Scientometrics
https://doi.org/10.1007/s11192-025-05516-7

Recommending university-industry collaboration: a topic-institution graph-based solution

Lu Huang[1,2] · Xiaoli Cao[3] · Hang Ren[4] · Guangchao Wang[1] · Yani Wang[5]

Received: 6 December 2024 / Accepted: 8 December 2025

Abstract

The collaboration between universities and enterprises synergistically foster the exchange of knowledge and technology. Research on university-industry collaboration (UIC) plays a pivotal role in promoting innovation, facilitating the transformation of scientific and technological achievements, and addressing "stuck neck" problems in technological advancement. This study proposes a topic-institution graph-based methodology for recommending UIC, applying graph analysis to extract multiple types of entities, cooperation information between institutions, semantic correlations between science, technology, and innovation (ST&I) topics, and co-occurrence relationships between institutions and ST&I topics. Firstly, the SciBERT model is applied to construct the science network and technology network, capturing the latent semantic correlations between paper keywords and patent keywords, respectively. Then, the Leiden community detection algorithm and multi-dimensional indicators are employed to identify promising ST&I topics within these networks. Next, a topic-institution graph is constructed, incorporating rich structural and semantic information between multiple entities. Finally, the heterogeneous graph attention network (HAN) model is applied to predict collaboration opportunities between universities and enterprises. Additionally, the interpretability of ChatGPT-4o helps provide insightful explanations for the recommended collaborations and validate the accuracy of results. An empirical study on artificial intelligence domain demonstrates the feasibility and reliability of the proposed methodology.

Keywords University-industry collaboration · Collaborators recommendation · ST&I topics · Graph embedding · Artificial intelligence

✉ Xiaoli Cao
cxl163990307@163.com

[1] School of Economics, Beijing Institute of Technology, Beijing, China

[2] Digital Economy and Policy Intelligentization Key Laboratory of Ministry of Industry and Information Technology, Beijing, China

[3] China Agricultural University Library, China Agricultural University, Beijing, China

[4] Beijing Institute of Technology, Zhuhai, China

[5] School of Global Governance, Beijing Institute of Technology, Beijing, China

Published online: 17 January 2026

Introduction

University-industry collaboration (UIC) is a crucial innovation mode, where universities and enterprises enhance their capabilities and foster synergistic innovation through the exchange of knowledge and technologies (O'Dwyer et al., 2023; Ollila, 2025). UIC can facilitate the orderly flow, integration, and sharing of innovative resources between universities and enterprises, enabling the transformation of scientific and technological achievements, while bridging the gap among science, technology, and industry (Chung et al., 2021; Foti et al., 2023; Wang & Huang, 2020). Therefore, UIC recommendation has become an important issue for university-industry-research collaborative innovation.

Nowadays, network analysis methods are mainly used to construct single-layer networks, such as collaboration networks and knowledge networks, for recommending university-industry collaboration (Wang et al., 2024a; Zhang & Chen, 2023). Collaboration networks capture the cooperative relationships between universities and enterprises, while knowledge networks uncover the connections between the research topics of institutions (Albats et al., 2022; Li et al., 2023). However, single-mode networks fail to capture the richness and complexity of interactions among various entities in real-world UIC (Guan et al., 2017; Lu et al., 2023). To address this problem, some scholars have integrated collaboration and knowledge networks into dual-layer networks for UIC recommendations (Chen et al., 2022a). Yet, dual-layer networks primarily focus on the relationships between layers (Cui & Zhu, 2024; Lian & Wang, 2024), making it difficult to effectively integrate collaboration information and knowledge associations. As a result, they fail to capture the complex global interactions between universities and enterprises. Therefore, this study constructs a topic-institution graph, revealing the multi-dimensional relationships between entities into a unified low-dimensional vector space (Kang et al., 2019), capturing the rich structural and semantic information (Noori et al., 2024; Park et al., 2023) and exploring potential collaboration opportunities between universities and enterprises through the global structure and higher-order connectivity of the graph.

Existing research primarily explores the knowledge interactions between universities and enterprises based on topic co-occurrences or International Patent Classification (IPC) code co-occurrences (Chang, 2018; Zhai et al., 2022). However, simple co-occurrence relationships struggle to capture the rich semantic information embedded within the knowledge, leading to an inaccurate depiction of the correlations between different knowledge entities (Huang et al., 2022a; Kodua-Ntim, 2023). The SciBERT model, a deep learning model based on BERT, is capable of extracting the latent semantic information between keywords within their context, thereby improving the accuracy of knowledge representation (Dinh et al., 2024; Thierry et al., 2023). Additionally, UIC should focus on promising science, technology, and innovation (ST&I) topics to ensure high-valued collaboration outcomes and foster a higher level of collaborative innovation between universities and industry (Reyes-Menendez et al., 2023; Xu et al., 2021a; Zhang et al., 2016). Therefore, developing an evaluation system to identify promising ST&I topics and exploring their correlations is crucial for successful UIC.

This study proposes a recommendation method for university-industry collaboration based on the topic-institution graph. This framework employs the graph analysis method to reveal complex information within the graph, including information on multiple types of entities, collaboration relationships between institutions, semantic associations between topics, and co-occurrence information between institutions and topics, enabling the effective prediction of UIC. The method begins by identifying promising ST&I topics with collaborative potential in the science network and technology network using the SciBERT model, the Leiden algorithm, and the multi-dimensional indicators, capturing the hidden semantic relationships between keywords and improving the accuracy of topic identification. Then, a topic-institution graph is constructed, which reveals the information of two types of nodes (ST&I topics and institutions), the semantic correlations between ST&I topics, the co-occurrence relationships between ST&I topics and institutions, and the co-authorship relationships between institutions. Finally, the heterogeneous graph attention network (HAN) model is applied to predict the future graph and then recommend UIC partners, fully considering both structural and semantic information in the graph. Furthermore, ChatGPT-4o is applied to provide explanations for potential future collaborations between universities and enterprises, and to validate the accuracy of UIC recommendation results.

The remainder of this study is organized as follows: "Related work" reviews the university-industry collaboration, ST&I topic identification, and graph embedding. "Methodology" section presents the research process and methods in detail. "Empirical study" describes a case study on the artificial intelligence field to recommend potential university-enterprise pairs. Validation tests are also conducted to demonstrate the feasibility of the proposed method. Then, the insights gained from the case study and limitations are discussed in "Discussion and conclusions" section.

Related work

University-industry collaboration

University-industry collaboration (UIC) is an essential mechanism for driving scientific and technological innovation (Lee & Miozzo, 2024; Steinmo & Rasmussen, 2018; Woolgar, 2007). In general, universities primarily engage in exploratory and creative scientific research activities, which form the foundation of scientific innovation, while enterprises play a crucial role in the integration of science, technology, and economics, which is the basis of technological innovation (Kang et al., 2019). UIC facilitates the bidirectional transfer of knowledge and technology, driving the transformation of scientific and technological achievements (Nsanzumuhire et al., 2021; Rossi et al., 2024).

To reap the benefits of collaborative partnerships, universities and enterprises should place greater emphasis on important factors for the success of cooperation, such as organizational trust, interaction channels, intellectual property management, and market conditions (Patnaik et al., 2022; Song et al., 2022). Particularly, proper partner selection is crucial for determining the success of UIC innovation and the alliance's performance (Chen et al., 2022b; Chung et al., 2021). One key consideration in evaluating potential collaboration partners is the relevance and applicability of their expertise (Gallagher et al., 2023; Rossi et al., 2024). In addition, not all potential partners possess the specific knowledge or technological capabilities that align with the needs of universities or enterprises (Li & Zhou, 2022; Wang et al., 2017). Therefore, universities and enterprises should carefully choose partners to ensure the success of collaborative innovation (Chung et al., 2021).

Numerous studies have been dedicated to recommending university-industry collaboration, which can be categorized into three approaches: subjective judgment-based methods (Roncancio-Marin et al., 2022), quantitative modeling-based methods (Gibson et al., 2019), and network analysis methods (Gallagher et al., 2023; Zhao et al., 2024). Subjective judgment-based approaches often rely on qualitative expert assessments, lacking quantitative analysis and resulting in subjective outcomes. Quantitative modeling-based approaches mainly focus on data statistics (Wang et al., 2016; Zhou et al., 2018), such as data envelopment analysis (DEA) method, optimization methods, and decision evaluation methods, without delving into an in-depth analysis of the research content and relationships of institutions. Network analysis methods mostly explore the collaboration between universities and enterprises using paper data and patent data (Chen et al., 2022b; Guan & Liu, 2016), revealing the collaboration characteristics and exploring the collaboration opportunities between universities and enterprises. Many studies construct single-layer networks (such as collaboration networks and knowledge networks) to recommend UIC (Albats et al., 2022; Wang et al., 2024a), while a few have integrated collaboration relationships and knowledge associations to build dual-layer networks for exploring collaboration opportunities (Chen et al., 2022a). However, single-layer networks fail to reflect the complex interactions among entities in real-world collaboration scenarios (Lu et al., 2023). Dual-layer networks, on the other hand, primarily capture the relationships between the collaboration layer and knowledge layer (Cui & Zhu, 2024), which leads to their inability to fully capture and predict the complex and dynamic global structural interactions between universities and enterprises within the network.

With the development of deep learning, many graph analysis methods based on neural networks have been increasingly applied in university-industry collaboration recommendation (Chung et al., 2021; Mao et al., 2020; Zeng et al., 2023). A graph is a mathematical structure used to model pairwise relationships between entities, where entities are represented as nodes and relationships as edges connecting them. Graphs offer a more comprehensive and rich representation of relationships, highlighting the complex structure and connectivity among multiple entities (Mei et al., 2022; Park et al., 2023). Compared to traditional network analysis methods, graph analysis techniques can more effectively excavate complex relationships between multiple entities, revealing potential correlations through the global structure and higher-order connectivity of the graph (Kang et al., 2019; Zhang et al., 2023).

ST&I topic identification

Identifying the promising and important science, technology, and innovation (ST&I) topics for UIC is a crucial issue (Libaers et al., 2017; Ran et al., 2020). ST&I topics represent the high-value research areas that are of significant interest to both universities and enterprises, with the potential to drive innovative outcomes and successful collaborations (Kang et al., 2019; Qian et al., 2024). Emerging and general-purpose topic identification has received significant attention in the ST&I field, indicating the increased emphasis of universities and enterprises on these types of topics (Wang, 2018; Zhang et al., 2021).

Most studies have indicated that emerging topics exhibit characteristics such as novelty, growth, and influence (Bianchini et al., 2022; Huang et al., 2021a), while general-purpose topics are typically associated with growth and fundamentality (Petralia, 2020; Zhang et al., 2021). However, most existing measures either rely solely on temporal novelty or focus on simple citation-based influence (Small et al., 2014; Yang et al., 2022b), which may not fully capture the semantic uniqueness, structural roles, and dynamic evolution of topics. To address these limitations, this study proposes an improved topic evaluation framework based on four enhanced indicators (novelty, fundamentality, growth, and impact) designed to capture both semantic, structural and dynamic dimensions. Novelty jointly considers temporal recency and semantic distinctiveness of a topic (Rotolo et al., 2015; Xu et al., 2021a). Fundamentality integrates multiple centrality measures to capture a topic's structural importance across the network (Zhang et al., 2021). Growth reflects dynamic structural expansion rather than publication volume alone (Petralia, 2020). Impact is measured via a modified PageRank approach, capturing both direct and indirect network influence (Xu et al., 2021b). These indicators are combined using entropy weighting to objectively identify ST&I topics with the highest potential for UIC.

In addition, it is crucial to thoroughly explore the correlations between the ST&I topics of universities and enterprises (Kim et al., 2020). Research indicates that universities and enterprises often have differing development goals and research focuses, which can lead to disagreements and friction during collaboration. These disparities may, in turn, affect the efficiency and outcomes of university-industry collaborative innovation (Huang et al., 2021b; Littleton et al., 2023). Therefore, this study incorporates the semantic correlations between ST&I topics into the factors considered for UIC, enhancing the accuracy of university-industry collaboration recommendations.

Graph embedding

Graph embedding is a graph analysis technique that represents graph-structured data as low-dimensional vectors, while preserving both the structural and semantic information inherent in the graph (Yang et al., 2022a). It has been widely applied in recommendation systems, technology innovation management, social network analysis, and bioinformatics (Bertram et al., 2023; Ding, 2022; Yi et al., 2022). Graph embedding methods mainly include three categories: (1) Matrix factorization methods, such as GraRep and HOPE, decompose adjacency matrices to capture graph structures but often struggle with scalability (Sun et al., 2023). (2) Random walk-based methods, like DeepWalk and Node2Vec, simulate random paths over graphs to learn embeddings that capture local and global graph features (Zhang & Tang, 2023). (3) Neural network-based approaches, such as Graph Convolutional Network (GCN), have gained popularity for their ability to model complex graph relationships (Wang et al., 2023). With the development of deep learning techniques, graph embedding based on the neural network has shown outstanding performance in various types of tasks, such as node classification and link prediction (Huang et al., 2022b; Wang et al., 2024a).

The heterogeneous graph attention network (HAN) model is a neural network-based graph embedding method, which is specifically designed for heterogeneous graphs including multiple types of nodes and edges (Zhang et al., 2019a). HAN leverages attention mechanisms to learn the importance of different nodes and edge types, effectively capturing both structural and semantic complexities in heterogeneous graphs (Park et al., 2023; Yang et al., 2024). Compared with other graph embedding methods, it leverages an attention mechanism to aggregate entity features, effectively capturing both rich semantic relationships and structural information within the graph (Mei et al., 2022). Therefore, HAN model can depict the associations between universities and enterprises by integrating research topics and collaboration information, enabling the prediction of potential university-industry collaborations.

Methodology

The proposed method is given in Fig. 1, which consists of two parts: (1) University-industry collaboration-oriented ST&I topics identification; and (2) University-industry collaboration prediction based on the HAN model. The HAN model extracts feature information from different types of entities (e.g., topic nodes, institution nodes) within the topic-institution graph, embedding them into a unified vector space. This approach effectively learns the enriched structural and semantic relationships between diverse nodes, thereby improving the accuracy of collaboration recommendations between universities and enterprises.

University-industry collaboration-oriented ST&I topics identification

Data collection and preprocessing

The dataset used in this study includes paper data from the Web of Science (WoS) and patent data from the Derwent Innovation Index (DII). After data pre-processing via ITGInsight,[1] paper and patent keywords are extracted and cleaned from the titles, abstracts, and author keywords, along with the corresponding universities, and enterprises.

Science network and technology network construction based on SciBERT model

The purpose of this section is to construct the science network and technology network using the SciBERT model. SciBERT (Beltagy et al., 2019) is a BERT-based pre-trained language model trained on a large corpus of scientific and technological texts, which can capture richer semantic and co-occurrence information within texts and obtain more precise semantic representations of paper and patent keywords (Dinh et al., 2024). In this study, SciBERT is used to generate word vectors corresponding to all paper and patent keywords.

Then, the semantic similarity between keywords is calculated using the cosine distance between keyword vectors. Finally, the science network G_s and technology network G_t are constructed based on semantic similarity as follows:

$$G_s = G(V_s, E_s, W_s) \tag{1}$$

$$G_t = G(V_t, E_t, W_t) \tag{2}$$

where V_s and V_t are the set of paper keywords and patent keywords, respectively; E_s and E_t denote the set of edges of the science network and technology network, respectively; W_s and W_t are the set of edge weights (semantic similarity) of science network and technology network, respectively.

[1] 1 ITGInsight is text mining and visualization software for bibliometric data, for more information see the official website at cn.itginsight.com.

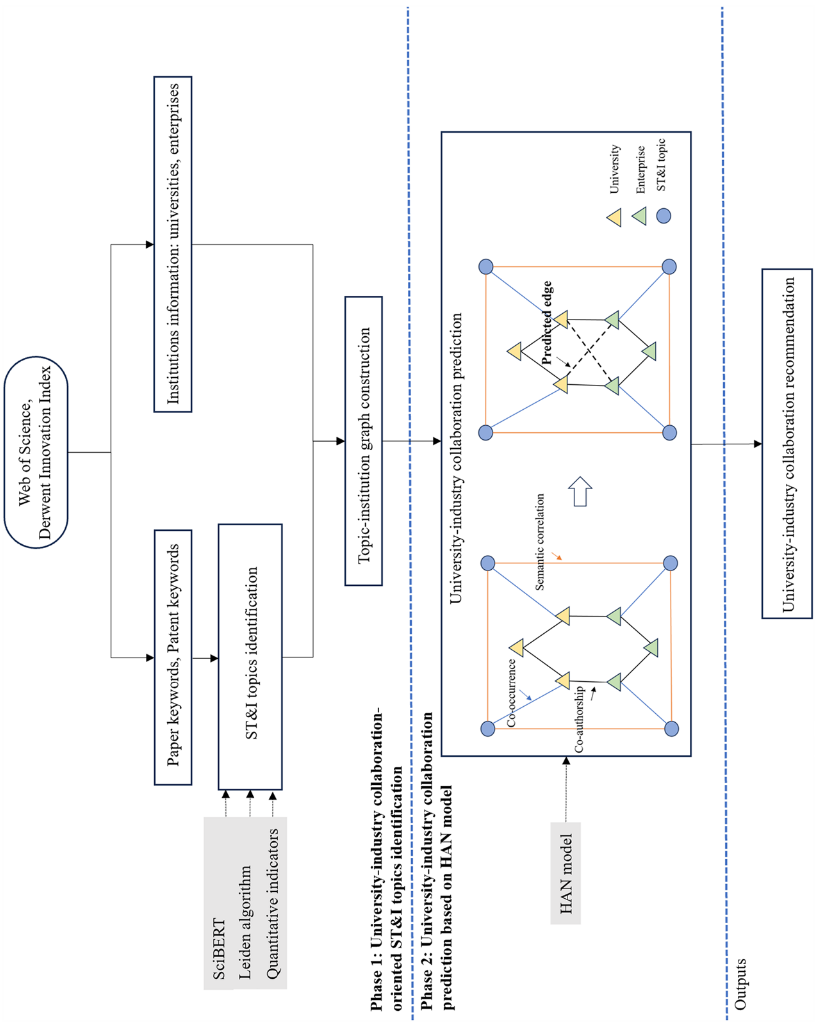

Fig. 1 Framework of recommending university-industry collaboration

ST&I topics identification based on quantitative indicators

To construct a semantically meaningful heterogeneous graph for university-industry collaboration, this study first identifies Science, Technology, and Innovation (ST&I) topics based on science and technology networks and community detection method. The overall process includes two stages: (1) topic community detection, and (2) ST&I topic identification.

(1) Topic community detection After constructing the science network and technology network, the Leiden community detection algorithm is applied to identify science topics (communities) and technology topics (communities). The Leiden algorithm is a modularity-based community detection method known for its efficiency and ability to ensure strong intra-community connections, thereby providing a more accurate representation of the community structure within a network (Bae et al., 2017). Hence, we use the Leiden algorithm to divide science communities and technology communities in the constructed science network and technology network. The output of this step is a collection of candidate science topics and technology topics, each represented as a group of closely related keywords.

(2) ST&I topic identification Following the study of Rotolo et al. (2015) and Xu et al. (2021b), we evaluate each topic (i.e., community) in the science network and technology network using four indicators: novelty, fundamentality, growth, and impact, aiming to identify science, technology, and innovation (ST&I) topics that are potentially high-value for future university-industry collaboration.

Novelty Traditional novelty measures typically rely on the time of emergence of topics (Huang et al., 2021a; Tu & Seng, 2012), potentially overlooking the semantic distinction between topics. To address this, we propose a dual-dimensional novelty indicator that jointly evaluates temporal novelty and content novelty (Xu et al., 2021a), and uses the entropy weighting method to weigh and measure the novelty of the topic. The former captures how recently a topic has emerged, while the latter uses cosine similarity to quantify the semantic dissimilarity between the current topic and all others. This integrated approach offers a more nuanced and informative representation of novelty. The calculation formulas for the temporal novelty (*Time_Novelty*) and content novelty (*Content_Novelty*) of topic t_c are as follows:

$$Time\_Novelty = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{h}\sum_{k=1}^{h} Y_k \tag{3}$$

$$Content\_Novelty = 1/(\frac{1}{m-1}\sum_{j=1}^{m-1} Cos\left(t_c, t_j\right)) \tag{4}$$

$$Cos\left(t_c, t_j\right) = \sum_{p=1}^{q}\sum_{i=1}^{n} Cos(i, p) \tag{5}$$

where t_j is a arbitrary topic of t_c, n is the number of paper or patent keywords in the topic t_c, y_i is the occurrence time of the keyword i in this topic; h is the number of paper or patent keywords in the science or technology network, Y_k is the occurrence time of the keyword k in the network; m is the number of topics in the network, $Cos(t_c, t_j)$ denotes the cosine similarity between topic t_c and t_j; q is the number of keywords in the topic t_j, $Cos(i,p)$ denotes the cosine similarity between keyword i in topic t_c and keyword p in topic t_j.

Fundamentality Traditional studies often rely on a single centrality measure (e.g., degree or betweenness) to capture the structural importance of topics (Dotsika & Watkins, 2017), which may overlook their multifaceted roles in knowledge networks. To address this limitation, we integrate three complementary indicators (degree centrality: local connectivity, closeness centrality: global accessibility, and betweenness centrality: bridging capability) to provide a more comprehensive representation of a node's importance in the network (Zhang et al., 2021). Subsequently, the entropy weight method is used to calculate the fundamental significance of the topics, more accurately reflecting the topic's foundational role in supporting diverse innovation activities.

Growth The indicator growth reveals the evolving trends of topics over time (Bianchini et al., 2022). Unlike prior works that assess growth based on publication count (Wang, 2018), we evaluate the growth of topics by using the growth rate of keyword node connections within the network. This method reflects not only the external expansion of a topic but also its internal knowledge complexity and structural evolution. The growth is computed as follows:

$$Growth = \frac{1}{n}\sum_{i=1}^{n} r_{i,t,t\prime} \tag{6}$$

$$r_{i,t,t\prime} = \frac{Edge_{i,t} - Edge_{i,t\prime}}{t - t\prime} \tag{7}$$

Where $r_{i,t,t'}$ represents the growth rate of connection numbers for the keyword i from time t' to t, n is the total number of keywords in this topic; $Edge_{i,t'}$ denotes the number of connections for keyword i at time t' (in the previous network), while $Edge_{i,t}$ represents the number of connections for keyword i at time t (in the current network).

Impact Rather than relying on simple citation or degree measures (Huang et al., 2021a), we assess topic impact using a PageRank-based approach, where each keyword's influence is recursively calculated based on the significance of its neighbors. This method captures both direct and indirect relational influence of topics within the heterogeneous network. The calculation formulas are as follows:

$$Impact = \frac{1}{n}\sum_{i=1}^{n} PR_i \tag{8}$$

$$PR_i = (1-d) + d \times \sum_{j \in L(i)} \frac{1}{|Count(j)|} PR_j \tag{9}$$

where PR_i is the PageRank value of the keyword i within the topic, n is the number of keywords in this topic; d is the damping factor, where $1-d$ represents the probability of randomly jumping to other keywords, usually taken as d = 0.85 (Gleich, 2015); $L(i)$ represents the set of all keywords that have a linking relationship with keyword i, and $Count(j)$ represents the number of keywords that have a linking relationship with keyword j.

Each topic's four indicators are normalized and weighted using the entropy weight method, and the final comprehensive score is calculated. In both science and technology networks, the top-20 topics with the highest comprehensive scores are selected as ST&I

topics, resulting in a total of 40 representative topics used in the subsequent graph modeling and recommendation tasks.

Measurement of correlations between ST&I topics

This section aims to analyze the correlation between ST&I topics in the science network and technology network, with a comprehensive view of the role of knowledge relevance in university-industry collaboration. Considering that each topic includes multiple keywords, this study incorporates the importance of different keywords within a topic into the measurement model to enhance the weights of key keywords, thereby more accurately capturing the potential semantic relationships between ST&I topics. Specifically, the text similarity metric SimDoc proposed by Maheshwari et al. (2017) is improved by weighting the keyword vectors using the PageRank values of the keywords. The topic vector is then obtained through weighted summation, and the correlation between ST&I topics in the science network and technology network is calculated based on cosine similarity. The correlation between ST&I topic t_i and t_j is formulated as follows:

$$Correlation\left(t_i, t_j\right) = Cos(Vec(t_i), Vec(t_j)) \tag{10}$$

$$Vec(t_i) = \frac{1}{P_i} \sum\nolimits_{k=1}^{P_i} (PR\left(V_k\right)\prime \times X_k) \tag{11}$$

$$Vec(t_j) = \frac{1}{Q_j} \sum\nolimits_{h=1}^{Q_j} (PR\left(V_h\right)\prime \times Y_h) \tag{12}$$

where $Vec(t_i)$ and $Vec(t_j)$ represent the vector of the ST&I topic t_i and t_j. $Cos(Vec(t_i), Vec(t_j))$ is the cosine similarity between $Vec(t_i)$ and $Vec(t_j)$. P_i and Q_j denote the total number of keywords in the topic t_i and t_j, respectively. X_k and Y_h represent the vector of keyword V_k and V_h within the topic t_i and t_j, respectively. $PR(V_k)'$ and $PR(V_h)'$ represent the range-normalized PageRank value of the keyword V_k and V_h, respectively.

Topic-institution graph construction

To capture the rich semantics and structural relationships among topics and institutions, we construct a heterogeneous graph G, referred to as the topic-institution graph. Referring to Fig. 1, the topic-institution graph is described as follows:

$$G = G(\mathrm{V}\{T, U, E\}, \mathrm{R}\left\{E_1, E_2, E_3\right\}, \mathrm{W}\{W_1, W_2, W_3\}) \tag{13}$$

where $\mathrm{V}\{T, U, E\}$ is a set of node types; T, U, and E element of the set denote ST&I topics, universities, and enterprises entity, respectively. $\mathrm{R}\{E_1, E_2, E_3\}$ is the set of relationship types; E_1 indicates the semantic correlation between ST&I topics; E_2 denotes the co-occurrence relationship between ST&I topics and institutions (universities and enterprises), representing the degree of correlation between topics and institutions; E_3 denotes the co-authorship relationship between institutions (universities and enterprises). $\mathrm{W}\{W_1, W_2, W_3\}$ is the set of weight types; W_1, W_2, and W_3 denote the weight of E_1, E_2, and E_3, which is defined as follows.

$$W_1(t_i, t_j) = Correlation(t_i, t_j) \tag{14}$$

$$W_2(t_i, I_j) = \sum_{p=1}^{n} Cooccur(k_p, I_j) \tag{15}$$

$$W_3(I_i, I_j) = Coauthor(I_i, I_j) \tag{16}$$

where $W_1(t_i, t_j)$ is the weight between ST&I topic t_i and t_j, $Correlation(t_i, t_j)$ is calculated by the formulas (10). $W_2(t_i, I_j)$ is the weight between ST&I topic t_i and institution (university or enterprise) I_j, n is the number of keywords in topic t_i, $Cooccur(k_p, I_j)$ is the co-occurrence frequency between keyword k_p in topic t_i and institution I_j. $W_3(I_i, I_j)$ is the weight between institution (university or enterprise) I_i and I_j, $Coauthor(I_i, I_j)$ is the the number of collaborations between institution I_i and I_j.

By combining topic semantic similarity, topic-institution relevance, and institutional collaboration strength, this graph structure captures multilevel heterogeneous relationships necessary for downstream HAN-based representation learning and recommendation.

University-industry collaboration recommendation

This part aims to recommend university-enterprise partners using the heterogeneous graph attention network (HAN) model. This study helps to fully consider the relevance of the research topics and cooperation information between universities and enterprises, thereby improving the accuracy of collaboration recommendations.

Constructing the HAN model

After constructing the topic-institution graph, the heterogeneous graph attention network (HAN) model is used to predict the future topic-institution graph, and identify collaborative opportunities for university-industry collaboration. HAN is well-suited for heterogeneous graphs because it learns both node features and semantic information from multiple types of nodes (ST&I topics and institutions) and edges (topic-topic, topic-institution, and institution-institution). By incorporating attention mechanisms at both node and semantic levels, HAN selectively emphasizes important neighbors and meta-path semantics, thus improving the interpretability of the learned node embeddings (Wang et al., 2019; Yang et al., 2024).

Firstly, the dataset was divided into training and testing subsets (Cao et al., 2023). The training set was used to build the recommendation model, while the testing set was used to evaluate predictive performance based on widely used metrics, including AUC, Accuracy, Precision, Recall, and F1 Score (Huang et al., 2021b).

Secondly, node embeddings were initialized using the node2vec algorithm (Grover & Leskovec, 2016), which preserves both local and global structural properties of the graph. These pre-trained embeddings served as input features for all nodes, providing a stable and structure-aware initialization for the HAN model.

Next, the HAN architecture was configured with two graph attention layers, each comprising (1) a node-level attention module that learns the importance of a node's neighbors under a given meta-path, and (2) a semantic-level attention module that aggregates node information across different meta-paths. The LeakyReLU function was used within the attention modules, while ReLU was applied elsewhere.

To effectively capture heterogeneous semantics, five representative meta-paths were designed to encode the structural and topical contexts relevant to university-industry collaboration:

U–T–E (University–ST&I Topic–Enterprise): connecting universities and enterprises through shared research topics;
T–U–E (ST&I Topic–University–Enterprise): topic-driven collaboration paths involving universities;
T–E–U (ST&I Topic–Enterprise–University): enterprise-initiated topic engagement pathways;
U–E–E (University–Enterprise–Enterprise): reflecting enterprise cluster engagement facilitated by universities;
U–U–E (University–University–Enterprise): representing multi-university cooperation linked to enterprise collaboration.

Then, during model training, the HAN iteratively updated node representations through forward and backward propagation: (1) The node-level attention module computed the relative importance of neighbors under each meta-path, generating meta-path-specific embeddings. (2) The semantic-level attention module aggregated these embeddings according to learned meta-path weights. (3) The output node representations were obtained and evaluated using a cross-entropy loss function. All parameters, including attention coefficients and fully connected layer weights, were updated via backpropagation until convergence.

Finally, model hyperparameters, such as learning rate, regularization strength, attention vector dimensionality, and the number of attention heads, were iteratively tuned. The configuration yielding the best performance on the test set was selected as the final model for UIC recommendation.

University-industry collaboration prediction based on the HAN model

The collaboration opportunities between universities and enterprises are predicted using the well-trained heterogeneous graph attention network model. First, the HAN model is used to generate vector representations for the university and enterprise nodes in the test set data. Cosine similarity is then applied to calculate the edge weights between the nodes, predicting the future topic-institution graph. Next, potential collaboration opportunities between universities and enterprises are identified in the future graph based on collaboration opportunity scores (edge weights), thereby recommending suitable collaboration partners for UIC. Finally, the top-K university-enterprise pairs $\left\{\left(u_1,e_1\right),\left(u_2,e_2\right),\cdots\left(u_K,e_K\right)\right\}$ with high collaboration opportunity scores are identified.

Empirical study

Artificial Intelligence (AI) is a multidisciplinary field that involves numerous universities and enterprises, forming a diverse and heterogeneous innovation network (Song, 2021). UIC exhibits higher quantity and quality, providing ample data support for the empirical analysis conducted in this study. Therefore, this study analyzed the university-industry collaboration recommendation in-depth in the AI domain to verify the effectiveness of the proposed method.

Identification of ST&I topics

The dataset used in this study comprises AI papers and patents retrieved from the Web of Science (WoS) and the Derwent Innovation Index (DII), respectively. On the one hand, following the study of Liu et al. (2021), the paper retrieval framework was established based on the core keywords identified in the AI field. On the other hand, referencing the study by Zhou et al. (2019), a patent retrieval framework was constructed from the aspects of technical support, infrastructure, algorithms and applications. Finally, 234,011 papers and 221,448 patents published between 2016 and 2024 were retrieved.

Then, following the study of Wang et al. (2021), the ITGInsight text mining software was used to extract and clean keywords and institution information by its NLP function and the term clumping process. As a result, 6000 science keywords and 6000 technology keywords were selected as keyword dataset, with a minimum word frequency of 2 and a maximum PC-Value. Additionally, 13,886 universities and 66,207 enterprises were extracted as the institution collection based on the publishing institutions of papers and patents, respectively.

Next, the science network and technology network were constructed, where nodes represent keywords and edges denote co-occurrence relationships between keyword pairs within the same document (paper or patent). To enhance the quality of the networks and ensure the robustness of subsequent community detection, we applied a pruning strategy based on semantic similarity (Zheng et la., 2025). The pruning thresholds for the science and technology networks were progressively set to 1.0, 0.95, 0.90, 0.85, 0.80, 0.75, 0.70, 0.65, 0.60, 0.55, and 0.5, and the optimal community structure was identified using modularity as the evaluation criterion. A higher modularity value indicates denser intra-community connections and sparser inter-community links.

As shown in Fig. 2, the modularity values of both the science and technology networks increase as the pruning threshold decreases, indicating clearer community structures with the removal of weak edges. Since an optimal modularity typically falls within the range of 0.3–0.7 (Newman, 2004), when the threshold is 0.60, the modularity values reach 0.6894 and 0.6683 for the science and technology networks, respectively, indicating better community partitioning performance. Although further decreasing the threshold (e.g., to 0.55 or 0.50) continues to raise modularity, it also introduces excessive weak or noisy associations between keywords. This can lead to overfitting of the network structure, distort true semantic relationships, and ultimately reduce the reliability and interpretability of the detected communities. Therefore, 0.60 was determined as the optimal pruning threshold, and edges with weights below this value were removed, retaining only those with semantic similarity greater than 0.60.

Following this, the Leiden community detection algorithm was applied to the pruned networks, resulting in 62 science topics (communities) and 56 technology topics (communities) in the AI field. The PageRank value for each keyword in the network was calculated, and the keyword with the highest PageRank value in each community was selected as the topic label. Subsequently, four indicators—novelty, fundamentality, growth, and impact—were calculated for each topic. To ensure consistency across the indicators, min–max normalization was applied, and the entropy weighting method was used to compute a

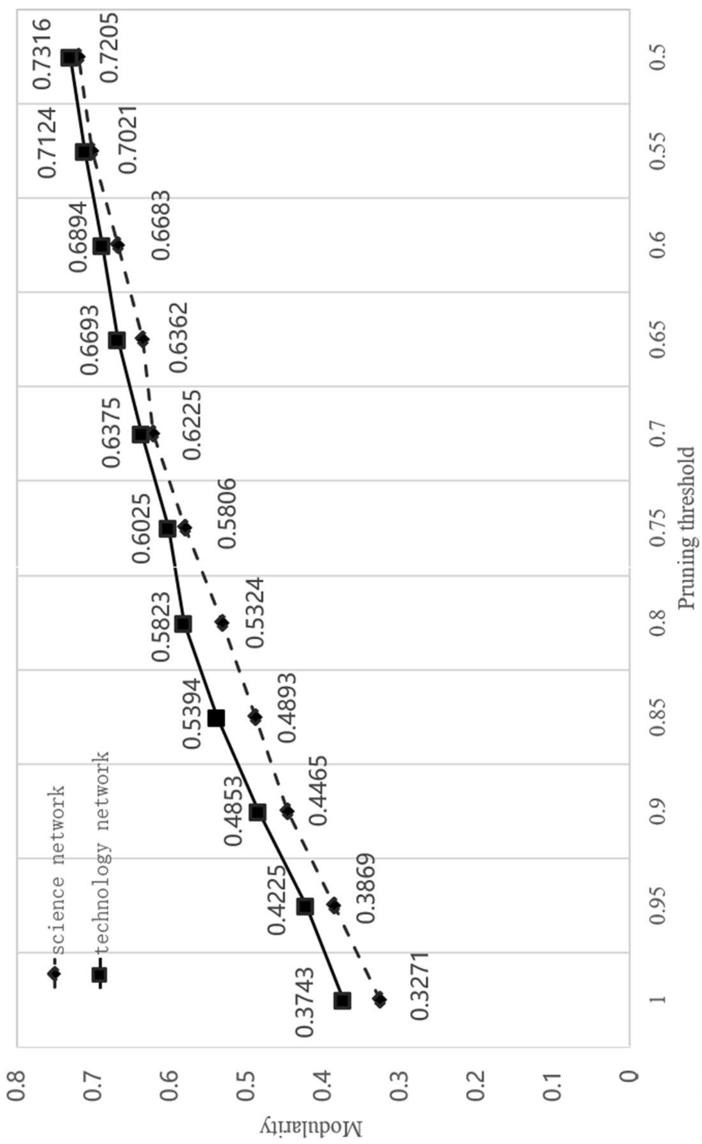

Fig. 2 Community detection results in science and technology networks

comprehensive score for each topic. The top 20 topics from the science network and technology network were selected as high-value ST&I topics, as presented in Table 1 and Table 2.

Table 1 shows that the ST&I topics in the science network primarily include core algorithms in the AI field (e.g., neural network, optimization algorithm, and classification method), key technological research areas(e.g., human–computer interaction, facial expression recognition, and visual image analysis), and core theories in AI interdisciplinary fields (e.g., financial data analysis, social network analysis, and numerical simulation). Table 2 shows that the ST&I topics in the technology network mainly cover key technologies in the AI field (e.g., computer vision, face recognition, and speech recognition), application scenarios (e.g., medical image, and smart device), and underlying infrastructures (e.g., computing device, and big data process).

From the distribution of topics in Table 1 and Table 2, common topics such as "neural network" "smart device" and "classification method" were identified. However, the ST&I topics in the science network and technology network also exhibit differences. Specifically, science topics are more focused on foundational theoretical research, such as human–computer interaction methods, neural network training algorithms, and model training algorithms, while technology topics emphasize applied research, including computer vision technology, computing devices, and big data processing technologies.

To further verify the necessity and unique contributions of the four indicators (novelty, fundamentality, growth, and impact), a two-step analysis was conducted. Firstly, a correlation matrix of the four indicators was calculated based on the indicator values of all

Table 1 Details of identified ST&I topics in the science network

No	ST&I topic in the science network	Novelty	Fundamentality	Growth	Impact	Total score
1	Neural network	0.3067	0.7625	0.7314	1.0000	0.6986
2	Human–computer interaction	0.4369	0.6352	1.0000	0.7371	0.6977
3	Facial expression recognition	0.6724	0.4352	0.8103	0.7862	0.6826
4	Engineering design	1.0000	0.5205	0.7436	0.2725	0.6338
5	Visual image analysis	0.4009	0.5098	0.6896	0.8302	0.6106
6	Adversarial learning	0.5874	0.5549	0.8910	0.3401	0.5860
7	Smart device	0.2762	0.7632	0.8325	0.3762	0.5465
8	Financial data analysis	0.5312	0.6210	0.9402	0.1279	0.5406
9	Statistical method	0.2596	0.8324	0.5283	0.5721	0.5378
10	Social network analysis	0.2362	0.6746	0.4359	0.8023	0.5359
11	Online learning	0.2715	0.7736	0.6459	0.4508	0.5231
12	Optimization algorithm	0.5563	0.4235	0.6530	0.3796	0.5018
13	Classification method	0.1159	0.7921	0.4431	0.5783	0.4719
14	Clustering algorithm	0.1220	0.6396	0.3725	0.7074	0.4570
15	Fuzzy mathematical model	0.0596	0.8781	0.8096	0.1638	0.4524
16	Numerical simulation	0.1328	0.8352	0.5826	0.2253	0.4243
17	Model training	0.0740	1.0000	0.2839	0.2736	0.3874
18	Random model	0.1204	0.8946	0.3368	0.2573	0.3841
19	Genetic algorithm	0.1448	0.4201	0.4769	0.4052	0.3569
20	Distance function	0.1032	0.6730	0.3953	0.2503	0.3416

Table 2 Details of identified ST&I topics in the technology network

No	ST&I topic in the technology network	Novelty	Fundamentality	Growth	Impact	Total score
1	Deep learning	0.8595	0.8979	0.6583	1.0000	0.8560
2	Computer vision	1.0000	0.7159	0.7433	0.8104	0.8243
3	Neural network	0.7269	0.8689	0.5864	0.8573	0.7592
4	Computing device	0.4281	1.0000	0.8892	0.7305	0.7480
5	Knowledge graph	0.7120	0.5489	0.6932	0.9020	0.7187
6	Face recognition	0.5201	0.5685	1.0000	0.7662	0.7097
7	Medical image	0.6543	0.6596	0.7913	0.6324	0.6828
8	Object detection	0.5092	0.7831	0.6709	0.6413	0.6449
9	Control system	0.4486	0.7101	0.7624	0.4009	0.5719
10	Smart device	0.6289	0.4726	0.6895	0.2538	0.5113
11	3D modeling	0.3496	0.6689	0.7687	0.2683	0.5029
12	Information processing system	0.3845	0.5829	0.6312	0.3089	0.4700
13	Speech recognition	0.4309	0.7301	0.3469	0.3859	0.4679
14	Predictive model	0.4310	0.4019	0.6206	0.3701	0.4543
15	Quality control	0.4303	0.4850	0.4993	0.3460	0.4378
16	Image process	0.2973	0.4702	0.4694	0.4978	0.4299
17	Big data process	0.1262	0.8701	0.4102	0.3420	0.4208
18	Feature extraction	0.2864	0.5498	0.3769	0.4362	0.4071
19	Classification method	0.3431	0.5856	0.2302	0.3520	0.3740
20	Threshold setting method	0.2913	0.5719	0.3164	0.2869	0.3607

topics, as shown in Table 3. It can be observed that all pairwise correlations are below 0.5, indicating that each indicator possesses distinct practical significance and captures a unique dimension of ST&I topic characteristics. Notably, the correlation between Novelty and Growth reaches 0.4852 (close to 0.5), which aligns with the logical inference that topics with higher novelty tend to attract more researcher attention, thereby fostering greater growth potential.

Secondly, an ablation study was performed to quantify the influence of each indicator on topic ranking outcomes. Specifically, for the 62 science topics and 56 technology topics, one indicator was removed at a time, and the remaining three indicators were recomputed using the entropy weighting method to generate new topic rankings. We then listed the rankings of the ST&I topics identified in Tables 1 and 2 under each ablation setting, as shown in Tables 4 and 5.

As illustrated in Tables 4 and 5, removing novelty, growth, or impact leads to the most significant fluctuations in topic rankings, while removing fundamentality results in relatively moderate variations. Moreover, eliminating any single indicator causes certain originally top-20 topics to be excluded from the ST&I topic group, confirming that each indicator captures irreplaceable and unique information. Collectively, the ablation experiment validates that all four indicators contribute nonredundant value to the evaluation framework, with novelty, growth, and impact playing more critical roles in identifying high-value ST&I topics.

Then, the correlation between ST&I topics in the science network and technology network was explored. Next, we constructed the 2016–2024 topic-institution graph by

Table 3 Correlation matrix

Indicators	Novelty	Fundamentality	Growth	Impact
Novelty	1.0000			
Fundamentality	−0.2671	1.0000		
Growth	0.4852	−0.0805	1.0000	
Impact	0.3781	0.0877	0.2293	1.0000

Table 4 Ablation experimental results in the science network

No	ST&I topic in the science network	Without novelty	Without fundamentality	Without growth	Without impact
1	Neural network	1	3	1	7
2	Human–computer interaction	2	2	3	3
3	Facial expression recognition	3	1	2	5
4	Engineering design	17	4	4	1
5	Visual image analysis	4	5	5	12
6	Adversarial learning	11	6	8	4
7	Smart device	5	13	17	6
8	Financial data analysis	15	8	19	2
9	Statistical method	7	14	7	14
10	Social network analysis	6	9	6	16
11	Online learning	8	15	9	9
12	Optimization algorithm	23	7	16	10
13	Classification method	9	17	13	18
14	Clustering algorithm	12	16	12	24
15	Fuzzy mathematical model	10	18	18	8
16	Numerical simulation	14	21	21	15
17	Model training	20	26	22	20
18	Random model	25	24	20	19
19	Genetic algorithm	26	20	25	25
20	Distance function	28	23	24	22

combining ST&I topic correlation relationships, co-occurrence relationships between topics and institutions, and co-authorship relationships between institutions. This graph includes two types of entities: ST&I topics and institutions (including universities and enterprises), and the descriptive statistics of the 2016–2024 topic-institution graph are given in Table 6.

Recommending university-industry collaboration

Given that the heterogeneous graph attention network (HAN) model will be used for recommending university-industry collaborations, the 2016–2024 topic-institution graph was divided into the training set and testing set to predict the future graph and identify potential collaboration opportunities for universities and enterprises. Firstly, the 2016–2022 graph was used as the training set to train the HAN model, and the 2023–2024 graph was used as

Table 5 Ablation experimental results in the technology network

No	ST&I topic in the technology network	Without novelty	Without fundamentality	Without growth	Without impact
1	Deep learning	2	2	1	2
2	Computer vision	5	1	2	1
3	Neural network	4	5	3	4
4	Computing device	1	8	5	3
5	Knowledge graph	6	3	4	9
6	Face recognition	3	4	8	8
7	Medical image	11	6	6	5
8	Object detection	7	10	7	11
9	Control system	19	12	9	14
10	Smart device	17	13	13	13
11	3D modeling	13	15	18	12
12	Information processing system	14	16	17	16
13	Speech recognition	15	20	12	17
14	Predictive model	18	14	23	21
15	Quality control	21	18	19	19
16	Image process	16	19	20	22
17	Big data process	12	21	14	18
18	Feature extraction	20	22	22	23
19	Classification method	24	23	16	25
20	Threshold setting method	23	25	24	24

the test set to evaluate the model's performance. The statistical information of the training set and test set is given in Table 7.

To provide stable and structure-preserving initial node features, we employed the node-2vec algorithm (Grover & Leskovec, 2016) to pre-train embeddings for all nodes, which capture both local and global structural information of the heterogeneous graph. These embeddings served as inputs to the HAN model, ensuring that the training process started from a meaningful representation space rather than random initialization.

During model training, the HAN effectively captured heterogeneous relationships by applying node-level and semantic-level attention mechanisms to extract and integrate important features from the topic–institution graph. The model architecture consisted of two attention layers, each with dual attention modules enabling message propagation across nodes and meta-paths.

To verify the robustness of the HAN model and identify the optimal hyperparameter configuration, we conducted a comprehensive parameter sensitivity analysis focusing on four key hyperparameters: learning rate, output channel dimension, number of attention heads, and training epochs. A grid search was implemented to systematically explore 36 parameter combinations, with each model evaluated using three performance metrics—AUC, Accuracy, and F1-Score. We then performed a statistical analysis of the average values of these evaluation metrics under each parameter setting and visualized the results using bar charts, as shown in Fig. 3.

As shown in Fig. 3, the HAN model's performance fluctuates to varying degrees across different hyperparameter settings. Among the tested parameters, the learning rate exerts the

Table 6 Descriptive statistics of the 2016–2024 topic-institution graph

Type		Number	Weights			
			Max	**Min**	**Mean**	**Stabdard deviation**
Node	ST&I topic	40	/			
	Institue-University	13,886	/			
	Institue-Enterprise	66,207	/			
Edge	ST&I topic-ST&I topic	1600	0.982	0.356	0.592	0.146
	University-University	95,268	1126	1	2.136	7.301
	Enterprise-Enterprise	15,965	395	1	1.332	2.350
	ST&I topic-University	67,452	38	1	3.384	3.486
	ST&I topic-Enterprise	49,360	161	1	4.137	6.331
	University-Enterprise	57,343	1865	1	2.352	12.624

Table 7 Statistical information of the experiment

Type		Training set	Test set
		Number	
Node	ST&I topic	40	40
	University	11,254	7845
	Enterprise	53,690	22,108
Edge	ST&I topic-ST&I topic	1600	1600
	University-University	58,230	50,246
	Enterprise-Enterprise	31,267	125,035
	ST&I topic-University	47,628	32,796
	ST&I topic-Enterprise	155,645	367,960
	University-Enterprise	34,261	31,650

most significant influence: as it increases from 0.001 to 0.1, the model's F1-Score, AUC, and Accuracy show distinct improvements, indicating high sensitivity to this hyperparameter. The output channel dimension and number of attention heads also have notable impacts on performance, with larger output channel dimensions (128) and more attention heads (8) consistently yielding better results. In contrast, the number of training epochs shows a relatively steady positive effect on performance, reflecting limited sensitivity once the model achieves adequate convergence. Based on the comprehensive analysis in Fig. 3, the optimal model is obtained when the learning rate is set to 0.1, the output channel dimension is 128, the number of attention heads is 8, and the training epochs is 100. Under this setup, the AUC, Accuracy, Precision, Recall, and F1-Score reach 0.7521, 0.8326, 0.5280, 0.6002, and 0.5618, respectively.

In addition, to enhance interpretability, we analyzed the learned node-level attention weights, semantic-level meta-path weights, and node embedding distribution.

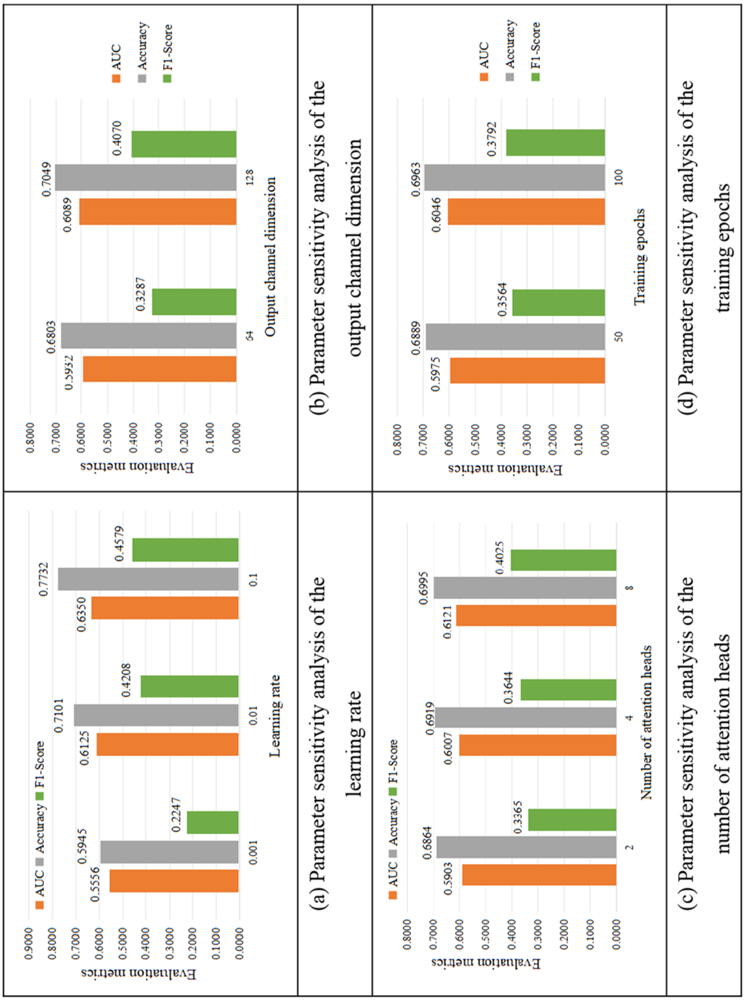

Fig. 3 Parameter sensitivity analysis results

Table 8 Statistical information of the node-level attention weights

Entity type pair	Max weight	Min weight	Mean weight
ST&I topic-ST&I topic	1	0	0.6903
University-University	1	0	0.7104
Enterprise-Enterprise	1	0	0.7143
ST&I topic-University	1	0	0.7201
ST&I topic-Enterprise	1	0	0.7184
University-Enterprise	1	0	0.7400

Table 9 Statistical information of the meta-path attention weights

Meta-path	Max weight	Min weight	Mean weight
University-ST&I topic-Enterprise	1	0	0.5923
ST&I topic-University-Enterprise	1	0	0.6779
ST&I topic-Enterprise-University	1	0	0.6511
University-Enterprise- Enterprise	1	0	0.6473
University-University- Enterprise	1	0	0.7186

(1) Node-level attention weights

The node-level attention mechanism captures the importance of neighboring nodes under a given meta-path. Table 8 reports the maximum, minimum, and mean attention weights across six key entity-type pairs.

These results demonstrate that heterogeneous neighbours (e.g., University-Enterprise links) received higher attention scores than homogeneous ones (e.g., University-University). This suggests that cross-entity collaboration signals, which often capture knowledge exchange pathways, are prioritized by the HAN model when generating predictions.

(2) Semantic-level attention weights

Semantic-level attention measures the relative contribution of different meta-paths. Table 9 presents the statistical distribution of attention weights across five representative meta-paths.

The highest semantic attention weights were assigned to the "University–University–Enterprise" (mean = 0.7186) and "ST&I Topic–University–Enterprise" (mean = 0.6779) meta-paths. This finding highlights the importance of academic collaboration history combined with topic-driven knowledge relevance in shaping potential university-enterprise partnerships.

(3) Node embedding distribution

To further interpret the learned representations, we randomly sampled 1,000 universities, 1,000 enterprises, and 40 ST&I topics (20 from the science network and 20 from the technology network) and visualized their embeddings using t-SNE (Maaten & Hinton, 2008).

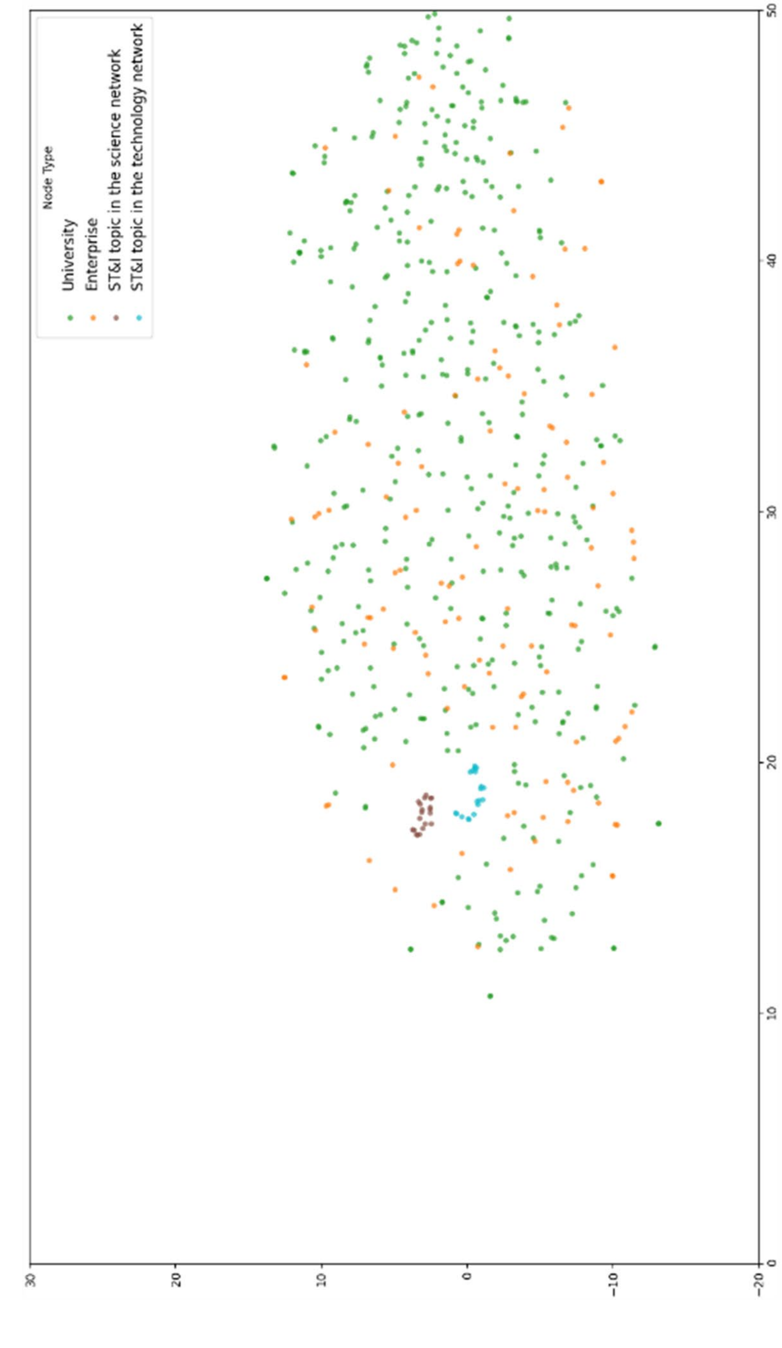

Fig. 4 T-SNE visualization of node embeddings for universities, enterprises, and ST&I topics

As shown in Fig. 4, nodes of different types exhibit distinguishable patterns in the embedding space. ST&I topics form relatively compact clusters, reflecting their semantic similarity as captured by the HAN model, whereas universities and enterprises are more broadly distributed yet show partial overlaps, indicating potential cross-type collaboration opportunities. This spatial proximity indicates that the HAN model effectively captures structural and semantic signals, thereby facilitating the identification of potential cross-type collaborations.

Next, the well-trained HAN model is used to predict the future graph and generate edge probabilities between universities and enterprises, representing the collaboration opportunity scores. Then, a university-enterprise collaboration list containing information on universities, enterprises, and collaboration scores was generated. This list is compared with the training dataset to filter out university-enterprise pairs that have not previously collaborated. Finally, the pairs are sorted based on collaboration scores, with top-ranked pairs being selected as candidates for university-enterprise collaboration recommendations. Here, take the top-10 university-enterprise pairs with the highest cooperation opportunity scores as an example, as shown in Table 10.

Furthermore, this study applied the advanced large language model ChatGPT-4o as an auxiliary tool for qualitative analysis to validate and interpret the recommended university–industry collaboration (UIC) pairs. The selection of ChatGPT-4o was based on the following considerations: (1) Multimodal analysis capability: ChatGPT-4, developed by OpenAI, processes both text and images and has been successfully applied in biomedical research (Saraiva et al., 2025; Turan et al., 2025), visual image analysis (Elyoseph et al., 2024), physics (Polverini et al., 2025), and social sciences (Polverini et al., 2025) for knowledge discovery and decision support. (2) Enhanced reasoning performance: ChatGPT-4o, as the latest release, provides faster response speed, improved reasoning ability, and superior multimodal integration compared to previous versions (Wei, 2024). (3) Suitability for complex analytical tasks: Its strong semantic understanding and natural language generation enable the production of logically consistent explanations, supporting nuanced decision-making in complex collaborative contexts (O'Leary, 2024). Therefore, ChatGPT-4o was selected as the qualitative analysis tool in this study.

To ensure analytical rigor, the prompt design and validation process for ChatGPT-4o followed a structured procedure: (1) Prompt design principles: Initial prompts were constructed to instruct ChatGPT-4o to assess the potential for future collaboration between specific university–enterprise pairs, explicitly considering three key factors: research alignment, geographical proximity, and prior cooperation history. (2) Iterative refinement: Prompts were iteratively refined through multiple testing rounds to maximize clarity, relevance, and explanatory depth in ChatGPT-4o's responses. (3) Final prompt selection: The final prompt format was chosen based on its ability to produce logically consistent, evidence-based, and contextually appropriate explanations. In this study, an exemplary prompt and its corresponding ChatGPT-4o response are presented in Fig. 5, directly captured from the interaction interface to ensure reproducibility and transparency of the analytical process.

In addition, this study compares the predicted UIC recommendations with real-world data by retrieving relevant publications, patents, and online information from 2023 to 2025. Table 11 shows the collaboration status of the top-5 university-enterprise pairs, demonstrating a high degree of consistency between the predicted results and actual data.

Table 10 Top-10 university-enterprise pairs

No	University	Enterprise	Collaboration opportunity score
1	National Yunlin University of science and technology	JBAO Technology Ltd	0.8946
2	Nanyang technological university	Yissum Research Development Co., Ltd	0.8923
3	Seoul national university	PAEAN Biotechnology Inc	0.8910
4	Aalen university	PVA TePla	0.8854
5	Institute of genetics & developmental biology, CAS	AgResearch Ltd	0.8823
6	Karlsruhe University of Applied Sciences	Physik Instrumente (PI) GmbH & Co. KG	0.8809
7	Sikkim manipal university	McKesson Corporation	0.8796
8	Polytechnic of Rijeka	Geal Pharma Co., Ltd	0.8774
9	SJB institute of technology	Sch Co., Ltd	0.8759
10	Centre for advanced studies	SAI.TECH Global Corporation	0.8742

Prompt

Please assess the potential for future collaboration between **National Yunlin University of Science and Technology and JBAO Technology Ltd**. Your response should consider:

1. Alignment in **research directions or technological focus** (based on typical research outputs of each party);

2**. Geographical proximity or regional innovation policies** that might promote collaboration;

3.**Possible future cooperation fields**;

4.**Risk points** that should be considered in cooperation;

5. Provide a reasoned, structured explanation (around 200-300 words) to support your assessment.

Assessment of Future Collaboration Potential between National Yunlin University of Science and Technology (YunTech) and JBAO Technology Ltd.

1. Research Alignment:

YunTech is recognized for its strengths in materials science, electrical engineering, and solar cell technologies, with numerous publications in thin-film deposition, semiconductor packaging, and photovoltaic device optimization. JBAO Technology Ltd. specializes in precision electroplating, microfabrication, and semiconductor packaging, with multiple patents related to copper electrodeposition and advanced materials integration. Notably, the two parties have co-authored research papers such as "The Impact of Substrate Temperature on the Adhesion Strength of Electroplated Copper on an Al-Doped ZnO/Si System", indicating strong alignment in applied research areas.

2. Geographical and Policy Support:

Located in Taiwan (YunTech) and mainland China (JBAO, Suzhou/Hsinchu), the two are regionally proximate within the Greater China innovation ecosystem. Taiwan's emphasis on cross-Strait industrial-academic collaboration and China's support for advanced manufacturing and green energy offer policy incentives to deepen such partnerships.

3. Potential Cooperation Fields:

Promising areas include:

- Next-generation photovoltaic materials and devices
- Advanced packaging and adhesion technologies
- Industrial-scale electrodeposition processes
- Joint patent development and commercialization

4. Risk Considerations:

- Cross-Strait political tensions may affect long-term collaboration continuity
- IP ownership and technology transfer mechanisms require careful legal structuring
- Potential asymmetry in commercialization interests

5. Conclusion:

Given their complementary technological expertise, successful history of joint publications, and supportive regional policy environments, YunTech and JBAO are well-positioned to deepen cooperation in emerging materials and clean energy technologies. However, clear governance structures and risk mitigation mechanisms—particularly around IP and political factors—are essential to ensure sustainable collaboration.

Fig. 5 Depiction of an exemplary prompt and the corresponding responses by ChatGPT-4o

Table 11 Relevant documentary proof of top-5 university-enterprise pairs

No	Relevant documentary proof
1	National Yunlin University of Science and Technology (YunTech) and JBAO Technology Ltd. co-authored publications on photovoltaic materials and thin-film engineering (Shan et al., 2023; Tseng et al., 2024). They focused on enhancing the adhesion strength and reliability of transparent conductive films for solar applications, and investigated electrodeposited copper films for heterojunction solar cells, thereby improving interfacial quality and mechanical stability
2	Nanyang Technological University (NTU) and Yissum Research Development Co., Ltd. jointly applied for patents on fluorine-free superhydrophobic coatings and electroactive bioadhesive compositions (Reches et al., 2025; Steele & Mandler, 2024). These technologies address environmentally friendly surface protection and biomedical tissue fixation, indicating active interdisciplinary collaboration
3	Seoul National University (SNU) and PAEAN Biotechnology Inc. collaborated on mitochondria-based therapies, including PN-101 for idiopathic inflammatory myopathy and mitochondrial pharmaceutical compositions for hereditary hearing impairment (Choi et al., 2025; Kim et al., 2025). Their joint efforts include peer-reviewed publications and co-filed patents demonstrating translational biomedical innovation
4	Aalen University and PVA TePla co-filed patents for advanced fiber-reinforced composite manufacturing techniques, introducing improved structural accuracy and residual stress control (Mindermann et al., 2025a, 2025b). These developments highlight practical cooperation in high-precision composite component fabrication
5	Institute of Genetics and Developmental Biology, CAS, and AgResearch Ltd. jointly published studies and patents on agricultural sustainability, including N_2O emission reduction via straw-based hydrogels and analyses of global agricultural carbon emission intensity, offering climate change mitigation solutions (Bai et al., 2024; Wang et al., 2024b)

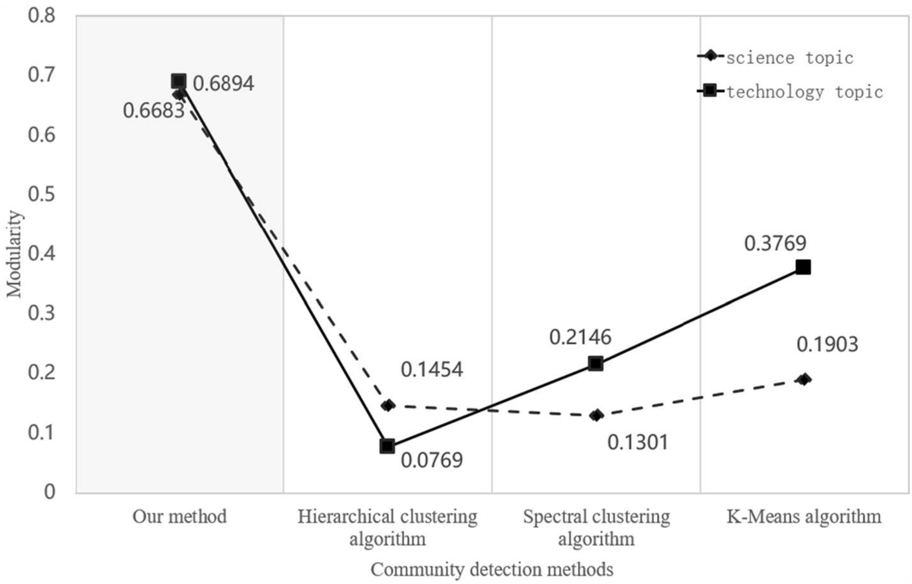

Fig. 6 Comparison of topic identification results

Validation

Verification of topic identification

In order to quantitatively validate the effectiveness of the Leiden community detection algorithm for topic identification in this study, we compared this method with hierarchical clustering, spectral clustering, and K-Means algorithms on both science network and technology network using modularity as the evaluation metric. Modularity is an indicator used to measure the effectiveness of community detection in networks. The higher the modularity, the more tightly connected the nodes within communities and the sparser the connections between communities, indicating better community detection performance (Traag et al., 2019). Therefore, we calculated the modularity for the clustering results obtained by the four methods to evaluate the effectiveness of the science topic and technology topic identification. The results are shown in Fig. 6.

It can be seen that the method used in this study outperforms the other three traditional topic identification algorithms in terms of modularity for scientific topic identification (0.6683) and technological topic identification (0.6894).

Verification of the trained recommendation model

In order to quantitatively evaluate the advantages of the recommendation method based on the heterogeneous graph attention network model, this section employs ten other recommendation algorithms to recommend university-industry collaborations, respectively, where AUC, Accuracy, Precision, Recall, and F1-Score were selected as evaluation metrics.

We categorized the baseline models into three groups.

(1) Traditional recommender system methods:

Content-based method (Huang et al., 2021b): recommends partners based on textual similarity.

CF (collaborative filtering) method (Molaei et al., 2021): relies on historical collaboration records.

(2) Classical graph embedding algorithms:

DeepWalk (Chanpuriya et al., 2021): learns network structure by simulating random walks on the graph, generating node embeddings that capture relational information.

Node2Vec (Grover & Leskovec, 2016): improves the Deepwalk method with a more flexible strategy of random walk.

GCN (graph convolutional network) model (Zhang et al., 2019b): captures node features and local structural information in homogeneous graphs.

(3) Advanced heterogeneous graph learning models:

HetGNN (heterogeneous graph neural network) model (Zhang et al., 2019a): learns structural and attribute information across multiple node and edge types.

GNN4TS (graph neural networks for time series) method (Jin et al., 2024): adopts a flexible graph neural network architecture capable of capturing complex heterogeneous dependencies and feature interactions.

HGCL (heterogeneous graph contrastive learning) algorithm (Chen et al., 2023): improves representation robustness by maximizing consistency under different graph views through contrastive learning strategies.

HHGT (hierarchical heterogeneous graph transformer) method (Zhu et al., 2025): employs a transformer-based architecture that hierarchically aggregates messages across metapaths and heterogeneous node types.

OmniSage (Badrinath et al., 2025): designs for large-scale multi-entity graphs, combining global and local context to achieve scalable and high-quality representations.

For each experiment, the 2016–2022 topic-institution graph was the training set, and the 2023–2024 graph was the testing set. To ensure the robustness of the model results, we conducted five independent experiments and took the average values of each metric as the final evaluation results. The comparison results are shown in Table 12.

From Table 12, it can be observed that our method achieved better results compared to the 10 baseline methods across all evaluation metrics, indicating its effectiveness and robustness in recommending university-industry collaborations. We made the following interpretations:

(1) Compared with traditional recommendation system models (Content-based and CF), our method achieved substantially higher Precision (0.5280 vs. 0.3653 and 0.4028) and Recall (0.6002 vs. 0.2546 and 0.3896), demonstrating that relying solely on textual features or historical collaborations is insufficient to capture the complex and evolving patterns of university-industry linkages.
(2) Compared with classical homogeneous graph embedding models (DeepWalk, Node2Vec, and GCN), which only exploit topological structures, our approach achieved higher AUC (0.7521 vs. 0.6259, 0.6992, and 0.3803) and F1-Score (0.5618 vs. 0.4540, 0.2279, and 0.2931), highlighting the necessity of incorporating heterogeneous node and edge attributes.
(3) Even against advanced heterogeneous graph learning models (HetGNN, GNN4TS, HGCL, HHGT, and OmniSage), which are designed to handle heterogeneous information, our method still shows superior performance (e.g., F1-Score = 0.5618 vs. 0.4642, 0.4784, 0.4784, 0.4889, and 0.5061). This demonstrates the advantage of our tailored

Table 12 The validation results of recommendation methods

Method	AUC	Accuracy	Precision	Recall	F1-Score
Our proposed method	**0.7521**	**0.8326**	**0.5280**	**0.6002**	**0.5618**
Content-based method	0.7043	0.8025	0.3653	0.2546	0.3001
CF	0.4390	0.4595	0.4028	0.3896	0.3961
Deepwalk	0.6259	0.8310	0.4156	0.5003	0.4540
Node2Vec	0.6992	0.8306	0.2091	0.2503	0.2279
GCN	0.3803	0.4310	0.3473	0.2536	0.2931
HetGNN	0.7296	0.8309	0.4308	0.5032	0.4642
GNN4TS	0.5680	0.7126	0.4580	0.5006	0.4784
HGCL	0.5678	0.8160	0.4579	0.5009	0.4784
HHGT	0.6472	0.8105	0.4698	0.5096	0.4889
OmniSage	0.7399	0.8256	0.5107	0.5015	0.5061

heterogeneous graph attention mechanism, which better captures the multi-entity, multi-relation characteristics of university-industry collaboration networks, yielding more balanced and robust recommendations.

Discussion and conclusions

This study proposed a university-industry collaboration (UIC) recommendation method using semantic analysis and graph analysis, effectively capturing rich information about multi-type entities, semantic associations, along with collaboration and co-occurrence relationships within the topic-institution graph. By integrating the SciBERT model with the Leiden community detection algorithm, topics within the science network and technology network were identified, improving the accuracy of semantic representation and topic extraction. Multi-dimensional indicators were used to identify promising ST&I topics, and the semantic correlations between these topics were calculated. Subsequently, a topic-institution graph was constructed, incorporating two types of entities (ST&I topics and institutions), along with complex relationships, including co-authorship connections between institutions, semantic correlations between ST&I topics, and co-occurrence relationships between institutions and ST&I topics. Finally, a heterogeneous graph attention network (HAN) model was employed to predict the future graph and identify potential collaboration opportunities between universities and enterprises, enhancing the precision of UIC recommendations. Further, ChatGPT-4o was utilized to provide explanations for the recommended university-industry collaborations and to validate the accuracy of the recommendation results, leveraging its strong reasoning and explanatory capabilities. Our framework provides technical intelligence on UIC, which has fully considered the research topics and cooperation information of institutions.

Theoretical contribution

The theoretical contributions of this study are as follows.

First, it extends the theory of university–industry collaborative innovation by revealing how semantic associations among scientific and technological knowledge interact with institutional collaboration structures to shape the formation and evolution of UIC networks. The proposed topic–institution graph, integrating ST&I topics, universities, enterprises, and their semantic and relational linkages, provides a novel analytical framework for representing and explaining heterogeneous collaboration mechanisms.

Second, it enriches knowledge transfer and innovation network theory by introducing a multi-stage topic identification approach that combines semantic representation learning (SciBERT), structural community detection (Leiden algorithm), and multidimensional indicators (novelty, fundamentality, growth, and impact). This approach deepens theoretical insight into how emerging and general-purpose knowledge domains can be systematically identified and linked to collaboration opportunities, thus clarifying the mechanisms of knowledge convergence and collaborative path formation.

Third, it contributes to innovation ecosystem governance theory by proposing a data-driven collaboration recommendation framework based on HAN and explainable AI. By integrating semantic, structural, and historical collaboration features, the framework offers a theoretical foundation for optimizing innovation resource allocation and enhancing knowledge flow efficiency, providing new insights into how scientific and industrial innovation can be deeply integrated within complex innovation ecosystems.

Practical implications

This study offers several practical implications for stakeholders involved in university-industry collaboration.

First, for universities, the proposed ST&I topic identification framework provides a data-driven approach to continuously monitor emerging and high-impact research areas. Such an approach enables universities to strategically align their research fields with industrial technological trajectories, thereby improving the efficiency of knowledge transfer and facilitating the establishment of sustainable and mutually beneficial partnerships with enterprises.

Second, for enterprises, the integration of semantic analysis and heterogeneous network-based recommendation methods allows firms to identify academic partners specializing in frontier and high-potential research topics at an early stage. This capability enhances access to cutting-edge scientific knowledge, reduces technological uncertainty, and supports evidence-based adjustments to R&D and innovation strategies, thereby improving market responsiveness and innovation agility.

Third, for policy-makers and innovation agencies, the use of multi-dimensional data sources (papers and patents) combined with an HAN-based predictive framework offers a robust tool for monitoring innovation trends and formulating targeted policy interventions. This supports more efficient resource allocation, the design of innovation programs, and the development of regional innovation clusters, ultimately promoting effective innovation ecosystem governance and accelerating the transformation of scientific and technological achievements.

Limitations and future works

Several limitations and future directions of research are summarized as follows. First, the innovation outcomes of UIC not only include papers and patents but also extend to projects, products, and corporate performance. Future research should use AI techniques to integrate multi-source heterogeneous data, such as products and corporate annual reports. Second, this study only includes data from the AI field between 2016 and 2024, broader empirical analyses should be conducted around future industries and strategic emerging industries to enhance the reliability and practicality of the methods. Additionally, future research will explore dynamic graph embedding methods, such as DySAT and DyHNet, to better capture the collaborative changes between universities and enterprises over time.

Acknowledgements This work was supported by the National Nature Science Foundation of China Funds [Grant No. 72274013 and No. 72371026].

Funding Funding was provided by National Natural Science Foundation of China,72274013,Lu Huang,72371026,Lu Huang

References

Albats, E., Alexander, A. T., & Cunningham, J. A. (2022). Traditional, virtual, and digital intermediaries in university-industry collaboration: Exploring institutional logics and bounded rationality. *Technological Forecasting and Social Change, 177*, 121470.

Badrinath, A., Yang, A., Rajesh, K., Agarwal, P., Yang, J., Chen, H., ... & Rosenberg, C (2025). OmniSage: Large Scale, Multi-Entity Heterogeneous Graph Representation Learning. *arXiv preprint* arXiv:2504.17811

Bae, S. H., Halperin, D., West, J. D., Rosvall, M., & Howe, B. (2017). Scalable and efficient flow-based community detection for large-scale graph analysis. *ACM Transactions on Knowledge Discovery from Data*, 11(3), 1-30.

Bai, Z., Zhang, N., Winiwarter, W., Luo, J., Chang, J., Smith, P., & Ma, L. (2024). Decline in carbon emission intensity of global agriculture has stagnated recently. *Proceedings of the National Academy of Sciences, 121*(34), e2317725121.

Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A pretrained language model for scientific text. *arXiv preprint* arXiv:1903.10676.

Bertram, N., Dunkel, J., & Hermoso, R. (2023). I am all EARS: Using open data and knowledge graph embeddings for music recommendations. *Expert Systems with Applications, 229*, 120347.

Bianchini, S., Müller, M., & Pelletier, P. (2022). Artificial intelligence in science: An emerging general method of invention. *Research Policy, 51*(10), 104604.

Cao, X., Chen, X., Huang, L., et al. (2023). Detecting technological recombination using semantic analysis and dynamic network analysis. *Scientometrics, 129*(11), 7385–7416.

Chang, S. H. (2018). A pilot study on the connection between scientific fields and patent classification systems. *Scientometrics, 114*(3), 951–970.

Chanpuriya, S., Musco, C., Sotiropoulos, K., & Tsourakakis, C. (2021). Deepwalking backwards: from embeddings back to graphs. In *International conference on machine learning* (pp. 1473–1483). PMLR.

Chen, M., Huang, C., Xia, L., Wei, W., Xu, Y., & Luo, R. (2023). Heterogeneous graph contrastive learning for recommendation. In *Proceedings of the sixteenth ACM international conference on web search and data mining* (pp. 544-552).

Chen, H., Jin, Q., Wang, X., & Xiong, F. (2022a). Profling academic-industrial collaborations in bibliometric-enhanced topic networks: A case study on digitalization research. *Technological Forecasting and Social Change, 175*, 121402.

Chen, H., Song, X., Jin, Q., & Wang, X. (2022b). Network dynamics in university-industry collaboration: A collaboration-knowledge dual-layer network perspective. *Scientometrics, 127*(11), 6637–6660.

Choi, B. Y., Kim, B. J., Kim, Y., Mira, P. A. (2025). Pharmaceutical composition containing mitochondria as active ingredient for preventing or treating hereditary hearing impairment (WO2025095735). *World Intellectual Property Organization*.

Chung, J., Ko, N., & Yoon, J. (2021). Inventor group identification approach for selecting university-industry collaboration partners. *Technological Forecasting and Social Change, 171*, 120988.

Cui, S., & Zhu, X. (2024). The information propagation mechanism of individual heterogeneous adoption behavior under the heterogeneous network. *Frontiers in Physics, 12*, 1404464.

Ding, Y. (2022). Correlation analysis model of social capital and innovation performance based on knowledge mapping. *Computational Intelligence and Neuroscience, 2022*(1), 2138200.

Dinh, T. N., Pham, P., Nguyen, G. L., & Vo, B. (2024). Enhancing local citation recommendation with recurrent highway networks and SciBERT-based embedding. *Expert Systems with Applications, 243*, 122911.

Dotsika, F., & Watkins, A. (2017). Identifying potentially disruptive trends by means of keyword network analysis. *Technological Forecasting and Social Change, 119*, 114–127.

Elyoseph, Z., Refoua, E., Asraf, K., Lvovsky, M., Shimoni, Y., & Hadar-Shoval, D. (2024). Capacity of generative AI to interpret human emotions from visual and textual data: Pilot evaluation study. *JMIR Mental Health, 11*, e54369.

Foti, L., Warwick, L., Lyons, E., Dhaliwal, S., & Alcorn, M. (2023). Knowledge transfer and innovation: Universities as catalysts for sustainable decision making in industry. *Sustainability, 15*(14), 11175.

Gallagher, C., Lusher, D., Koskinen, J., Roden, B., Wang, P., Gosling, A., & Simpson, G. (2023). Network patterns of university-industry collaboration: A case study of the chemical sciences in Australia. *Scientometrics, 128*(8), 4559–4588.

Gibson, E., Daim, T. U., & Dabic, M. (2019). Evaluating university industry collaborative research centers. *Technological Forecasting and Social Change, 146*, 181–202.

Gleich, D. F. (2015). PageRank beyond the web. *SIAM Review, 57*(3), 321–363.

Grover, A., & Leskovec, J. (2016). Node2vec: Scalable feature learning for networks. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 855–864.

Guan, J., & Liu, N. (2016). Exploitative and exploratory innovations in knowledge network and collaboration network: A patent analysis in the technological feld of nano-energy. *Research Policy, 45*, 97–112.

Guan, J., Yan, Y., & Zhang, J. J. (2017). The impact of collaboration and knowledge networks on citations. *Journal of Informetrics, 11*, 407–422.

Huang, L., Chen, X., Ni, X., Liu, J., Cao, X., & Wang, C. (2021a). Tracking the dynamics of co-word networks for emerging topic identification. *Technological Forecasting and Social Change, 170*, 120944.

Huang, L., Chen, X., Zhang, Y., Wang, C., Cao, X., & Liu, J. (2022a). Identification of topic evolution: network analytics with piecewise linear representation and word embedding. *Scientometrics, 127*(9), 5353–5383.

Huang, L., Chen, X., Zhang, Y., Zhu, Y., Li, S., & Ni, X. (2021b). Dynamic network analytics for recommending scientific collaborators. *Scientometrics, 126*(11), 8789–8814.

Huang, Z., Tang, Y., & Chen, Y. (2022b). A graph neural network-based node classification model on class-imbalanced graph data. *Knowledge-Based Systems, 244*, 108538.

Jin, M., Koh, H. Y., Wen, Q., Zambon, D., Alippi, C., Webb, G. I., & Pan, S. (2024). A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2024.3443141

Kang, J., Lee, J., Jang, D., & Park, S. (2019). A methodology of partner selection for sustainable industry-university cooperation based on LDA topic model. *Sustainability, 11*(12), 3478.

Kim, H. J., San Kim, T., & Sohn, S. Y. (2020). Recommendation of startups as technology cooperation candidates from the perspectives of similarity and potential: A deep learning approach. *Decision Support Systems, 130*, 113229.

Kim, J. Y., Kang, Y. C., Kim, M. J., Kim, S. U., Kang, H. R., Yeo, J. S., & Lee, E. Y. (2025). Mitochondrial transplantation as a novel therapeutic approach in idiopathic inflammatory myopathy. *Annals of the Rheumatic Diseases, 84*(4), 609–619.

Kodua-Ntim, K. (2023). Narrative review on open access institutional repositories and knowledge sharing in South Africa. *Journal of the Association for Information Science and Technology, 74*(9), 1118–1123.

Lee, H. F., & Miozzo, M. (2024). Beyond complementarity and substitutability? understanding relational governance and formal contracts in university-industry collaborations for innovation. *Technovation, 138*, 103100.

Li, G., Siddharth, L., & Luo, J. (2023). Embedding knowledge graph of patent metadata to measure knowledge proximity. *Journal of the Association for Information Science and Technology, 74*(4), 476–490.

Li, T., & Zhou, X. (2022). Research on the mechanism of government–industry–university–institute collaborative innovation in green technology based on game–based cellular automata. *International Journal of Environmental Research and Public Health, 19*(5), 3046.

Lian, C., & Wang, J. (2024). Multi-actor cooperation for emergency supply support: A simulation of behavior diffusion based on social networks. *Natural Hazards, 120*(2), 1241–1262.

Libaers, D. (2017). Time allocations across collaborations of academic scientists and their impact on efforts to commercialize novel technologies: Is more always better? *R&D ManagemeNt, 47*(2), 180–197.

Littleton, C., Townsin, L., & Beilby, J. (2023). The motivations of stakeholders when developing university industry collaborations in an Australian university: Three case studies. *Journal of Higher Education Policy and Management, 45*(5), 481–494.

Liu, N., Shapira, P., & Yue, X. (2021). Tracking developments in artificial intelligence research: Constructing and applying a new search strategy. *Scientometrics, 126*(4), 3153–3192.

Lu, Y., Yuan, M., Liu, J., & Chen, M. (2023). Research on semantic representation and citation recommendation of scientific papers with multiple semantics fusion. *Scientometrics, 128*(2), 1367–1393.

Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research, 9*, 2579–2605.

Maheshwari, G., Trivedi, P., Sahijwani, H., Jha, K., Dasgupta, S., & Lehmann, J. (2017). Simdoc: topic sequence alignment based document similarity framework. In *Proceedings of the 9th Knowledge Capture Conference* (pp. 1-8).

Mao, C., Yu, X., Zhou, Q., Harms, R., & Fang, G. (2020). Knowledge growth in university-industry innovation networks–results from a simulation study. *Technological Forecasting and Social Change, 151*, 119746.

Mei, G., Pan, L., & Liu, S. (2022). Heterogeneous graph embedding by aggregating meta-path and metastructure through attention mechanism. *Neurocomputing, 468*, 276–285.

Mindermann, P., Müllner, R., Hoffrogge, P. (2025a). Method for producing coreless fibre composite components and winding pin for a fibre composite component (EP23188451.1). *European Patent Office.*

Mindermann, P., Müllner, R., & Hoffrogge, P. (2025b). *Procedure for the production of core fiber composite components and wrapping pencil for a fiber composite component (DE502023000698D1)*.

Molaei, S., Havvaei, A., Zare, H., & Jalili, M. (2021). Collaborative deep forest learning for recommender systems. *IEEE Access, 9*, 22053–22061.

Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E, 69*(2), 026113.

Noori, A., Balafar, M. A., Bouyer, A., & Salmani, K. (2024). Review of heterogeneous graph embedding methods based on deep learning techniques and comparing their efficiency in node classification. *Social Network Analysis and Mining, 14*(1), 17.

Nsanzumuhire, S. U., Groot, W., Cabus, S. J., & Bizimana, B. (2021). Understanding the extent and nature of academia-industry interactions in Rwanda. *Technological Forecasting and Social Change, 170*, 120913.

O'Dwyer, M., Filieri, R., & O'Malley, L. (2023). Establishing successful university–industry collaborations: Barriers and enablers deconstructed. *The Journal of Technology Transfer, 48*(3), 900–931.

O'Leary, D. E. (2024). Do ChatGPT 4o, 4, and 3.5 generate "similar" ratings? Findings and implications. *IEEE Intelligent Systems, 39*(5), 78–81.

Ollila, S. (2025). In-between identity work: Transcending boundaries in university-industry collaboration. *Technovation, 139*, 103128.

Park, J., Jeong, S., Lee, B. S., & Lim, S. (2023). MIGTNet: Metapath instance-based graph transformation network for heterogeneous graph embedding. *Future Generation Computer Systems, 149*, 390–401.

Patnaik, S., Pereira, V., & Temouri, Y. (2022). Intra-organisational dynamics as 'dark side'in inter-organizational relationships: Evidence from a longitudinal investigation into a university-industry collaboration. *Technological Forecasting and Social Change, 174*, 121259.

Petralia, S. (2020). Mapping general purpose technologies with patent data. *Research Policy, 49*(7), 104013.

Polverini, G., Melin, J., Önerud, E., & Gregorcic, B. (2025). Performance of ChatGPT on tasks involving physics visual representations: The case of the brief electricity and magnetism assessment. *Physical Review Physics Education Research, 21*(1), 010154.

Qian, M., Zhao, M., Yang, J., Yang, G., Xu, J., & Cheng, X. (2024). A novel approach to enterprise technical collaboration: Recommending R&D partners through technological similarity and complementarity. *Journal of Informetrics, 18*(4), 101571.

Ran, C., Song, K., & Yang, L. (2020). An improved solution for partner selection of industry-university cooperation. *Technology Analysis & Strategic Management, 32*(12), 1478–1493.

Reches, M., Lee, P. S., & Hu, T. (2025). Superhydrophobic coating (WO2025094171A1). *World Intellectual Property Organization*.

Reyes-Menendez, A., Clemente-Mediavilla, J., & Villagra, N. (2023). Understanding STI and SDG with artificial intelligence: A review and research agenda for entrepreneurial action. *Technological Forecasting and Social Change, 196*, 122785.

Roncancio-Marin, J., Dentchev, N., Guerrero, M., Díaz-González, A., & Crispeels, T. (2022). University-Industry joint undertakings with high societal impact: A micro-processes approach. *Technological Forecasting and Social Change, 174*, 121223.

Rossi, F., De Silva, M., Pavone, P., Rosli, A., & Yip, N. K. (2024). Proximity and impact of university-industry collaborations A topic detection analysis of impact reports. *Technological Forecasting and Social Change, 205*, 123473.

Rotolo, D., Hicks, D., & Martin, B. R. (2015). What is an emerging technology? *Research Policy, 44*(10), 1827–1843.

Saraiva, M. M., Ribeiro, T., Agudo, B., Afonso, J., Mendes, F., Martins, M., & Macedo, G. (2025). Evaluating ChatGPT-4 for the Interpretation of Images from several diagnostic techniques in gastroenterology. *Journal of Clinical Medicine, 14*(2), 572.

Shan, J., Shan, C. H., Huang, C., Wu, Y. P., Lia, Y. K., & Chen, W. J. (2023). Study of microstructure and mechanical properties of electrodeposited Cu on silicon heterojunction solar cells. *Metals, 13*(7), 1223.

Small, H., Boyack, K. W., & Klavans, R. (2014). Identifying emerging topics in science and technology. *Research Policy, 43*(8), 1450–1467.

Song, K. (2021). A selection method for industry-university cooperation from the perspective of patentometrics. *Library Tribune, 41*(11), 19–27.

Song, Y., Sahut, J. M., Zhang, Z., Tian, Y., & Hikkerova, L. (2022). The effects of government subsidies on the sustainable innovation of university-industry collaboration. *Technological Forecasting and Social Change, 174*, 121233.

Steele, T., & Mandler, D. (2024). Electroactive bioadhesive compositions (US11964071B2). *America.*

Steinmo, M., & Rasmussen, E. (2018). The interplay of cognitive and relational social capital dimensions in university-industry collaboration: Overcoming the experience barrier. *Research Policy, 47*(10), 1964–1974.

Sun, Z., Wang, L. E., & Sun, J. (2023). A multi-scale graph embedding method via multiple corpora. *Neurocomputing, 540*, 126192.

Thierry, N., Bao, B. K., & Ali, Z. (2023). RAR-SB: Research article recommendation using SciBERT with BiGRU. *Scientometrics, 128*(12), 6427–6448.

Traag, V. A., Waltman, L., & Van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports, 9*(1), 1–12.

Tseng, J. Y., Chen, W. J., & Chen, P. H. (2024). The impact of substrate temperature on the adhesion strength of electroplated copper on an Al-doped ZnO/Si system. *Materials, 17*(20), 4953.

Tu, Y. N., & Seng, J. L. (2012). Indices of novelty for emerging topic detection. *Information Processing & Management, 48*(2), 303–325.

Turan, E. I., Baydemir, A. E., Balıtatlı, A. B., & Şahin, A. S. (2025). Assessing the accuracy of ChatGPT in interpreting blood gas analysis results ChatGPT-4 in blood gas analysis. *Journal of Clinical Anesthesia, 102*, 111787.

Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P., & Yu, P. S. (2019). Heterogeneous graph attention network. In *The world wide web conference* (pp. 2022-2032).

Wang, J., Xu, B., & Zu, Y. (2021). Deep learning for aspect-based sentiment analysis. In *2021 international conference on machine learning and intelligent systems engineering (MLISE)* (pp. 267–271). IEEE.

Wang, C. N., Nguyen, X. T., & Wang, Y. H. (2016). Automobile industry strategic alliance partner selection: The application of a hybrid DEA and grey theory model. *Sustainability, 8*(2), 173.

Wang, J., Li, S., Chen, J., & Ding, S. (2024a). Study on the evolution of multi-level collaborative innovation networks in China's cloud manufacturing industry. *Technology Analysis & Strategic Management., 37*(12), 1–22.

Wang, L., & Huang, Z. (2020). Research on the synergetic innovation between pharmaceutical enterprises and scientific research institutions based on the quantum game. *IEEE Access, 8*, 63718–63724.

Wang, Q. (2018). A bibliometric model for identifying emerging research topics. *Journal of the Association for Information Science and Technology, 69*(2), 290–304.

Wang, Q., Ma, J., Liao, X., & Du, W. (2017). A context-aware researcher recommendation system for university-industry collaboration on R&D projects. *Decision Support Systems, 103*, 46–57.

Wang, W., Wang, X., Qin, S., Yuan, J., Luo, J., Bai, Z., & Ma, L. (2024b). Modifying functional groups of straw-based hydrogel provide soil N_2O mitigation potential by complete denitrification. *Chemical Engineering Journal, 500*, 157442.

Wang, W., Zhang, G., Han, H., & Zhang, C. (2023). Correntropy-induced Wasserstein GCN: Learning graph embedding via domain adaptation. *IEEE Transactions on Image Processing, 32*, 3980–3993.

Wei, X. (2024). Evaluating chatGPT-4 and chatGPT-4o: Performance insights from NAEP mathematics problem solving. *Frontiers in Education, 9*, 1452570.

Woolgar, L. (2007). New institutional policies for university-industry links in Japan. *Research Policy, 36*(8), 1261–1274.

Xu, H., Winnink, J., Yue, Z., Zhang, H., & Pang, H. (2021a). Multidimensional Scientometric indicators for the detection of emerging research topics. *Technological Forecasting and Social Change, 163*, 120490.

Xu, S., Hao, L., Yang, G., Lu, K., & An, X. (2021b). A topic models based framework for detecting and forecasting emerging technologies. *Technological Forecasting and Social Change, 162*, 120366.

Yang, D., Qu, B., Yang, J., Wang, L., & Cudre-Mauroux, P. (2022a). Streaming graph embeddings via incremental neighborhood sketching. *IEEE Transactions on Knowledge and Data Engineering, 35*(5), 5296–5310.

Yang, J., Lu, W., Hu, J., & Huang, S. (2022b). A novel emerging topic detection method: A knowledge ecology perspective. *Information Processing & Management, 59*(2), 102843.

Yang, M., Li, Z., Gao, Y., He, C., Huang, F., & Chen, W. (2024). heterogeneous graph attention networks for depression identification by campus cyber-activity patterns. *IEEE Transactions on Computational Social Systems, 11*(3), 3493–3503.

Yi, H. C., You, Z. H., Huang, D. S., & Kwoh, C. K. (2022). Graph representation learning in bioinformatics: Trends, methods and applications. *Briefings in Bioinformatics, 23*(1), bbab340.

Zeng, X., Xing, Z., & Zhu, J. (2023). Development evaluation of China's regional innovation during the deep integration process of industry-university research institute cooperation network-a perspective of change speed. *Technology Analysis & Strategic Management, 36*(12), 1–18.

Zhai, D., Zhai, L., Li, M., et al. (2022). Patent representation learning with a novel design of patent ontology: Case study on PEM patents. *Technological Forecasting and Social Change, 183*, 121912.

Zhang, C., Song, D., Huang, C., Swami, A., & Chawla, N. V. (2019a). Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 793-803).

Zhang, S., Tong, H., Xu, J., & Maciejewski, R. (2019b). Graph convolutional networks: A comprehensive review. *Computational Social Networks, 6*(1), 1–23.

Zhang, Y., & Chen, X. (2023). Empirical analysis of university–industry collaboration in postgraduate education: A case study of Chinese universities of applied sciences. *Sustainability, 15*(7), 6252.

Zhang, Y., & Tang, M. (2023). A theoretical analysis of DeepWalk and Node2vec for exact recovery of community structures in stochastic blockmodels. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 46*(2), 1065–1078.

Zhang, Y., Wu, M., Miao, W., Huang, L., & Lu, J. (2021). Bi-layer network analytics: A methodology for characterizing emerging general-purpose technologies. *Journal of Informetrics, 15*(4), 101202.

Zhang, Y., Wu, M., Zhang, G., & Lu, J. (2023). Stepping beyond your comfort zone: Diffusion-based network analytics for knowledge trajectory recommendation. *Journal of the Association for Information Science and Technology, 74*(7), 775–790.

Zhang, Y., Zhang, G., Chen, H., et al. (2016). Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research. *Technological Forecasting and Social Change, 105*, 179–191.

Zhao, Y., Yongquan, Y., Jian, M., Lu, A., & Xuanhua, X. (2024). Policy-induced cooperative knowledge network, university-industry collaboration and firm innovation: Evidence from the greater bay area. *Technological Forecasting and Social Change, 200*, 123143.

Zheng, X., Ma, Y., Xi, T., Zhang, G., Ding, E., Li, Y., Ji, R. (2025). An information theory-inspired strategy for automated network pruning. *International Journal of Computer Vision*, 1-28.

Zhou, F., Wang, X., Lim, M. K., He, Y., & Li, L. (2018). Sustainable recycling partner selection using fuzzy DEMATEL-AEW-FVIKOR: A case study in small-and-medium enterprises (SMEs). *Journal of Cleaner Production, 196*, 489–504.

Zhou, X., Huang, L., Zhang, Y., & Yu, M. (2019). A hybrid approach to detecting technological recombination based on text mining and patent network analysis. *Scientometrics, 121*, 699–737.

Zhu, Q., Zhang, L., Xu, Q., Liu, K., Long, C., & Wang, X. (2025, March). HHGT: hierarchical heterogeneous graph transformer for heterogeneous graph representation learning. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining* (pp. 318-326).