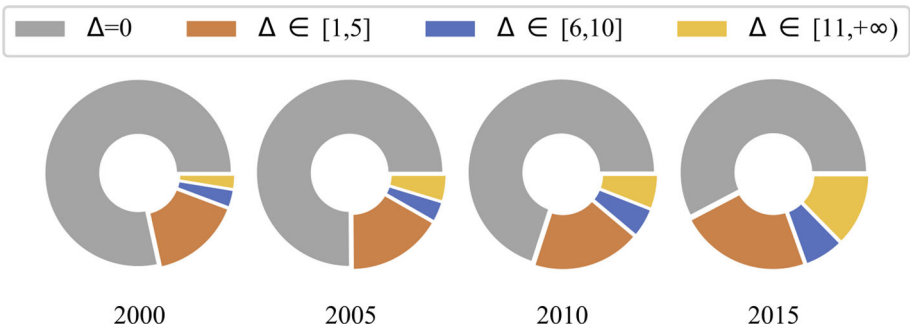


Fig. 4 Difference between inflows and outflows

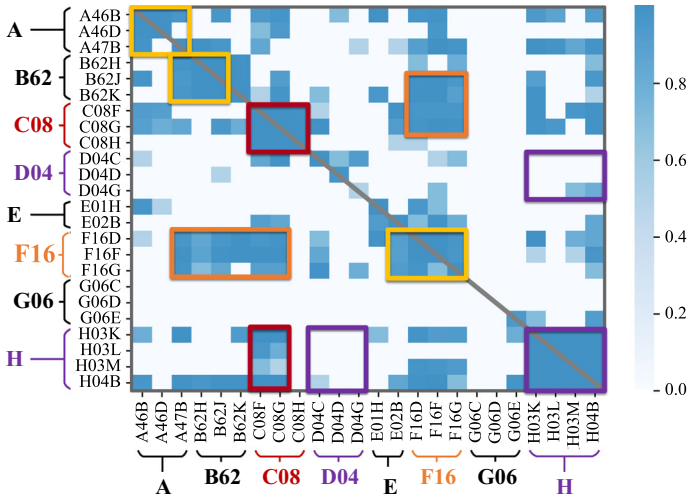


Fig. 5 Flow intensity between technologies

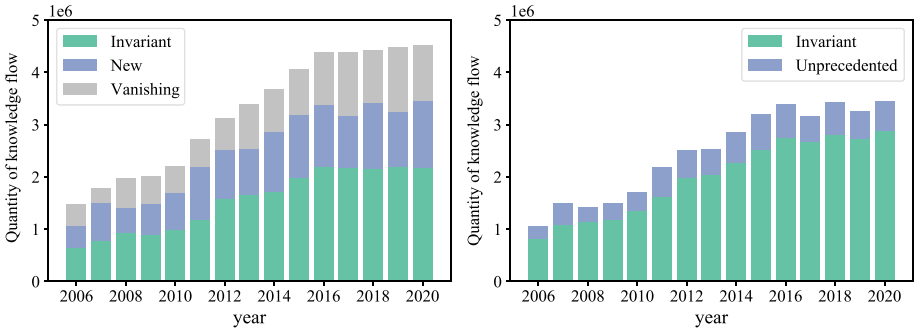
3.3.2 Technological relations under hierarchical structure

It is noteworthy that technologies can be connected not only by objective patent citations but also through hierarchical structures of CPC. Therefore, we make a flow analysis between different CPC Subgroups. The result is shown in Fig. 5, where the color of each block represents the intensity of TKF flows, and a deeper color indicates a stronger connection. Here are two remarkable observations:

- In general, there exists clear region clustering in this figure, meaning that the mutual interactions between technologies in the same category are stronger and these technologies are more likely to flow to the same node (or absorb knowledge from the same node). For example, "C08F," "C08G" and "C08H" belong to the same category "C08," and we can see a darker part of "C08" falls on the main diagonal, indicating it tends to interact more with itself. Also, "F16D," "F16F" and "F16G" belong to the same category "F16," they are more likely to spread to the same node (or absorb knowledge from the same node). "H03K," "H03L" and "D04C," "D04D" belong to different classes, and there are few flows among them.
- Microscopically, apart from the large dark areas, there are also many individual small dark areas on the map. This demonstrates that, although technologies belonging to the same category share a high degree of similarity, each node has its own characteristics. For example, in the clustering of category "A" and the intersection of category "H" and category "D04," they all have some small areas that are different from the majority of areas.

3.3.3 Dynamic changes of technological knowledge flows

Figure 6a shows the annual flow changes of TKF from 2006 to 2020. Here we can find that more than one-third of the knowledge flows are newly produced compared with last year. In contrast, almost the same number of flows appeared last year but vanished the following



(a) Annual changes for TKFs compared with the last year.

(b) Dynamic changes for TKFs compared with previous years.

Fig. 6 Dynamic changes for technological knowledge flows

year. In addition, there are nearly two thirds of flows have been available in both years. Figure 6b shows dynamic changes in TKF from 2006 to 2020. It is clear that nearly one-seventh of knowledge flows each year are completely new. In contrast, almost six-sevenths of technologies exist consistently. Therefore, we can draw a conclusion from the figures that TKF is consistently changing with time, and these dynamics are not dramatic, but relatively stable.

The preceding fascinating observations can be instructive and meaningful, allowing us to summarize the following instructions for future TKF prediction:

- Double-faced aspects of technologies. When predicting the knowledge flow between two technologies, we need to consider both diffusion characteristics (diffusion capability of the source technology) and absorption characteristics (absorption capability of the target technology).
- Technological relations under hierarchical structure. Technologies belonging to the same category have stronger similarities, but at the same time, the uniqueness of technology nodes cannot be ignored.
- Dynamic changes of TKF. Technological knowledge flows keep evolving consistently, so we need to model dynamic flows among them and pay more attention to the prediction of brand new flows.

4 Problem formulation

Formally, we assume that there are N technologies. In a specific time window t, several patents have been published. According to Sect. 3.2, we can extract technological knowledge flows from patent citations. Then, in time window t, a TKF graph can be constructed, denoted by E^t. Here, each node denotes one technology and edge E^t_{ij} denotes the edge from technology i to j that equals 1 if there exists flow from technology i to j and 0 otherwise. Next, we formulate the problem as follows:

Definition 1 (*TKF Forecasting*) Given the TKF graphs from time window 1 to T, i.e., $\{E^1, \ldots, E^T\}$, and both static and dynamic features of technologies. Our target is to predict the existence of edges in the graph E^{T+1}.

5 HIMTKF framework

In this section, we will introduce technical details of our HIMTKF framework for the TKF forecasting problem. As illustrated in Fig. 7, this framework is composed of four modules, namely high-order interaction module (HOI), co-occurrence module (CO), improved hierarchical delivery module (IHD) and technological knowledge flow tracing module (TFT), respectively. First of all, in order to learn the intra-hierarchical representation of nodes, we use HOI and CO to learn the representation of technology nodes at the subgroup level. Then, the IHD plays the function of exploiting the hierarchical structures inherent in technology categories without neglecting the personal characteristics of technologies in the same category. After that, we employ TFT to simulate the dynamic interactions of technologies.

5.1 Input layer

As mentioned before, a technology often acts in two roles, having diffusion ability as a source node and absorption ability as a target node. Therefore, we extract characteristics for these two aspects, respectively, which are listed below:

- The number of patents included in each technology in the current year.
- The number of patents owned by the technology as a percentage of all patents in the current year.
- Technology growth rate [54] refers to the ratio of granted number of a technology to the total granted number of this technology in the past years.
- Technology diffusion (absorption) amount means the number of flows diffused (absorbed) by a technology.
- Technology diffusion (absorption) ratio [16] represents the proportion of diffusion (absorption) of a technology field in the total diffusion (absorption) of all technologies.
- Technology diffusion (absorption) growth rate refers to the ratio of the number of flows diffused (absorbed) of a technology field to the total number of flows diffused (absorbed) of this technology filed in the past years.

Let $\mathbf{d}_i^t \in \mathbb{R}^D$ and $\mathbf{a}_i^t \in \mathbb{R}^D$ denote the diffusion and absorption embedding of technology U_i in time window t. We take $\mathbf{d}^t = [\mathbf{d}_1^t, \mathbf{d}_2^t, \ldots, \mathbf{d}_N^t]$ and $\mathbf{a}^t = [\mathbf{a}_1^t, \mathbf{a}_2^t, \ldots, \mathbf{a}_N^t]$ as the input for the following modules.

5.2 High-order interaction module (HOI)

In the process of TKF, the interactions of technologies are reciprocal, and the interaction between any two technologies may also have an impact on other technologies. In order to model the interaction between technologies while considering high-level influences, we build HOI.

As mentioned before, an edge in TKF graph can be formulated as the interaction of the source node's diffusion ability and the target node's absorption ability. Therefore, we follow the idea of an interactive graph neural network to update the double-faced representation of technologies [55]. To be more precise, the diffusion capacity of a technology domain is updated by aggregating its neighbors' absorption capacity, and vice versa. With regard to both efficiency and effect, we choose the simple weighted sum operation as the operator in this case. The aggregation process is illustrated as follows:

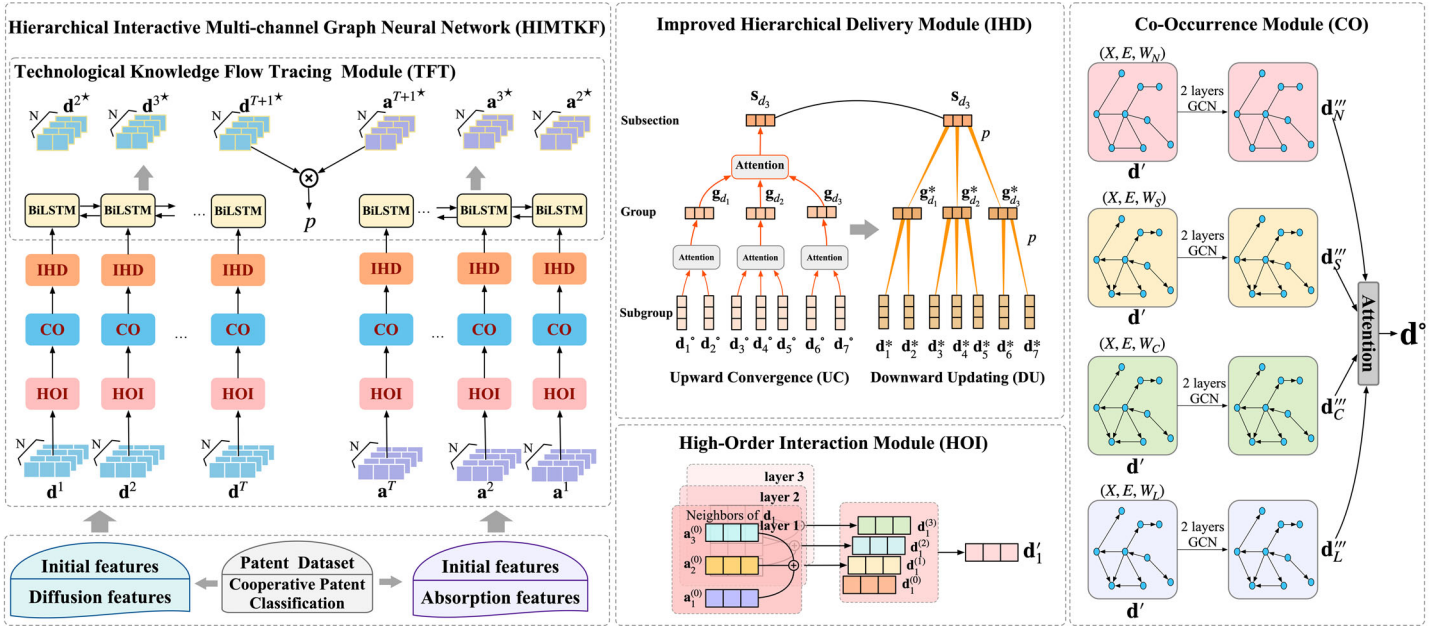


Fig. 7 Our HIMTKF framework is composed of four modules: high-order interaction module (HOI), co-occurrence module (CO), improved hierarchical delivery module (IHD) and technological knowledge flow tracing module (TFT). The input layer is represented by the lower left corner

$$\mathbf{d}_i^{(k+1)} = \sum_{j \in \mathcal{N}_d(i)} \frac{1}{\sqrt{|\mathcal{N}_d(i)|}\sqrt{|\mathcal{N}_a(j)|}} \mathbf{a}_j{}^{(k)},$$
$$\mathbf{a}_i^{(k+1)} = \sum_{j \in \mathcal{N}_a(i)} \frac{1}{\sqrt{|\mathcal{N}_a(i)|}\sqrt{|\mathcal{N}_d(j)|}} \mathbf{d}_j{}^{(k)}, \tag{1}$$

where $\mathbf{d}_i^{(k)}$ and $\mathbf{a}_i^{(k)}$ represent the diffusion and absorption embedding of technology U_i. And $\mathcal{N}_d(i)$ and $\mathcal{N}_a(i)$ denote out-neighbors and in-neighbors set of technology U_i. The superscript k denotes the k-th layer, and the larger layer can capture the higher-order influence.

After K layers of aggregation, we further combine the embeddings obtained at each layer to form refined representations of the diffusion and absorption abilities for technologies:

$$\mathbf{d}_i' = \sum_{k=0}^{K} \alpha_k \mathbf{d}_i^{(k)}, \quad \mathbf{a}_i' = \sum_{k=0}^{K} \alpha_k \mathbf{a}_i^{(k)}, \tag{2}$$

where $\alpha_k \geq 0$ denotes the importance of the k-th layer embedding for the refined embedding.

5.3 Co-occurrence module (CO)

As mentioned above, the phenomenon of co-occurrence implies that technologies appear in the same patent. That is to say, these technologies are more likely to absorb knowledge from the same technology node at that time, indicating that they have similar absorption capabilities. Furthermore, if this patent is cited in the future, then these technologies will spread their knowledge to the same technology nodes, this in turn reflects their similarities in diffusion capabilities. In addition, according to [56], three indexes, i.e., "support," "confidence" and "lift," are used to represent relationships between technology classes as follows:

$$\text{Support}(X \to Y) = P(X \cap Y) = \frac{N(X \cap Y)}{N(O)},$$
$$\text{Confidence}(X \to Y) = \frac{P(X \cap Y)}{P(X)} = P(Y \mid X),$$
$$\text{Lift}(X \to Y) = \frac{P(X \cap Y)}{P(X)P(Y)} = \frac{P(Y \mid X)}{P(Y)},$$

where $N(O)$ denotes the number of all occurrences and $N(X \cap Y)$ means the number of co-occurrences of technology X and technology Y.

Therefore, we build co-occurrence module (CO) to better exploit the co-occurrence information between technologies in order to gain a more precise understanding of their diffusion and absorption capacities. We extend the original technological knowledge flow network by adding four new types of edges to represent the co-occurrences of technologies. These edges have same nodes, but different weights and different properties. The first type of edge is undirected, and its weight is equal to the number of times two nodes occur together, denoted by $N(X \cap Y)$. And the remaining types of edges are directed, with their weights denoted by $\text{Support}(X \to Y)$, $\text{Confidence}(X \to Y)$ and $\text{Lift}(X \to Y)$, respectively. Inspired by [39], we use a multi-channel graph convolutional network to model the co-occurrence relationship between technologies.

First of all, we model the representations of diffusion and absorption features of technology nodes on these four topological graphs (X, E, W_N), (X, E, W_S), (X, E, W_C) and (X, E, W_L), where X denotes nodes, E means edges, and W denotes four different types of

weights. Take the diffusion characteristics of nodes on graph (X, E, W_S) as an example. We utilize a simple two-layer GCN [36] to extract the diffusion characteristics of nodes from topological graph (X, E, W_S) as follows:

$$\begin{aligned} \mathbf{d}_S'' &= \mathrm{ReLU}\left(\hat{\mathbf{D}_S}^{-1/2}\hat{\mathbf{A}}_S\hat{\mathbf{D}}_S^{-1/2}\mathbf{d}'\boldsymbol{\Theta}_{S1}\right), \\ \mathbf{d}_S''' &= \hat{\mathbf{D}_S}^{-1/2}\hat{\mathbf{A}}_S\hat{\mathbf{D}_S}^{-1/2}\mathbf{d}''\boldsymbol{\Theta}_{S2}, \end{aligned} \tag{3}$$

where $\hat{\mathbf{A}}_S = \mathbf{A}_S + \mathbf{I}$ denotes the adjacency matrix with inserted self-loops, $\hat{D}_{S_{ii}} = \sum_{j=0} \hat{A}_{S_{ij}}$ means its diagonal degree matrix and $\boldsymbol{\Theta}_{S1} \in \mathbb{R}^{D*D_1}$, $\boldsymbol{\Theta}_{S2} \in \mathbb{R}^{D_1*D_2}$ are the transform matrix. And the values of elements in $\mathbf{A}$ represent the edge weights Support$(X \cap Y)$. Then we get four different absorption features and diffusion features of nodes. They are $\mathbf{d}_N'''$, $\mathbf{d}_S'''$, $\mathbf{d}_C'''$, $\mathbf{d}_L'''$ and $\mathbf{a}_N'''$, $\mathbf{a}_S'''$, $\mathbf{a}_C'''$, $\mathbf{a}_L'''$, respectively.

Next, considering that the node representations formed by these various types of edges may differ in importance from the final node representations, we use the attention mechanism to assign them appropriate importance. Here we focus on node i, where its embedding in the diffusion feature $\mathbf{d}_S'''$ is $\mathbf{d}_{S_i}''' \in \mathbb{R}^{1*D_2}$. Its weight to the whole diffusion feature $\mathbf{d}_i^\circ$ can be calculated as follows:

$$w_{S_i} = (\mathbf{q}_S)^T \cdot \tanh\left(\mathbf{W}_S \cdot (\mathbf{d}_{S_i}''')^T + \mathbf{b}_S\right). \tag{4}$$

Here $\mathbf{q}_S \in \mathbb{R}^{h*1}$ is the attention vector, $\mathbf{W}_S \in \mathbb{R}^{h*D_2}$ is the transform matrix, $(\mathbf{d}_{S_i}''')^T$ is the transpose matrix of $\mathbf{d}_{S_i}'''$ and $\mathbf{b}_S \in \mathbb{R}^h$ is the bias vector. Similarly, we can get the attention weights w_{N_i}, w_{C_i} and w_{L_i} for node i in diffusion features $\mathbf{d}_N'''$, $\mathbf{d}_C'''$, $\mathbf{d}_L'''$, respectively. After that, we normalize the weight with softmax function, taking the calculation of $_{S_i}$ as an example:

$$\alpha_{S_i} = \frac{\exp\left(w_{S_i}\right)}{\exp\left(w_{N_i}\right) + \exp\left(w_{S_i}\right) + \exp\left(w_{C_i}\right) + \exp\left(w_{L_i}\right)}, \tag{5}$$

where $\alpha_{S_i} \geq 0$ denotes the importance of the corresponding embedding, i.e., larger α_{S_i} implies the corresponding embedding is more important. Then, we can get $\alpha_S = [\alpha_{S_1}, \alpha_{S_2}, ..., \alpha_{S_N}]$ as the attention weight of $\mathbf{d}_S'''$ and $\boldsymbol{\alpha}_N, \boldsymbol{\alpha}_C, \boldsymbol{\alpha}_L$ likewise.

Finally, we can combine four diffusion features to get the whole diffusion feature $\mathbf{d}^\circ$:

$$\mathbf{d}^\circ = \boldsymbol{\alpha}_N \cdot \mathbf{d}_N''' + \boldsymbol{\alpha}_S \cdot \mathbf{d}_S''' + \boldsymbol{\alpha}_C \cdot \mathbf{d}_C''' + \boldsymbol{\alpha}_L \cdot \mathbf{d}_L'''. \tag{6}$$

And under the same operation, we can obtain the whole absorption feature $\mathbf{a}^\circ$.

5.4 Improved hierarchical delivery module (IHD)

After completing the above two modules on modeling intra-hierarchical node representations, we turn our attention to node representations based on hierarchical structure.

As mentioned in Sect. 3, technologies can also be connected through the hierarchical structure of the CPC system, which reflects some intrinsic connections among technologies. Therefore, we are inspired to leverage the hierarchical relationships among technologies to generate more comprehensive representations of technologies for more accurate forecasting of TKF. Since the upper category has an inclusive relationship with the lower category, we can aggregate the information from the lower category to represent the upper category. On the other hand, the lower category can also be supplemented by the upper category. However, it is worth noting that we cannot generalize all the lower nodes without considering their uniqueness. We need to design the model according to their uniqueness and selectively aggregate

the information from the upper nodes. To this end, based on hierarchical delivery module (HD) proposed in HighTKF [18], we build improved hierarchical delivery module (IHD) to describe the mutual influence between upper and lower categories. As shown in Fig. 7, IHD contains two phases, namely upward convergence (UC) and downward updating (DU).

In this paper, each section in the CPC system can construct a dataset, as shown in Fig. 1. So we mainly focus on the hierarchical structure among three levels, which consists of subgroup, group and subsection, respectively. The Upward Convergence phase is aimed to aggregate the information from the lower level to the upper level (i.e., subgroup to group and group to subsection). Here we take the UC from subgroup to group as an example. For the diffusion representations of nodes at group level, we aggregate the diffusion representations of the corresponding child nodes at subgroup level. To distinguish the importance of different child nodes, we apply the attention mechanism [57] to automatically learn weights for child nodes. Let w_{ij} denotes the weight of subgroup node j to the group node i, which is calculated as follows:

$$w_{ij} = \frac{(\mathbf{q}_i)^T}{\mid g_i \mid} \cdot \tanh\left(\mathbf{W}_u \cdot (\mathbf{d}_j^{\circ})^T + \mathbf{b}_u\right), \tag{7}$$

where $\mathbf{q}_i \in \mathbb{R}^{h_i * 1}$ is the attention vector, $\mathbf{W}_u \in \mathbb{R}^{h_i * D_2}$ is the transform matrix, $\mathbf{b}_u \in \mathbb{R}^{h_i}$ is the bias vector, and $\mid g_i \mid$ is the number of lower-level nodes included in the upper-level technology g_i. Further, we normalize the weight with softmax function:

$$\beta_{ij} = \frac{\exp\left(w_{ij}\right)}{\sum_{k=1}^{|g_i|} \exp\left(w_{ik}\right)}. \tag{8}$$

Then, we combine the weight to get the final embeddings of diffusion features for the upper-level technology node. And considering that the upper node contains more information, we broaden its dimension using a linear function:

$$\mathbf{g}_{d_i} = \sum_{j=1}^{|g_i|} \beta_{ij} (\mathbf{W}_g \cdot (\mathbf{d}_j^{\circ})^T + \mathbf{b}_g), \tag{9}$$

where $\mathbf{W}_g \in \mathbb{R}^{D_g * D_2}$ is the transform matrix, $\mathbf{b}_u \in \mathbb{R}^{D_g}$ is the bias vector. Similarly, we can get the absorption representation for technology i in group level, denoted by $\mathbf{g}_{a_i}$. Following that, by conducting upward convergence from group to subsection, we can get the diffusion and absorption representation for technology k in the subsection level, denoted as $\mathbf{s}_{d_k}$ and $\mathbf{s}_{a_k}$.

In terms of the Downward Updating phase, it is targeted at using the upper-level information to update lower-level nodes (i.e., subsection to group and group to subgroup) while taking into account the personalized characteristics of technologies in the lower level. Take the update process from subsection to group as an example. To update the diffusion representations of nodes at the group level, we first convert their parent nodes to the same metric as child nodes using a linear transformation:

$$\mathbf{s}'_{d_k} = \mathbf{W}_s \cdot \mathbf{s}_{d_k} + \mathbf{b}_s, \tag{10}$$

where $\mathbf{W}_s \in \mathbb{R}^{D_g * D_s}$ is the transform matrix and $\mathbf{b}_s \in \mathbb{R}^{D_g}$ is the bias vector.

Then, at the subsection level, we concatenate the diffusion representations of current nodes and their parent nodes. To capture the node's uniqueness, we measure the influence of the upper node on the lower node using the ratio of the number of patents owned by the

technology node to the total number of patents owned by the parent node. Then a nonlinear transformation is adopted to obtain an enhanced diffusion representation as follows:

$$\mathbf{g}_{d_i}^{*} = \sigma\left(\mathbf{W}_d \cdot [\mathbf{g}_{d_i}; p \cdot \mathbf{s}'_{d_k}] + \mathbf{b}_d\right), \tag{11}$$

where $\mathbf{W}_d \in \mathbb{R}^{D_g * 2D_g}$ is the transform matrix, p denotes the proportion of the number of patents owned by technology i in group level to the number of patents of technology k in subsection level, $\mathbf{b}_d \in \mathbb{R}^{D_g}$ is the bias vector, σ is an activation function. Similarly, we can get the enhanced absorption representation for technology i in group level, denoted by $\mathbf{g}_{a_i}^{*}$. By conducting the downward updating process from group to subgroup, we can obtain the enhanced diffusion and absorption representation for technology j in subgroup level, denoted by $\mathbf{d}_j^{*}$ and $\mathbf{a}_j^{*}$, respectively.

Till now, we can obtain the diffusion and absorption representations of technologies in each time window. Next, we will use them to predict future node representations.

5.5 Technological knowledge flow tracing Module (TFT)

As mentioned in Fig. 6, technological knowledge flow keeps evolving consistently, so we need to take its temporal characteristics into consideration. Therefore, in this part, we employ bidirectional long short-term memory (BiLSTM) to model the dynamic feature evolution, which is a combination of LSTM [58, 59] in both directions. Formally, given the diffusion embeddings from time 1 to time T, i.e., $\mathbf{d} = [\mathbf{d}^1, \mathbf{d}^2, \ldots, \mathbf{d}^T]$. The forward and reverse LSTM outputs in time t can be shown as:

$$\begin{aligned} \overrightarrow{\mathbf{h}}^{t} &= \mathrm{LSTM}\left(\mathbf{d}^{t*}, \overrightarrow{\mathbf{h}}^{t-1}\right); \\ \overleftarrow{\mathbf{h}}^{t} &= \mathrm{LSTM}\left(\mathbf{d}^{t*}, \overleftarrow{\mathbf{h}}^{t-1}\right). \end{aligned} \tag{12}$$

Then $\overrightarrow{\mathbf{h}}^{t}$ and $\overleftarrow{\mathbf{h}}^{t}$ are concatenated together to get hidden layer output of BiLSTM in time t:

$$\mathbf{h}^{t} = \left[\overrightarrow{\mathbf{h}}^{t}; \overleftarrow{\mathbf{h}}^{t}\right]. \tag{13}$$

In this way, BiLSTM updates the feature embeddings and output the diffusion features $\mathbf{d}^{t\star} = \mathbf{h}^{t}$ in time t. Analogously, we can obtain absorption features $\mathbf{a}^{t\star}$ in time t as the same operations above. After the above modules, we can acquire the diffusion embedding $\mathbf{d}^{T+1\star} = \left[\mathbf{d}_1^{T+1\star}, \mathbf{d}_2^{T+1\star}, \cdots, \mathbf{d}_N^{T+1\star}\right]$ and absorption embedding $\mathbf{a}^{T+1\star} = \left[\mathbf{a}_1^{T+1\star}, \mathbf{a}_2^{T+1\star}, \cdots, \mathbf{a}_N^{T+1\star}\right]$ of technologies in time $T+1$.

The feature representations of the source technology and target technology must then be transformed into the probability of an edge between them. Because technological knowledge flow refers to the directional flow of knowledge from source technology to target technology, it reflects the source technology's knowledge diffusion ability and the target technology's knowledge absorption ability. As a result, the knowledge flow probability from source technology to target technology can be represented using the diffusion characteristic of source technology and the absorption characteristic of target technology. According to the common practice of using a neural network to perform a link prediction task [60], the prediction probability of technological knowledge flow in time $t+1$ can be calculated as:

$$\hat{p}_{ij} = \sigma\left((\mathbf{d}_i^{T+1^\star})^T \cdot \mathbf{a}_j^{T+1^\star}\right). \tag{14}$$

Here, $(\mathbf{d}_i^{T+1^\star})^T$ is the transpose matrix of $\mathbf{d}_i^{T+1^\star}$ and σ is the sigmoid function.

5.6 Loss function

In this subsection, we focus on the construction of a novel loss function during the training process. From Fig. 6, we observe that every year almost one-seventh of knowledge flows are unprecedented. These brand new flows reveal the new trends of technologies [61] and may bring a totally new era of technology. However, these entirely new flows are difficult to predict because they have never been connected before, and general methods struggle to accurately predict such a large number of flows. Due to the significant of the unprecedented flows, a well-designed objective function is still required to train the HIMTKF model. For the purpose of predicting new edges more accurately without sacrificing the accuracy of all edges, we define a loss function for emerging edges in each time window as follows:

$$\mathcal{L}_e = \sum_{(i,j)\in F_1} p_{ij} \cdot \log \hat{p}_{ij}, \tag{15}$$

where F_1 denotes the set of emerging edges and p_{ij} represents whether there is a knowledge flow from i to j. Besides, the loss function for global edges, which includes both those that exist each year and those that do not exist, and can be expressed as follows:

$$\mathcal{L}_g = \sum_{(i,j)\in F_2} \left[p_{ij} \cdot \log \hat{p}_{ij} + \mathbb{E}_{k_n \sim P_n(k)} \left(1 - p_{ik_n}\right) \cdot \log\left(1 - \hat{p}_{ik_n}\right)\right], \tag{16}$$

where F_2 contains all edges and σ is the sigmoid function, P_n is a negative sampling distribution. Thus, the total loss of the training process contains two parts:

$$\mathcal{L} = (1-\alpha)\mathcal{L}_e + \alpha\mathcal{L}_g, \tag{17}$$

where α controls the proportions of different objectives.

6 Experiments

6.1 Experimental setup

6.1.1 Dataset description

Using the CPC system, we conduct experiments on the USPTO dataset. Specifically, we chose two sections in CPC, namely "F" (Mechanical engineering; Lighting; Heating; Weapons; Blasting) and "H" (Electricity), as well as a set of 50,000 randomly selected technology nodes from the CPC Subgroup for our experiments. Each dataset contains all the patents and their citations belonging to the respective class granted from 1991 to 2020. Some important statistics are listed in Table 1.

Table 1 Statistics of the Datasets from 1991 to 2020

Dataset	Network information			Hierarchical classification		
	#Nodes	#Links	AND	#Groups	#Subsections	#Sections
F	27,551	4,890,188	177	103	18	1
H	38,871	46,750,633	1202	51	5	1
Random data	50,000	9,484,957	189	656	128	9

AND Average node degree

6.1.2 Benchmark methods

We compared HIMTKF and HighTKF with several state-of-the-art methods, which can be divided into two categories: static methods (Random, MF) and dynamic methods (DySAT, GCN+LSTM, GraphSage+LSTM) as follows:

(1) *Random*. We assume that the technological knowledge flow is a random process, meaning that the probability of knowledge flow between arbitrary two technologies is 0.5.
(2) *MF* [62]. This method predicts the probability of TKF by factoring TKF matrix and obtain the latent vectors of technologies.
(3) *DySAT* [43]. It is an up-to-date network embedding method for link prediction on dynamic graphs, which could capture both structural properties and temporal evolutionary patterns.
(4) *GCN+LSTM*. We use GCN [36] to obtain representations of nodes in each year and then feed them sequentially into LSTM [58] to generate node embeddings in the future.
(5) *GraphSage+LSTM*. GraphSage [38] is an effective network embedding method for large-scale graph embedding. Here we combine it with LSTM to get the future node representations.

Finally, to further validate the performance of each component in our model, we also design some simplified variants of HIMTKF as follows:

(6) *HIMTKF-HOI* excludes the High-Order Interaction Module from HIMTKF. That is to say, it does not consider the high-order network relationship between technologies.
(7) *HIMTKF-CO* excludes the Co-Occurrence Module from HIMTKF, indicating that it ignores the co-occurrence relationships between technologies.
(8) *HIMTKF-IHD* excludes the Improved Hierarchical Delivery Module from HIMTKF, meaning that it does not take hierarchical relationships among technologies into consideration.
(9) *HIMTKF-Loss* removes the $\mathcal{L}_e$ loss function in the objective function, implying that it neglects the importance of unprecedented flows.
(10) *HIMTKF-Single* combines the diffusion feature and absorption feature into one feature as the input of HIMTKF.

6.1.3 Evaluation metrics

To comprehensively evaluate the performance of our model, we adopt various metrics from different perspectives, i.e., Precision, Recall, F1-score, AUC and AP [63]. Among these metrics, F1, Precision and Recall are dependent on the threshold of probability [64]. It is worth noting that all of the aforementioned indicators are aimed at forecasting all flows, regardless of whether they have existed previously or not. However, from the perspective of

technological knowledge flow prediction, we prefer to see technological knowledge flows that have never appeared before [24], which accounts for one-seventh of the total flows shown in Fig. 6b. Accordingly, we design a special evaluation metric called Unprecedented Right (UR) to evaluate the prediction performance for these unprecedented edges as follows:

$$UR = \frac{\#positive\ unprecedented\ edges}{\#all\ unprecedented\ edges}. \tag{18}$$

6.1.4 Hyper-parameter settings

The parameters in our model are all initialized using Xavier [65] initialization. The citation threshold of TKF is set to 5 and the number of orders in HOI and layers of LSTM are set to 3 and 2, respectively. In the process of model training, we use Adam optimizer [66] to optimize the parameters and initialize the learning rate to 0.001. During the training process, when the loss function stops improving after 20 steps, the learning rate is reduced by a factor of 0.5 to a minimum value of 0.00001. We set $\alpha = 0.65$, $\alpha = 0.5$ and $\alpha = 0.5$ in loss function for dataset F, H and Random Data, respectively. The baseline parameters are configured similarly to our method and are all optimized to ensure fair comparisons. All experiments are conducted on a Linux server with eight 2.60 GHz Intel (R) Xeon (R) E5-2960 CPUs and a NVIDIA Tesla V100 PCIe 16 GB GPU.

6.2 Experimental results

6.2.1 Overall performance

In this part, we evaluate the overall performance of HIMTKF, HighTKF and other baselines in terms of TKF forecasting. In the experiments, each model is trained on snapshots of each year from 2006 to 2018. More specifically, the input contains snapshots from 2006 to 2017, and the forecast target is TKF in 2018. The validation set is from 2007 to 2019, and the test set is from 2008 to 2020. All results are shown in Table 2.

We can deduce several significant conclusions from this table. To begin, HIMTKF and HighTKF outperform all baselines on the majority of evaluation metrics, demonstrating our models' effectiveness. Especially on the UR metric, HIMTKF and HighTKF perform significantly better than other baselines, indicating the model's validity in predicting unprecedented flows. Second, the performances of dynamic models (3–5) are better than that of static models (1–2) on most of evaluation metrics, especially on the index UR, which proves the significance of taking sequential features into account.

As for the comparison among the HIMTKF, HighTKF and variants of HIMTKF, we can easily observe that HIMTKF outperforms HighTKF and other variants on most metrics, thereby proving the effectiveness of each module of HIMTKF. In addition, we find that some variants of HIMTKF performs better than HighTKF in some evaluation metrics such as HIMTKF-CO on all datasets and HIMTKF-IHD on Random Data, showing that CO and IHD are helpful in improving the effect of the model. Moreover, the performance on UR of HIMTKF-Loss is significantly lower than that of other variants, indicating that the addition of $\mathcal{L}_e$ loss function can effectively improve the prediction accuracy of the unprecedented edges.

Table 2 Overall performance evaluation (%)

Method		Static		Dynamic			Variant					Our model	
		Random	MF	GCN +LSTM	GraphSage +LSTM	DySAT	HIMTKF -HOI	HIMTKF-CO	HIMTKF-HD	HIMTKF -Loss	HIMTKF -Single	HighTKF	HIMTKF
F	AUC	50.06	88.83	92.29	90.82	88.5	94.76	97.33	94.44	94.41	94.25	96.78	**98.01**
	AP	50.03	88.94	91.91	90.36	87.81	94.91	97.39	94.53	94.69	94.12	96.99	**98.09**
	F1	50.01	88.01	85.54	86.2	80.45	87.93	**91.42**	87.36	84.5	87.73	89.22	91.29
	precision	50.09	88.02	82.96	83.61	80.45	84	87.86	83.2	80.48	83.12	85.39	**88.33**
	Recall	50.05	87.99	88.29	88.95	80.45	92.25	**95.27**	91.94	88.95	92.88	93.4	94.46
	UR	50.16	65.63	69.48	71.78	75.69	88.98	88.17	90.05	77.69	86.89	86.23	**92.22**
H	AUC	50.00	88.97	91.39	91.78	92.36	95.12	95.58	93.86	94.23	94.02	95.17	**96.58**
	AP	50.02	88.96	90.91	91.62	92.46	95.07	96.02	93.52	94.18	93.92	95.2	**96.63**
	F1	50.04	85.78	85.46	85.78	86.27	88.05	87.7	86.56	83.74	86.13	87.25	**89.97**
	precision	49.95	**86.65**	79.94	83.97	86.27	85.85	84.14	83.86	80.49	82.17	83.47	86.49
	Recall	49.99	84.93	88.41	87.68	86.27	90.37	91.56	89.45	87.26	90.48	91.38	**93.75**
	UR	49.98	70.95	73.62	72.98	75.53	82.52	83.21	83.64	76.08	82.32	81.68	**85.19**
Random data	AUC	49.99	86.54	90.87	92.65	91.4	95.59	96.64	96.91	96.36	93.73	96.37	**97.7**
	AP	50.01	86.69	91.07	92.24	90.74	95.44	96.63	96.82	96.3	93.79	96.61	**97.24**
	F1	50.06	87.36	82.98	82.63	83.08	87.71	88.08	90.7	90.36	88.18	84.3	**91.8**
	precision	50.05	85.24	80.92	81.55	83.61	81.52	81.98	87.01	**88.66**	88.42	78.92	88.29
	Recall	50.05	89.59	85.14	84.68	83.61	94.93	95.16	94.71	92.13	87.94	90.47	**95.6**
	UR	50.01	69.75	71.12	70.59	74.16	87.18	88.33	87.21	79.7	78.18	82.54	**89.56**

The bold values indicates the best performance for comparison

6.2.2 Performance in different periods

In order to verify the performance of our HIMTKF model in detail, here we evaluate the performance of HIMTKF in different periods [67]. More specifically, we divide the data from 1991 to 2020 into three periods (i.e., 1991–2000, 2001–2010, 2011–2020), and conduct experiments, respectively. For better illustration and effective analysis, we only compare HIMTKF with HighTKF and four well-performing baselines (Models (2–5)). And the results are illustrated in Figs. 8, 9 and 10. From the figures, we can find that HIMTKF significantly outperforms HighTKF and all baseline methods, which further demonstrates the effectiveness of our method. Furthermore, the performance of HIMTKF is basically consistent with the overall performance in Table 2, which reflects the stability of it in different time periods.

In addition, as time goes by, the performances of most models show a slight decline in general. It may come from the reason that the interactions between technologies become more frequent over time, as illustrated by the increasing number of unprecedented edges in Fig. 6. Because of these massive new edges, the difficulty of prediction increases, resulting in a drop in prediction accuracy.

6.2.3 Multi-step forecasting

Moreover, we perform a multi-step prediction on our model. As mentioned before, using characteristic representations of previous T years, we can predict characteristics of year $T + 1$, so as to obtain flows of technological knowledge in year $T + 1$. Analogously, taking characteristics from year 2 to year $T + 1$ as inputs of TFT module, we can get node characteristics and technological knowledge flows in year $T + 2$. Through this continuous prediction method, we evaluate the performance of HIMTKF in multi-step prediction on three datasets.

More specifically, we use the data from 2001 to 2017 to predict next three years, i.e., 2018, 2019 and 2020. As shown in Fig. 11, HIMTKF performs good stability on multi-step prediction task with values of most evaluation metrics descend gently and steadily. For the metric UR, we notice that it has a more obvious downward trend compares with others. Even so, its worst-case result in the third step prediction is still higher than 80%. Therefore, HIMTKF can accurately predict TKF in the future.

6.2.4 Parameter sensitivity

In this part, we investigate the sensitivity of model parameters. In more detail, we look at the effect of weight parameter α in loss function [Eq. (17)], as illustrated in Fig. 12.

In HIMTKF, the trade-off parameter α plays a crucial role which balances the contribution from different influences of loss function $\mathcal{L}_e$ and $\mathcal{L}_g$. From this figure, we can find that, with the increasing of α, AUC remains essentially unchanged and Precision and F1 tend to increase gradually, while UR and Recall decrease gradually. These indicate that the decreasing of the proportion of $\mathcal{L}_e$ will result in an inaccuracy in predicting UR. To ensure best performance, we choose $\alpha = 0.65$ in dataset F and $\alpha = 0.5$ in dataset H and Random Data.

6.2.5 Case Study

Representations for diffusion and absorption features To further illustrate what the generated diffusion and absorption features are going to reveal, we provide a detailed illustration

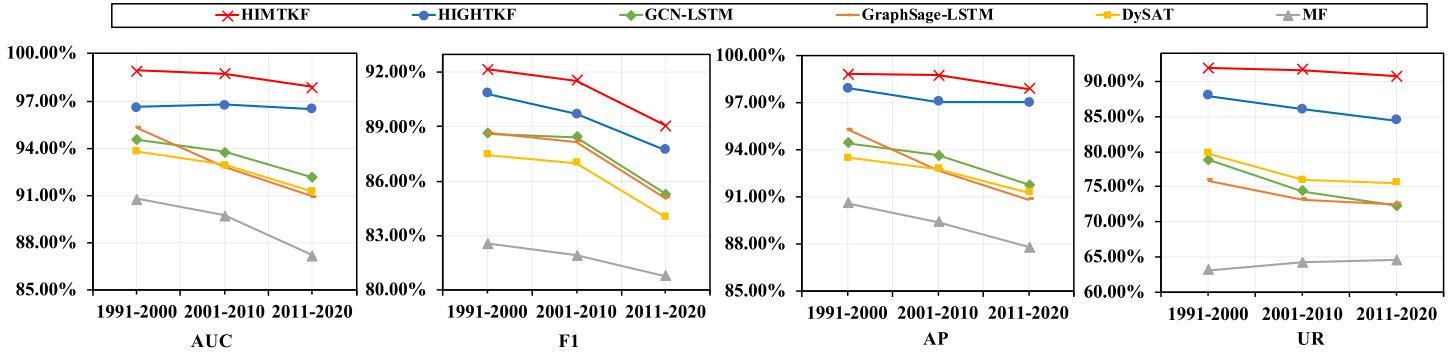


Fig. 8 Performance in different periods on dataset F (Mechanical Engineering)

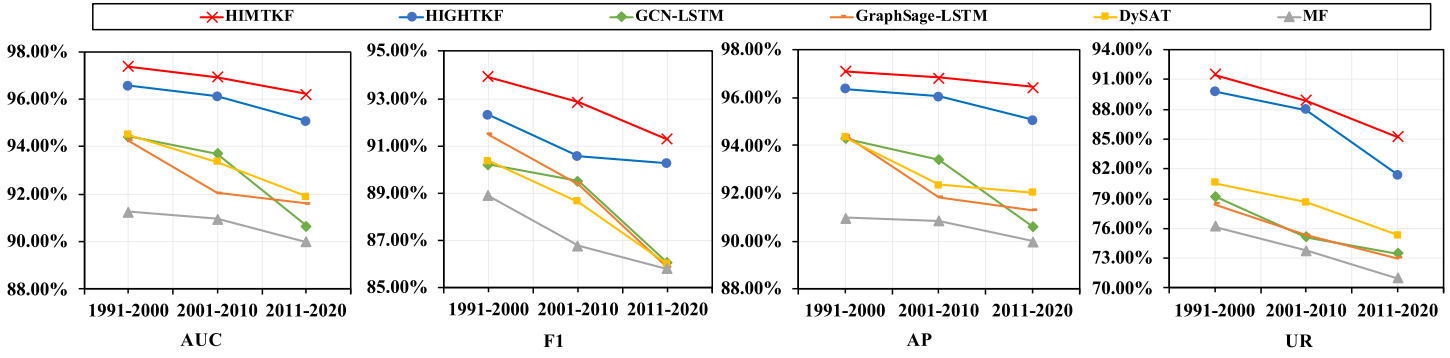


Fig. 9 Performance in different periods on data set H (Electricity)

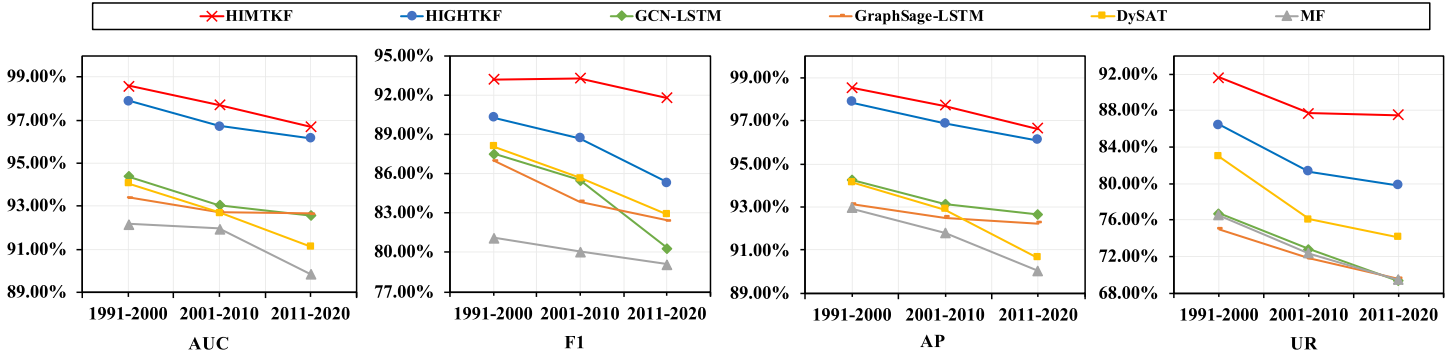


Fig. 10 Performance in different periods on Random Data

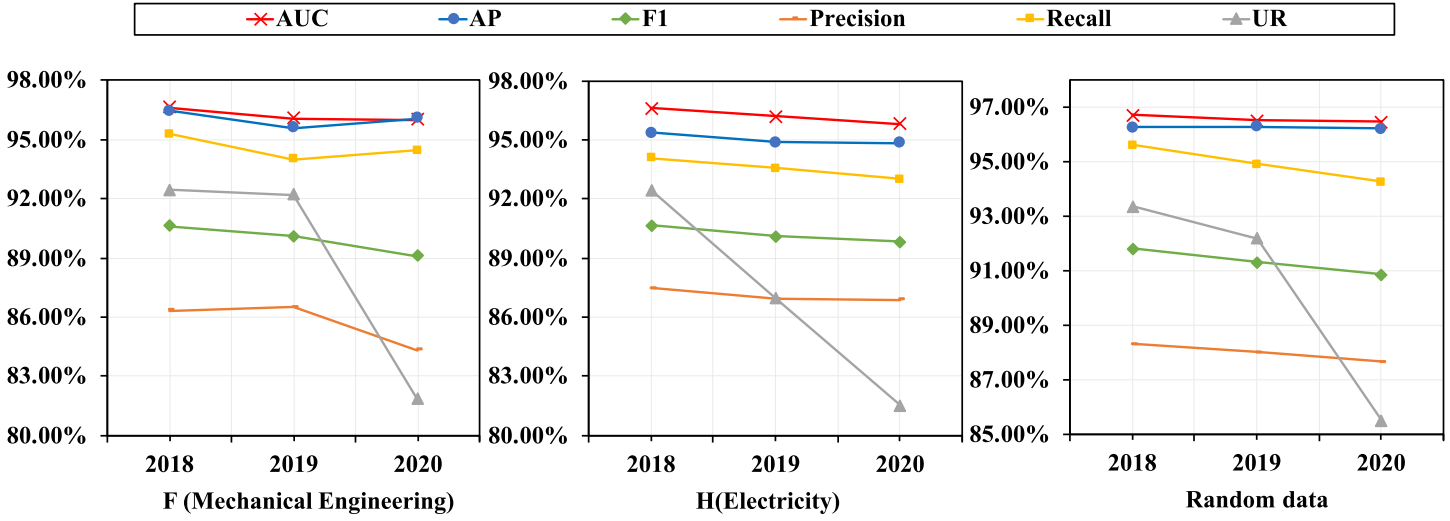


Fig. 11 Multi-step forecasting on three datasets

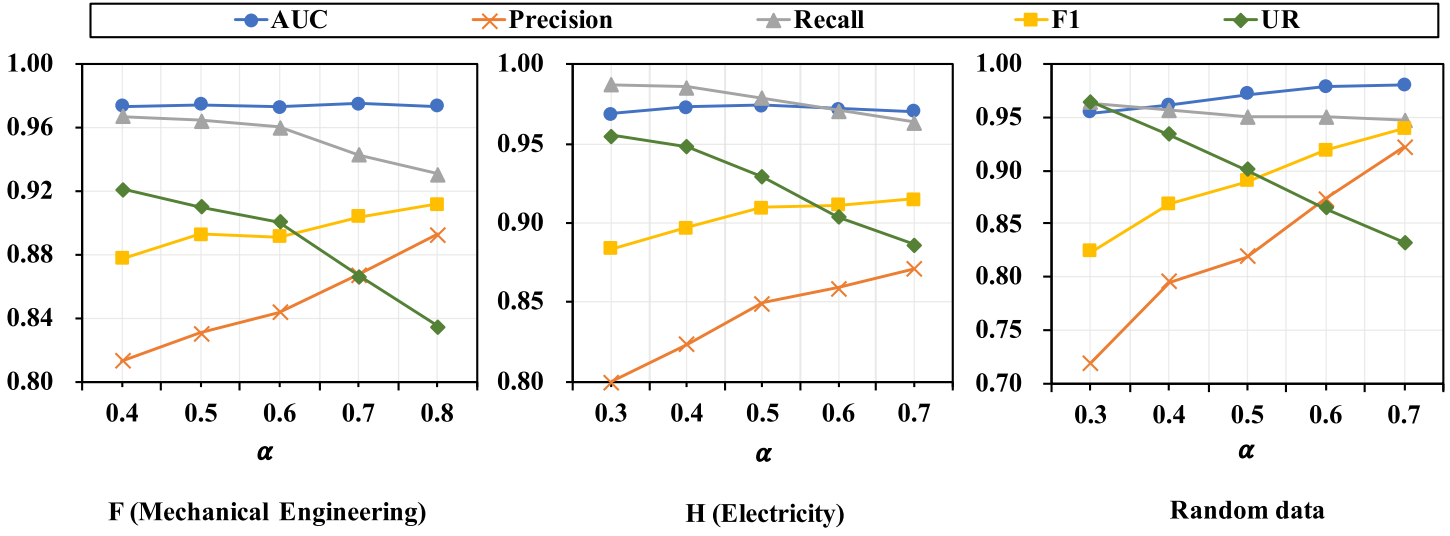
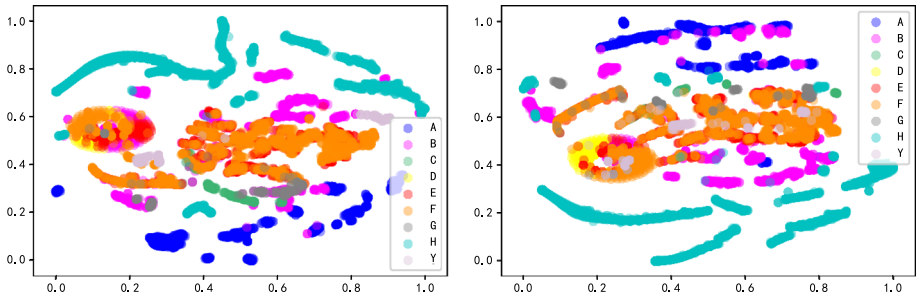


Fig. 12 Comparison on different weights of loss function



(a) Diffusion Features **(b)** Absorption Features

Fig. 13 Visualization of the learned t-SNE transformed representations

and analysis in this part. In more detail, we project the diffusion and absorption features of technologies in 2021 to a two dimensional space with the widely used visualization tool t-SNE.

As shown in Fig. 13, the nodes represent the diffusion features and absorption features of technologies, and the colors indicate the categories to which they belong. As shown in this figure, the features of nodes belonging to the same category are relatively similar, demonstrating that our model is capable of effectively learning the hierarchical similarity in CPC architecture. Meanwhile, we have noticed some outliers that are not quite as close to the nodes in the same category. This phenomenon implies that HIMTKF not only considers the hierarchical similarity of technologies, but also protects the personalized characteristics of technologies.

Moreover, we observe some overlaps between technologies classified in different categories, such as Section D (Textiles; Paper), Section E (Fixed Constructions) and Section F (Mechanical Engineering), indicating that there is a trend of convergence between these types of technologies. Also, by comparing Fig. 13a, b, we can observe that although the diffusion and absorption characteristics of the technologies are slightly different, the overall structures are obviously similar. This finding explains why HIMTKF with a single feature (HIMTKF-Single) can still achieve a good performance, as illustrated in Table 2.

Unprecedented TKF in the future The unprecedented flow of technological knowledge in the future is a key topic of concern. In this part, we make a multi-step prediction to obtain the TKF in 2022. We extract some representative technological knowledge flows and list them in Table 3. We observe that CPC code B29L2031/7546 may flow to many electric communication related technologies in the future. There are also many useful findings through multi-step prediction of HIMTKF, which can help companies determine strategic direction to maximize research and development (R&D) performance in the future.

7 Conclusion

In this paper, we have presented a focused study on technological knowledge flow prediction based on HighTKF proposed by the previous work [18]. Though HighTKF could effectively deal with the problem of predicting technological knowledge flow in the future, it is unable

Table 3 Technological knowledge flows in 2022

CPC code	Description of technology
A01G25/167	Watering gardens, fields, sports grounds or the like—Control of watering-Control by humidity of the soil itself or of devices simulating soil or of the atmosphere; Soil humidity sensors
A61B2017/00398	Surgical instruments, devices or methods, e.g., tourniquets—Details of actuation of instruments, e.g., relations between pushing buttons, or the like, and activation of the tool, working tip, or the like-using powered actuators, e.g., stepper motors, solenoids
A01G7/04	Botany in general-Electric or magnetic; or acoustic; treatment of plants for promoting growth
G01S1/7038	Beacons or beacon systems transmitting signals having a characteristic or characteristics capable of being detected by non-directional receivers and defining directions, positions, or position lines fixed relatively to the beacon transmitters; Receivers cooperating there—with using electromagnetic waves other than radio waves-Details-Transmitters-Signal details
G05B2219/23363	Program-control systems-Pc systems-Pc programming-Barcode
F21S8/035	Lighting devices intended for fixed installation of surface-mounted type—the surface being a wall or like vertical structure, e.g., building facade—by means of plugging into a wall outlet, e.g., night light
B29L2031/7546	Other particular articles-Medical equipment; Accessories therefor—Surgical equipment
H04M2201/36	Electronic components, circuits, software, systems or apparatus used in telephone systems-Memories
B29L2031/7546	Other particular articles-Medical equipment; Accessories therefor—Surgical equipment
H04Q2213/1309	Indexing scheme relating to selecting arrangements in general and for multiplex systems-Apparatus individually associated with a subscriber line, line circuits
B29L2031/7546	Other particular articles—medical equipment; Accessories therefor—Surgical equipment
H04Q2213/13405	Indexing scheme relating to selecting arrangements in general and for multiplex systems—dual frequency signaling, DTMF

to exploit co-occurrence relationships between technologies or to take into account unique characteristics of different technology nodes at the same level. Therefore, we have proposed a hierarchical interactive multi-channel graph neural network for forecasting technological knowledge flow (HIMTKF) to address the aforementioned issue. In detail, we first conduct a data analysis of the intrinsic characteristics of TKF, and identified many potential influencing factors, including double-faced aspects of technology nodes, multiple complex relationships among different technologies, and dynamics of the TKF process. On the basis of these findings, we develop four components to learn two distinct representations of each technology for final prediction, namely High-Order Interaction Module (HOI), Co-occurrence Module (CO), Improved Hierarchical Delivery Module (IHD) and Technological Knowledge Flow Tracing Module (TFT). HOI and CO are aimed to model high-order network relationships and co-occurrence relationships between technologies inside the same level of the hierarchy; IHD is intended to model hierarchical relationships among technologies without neglecting personalities of technologies; TFT is attempted to capture the dynamic evolution of techno-

logical features through the use of the aforementioned relationships. Additionally, a hybrid loss function and a new evaluation metric are introduced for better predicting unprecedented flows between technologies. Finally, extensive experiments performed on a real-world patent database can clearly validate the effectiveness of our approaches for TKF forecasting. We hope this work could lead to additional future investigations.

Acknowledgements This research was partially supported by grants from the National Natural Science Foundation of China (Grants No. U20A20229, 61922073 and 71802068).

References

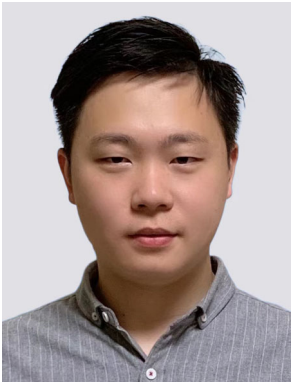
1. Aaldering LJ, Leker J, Song CH (2019) Competition or collaboration?—analysis of technological knowledge ecosystem within the field of alternative powertrain systems: a patent-based approach. J Clean Prod 212:362–371
2. Zhang L, Zhu H, Xu T, et al (2019) Large-scale talent flow forecast with dynamic latent factor model. In: The world wide web conference, pp 2312–2322
3. Song CH, Elvers D, Leker J (2017) Anticipation of converging technology areas—a refined approach for the identification of attractive fields of innovation. Technol Forecast Soc Chang 116:98–115
4. Song K, Kim K, Lee S (2017) Discovering new technology opportunities based on patents: text-mining and f-term analysis. Technovation 60–61:1–14
5. Abramo G, D'Angelo CA, Di Costa F (2020) The role of geographical proximity in knowledge diffusion, measured by citations to scientific literature. J Inform 14(1):101010
6. Emmanuel D, Megan M (2005) How well do patent citations measure flows of technology? Evidence from French innovation surveys. Dev Comput Syst 14(5):375–393
7. Zhang L, Li L, Li T (2015) Patent mining: A survey. ACM SIGKDD Explorations Newsl 16(2):1–19
8. Liu Y, Wu H, Huang Z et al (2020) Technical phrase extraction for patent mining: a multi-level approach. In: ICDM. IEEE, Sorrento, pp 1142–1147
9. Magerman T, Looy BV, Song X (2010) Exploring the feasibility and accuracy of latent semantic analysis based text mining techniques to detect similarity between patent documents and scientific publications. Scientometrics 82(2):289–306
10. Rui LI (2010) On the framing of patent citations and academic paper citations in reflecting knowledge linkage: a discussion of the discrepancy of their divergent value-orientations. Chin J Libr Inf Sci 3:37–45
11. Chen L (2017) Do patent citations indicate knowledge linkage? The evidence from text similarities between patents and their citations. J Informetr 11(1):63–79
12. Wu H, Zhang K, Lv G et al (2019) Deep technology tracing for high-tech companies. In: ICDM. IEEE, New York, pp 1396–1401
13. Kim J, Magee CL (2017) Dynamic patterns of knowledge flows across technological domains: empirical results and link prediction. SSRN Electron J
14. Alcacer J, Gittelman M (2006) Patent citations as a measure of knowledge flows: the influence of examiner citations. Rev Econ Stat 88(4):774–779
15. Goodman CM (1987) The Delphi technique: a critique. J Adv Nurs 12(6):729–734
16. Ko N, Yoon J, Seo W (2014) Analyzing interdisciplinarity of technology fusion using knowledge flows of patents. Expert Syst Appl 41(4):1955–1963
17. Smojver V et al (2020) Exploring knowledge flow within a technology domain by conducting a dynamic analysis of a patent co-citation network. J Knowl Manag **25**
18. Liu H, Wu H, Zhang L et al (2021) Technological knowledge flow forecasting through a hierarchical interactive graph neural network. In: ICDM. IEEE, Auckland, New Zealand, pp 389–398
19. Acemoglu D, Akcigit U, Kerr WR (2016) Innovation network. Proc Natl Acad Sci 113(41):11483–11488
20. Porter A, Cunningham S (2006) Tech mining: exploiting new technologies for competitive advantage. Technol Forecast Soc Chang 73:91–93
21. Harb YA, Abu-Shanab E (2020) A descriptive framework for the field of knowledge management. Knowl Inf Syst 62(12):4481–4508
22. Verhaegen PA, D'Hondt J, Vertommen J et al (2009) Relating properties and functions from patents to TRIZ trends. CIRP J Manuf Sci Technol 1(3):126–130
23. Cho Y, Kim E, Kim W (2015) Strategy transformation under technological convergence: evidence from the printed electronics industry. Soc Sci Electron Publ 674(67):106–131

24. Park I, Yoon B (2018) Technological opportunity discovery for technological convergence based on the prediction of technology knowledge flow in a citation network. J Informetr 12(4):1199–1222
25. Lee J, Kim C, Shin J (2017) Technology opportunity discovery to R&D planning: key technological performance analysis. Technol Forecast Soc Chang 119:53–63
26. Lü L, Zhou T (2011) Link prediction in complex networks: a survey. Physica A 390(6):1150–1170
27. Zhou T et al (2009) Predicting missing links via local information. Eur Phys J B 71(4):623–630
28. Adamic LA, Adar E (2003) Friends and neighbors on the web. Soc Netw 25(3):211–230
29. Sasaki H, Sakata I (2020) Identifying potential technological spin-offs using hierarchical information in international patent classification. Technovation 102192
30. Lee WS, Han EJ, Sohn SY (2015) Predicting the pattern of technology convergence using big-data technology on large-scale triadic patents. Technol Forecast Soc Chang 100:317–329
31. Yang C, Huang C, Su J (2018) An improved SAO network-based method for technology trend analysis: a case study of graphene. J Informetr 12(1):271–286
32. Yoon B, Park Y (2004) A text-mining-based patent network: analytical tool for high-technology trend. J High Technol Manag Res 15(1):37–50
33. Gori M, Monfardini G, Scarselli F (2005) A new model for learning in graph domains. In: IJCNN, vol. 2. IEEE, Montreal, QC, Canada, pp 729–734
34. Estrach JB, Zaremba W, Szlam A et al (2014) Spectral networks and deep locally connected networks on graphs. In: ICLR
35. Defferrard M, Bresson X, Vandergheynst P (2016) Convolutional neural networks on graphs with fast localized spectral filtering. In: NeurIPS, pp 3844–3852
36. Kipf TN, Welling M (2017) Semi-supervised classification with graph convolutional networks. In: ICLR
37. Velikovi P, Cucurull G, Casanova A et al (2018) Graph attention networks. In: ICLR
38. Hamilton WL, Ying R, Leskovec J (2017) Inductive representation learning on large graphs. In: NeurIPS, Long Beach, CA, USA, pp 1025–1035
39. Wang X, Zhu M, Bo D et al (2020) AM-GCN: adaptive multi-channel graph convolutional networks. In: ACM SIGKDD. Association for Computing Machinery, New York, pp 1243–1253
40. Guo X, Zhao L, Homayoun H et al (2021) Deep graph transformation for attributed, directed, and signed networks. Knowl Inf Syst 63(6):1305–1337
41. Schlichtkrull M, Kipf TN, Bloem P et al (2018) Modeling relational data with graph convolutional networks. In: European semantic web conference. Springer, Cham, pp 593–607
42. Toujani R, Akaichi J (2019) An approach based on mixed hierarchical clustering and optimization for graph analysis in social media network: toward globally hierarchical community structure. Knowl Inf Syst 60(2):907–947
43. Sankar A, Wu Y, Gou L et al (2020) Dysat: deep neural representation learning on dynamic graphs via self-attention networks. In: WSDM, pp 519–527
44. Zhang J, Li M, Gao K et al (2021) Word and graph attention networks for semi-supervised classification. Knowl Inf Syst 63(11):2841–2859
45. Li W, Xiao X, Liu J et al (2020) Leveraging graph to improve abstractive multi-document summarization. arXiv preprint arXiv:2005.10043
46. Wang H, Lian D, Tong H et al (2021) Hypersorec: exploiting hyperbolic user and item representations with multiple aspects for social-aware recommendation. ACM Trans Inf Syst (TOIS) 40(2):1–28
47. Mauw S, Ramírez-Cruz Y, Trujillo-Rasua R (2019) Conditional adjacency anonymity in social graphs under active attacks. Knowl Inf Syst 61(1):485–511
48. You H et al (2017) Development trend forecasting for coherent light generator technology based on patent citation network analysis. Scientometrics 111(1):297–315
49. Clough JR, Gollings J, Loach TV et al (2015) Transitive reduction of citation networks. J Complex Netw 3(2):189–203
50. Liu Q, Wu H, Ye Y et al (2018) Patent litigation prediction: a convolutional tensor factorization approach. In: IJCAI. AAAI Press, Stockholm
51. Lobo J et al (2019) Sources of inventive novelty: two patent classification schemas, same story. Scientometrics 120(1):19–37
52. Kapoor R, Karvonen M, Ranaei S et al (2015) Patent portfolios of European wind industry: new insights using citation categories. World Patent Inf 41:4–10
53. Liu Q, Ge Y, Li Z et al (2011) Personalized travel package recommendation. In: 2011 IEEE 11th international conference on data mining. IEEE, Vancouver, pp 407–416
54. Ernst H (1999) Evaluation of dynamic technological developments by means of patent data. In: The dynamics of innovation. Springer, Berlin, pp 103–132
55. He X, Deng K, Wang X et al (2020) Lightgcn: simplifying and powering graph convolution network for recommendation. In: ACM SIGIR, pp 639–648

56. Jeon J, Suh Y (2019) Multiple patent network analysis for identifying safety technology convergence. Data Technol Appl
57. Shi C, Han X, Song L et al (2021) Deep collaborative filtering with multi-aspect information in heterogeneous networks. IEEE Trans Knowl Data Eng 33(4):1413–1425
58. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
59. Liu Q, Huang Z, Yin Y et al (2021) EKT: exercise-aware knowledge tracing for student performance prediction. IEEE Trans Knowl Data Eng 33(1):100–115
60. Kipf TN, Welling M (2016) Variational graph auto-encoders. In: Bayesian deep learning workshop, NeurIPS (2016)
61. Caviggioli F (2016) Technology fusion: identification and analysis of the drivers of technology convergence using patent data. Technovation 55–56:22–32
62. Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788–791
63. Yu R, Liu Q, Ye Y et al (2020) Collaborative list-and-pairwise filtering from implicit feedback. IEEE Trans Knowl Data Eng
64. Hu W, Gao J, Li B et al (2020) Anomaly detection using local kernel density estimation and context-based regression. IEEE Trans Knowl Data Eng 32(2):218–233
65. Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feed forward neural networks. J Mach Learn Res 9:249–256
66. Kingma DP, Ba J (2015) Adam: a method for stochastic optimization. In: ICLR
67. Zhao H, Liu Q, Zhu H et al (2018) A sequential approach to market state modeling and analysis in online P2P lending. IEEE Trans Syst Man Cybern Syst 48(1):21–33

Huijie Liu received the B.S. degree in statistics from the University of Science and Technology of China (USTC), Hefei, China, in 2019. She is currently pursuing the Ph.D. degree in the Anhui Province Key Laboratory of Big Data Analysis and Application (BDAA), School of Computer Science and Technology, University of Science and Technology of China (USTC), Hefei, China. Her current research interests include network embedding and its application for knowledge discovery and data mining.

Han Wu received the B.E. degree in software engineering from the Dalian University of Technology, Dalian, China, in 2015. She is currently pursuing the Ph.D. degree in data mining with the School of Computer Science and Technology, University of Science and Technology of China, Hefei, China. She has published several papers in refereed conference proceedings, such as IJCAI, ICDM, SIGIR and KDD. Her current research interests include data mining in business intelligence, with a focus on patent analysis such as patent litigation prediction and patent technology tracing.



Le Zhang is currently a researcher at Baidu Research. He received his B.E. degree in Software Engineering from Dalian University of Technology in 2016. And he received his Ph.D. degree in Computer Science from University of Science and Technology of China (USTC) in 2022. His research interests include graph representation learning and its application in business intelligence.

Runlong Yu received the B.E degree in computer science and technology from the University of Science and Technology of China (USTC), Hefei, China, in 2017. He is currently working toward a PhD degree in the Anhui Province Key Laboratory of Big Data Analysis and Application (BDAA), School of Computer Science and Technology, University of Science and Technology of China (USTC), Hefei, China. His research interests include recommender systems, data mining and computational intelligence. He was the recipient of the KDD CUP'19 Special Award. He was also the recipient of the China National Scholarship, in 2019.

Ye Liu received the B.E. degree in Electronic Information Engineering from the University of Science and Technology of China (USTC), Hefei, China, in 2019. He is currently working toward the Ph.D. degree in the Anhui Province Key Laboratory of Big Data Analysis and Application (BDAA), School of Data Science, University of Science and Technology of China (USTC), Hefei, China. His research interests include data mining, natural language processing and knowledge graph. He has published several papers in refereed conference proceedings, such as ICDM, SIGIR and IJCAI. He was also the recipient of China National Scholarship, in 2016.

Chunli Liu received the PhD degree from Sun Yat-sen University, Guangzhou, China, in 2017. She is currently an assistant professor with School of Management, Hefei University of Technology, Hefei, China. Her research interests include data mining, Fintech and e-government. Her works appeared in several major international conferences and journals, including IJCIM, JMSE, etc.

Minglei Li receives the PhD degree in Computer Science in 2017 from the Hong Kong Polytechnic University, Hong Kong, China. He is now working in Huawei Cloud Computing Technologies Co. Ltd. His research interests include natural language processing, multimodal analysis, emotion analysis and human–computer interaction.

Qi Liu received the PhD degree from the University of Science and Technology of China (USTC), Hefei, China, in 2013. He is currently a professor with the Anhui Province Key Laboratory of Big Data Analysis and Application (BDAA), School of Computer Science and Technology, University of Science and Technology of China (USTC), Hefei, China. His research interests include data mining, machine learning, and recommender systems. He was the recipient of KDD'18 Best Student Paper Award and ICDM'11 Best Research Paper Award. He was also the recipient of China Out-standing Youth Science Foundation, in 2019.

Enhong Chen received the PhD degree from the University of Science 1187 and Technology of China (USTC), Hefei, China, in 1996. He is currently a director (professor) of the Anhui Province Key Laboratory of Big Data Analysis and Application (BDAA), University of Science and Technology of China (USTC), Hefei, China. He is an executive dean of the School of Data Science, USTC, and vice dean of the School of Computer Science and Technology, USTC. He is a CCF fellow. His research interests include data mining, machine learning and recommender systems. He has published more than 200 refereed international conference and journal papers. He was the recipient of KDD'08 Best Application Paper Award and ICDM'11 Best Research Paper Award.

Authors and Affiliations

Huijie Liu[1,2] · Han Wu[1,2] · Le Zhang[3] · Runlong Yu[1,2] · Ye Liu[1,2] · Chunli Liu[4] · Minglei Li[5] · Qi Liu[1,2] · Enhong Chen[1,2]

Huijie Liu
lhj33@mail.ustc.edu.cn

Han Wu
wuhanhan@mail.ustc.edu.cn

Le Zhang
zhangle0202@gmail.com

Runlong Yu
yrunl@mail.ustc.edu.cn

Ye Liu
liuyer@mail.ustc.edu.cn

Chunli Liu
liuchunli@hfut.edu.cn

Minglei Li
liminglei29@huawei.com

Qi Liu
qiliuql@ustc.edu.cn

1 Anhui Province Key Laboratory of Big Data Analysis and Application, School of Data Science & School of Computer Science and Technology, University of Science and Technology of China, Hefei, China

2 State Key Laboratory of Cognitive Intelligence, Hefei, China

3 Baidu Research, Beijing, China

4 School of Management, Hefei University of Technology, Hefei, China

5 Huawei Cloud Computing Technologies Co., Ltd, Shenzhen, China