Hybrid Topology Representation and deep Semantics with Temporal Weighting for Link Prediction in Co-word Network

Yuhan Yang, Jiahui Wang, Xiaojuan Zhang
Department of Information Management, Sichuan University, Chengdu, China

Abstract
Link prediction in co-word network is a valuable tool for researchers to understand the evolution trend and detect new research directions. It has attracted interest in both academic and industrial fields. Current link prediction methods in co-word networks face two key limitations: they rely on traditional topological features that fail to integrate local and global structures, often misjudging local links or missing distant connections, and on semantic features from conventional topic models ignore syntax and context, hindering meaningful semantic discrimination. To improve the performance of link prediction in co-word networks, this study proposes a new approach: **H**ybrid **T**opology Representation and Deep **S**emantic features with **T**emporal Weighting (HTST) for Link Prediction in Co-word Network, dynamically integrating hybrid topological representations and deep semantic information of words. Specifically, we leverage GAT and Node2Vec, which capture both local and global structure, to form topological features. To mine the deep semantics hidden in words, we first feed SAO-annotated text into SciBERT to generate context-aware embeddings enriched with syntactic structure, and enhance semantic embeddings of each node using GCN-based co-occurrence modeling to capture more linear and non-linear higher-order semantic relationships among nodes. Finally, we integrate the proposed features using a designed time-weighted model that incorporates word novelty. To evaluate the effectiveness of our model, we conduct extensive experiments across four disciplines. Experimental results demonstrate that our model consistently outperforms three baselines on five evaluation metrics, confirming its robustness and generalizability. Additionally, we apply diverse validation approaches and present an in-depth case study to support our method.

Keywords Co-word network; Link prediction; Topological features; Deep semantic mining; Time weighting

Introduction

The scientific knowledge, especially in Science and Technology, is of great value to the government, enterprise, and individual scholars (Small et al., 2014; Wang & Waltman, 2016). In the current era of information explosion, the massive growth of academic resources has led to an increasing accumulation of scientific knowledge. It is extremely hard to catch knowledge structures and the research hotspots of a specific discipline. Many researchers have already devoted themselves to addressing this challenging issue and co-word network analysis technique is widely used (Behrouzi et al., 2020; Huang et al., 2021). As a classical content analysis technique in the bibliometric area, co-word analysis mainly consists of the following two subtasks. First, a co-word network, which is composed of nodes and edges, is constructed for a particular domain. The nodes of the network generally are derived from subject words or keywords of literatures, which are considered as basic knowledge units of literatures. The edges of the network represent the co-occurrence relationship between words in a literature. Second, the techniques such as co-occurrence-based techniques in bibliometrics and graph related measurements, are employed in the co-word network to discern the complex relationships among different knowledge units (Cao et al., 2024). The co-word network analysis is limited to reveal the implicit knowledge structure as well as evolution law of scientific knowledge for a given discipline. With the increasing competition in various fields, more and more governments, enterprises, or individual scholars desire to master the research trend of a specific domain in advance to promote innovation and win competitive advantages in their own field. To better cater to these needs, scholars have shifted to make utilization of link prediction method to forecast the formation of new links between two nodes in the future for a co-word network, aiming at deducing the development trend of a given discipline.

The link prediction in co-word network is always dealt with using machine learning-based approaches (Martínez et al., 2016; Wang & Waltman, 2016; Behrouzi et al., 2020; Xiong et al., 2022), which convert the link prediction problem into a binary classification task. The quality of the node pair vectors determines the predictive performance of these methods, and how to accurately represent features of node pairs has become a key issue in such research fields. Particularly, the features can be extracted from network structures or attributes of nodes. Node attributes-based approaches majorly mine the features of node pairs through the attributions of node themselves. As obtaining node features and judging their validity is a problem, research activities mainly focus on typological structure-based features (Choudhury & Uddin, 2016; Behrouzi et al., 2020). The acquisition of these features is mainly achieved by utilizing some network structure-based similarity indicators to calculate the similarity between two nodes in co-word network, and the neighborhood-based（Newman, 2001; Damic & Adar, 2003; Yao et al., 2023）, path-based (Katz, 1953; Xu et al., 2016) and random walk-based（Fouss et al., 2007） indicators are popularly used in previous studies. Such metrics are prone to a limited perspective, typically causing an overestimation of minor local links or a neglect of meaningful yet distant potential connections.

In addition to structural features, previous studies have incorporated semantic attributes to capture the rich contextual meaning of words within the network (Wang et al., 2016). While topological features primarily reflect direct word co-occurrences, they often fail to reveal connections between semantically similar but topologically distant words (Tang et al., 2015; Yao et al., 2018). Semantic information helps bridge this gap by linking words across different subtopics based on shared meanings or topic distributions (Gao & Huang, 2018). While some studies have incorporated semantic features using topic models like Latent Dirichlet Allocation (LDA), such approaches rely on a bag-of-words assumption that ignores word order or syntactic structure, limiting their ability to learn nuanced, context-sensitive semantics (Blei et al., 2003; Xiong et al., 2022). More recent methods introduce contextualized embeddings such as SciBERT into link prediction tasks (Beltagy et al., 2019; Cai et al., 2022). However, these models primarily extract linear semantic associations between words from sequential patterns, neglecting their syntactic roles and cross-sentential semantic relationships. This leads to two main issues. First, the models may over-rely on word order and co-occurrence statistics, spuriously associating frequently co-occurring but semantically unrelated words. Furthermore, their inability to model cross-sentential semantic relationships confines understanding to isolated sentences, preventing the construction of a coherent discourse representation and thus undermining accurate lexical semantic understanding in complex texts.

To address these challenges, this paper targets at extracting novel topological and deep semantic features of co-word network, and these features are combined by temporal-based weighting . The method is accordingly designated as “HTST” (**H**ybrid **T**opological and Deep-**S**emantic Features with **T**emporal-based Weighting). Compared with previous work, the significant contribution of our work can be summarized as follows:

(1) Our study introduces a novel framework for link prediction in co-word networks that can effectively integrate hybrid structural features and deep semantic features, with these features being time-based weighted. In particular, the temporal weighting method incorporates word novelty into the calculation.

(2) To mine the deep semantics hidden in words of co-word network, we develop a novel approach. It first derives semantic features from SAO triple-based SciBERT embeddings and refines them with the Graph Convolutional Networks(GCN) model, thereby generating context-aware embeddings enriched by syntactic structure and non-linear, higher-order semantic relationships among words.

(3) To assess the robustness and generalizability of our model, we conduct numerous experiments on multiple datasets originating from various disciplines using multiple classification algorithms. Furthermore, we carry out domain-specific case studies to comprehensively evaluate our method.

The remaining parts of this paper are organized as follows. In the “Related Work”section, we briefly dicuss the recent studies related to our research. In the “Methods” section, we introduce our methodology, which comprises three key components: topological feature extraction, deep semantic mining of terms, and temporal weighting of the derived features. In the “Experimental Setups” and “Experimental results” sections, we detail the experimental setup and present the results, respectively. Moreover, we perform domain-specific case studies to comprehensively evaluate our method in the section “Case Study”. Finally, we conclude the paper with a summary and suggest potential directions for future research.

Related Work

Link prediction has been the subject of extensive research due to its wide applicability. Various technologies have been proposed for link prediction. To date, no published work has explored both structural and semantic characteristics for link prediction in co-word networks. The recent studies related to this research encompass the following two aspects:

(1) machine learning-based link prediction with structural features;

(2) methods for constructing semantic representations of words in co-word networks.

Machine learning-based link prediction with network structural features

Recently, machine learning-based link prediction has been increasingly favored by researchers in the field of link prediction. In this study, we summarize related work that uses network structures as features to perform link prediction tasks. Initially, many researchers focused on taking advantage of traditional topological similarity-based indicators for supervised learning models. The underlying assumption in these methods is that the likelihood of a link forming between two nodes can be determined by analyzing their structural proximity, often captured by various similarity metrics. For instance, Liben-Nowell and Kleinberg (2003) systematically formalized the link prediction problem and evaluated classic topological similarity measures such as Common Neighbors (CN), Jaccard, and Preferential Attachment (PA) for predicting links in social networks. Furthermore, Kumari et al. (2022) constructed feature vectors using path-based similarity measures, such as Katz and SimRank, to feed into supervised classifiers for link prediction. Ma et al. (2024) developed a framework integrating a keyword citation network with three feature dimensions, including network topological analysis alongside altmetrics indicators, to identify weak signals of emerging topics. Shang et al. (2025) introduced TriHetGCN, which incorporates explicit topological features such as triadic closure and degree heterogeneity to enhance link prediction performance.

The proliferation of deep learning techniques for graph analysis has led to the development of graph embedding models, including DeepWalk, Node2Vec, and GCN. These models learn low-dimensional vector representations that encode structural information, node relations, and local subgraph patterns, thereby inaugurating new paradigms for topological information extraction. The emergence of these techniques has naturally motivated scholars to explore their potential for solving link prediction problems. For instance, Huo et al. (2025) developed a graph neural network (GNN)-based approach for cross-disciplinary topic link prediction. Choudhary and Kumar (2025) presented a hybrid model integrating GCNs and long short-term memory (LSTM) networks to enhance link prediction in dynamic social networks. Chatterjee et al. (2025) introduced a transfer learning framework for temporal link prediction by incorporating a structure-mapping module that learns to align topological features with memory-based embeddings across evolving graphs. Cao et al. (2024) designed a dynamic method for detecting technological recombination using semantic analysis and dynamic network analysis, which applied a dynamic word embedding model to generate word vectors. Liu et al. (2025) proposed the Higher-order structure Temporal Graph Neural Network (HTGN) method, a temporal graph model that captures high-order structural dynamics for improved link prediction.

Additionally, some scholars have exploited machine learning-based methods to perform link prediction in co-word networks. For example, Choudhury and Uddin (2016) attempted to exploit supervised learning models to address the task of link prediction in co-word networks, where the features between two nodes were extracted based on network topological similarity indicators and their temporal evolutionary information. Behrouzi et al. (2020) used two classical topology-based link prediction methods (i.e., Jaccard Coefficient and Resource Allocation) and a new model developed based on the clustering coefficient, a centrality measure, and community information to calculate the scores between nodes, which were used as features to be input into classifiers to predict future links in a co-word network. Huang et al. (2021) employed three traditional similarity indicators, such as CN and SimRank, to reveal the dynamic changes in co-word networks, and further exploited machine learning algorithms to fit the prediction metrics to promote the improvement of predictive performance. Xiong et al. (2022) combined the semantic features of nodes with exsiting neigbor-based topological similarity features (e.g., CN, AA, PA,Salton,weighted CN and weighted AA) to realize link prediction in a co-word network.

To sum up, link prediction in a co-word network was achieved by leveraging features from traditional similarity indicators. However, few studies have focused on mining more effective and deeper relationships between nodes in co-word networks. To the best of our knowledge, this study is the first to pioneer the application of graph embeddings and deep semantic mining methods to overcome this shortcoming.

Semantic representation methods of keywords in co-word analysis

Although research on mining semantics hidden in words remains relatively scarce in link prediction of co-word networks, extensive exploration has been conducted in co-word analysis for revealing the implicit knowledge structure or the evolution law of scientific knowledge for a given discipline. Based on the different techniques employed for capturing semantics of words, such research can be categorized into the following three types: external knowledge-based, topic detection-based, and embedding learning-based methods.

External knowledge-driven methods incorporate semantic information from the outside, e.g., by introducing external knowledge bases (e.g., WordNet and DBpedia), expert knowledge and domain dictionary(e.g., MeSH and OpenGALEN), etc. For example, Wang et al. (2011) proposed a semantic co-word analysis framework that integrated a topic ontology to embed expert-defined conceptual hierarchies, effectively mitigating issues of synonymy, polysemy, and missing conceptual links in traditional co-word networks. Feng et al. (2017) represented the co-occurring words as ontological concepts and determined their conceptual affinity by their positions within the ontology. Zhao et al. (2018) introduced semantic-enhanced full-text co-occurrence networks, where co-occurring terms were annotated with relational information from ConceptNet to capture nuanced semantic associations. To address the limitation that some co-occurring word pairs lack direct semantic relations in ConceptNet, Zhao et al. (2023) introduced a relation annotation method which connected them via an intermediate word, and they identified six semantic relations in ConceptNet, including RelatedTo, IsA, Synonym, HasContext, Antonym, and MannerOf, as key factors directly contributing to word co-occurrence.

Topic detection-based methods leverage models like topic detection methods (e.g., Latent Dirichlet Allocation (LDA), Hierarchical Dirichlet Process (HDP), Non-negative Matrix Factorization (NMF) to generate semantic representations of textual features. Latent topic models capture semantic structures by uncovering hidden thematic patterns in large text corpora. These models assume that semantically related words tend to exhibit similar probabilistic distributions across latent topics, and that documents can be represented as mixtures of such topics. For instance, Zhou et al. (2018) developed a method for semantically measuring co-word associations by combining LDA and Word2Vec. Zhang et al. (2019) proposed HDP–GTM, a nonparametric topic model that integrates HDP with graph structure to flexibly infer topic numbers and better capture semantic relations. Wang et al. (2021) proposed the DWGTM (Dual Word Graph Topic Model), which incorporated both word co-occurrence and semantic correlation graphs to extract more coherent and semantically meaningful topics from short texts. Austin et al. (2022) further extended topic modeling to community discovery by constructing topic–word networks, enabling the detection of flat and hierarchical topic structures with enhanced interpretability and semantic consistency. Si et al. (2022) introduced FedNMF, a federated NMF-based topic modeling framework, and shows that mutual-information extension improves classification performance under heterogeneous data.

Embedding Learning-based approaches initially used neural network to learn the relationship between a word and its surrounding words, mapping each word to a low-dimensional dense vector. Ultimately, the distance and direction of these vectors in space can capture the semantic and syntactic similarities between words. Techniques such as word embeddings (e.g., Word2Vec) are popularly used (Liu et al., 2023). A primary limitation of such methods is their use of context-independent embeddings, resulting in a single static representation for each word. This inherently prevents them from handling polysemy, capturing long-range dependencies, or modeling complex syntactic structures effectively (Matthews et al., 2024). The introduction of pre-trained contextualized language models (e.g., BERT) has led to a paradigm shift in semantic mining, achieving markedly higher effectiveness than earlier techniques (Qiu et al., 2020; Rogers et al., 2023). -The methods(i.e., topic-based methods, Word2vec and BERT) mentioned above mainly explore semantic associations between words from sequential patterns, and they are hard to capture semantic associations across different sequences. To overcome this limitation, graph-embedding methods(e.g., DeepWalk, Node2Vec, Struc2Vec) are applied aiming to capture semantic relationships between words through mining higher-order topological patterns among nodes (Liu et al., 2023; Ling et al., 2024). Representative studies that utilize graph-embedding methods to explore semantic associations in co-word networks include: Zhu &Zhang (2020) proposed a novel method that leverages meta-paths within a subject knowledge network. This approach involves constructing the network to identify word-to-word meta-paths, followed by employing the HeteSim algorithm to compute the semantic relatedness between words along each path. Liu et al. (2023) developed the HIE, a hierarchical embedding model that jointly encodes entities and relations across multiple semantic levels, enabling fine-grained representation of conceptual proximity and improving interpretability in semantic network analysis. Ling et al. (2024) proposed a model Link2Doc, which integrated textual semantics into edge-labeled graphs using graph neural networks and large language models to enhance link prediction and topic association. While graph-based methods can uncover indirect and semantically important relationships, a key limitation is their tendency to represent a word as a single node. This approach is prone to collaps its diverse meanings across different contexts.

In summary, while existing research in co-word analysis demonstrates substantial progress in modeling both linear and non-linear semantic relationships between words. However, these studies often overlook the fine-grained syntactic structures between words and fail to integrate contextual relationships within a unified framework. To address this issue, this study attempts to uncover semantic relationships between words by comprehensively incorporating syntactic structures along with contextual patterns, hence providing new research insights for such studies.

Methods

Formally, our task is defined as follows. First, a co-word network $G(V, E)$ is constructed from a literature set M, published during the period $[T1, T2]$. In this network, keywords from the literature serve as nodes (V), and their co-occurrence within the same document forms undirected edges (E), with edge weights representing the co-occurrence count. The objective of link prediction is to forecast new keyword co-occurrences (i.e., new links) likely to emerge in the subsequent period $[T2, T3]$, based on the established network $G(V, E)$ (Liben-Nowell & Kleinberg, 2003). The main framework of the our method is presented in Fig.1. The framework consists of three key components: topological feature extraction, deep semantic mining of terms, and temporal weighting of the derived features. Each of these components will be elaborated on in the subsequent subsections. We highlight the key innovative components of our method using solid-line boxes.

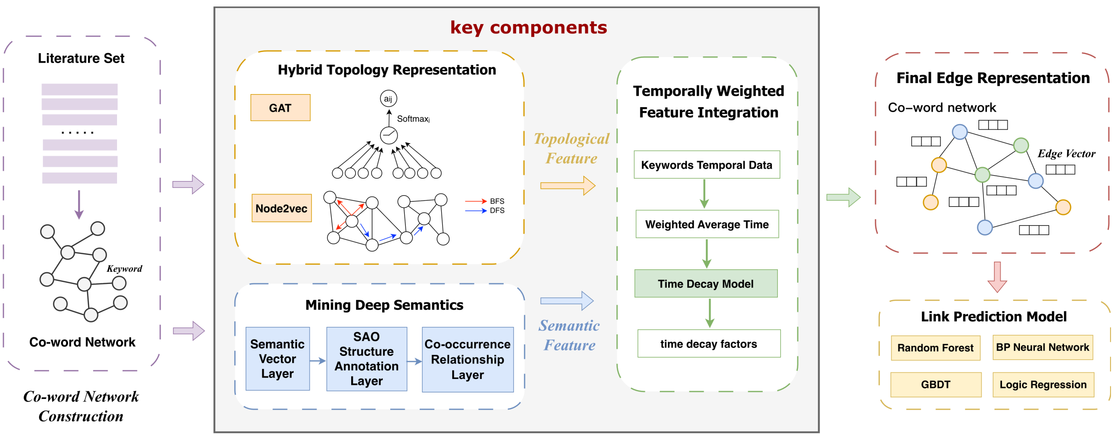


Fig.1 The framework of our method (HTST)

Capturing topological features

To fully characterize complex network structures, both the Graph Attention Networks (GAT) and the Node2Vec model are applied in our work. GAT is a type of graph neural network that incorporates an attention mechanism, enabling it to adaptively compute importance weights for the neighbors of each node in a graph, focusing more on local structure (Velickovic et al., 2018). The GAT model offers the following advantages for co-word network analysis: (1) it introduces an attention mechanism that adaptively assigns varying importance weights to neighboring nodes, and this capability can effectively capture more nuanced and significant topological relationships between co-occurring terms. (2) it aggregates neighbor features to generate refined node representations, which highlight the most significant co-occurrence relationships. This allows GAT to produce highly informative node representations, which are crucial for predictive tasks like link prediction that demand a deep understanding of intricate network structures.

On the other hand, Node2Vec is a random walk-based graph embedding algorithm, which makes it helpful to represent global structure (Grover & Leskovec, 2016). Node2Vec first generates biased random walks that interpolate between breadth-first and depth-first sampling strategies, producing node sequences. Then, these sequences are then used to train a Skip-Gram model to learn node embeddings. This biased random walk strategy equips the Node2Vec model with two essential capabilities for co-word network analysis: (1) it flexibly balances BFS and DFS strategies through adjustable parameters, enabling it to sample node neighborhoods that reflect both local homophily and broader structural roles in the co-word network. (2) By applying the skip-gram model to the sampled node sequences, Node2Vec captures latent semantic and structural relationships, producing embeddings that encode both proximity and functional similarity between nodes.

It can be concluded from the above discussion that since GAT and Node2Vec excel at capturing local and global information of a network respectively, integrating them can contribute to a comprehensive mining of network features. For this purpose, the final topological representation of each node is obtained by concatenating its GAT and Node2Vec vectors, which are generated individually.

Mining Deep Semantics for Each Node

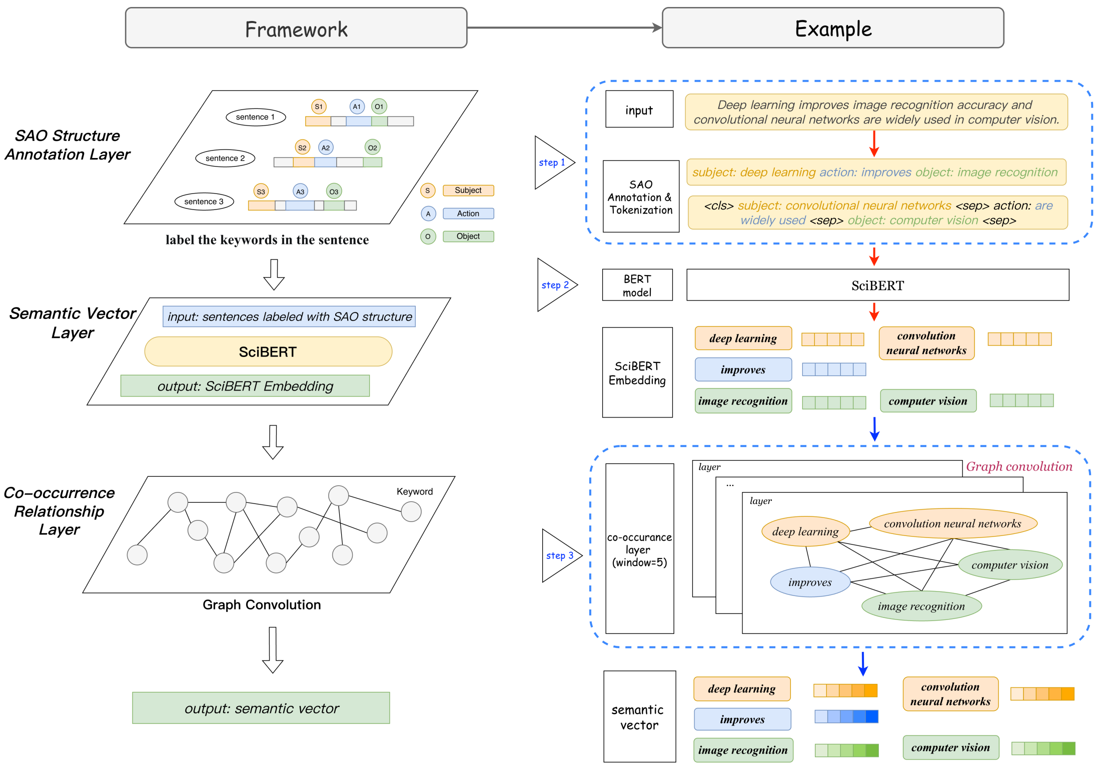
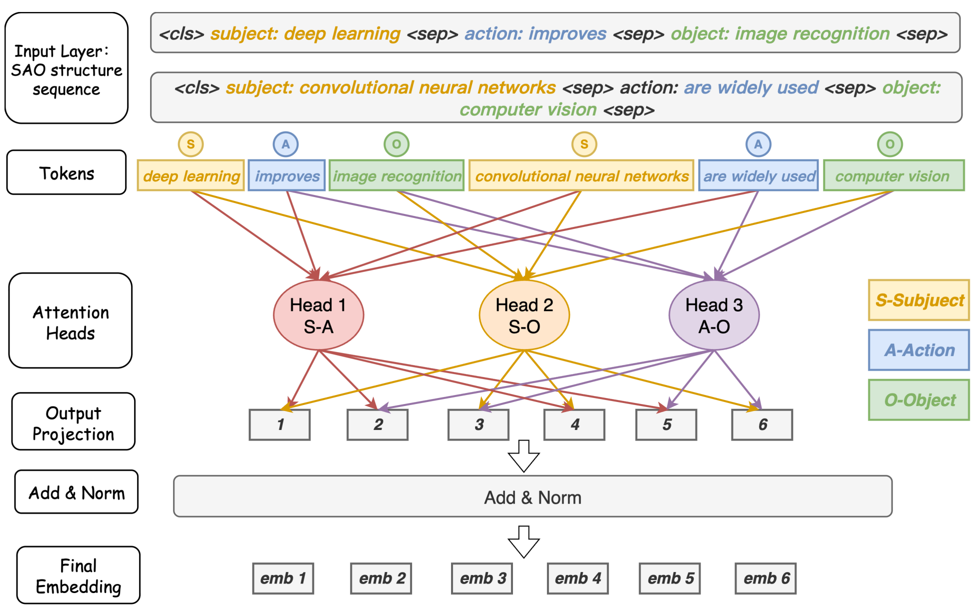


Fig.2 The framework of mining deep semantics for words

Currently, SciBERT represents the dominant approach for semantic understanding in scientific literature. However, such methods may over-rely on word order and co-occurrence statistics while neglecting crucial semantic roles (e.g., agent, theme, and instrument). This limitation can consequently restrict their capacity for cross-sentential semantic understanding. Consider the example: "The algorithm proposed by the team from Stanford, which incorporates several novel techniques, significantly accelerates the computation of complex models", SciBERT's attention mechanism, while theoretically capable of linking "algorithm" and "accelerates", may have its attention diverted by intermediate words like "Stanford" or "techniques". Explicit identification of "algorithm" as the Agent and "computation" as the Theme would enable more stable long-distance semantic connections, making the model resilient to complex clause interference. Therefore, we argue that for generating word embeddings with SciBERT, it is essential to incorporate the explicit semantic roles of words within their context into the model's calculations.

Prior research has predominantly employed the Subject–Action–Object (SAO) framework to identify semantic roles. In this framework, the Subject generally corresponds to the Agent (the doer), the Action to the predicate or event, and the Object to the Theme (the entity acted upon). Notably, the SAO framework explicitly captures action–argument relationships in language and provides an interpretable grammatical and logical structure for analyzing word co-occurrence. Previous studies have further established its utility in modeling semantic information and syntactic dependencies within scientific literature (Guo et al., 2016; Yang et al., 2017). Accordingly, this study seeks to apply the SAO framework to enhance SciBERT by equipping it with a standardized "Subject-Action-Object" triple structure, thereby transforming unstructured text into an explicit semantic network. This enables the model to directly capture the precise predicate logic often overlooked in traditional contextual modeling. After obtaining semantic vectors for words using the enhanced SciBERT, this study further employs the GCN model to refine the semantic feature of each word, aiming to uncover deeper-level semantic relationships between terms. Concretely, our proposed framework for mining deep semantic of words is presented illustrated in Fig. 2. It comprises three layers: the SAO Structure Annotation Layer, the Semantic Vector Layer, and the Co-occurrence Relationship Layer. The detailed implementation of each layer will be described subsequently.

SAO Structure Annotation Layer. In this layer, we first employ the Stanford CoreNLP OpenIE module[1] to extract SAO triples (Subject-Action-Object) from the sentences. Specifically, we enhance the input representation by introducing descriptive prefixes such as "subject:", "action:", and "object:", which provide strong syntactic cues for the pre-trained model. Subsequently, each article’s abstract is then processed by a tokenizer, which breaks the text into subword units that can accurately represent scientific terminology. This process yields a tokenized sequence that can be directly fed into the model. During tokenization, a special [CLS] token is added at the beginning of the sequence to capture the overall context, while [SEP] tokens are inserted to separate different text segments. As illustrated in Fig.2, the sentence "Deep learning improves image recognition accuracy" is structured as [CLS] subject: deep learning action: improves object: image recognition



[1] https://stanfordnlp.github.io/CoreNLP/

accuracy [SEP]. This extraction process identifies key phrasal elements functioning as subjects, actions (predicates), and objects within the text.

Fig.3 Semantic embedding generation from SAO structure sequence via SciBERT

Semantic Vector Layer. The objective of this layer is to use the annotated SAO structure to guide SciBERT in generating token embeddings that capture both syntactic patterns and contextual relationships. As illustrated in Fig. 3, SciBERT functions as a structured semantic encoder in our framework. The input sequence is first enhanced with syntactic signals: we insert special tokens ([S], [A], [O]) immediately before each subject, action, and object phrase, serving as explicit role indicators for the model. These annotated sequences are then mapped into dense vectors and processed by Transformer layers. To explicitly encode syntactic interactions, we organize the multi-head self-attention into three functional groups. Heads 1–4 examine subject–action interactions, heads 5–8 represent subject–object associations, and heads 9–12 capture action–object relationships. For each attention head h, the self-attention output is computed as Eq.(1)

$$\mathrm{Attn_h}(Q,K,V) = \mathrm{softmax}\left(\frac{Q_h K_h^{\mathrm{T}}}{\sqrt{d}}\right) V_h \tag{1}$$

where Q_h, K_h, and V_h are head-specific projections. Because each token is annotated with an SAO role, the query–key similarity reflects both word meaning and role information. As training proceeds, heads within the same group learn projection parameters that assign higher attention weights to token pairs matching their target role relations. This leads to stable head-level specialization, allowing different heads to encode distinct Subject–Action, Subject–Object, and Action–Object interactions within the same layer. Each attention calculation is followed by an output projection and an Add-&-Norm step, ensuring stable residual refinement of the contextual representations. Finally, by aggregating the outputs from these functionally specialized heads, the model produces the final contextual embeddings that encode the precise syntactic roles and semantic dependencies of the SAO-structured input.

Co-occurrence Relationship Layer. For a Transformer model like SciBERT, although its internal self-attention mechanism can capture relationships between all words in a sequence, its basic input and modeling unit remains a linear token sequence. It understands semantics by learning patterns of word co-occurrence in large volumes of text. However, this understanding is largely confined to highly similar sequential patterns and cannot establish explicit word-level connections across sentences. To mitigate this limitation，we further introduce a Co-occurrence Relationship Layer that builds a weighted keyword co-occurrence graph to identify potential semantic relations between words across sentence boundaries.Operationally, this layer first constructs a weighted keyword co-occurrence graph. Then, abstracts are segmented into sentences, and a sliding window (e.g., size five) is employed to identify proximal keyword co-occurrences, a strategy intended to mitigate spurious edges arising from excessive sentence length. Undirected edges are established between keywords co-occurring within this window, with edge weights assigned proportionally to their co-occurrence frequency. Subsequently, this co-occurrence graph serves as the structural basis for refining the preliminary keyword embeddings obtained from the Semantic Vector Layer. The GCN (Kipf & Welling, 2017) is utilized for this refinement. The GCN operates by leveraging the graph's connectivity and edge weights to iteratively aggregate information from a keyword's neighbors to update its node representation.The core propagation mechanism of Graph Convolutional Networks (GCNs) updates each keyword vector using the following Eq.(2):

$$H^{(l+1)} = \sigma(\widetilde{D}^{-\frac{1}{2}}\tilde{A}\widetilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}) \tag{2}$$

Here, $H^{(l)}$ denotes the matrix of all keyword feature vectors at layer l, where each row corresponds to the representation of a specific keyword. and $W^{(l)}$ is a trainable weight matrix that applies a linear transformation to the aggregated features, allowing the model to learn task-specific feature projections. $\tilde{A}$ represents the adjacency matrix of the co-occurrence graph with added self-loops, ensuring that each keyword retains its own feature information during updates. $\widetilde{D}$ is the degree matrix used to normalize the adjacency matrix by accounting for the total connection strength of each keyword (including self-loops). The normalized term $\widetilde{D}^{-\frac{l}{2}}\tilde{A}\widetilde{D}^{-\frac{l}{2}}$ serves to balance the influence of neighboring nodes, preventing nodes with high degrees from dominating the aggregation process. Overall, this operation updates each keyword vector by aggregating information from itself and its co-occurring neighbors. Through multiple layers of propagation, the model captures both local and higher-order semantic associations, leading to refined vector representations where frequently co-occurring keywords are embedded closer in the semantic space.

The GCN refines keywords' semantic embeddings by leveraging co-occurrence relationships, addressing the limitations of relying solely on explicit structures. As defined in Eq. (2), it iteratively aggregates and normalizes feature vectors from neighboring keywords, with edge weights proportional to co-occurrence frequency. Thus, frequently co-occurring keywords exert stronger mutual influence, pulling their vectors closer in semantic space and embedding implicit associations derived from co-occurrence statistics. To enhance robustness, average pooling is applied to combine vectors across different contexts into a unified representation. This message aggregation mechanism ensures that frequently co-occurring terms are positioned in closer proximity, forming a complementary semantic representation that works in tandem with the explicit logic captured by the SAO Structure Annotation Layer.

The final generation of edge representation based on time weighting

After generating the semantic vector and topological structure vector for each node, the node-pair vectors will then be constructed. It should be noted that the features employed in this study, including topological structure and semantics are predicated on the assumption that past connection patterns will persist into the future. These features have limited capability in predicting breakthrough and non-continuous changes. Considering that a new word or a revitalized old one often acts as a "seed" or "catalyst" for emerging research frontiers (Shi & Evans, 2023; Yang et al., 2024), this study postulates that a highly novel word possesses greater "link growth space". Therefore, when predicting the formation of an edge between two nodes, we further incorporate the novelty of the nodal words. In particular, we apply a time decay function that assigns a weight to each keyword and the final node-pair vectors are determined based on the following Eq.(3):

$$Edge\ Weight_{(i,j)} = \left(w_i * tv_i + w_j * tv_j\right) * cos\left(sv_i, sv_j\right) \tag{3}$$

Here, sv_1 and sv_2 are semantic vectors generated by in the section "Mining Deep Semantics for Each Node". $cos\left(sv_i, sv_j\right)$ is the cosine similarity between the sv_1 and sv_j. tv_i and tv_j represent the initial vectors derived from the network topology. w_1 and w_2 denote the time decay factors of two nodes. Specifically, we use the exponential time decay model and it is computed as,

$$w = e^{-\lambda(t_{current} - t_{w_i})} \tag{4}$$

where λ is the decay coefficient (set to 0.8 in our case), and $t_{current}$ represents the most recent year in the dataset. t_{w_i} is the timestamp associated with the word w_i. In other application scenarios (Sugiyama & Kan,2010; Hua et al., 2020), t_{w_i} here generally represents the time when an object first appears. If the this method is employed to represent lexical novelty, it can only record only the birth of a word, being inherently incapable of accounting for its sustained influence, periods of dormancy, or subsequent revival within the timeline. Therefore, this method cannot distinguish between an archaic word that appeared only once briefly in history and a perennial classic that has maintained activity to the present day, erroneously categorizing both as equally obsolete.To tackle this issue, t_{w_i} is defined as Eq. (5):

$$\bar{t} = \sum_{t=t_{start}}^{t_{\text{current}}} \frac{f_t(w_i) * t}{\sum_{\text{t}=t_{start}}^{\text{t}=t_{\text{current}}} f_t(w_i)} \tag{5}$$

where t_{start} denotes the earliest available year in the dataset. $f_t(w_i)$denotes the total frequency of a word w_i in the dataset in a given year t. $\sum_{\text{t}=t_{start}}^{\text{t}=t_{\text{current}}} f_t(w_i)$ denotes the aggregate count of occurrences for the word w_i within the time interval between t_{start} and $t_{current}$.

Experimental Setups

This section outlines the experimental setup used to evaluate our proposed models. We begin by introducing the dataset construction, then provide a detailed explanation of the evaluation metrics and baseline methods. Finally, we present additional experimental settings and parameters.

Dataset construction

In this section, we first describe the data acquisition in detail, and then the pre-processing of data, selection of the scope of keywords set as well as classification dataset construction are followed in turn.

Data acquisition and pre-processing. To illustrate the applicability and flexibility of our proposed method, this study collects data from multiple disciplines for empirical analysis. Specifically, we randomly selected four distinct disciplines: Information Science & Library Science (ISLS), Law, Economics, and Oceanography (Ocean), as our experimental subjects from the subject category directory of the Web of Science Core Collection. For data collection in each discipline, this study is guided by Bradford's Law (Egghe & Rousseau, 1990). This law holds that a discipline's core journals represent its research frontier and concentrate its authoritative literature, which accurately captures the field's research hotspots and knowledge structure. This principle is operationalized by selecting the core journals of each discipline as the data source, instead of all the literature published in that field. We used the Journal Citation Reports (JCR)[3] to identify the core journals. For each chosen discipline, the top 10 journals from the JCR rankings in year 2024 are selected. These selected journals are listed in Table 1. Following the identification of core journals for each discipline, we gathered raw data for each discipline by searching documents for each selected journal within that discipline from the core collection of Web of Science[4]. When collecting the data for a given journal, the publication time interval is limited to between 2014 to 2023 (including 2014 and 2023, with each year spanning from January 1 to December 31). The total number of raw literature obtained for each of four disciplines is presented in Fig.4. It should be noted that only the abstracts of the collected documents are used as the corpus for our predictive analysis.

Table 1 Top 10 journals in each discipline ranked by Journal Impact Factor (JIF) in the JCR

No	ISLS	LAW	Economics	Ocean
1	European Journal of Information Systems	European Journal of Psychological Applied to Legal Context	Journal of Economics	Annual Review of Marine Science
2	Information & Management	Yale Law Journal	American Economic Review	Limnology and Oceanography Letters
3	Information Process and Management	Stanford Law Review	Journal of Financial Economics	Ocean & Coastal Management
4	International Journal of Information Science	Internet Policy Review	Quarterly Journal of Economics	Ocean Engineering
5	Telematics and Informatics	Harvard Law Review	American Economic Review-Insights	Applied Ocean Research
6	Government Information Quarterly	Columbia Law Review	World Bank Research Observer	Ocean Science
7	Journal of Management Information Systems	Computer Law & Security Review	Journal of Economic Literature	Limnology and Oceanography
8	MIS Quarterly	Regulation & Governance	Energy Economics	IEEE Journal of Oceanic Engineering
9	Journal of Computer Mediated Communication	Artificial Intelligence and Law	Energy Policy	Progress in Oceanography
10	Journal of Strategy Information Systems	Journal of Legal Analysis	Transportation Research part E-Logistics and Transportation Review	Journal of Geophysical Research-Oceans

To clean the noise for the collected raw data, the following data preprocessing steps are performed. We initially removed blank and redundant entries. Subsequently, abstracts are preprocessed to eliminate stop words and tokenization using the Natural Language Toolkit (NLTK)[6] library. Besides, we further removed and replaced special characters, punctuation, and digits contained within the words. After the data preprocessing, the data from the year 2023 serves for

[3] https://jcr.clarivate.com/

[4] www.webofscience.com

[6] https://www.nltk.org

predicting stage for each individual discipline, with the remaining data used for the training stage.

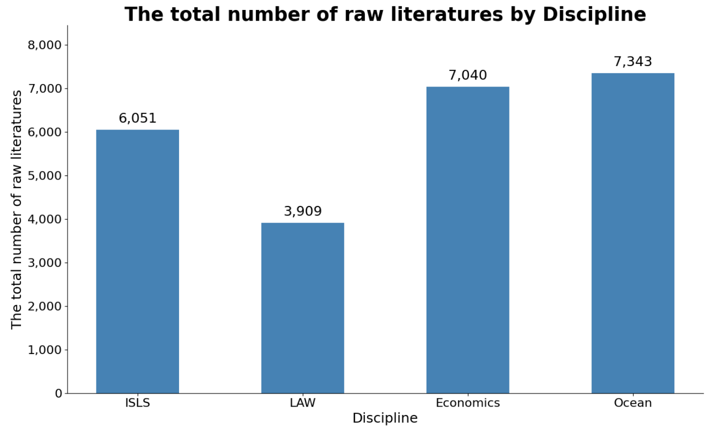


Fig.4 The total number of raw literatures obtained in each of four disciplines

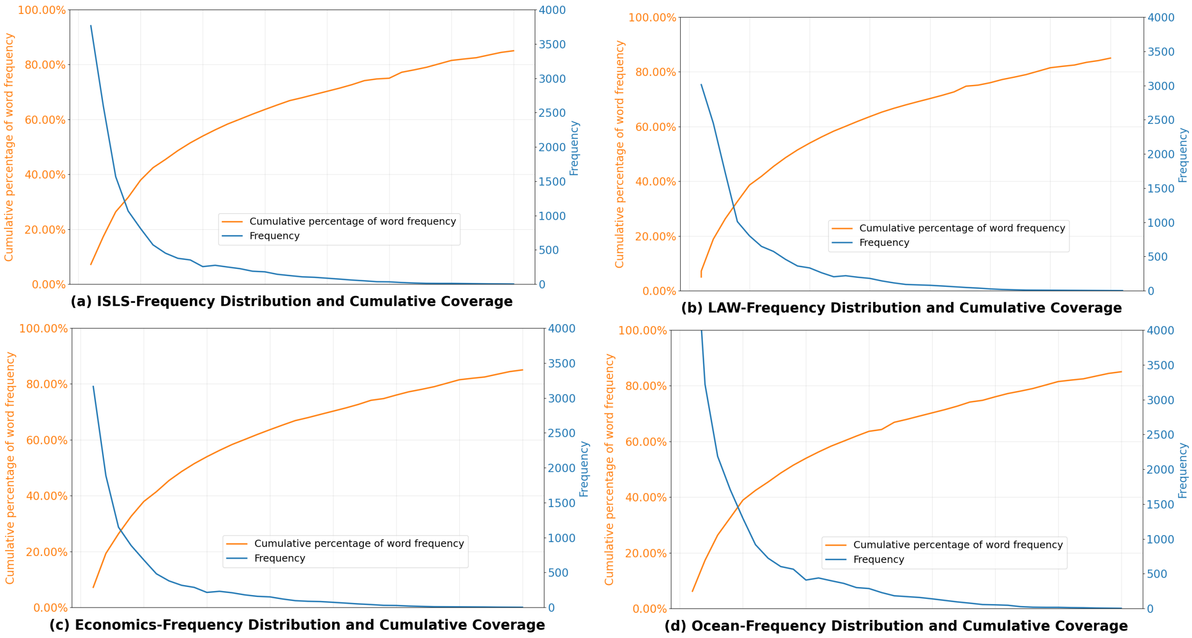


Fig.5 The change of word frequency and cumulative word frequency percentage across four disciplines

Keywords selection. Due to limited computational resources, it is impractical to construct a co-word network using all keywords extracted for each discipline. Moreover, keywords that present only one or two articles are supposed to be unable to successfully attract the attention of other authors, and they are considered unable to completely represent the research core and direction of the discipline (Liu et al., 2016). Consequently, to ensure the quality of the co-word network and mitigate computational expense, representive keywords should be selected. To serve this purpose, we adopted a technique from Liu et al. (2016) that defines word sets using metrics of word frequency, word quantity, and cumulative word frequency ratio. This approach is consistent with the "vocabulary coverage principle" in information retrieval and computational linguistics (Quispe et al., 2021; Gu et al., 2023), as it operates on the premise that a small subset of high-frequency words can represent most of a corpus's content. We visualized the change of word frequency and cumulative word frequency ratio with the number of words for each discilpline, as shown in Fig.5 . An examination of the four figures demonstrates a consistent long-tail distribution across ISLS, Law, Economics, and Oceanography, where a limited set of keywords accounts for the majority of occurrences. While overall vocabulary sizes differ, the frequency-coverage relationship is uniform, characterized by a rapidly increasing cumulative curve that eventually plateaus. It is found that in each discipline, keywords with a frequency of 5 or more account for about 80% of the cumulative total. This means that these high-frequency keywords (frequency >5) can effectively represent the core information of each field, leading us to determine that the optimal value for K should be around 5. At this threshold, the final number of keywords retained for each discipline is recorded in Table 2.

Table 2 Keyword counts per discipline at threshold K=5

Discipline	Quantity of keywords
ISLS	1,042
LAW	1,017
Economics	1,021
Ocean	1,535

After preprocessing the data, we constructed the co-word network. Each unique keyword is taken as a node, and an edge is added between two nodes when the corresponding keywords co-occurred in the same document. The properties of the resulting network are summarized in Table 3.

Table 3 Structural properties of each co-word network constructed for each discipline within the training set.

Discipline	Nodes	Edges	Density	Avg. Degree	Clustering Coefficient
ISLS	1,042	10,436	0.0192	20.03	0.6143
LAW	1,017	4,197	0.0081	8.25	0.6421
Economics	1,021	6,602	0.0126	12.94	0.5969
Ocean	1,535	23,630	0.0200	30.80	0.5721

Table 4 depicts that the topology parameters in the various networks has diverse values, which demonstrates that our established network covers a wide range of structural properties, including various densities, average clustering coefficients, and average degree. This finding supported the testing and verification of the performance of the proposed network method with various topological structures in later experiments.

Moreover, the keyword pairs without co-occurrence relationships in this network are used as research samples. These samples are then divided based on the co-occurrence relationships observed during the testing phase, yielding positive samples (i.e., those that form connections during the testing phase) and negative samples (i.e., those that do not form connections during the testing phase). The Statistics of constructed classification dataset for each discipline is summarized in Table 4. It is evident that the number of positive samples in this study is significantly smaller than the number of negative samples. Such an imbalance between positive and negative samples can easily lead to overfitting, causing the prediction results to be biased toward the category with more samples. To address this problem, we adopt balanced sampling to maintain a 1:1 ratio between the two classes for all subsequent experiments.

Table 4 Statistics of constructed classification dataset

Dataset	Positive sample	Negative sample	Sample ratio
ISLS	1,0436	273,445	1:26.20
LAW	4,197	123,568	1:29.44
Economics	6,602	291,776	1:44.20
Ocean	2,3630	466,915	1:19.76

Settings and Parameters

The parameters of the used models, namely, GAT, Node2Vec, SciBERT, and GCN models are reported in Table 5. Moreover, the classifiers applied and their corresponding parameter settings are listed in Table 6.

Table 5 Parameter setted for our applied models

Models	Parameters	Values
GAT	Input feature dimension	16
	Hidden layer dimension	128
	Output feature dimension	10

	Number of attention heads (layer 1)	5
	Number of attention heads (layer 2)	1
	Activation function	ReLU
	Dropout rate	0.5
Node2Vec	Embedding dimension	50
	Random walk length	10
	Number of random walks per node	30
	Number of parallel workers	10
	Context window size	10
	Return parameter (p)	1
	In–out parameter (q)	1
SciBERT	Maximum input length	512
	Hidden layer dimension	768
	Pretrained model	allenai/scibert_scivocab_uncased
	Number of transformer layers	12
	Number of attention heads	12
GCN	Input feature dimension	768
	Hidden layer dimension	128
	Output feature dimension	64
	Number of layers	2
	Activation function	ReLU
	Dropout rate	0.5
	Normalize	True
	Residual connection coefficient (alpha)	0.6

Table 6 Parameters settings for each classifier used in our work

Models	Key Parameters	Values
Random Forest (RF)	Number of trees	100
	Maximum depth	10
	Minimum samples per split	2
	Minimum samples per Leaf	1
	Criterion	gini
	Bootstrap	True
	Random state	42
Gradient Boosting (GBDT)	Learning rate	0.1
	Number of estimators	100
	Maximum depth	3
	Subsample ratio	0.8
	Criterion	friedman_mse
	Random state	42
Logistic Regression (LR)	Solver (optimization algorithm)	liblinear
	Maximum iterations (epochs)	100
	Regularization type	L2
	Random state	42
BP Neural Network	Hidden layer sizes	100
	Maximum iterations	500
	Activation function	ReLU
	Solver (optimization algorithm)	Adam
	Learning rate	0.001
	Batch size	32
	Random state	42

Evaluation Metrics

To ensure a comprehensive assessment of prediction quality of our proposed model, the evaluation metrics (i.e., Accuracy, AUC, F1-score, Precession and Recall) widely used for link prediction in network are leveraged in our work (Xiong et al., 2022; Liu et al., 2023; Li et al., 2025).

Accuracy reflects the proportion of correctly predicted links in the network, providing an overall measure of correctness. It is calculated as,

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

where TP represents the count of true positive predictions (correctly identified links), TN denotes true negatives (correctly identified non-links), FP indicates false positives (non-links wrongly predicted as links), and FN signifies false negatives (links wrongly predicted as non-links).

AUC (Area Under Curve) measures the model's ability to distinguish between existing and non-existing links. It is formally defined as,

$$\text{AUC} = \frac{n\prime + 0.5 * n''}{n} \tag{7}$$

where n' is the number of times a positive sample is ranked higher than a negative sample, n'' is the number of times they receive the same score, and n is the total number of comparisons. AUC values range from 0.5 to 1, where a higher value indicates better model performance.

Precision measures the proportion of correctly predicted positive instances (links) among all instances predicted as positive, and it is computed as,

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

A higher Precision indicates that the model is effective at predicting true links and avoiding false positive predictions.

Recall measures the proportion of correctly predicted positive instances (links) among all actual positive instances. It is calculated as,

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

F1-score is the harmonic mean of precision and recall, providing a balanced evaluation of model performance, particularly when dealing with imbalanced datasets. It is calculated as,

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{10}$$

where the calculations of Precision and Recall are referred to Eq. (8) and Eq. (9), respectively. A higher F1-score indicates that the model achieves a good balance between identifying positive samples and minimizing false positives.

Baseline methods

To validate the superiority of our proposed method, we compare our approach with the following six comparative baselines. These baseline approaches are also evaluated using the aforementioned datasets and metrics:

(1) LPSI (Xiong et al., 2022). This method integrated both semantic and topological features of network nodes to improve prediction efficiency. For semantic representation, the LDA model is employed to extract topic distributions from keywords. Meanwhile, nine existing neighbor-based topological metrics (e.g., CN, AA, PA ,Salton, weighted CN and weighted AA) are selected to capture the topological properties of the network. Finally, various classifiers are applied to predict potential links based on these combined semantic and topological features..

(2) HTST-G. This method is designed to evaluate the effectiveness of integrating the Graph Attention Network (GAT) technique into our framework for capturing network structure. Notably, the node topological representations in this approach are constructed without relying on GAT. Instead, Node2Vec is independently utilized to extract structural features from the nodes. The subsequent steps remain consistent with the framework described in the "Method" section.

(3) HTST-N. In this method, node topological features are generated without the use of Node2Vec. Instead, GAT alone is adopted to encode the structural information of nodes. The remaining pipeline remains consistent with the "Method" section,.

(4) HTST-S. To illustrated the benefit of applying SAO structure in our task, the following ablation experiment is designed. The acquisition of semantic features for keywords in the network is accomplished by directly employing SciBERT on abstracts, while leaving the remainder of the methodological framework unchanged.

(5) HTST-C. To verify the effectiveness of the GCN model for capturing the implicit semantics, the following experiment is conducted. After generating word embeddings using the SciBERT model integrated with the SAO structure, these vectors are directly applied in Eq. (3) for computation.

(6) HTST-T. The experiment is designed to verify the efficacy of applying temporal weighting to the proposed features. Following the linear fusion formula defined in Eq.(3), we set the two time decay factors, w_1 and w_2 equally to 1, thus effectively eliminating the temporal dimension's impact on the experiment.

Experimental results

In this section, we design a series of experiments to demonstrate our model's performance, including comparisons with baseline models, and analyses across various sample proportion .

Model Comparison

In this task, we seek to compare the performances of our proposed method with those of the six selected baselines(i.e., LPSI, HTST-G, HTST-N, HTST-S, HTST-C and HTST-T) in the constructed networks within each discipline. All the models are evaluated using the metrics mentioned earlier. The comparative results of models are presented in Figs.6-9.

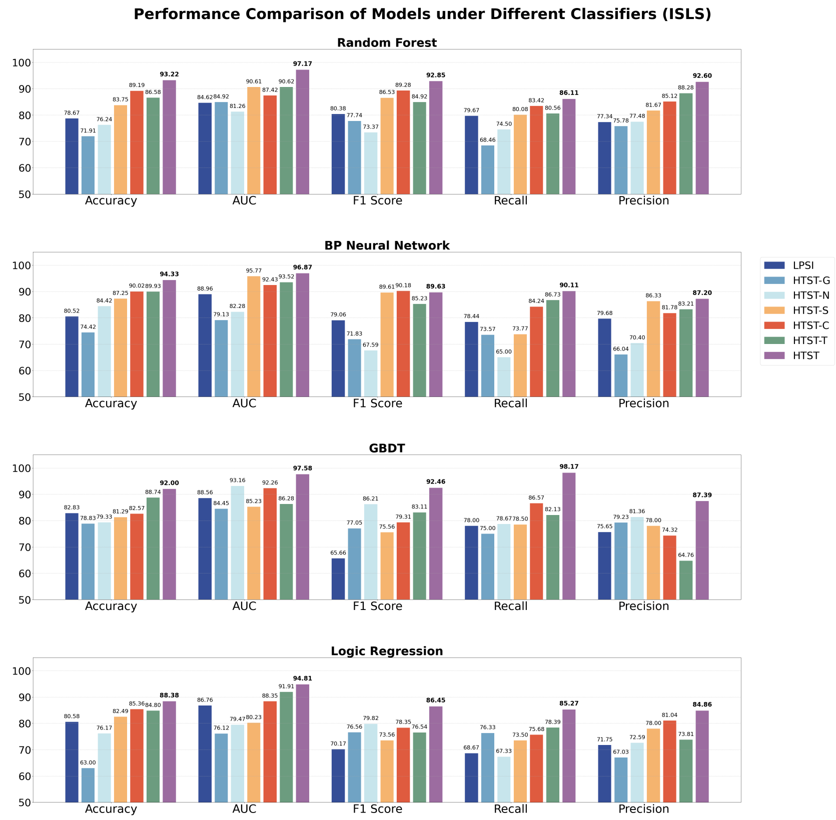


Fig.6 Experimental results in ISLS

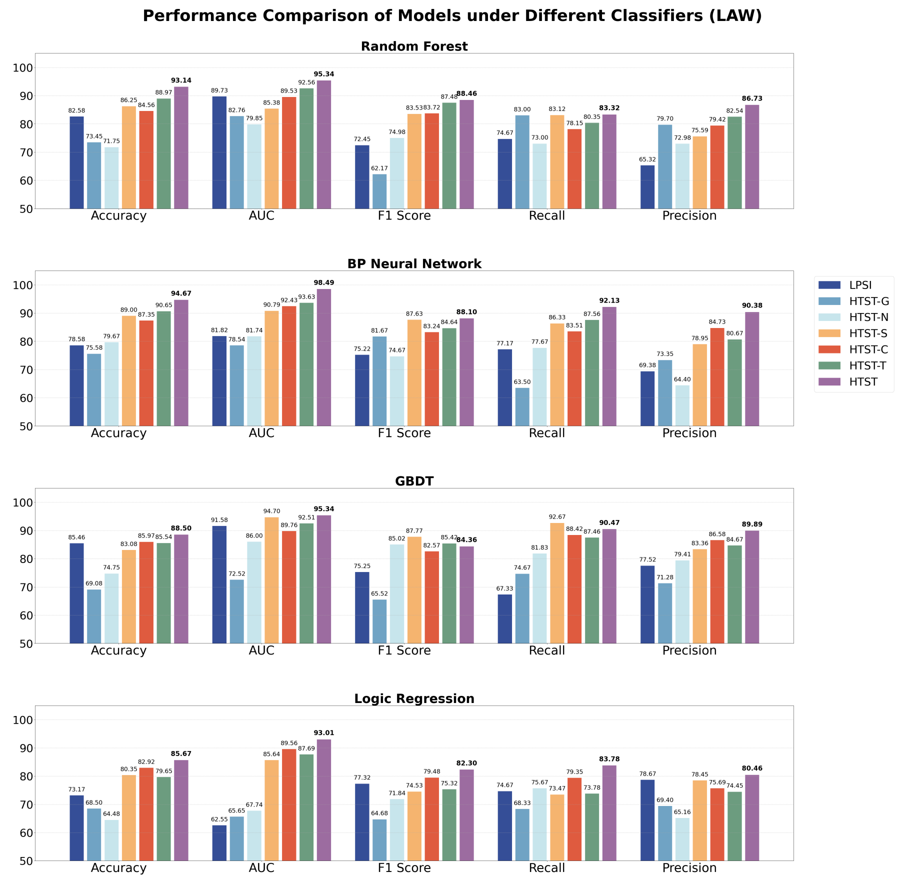


Fig.7 Experimental results in LAW

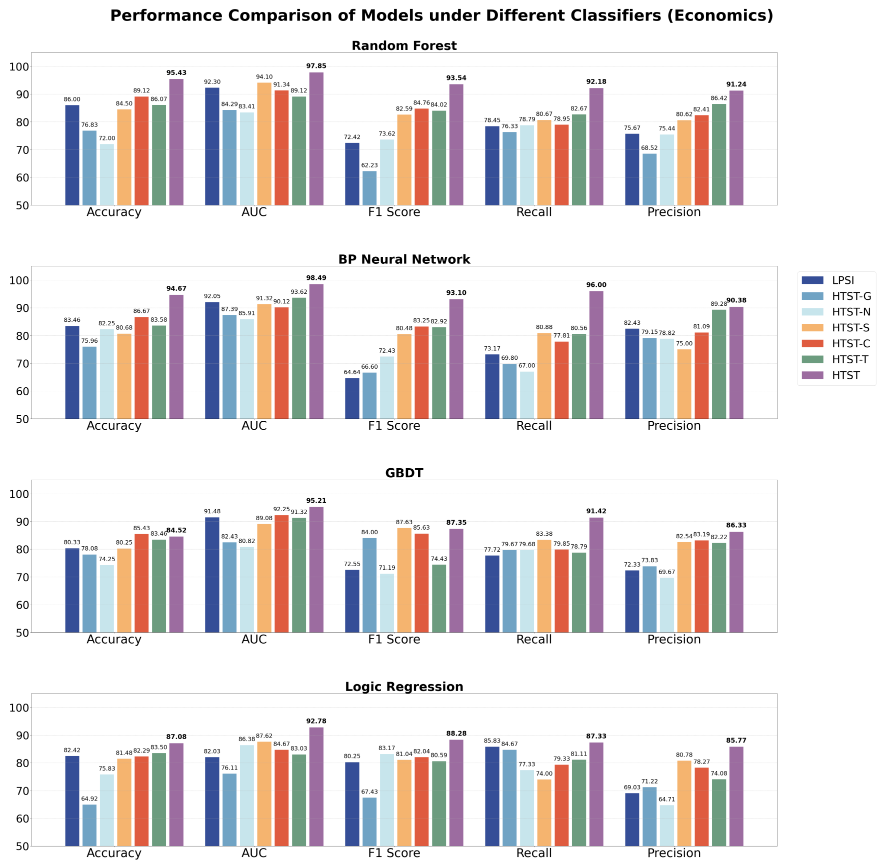


Fig.8 Experimental results in Economics

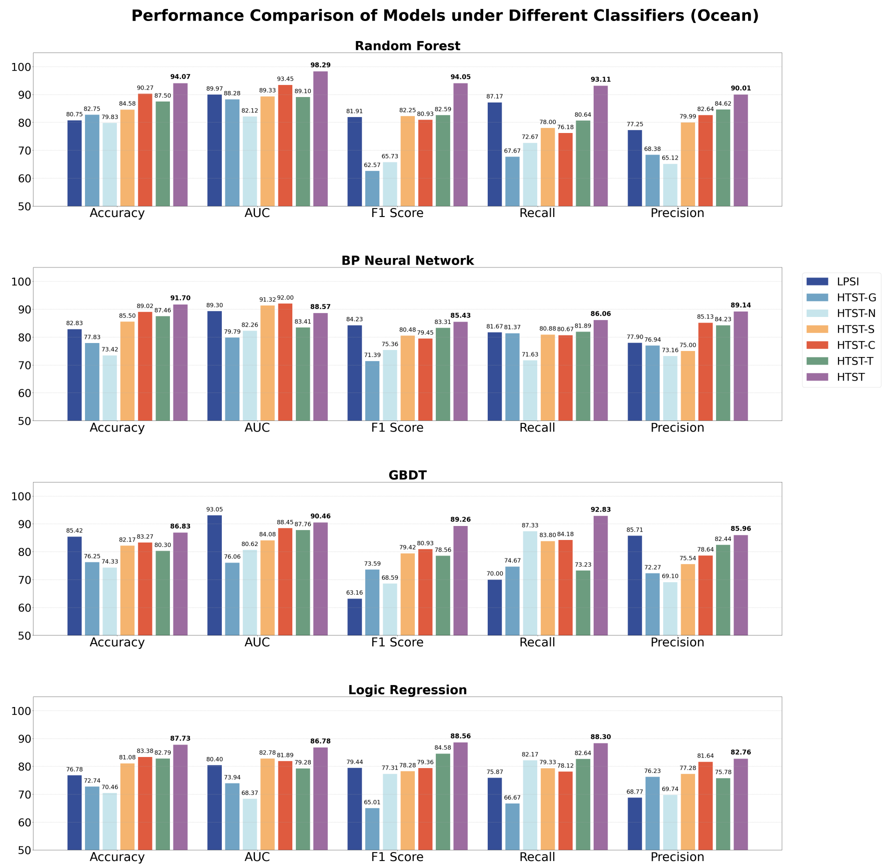


Fig.9 Experimental results in Ocean

As shown in Figs.6-9, our method HTST consistently achieves better performance than all baseline models across the five evaluation metrics in different disciplines. The main reason for this result may be that HTST effectively captures both the neighborhood structure and the semantic relationships between nodes. Given the data in Table 4 shows that our constructed networks have different sizes, we can conclude that that HTST exhibits excellent performance in co-word networks of various sizes and can significantly improve the link prediction effect with better robustness to network sizes. In addition, we can see that the LPSI consistently outperforms methods HTST-G and HTST-N individually; however, it is consistently outperformed by methods HTST-S, HTST- C and HTST-T.

From the Figs.6-9, we can also find that methods HTST-G, HTST-N, HTST-S, HTST-C and HTST-T all exhibit degraded performance compared to method HTST across various evaluation metrics in different disciplines. Therefore, we can confidently conclude that each component of our method is effective. Moreover, a performance comparison of the five methods (HTST-G, HTST-N, HTST-S, HTST-C and HTST-T) reveals that both methods HTST-G and HTST-N always demonstrate lower performance than methods HTST-S, HTST-C and HTST-T. These results suggest that the contribution of local and global topological structures in our method outweighs that of other components, such as the SAO structure for capturing semantics, GCN for semantic enhancement, and temporal weighting.

From Figs.6-9, we can also see that the Random Forest classifier consistently outperforms the other classifiers across most evaluation metrics, with particularly strong results in accuracy and AUC. This finding is consistent with those reported in previous research (Xiong et al., 2022), which demonstrated the excellent predictive performance of RF in the field of co-word networks. This is principally due to the fact that Random Forest combines the results of many decision trees, so it avoids the overfitting problem of a single tree and makes the model more reliable on high-dimensional data. In addition, the BP neural network performs nearly as well as Random Forest, particularly in terms of Recall. This is mainly because the BP network, with its multilayer structure and nonlinear activation functions, can capture complex feature interactions and subtle patterns in the data, which enhances its ability to correctly identify positive samples.

Impact of Sample Proportion on Model Performance

Given that positive and negative samples are inherently imbalanced in real-world network environments, with positives often being far fewer than negatives, a competent link prediction model must achieve good results not only with balanced samples but also under conditions of extreme imbalance. Only such a model possesses practical application value. Accordingly, to test whether our approach can maintain reliable predictions under different sample distributions, we examine how its performance changes with variations in the ratio of positive samples. To comprehensively investigate the impact of varying positive-to-negative sample ratios on model efficiency, we conducted extensive experiments across the four aforementioned datasets. In each discipline, we employ Random Forest (RF) as the classifier, and the outcomes are assessed using two metrics (Accuracy and AUC) against the baseline LPSI (red lines). The comparison is illustrated in Fig.10, where the x-axis denotes the positive sample ratio and the y-axis represent the corresponding performance values. Each curve illustrates how the model's predictive performance on a given dataset varies with changes in the proportion of positive samples.

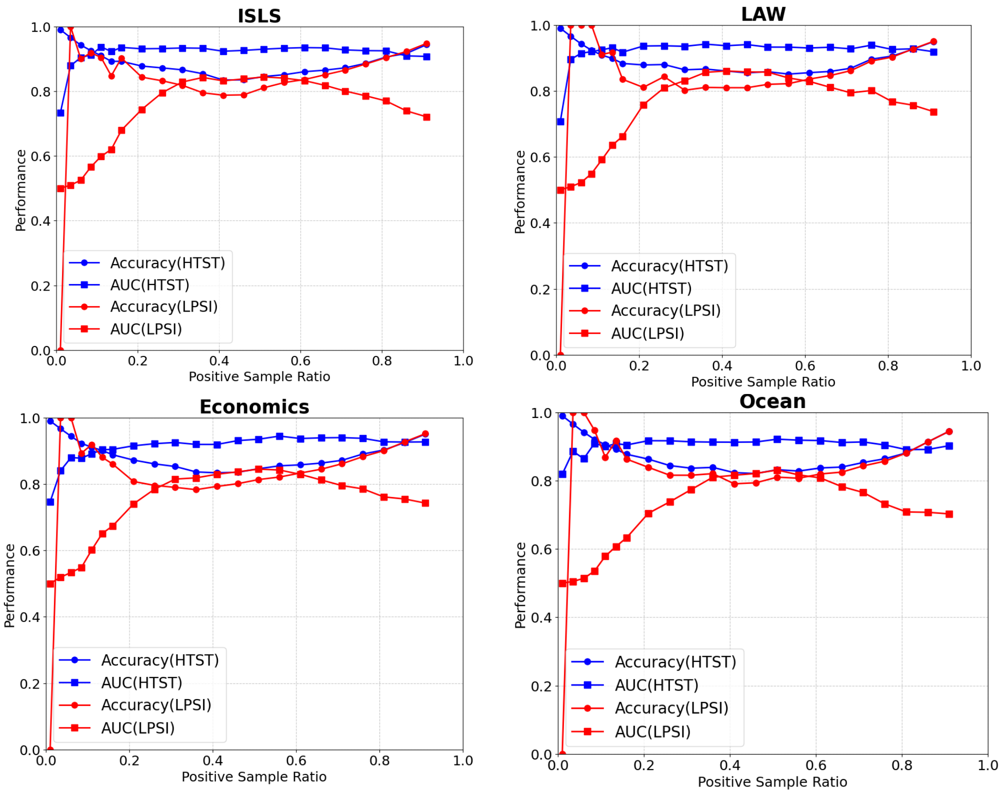


Fig. 10 Model performance under different sample ratios

As demonstrated in Fig.10, the performance of both methods is influenced by changes in the positive sample ratio. From the overall trends, accuracy initially decreases before increasing, while its AUC shows the opposite trend, first rising then falling. In comparison, LPSI is more sensitive to sample proportion variations than HTST, particularly with unstable performance in imbalanced data distributions. In contrast, our proposed HTST method demonstrates superior robustness. Its performance curves are consistently more stable across all datasets, with both Accuracy and AUC remaining at a high level regardless of the positive sample ratio. This is especially evident in the 0-0.2 range of the positive sample ratio, which is common in real-world network environments. For example, in the Economics dataset, when the positive sample ratio is below 0.1, LPSI's AUC approaches 0, while HTST's AUC remains stable above 0.8. In summary, our HTST method is not only superior to LPSI in overall performance but, more importantly, it exhibits exceptional robustness and practicality in diverse sample distributions, especially under extremely imbalanced conditions, which is critical for real-world link prediction tasks.

Case study

In this section, we further analyze the proposed method to substantiate its accuracy and reliability. To this end, we validate our model using literatures from the Information Science and Library Science (ISLS), an interdisciplinary field focused on how information is generated, organized, retrieved, and disseminated (Huang et al., 2021; Liu et al., 2024). As mentioned previously, we train the model on historical data (2014-2023) and evaluate its predictive performance on more recent data (2024). We first employ both empirical-based validation and expert-based evaluation, integrating objective and subjective perspectives to verify the prediction accuracy of our method. In addition, visualization techniques are employed to conduct comparative analysis of the prediction results from different methods.

Validation

Empirical-based validation

This method is crucial because it allows us to objectively assess whether the predicted keyword links correspond to actual co-occurrences reported in the most recent scientific publications, consequently grounding our predictions in real-world evidence (Huang et al., 2024). Note that our analysis is grounded in data from 2014 to 2023, with forecasts for **co-occurring word pairs** targeted at 2024. From the potential co-occurring keyword pairs predicted by our method for 2024, we randomly selected 10 pairs and attempted to empirically validate their co-occurrence relationships using literature published in that year. The validation results are recorded in the Table 7 below.

Table 7 Empirical-based validation of keyword pairs

No	Keyword 1	Keyword 2	Relevant documentary proof
1	Policy making	Open data	Cole et al. (2024) investigate the role of open science in policy making, highlighting how open data practices can enhance transparency and participatory decision-making.
2	Firm performance	Artificial intelligence	Wang et al. (2024) analyze the impact of artificial intelligence on firm performance, finding that AI adoption promotes innovation and environmental efficiency.
3	Blockchain	Records Management	Tahir et al. (2024) propose a blockchain-based healthcare records management framework that enhances security, privacy, and interoperability while addressing the limitations of centralized EHR systems.
4	ChatGPT	Academic Libraries	Hussain (2024) reviews the application of ChatGPT in academic libraries, showing its potential to serve as a virtual librarian that reduces staff workload while enhancing personalized user assistance.
5	Diffusion innovation	Social influence	Miranda et al. (2024) examine the diffusion of innovation through indirect social influence, showing that adoption is significantly affected by extended social connections.
6	User Acceptance	Technology perception	Sun and Zhang (2024) model the spread of risk perception in emerging technologies, shedding light on how user acceptance evolves through social contagion in networks.
7	Digitalization	Sustainable performance	Liang et al. (2024) study how digitalization capabilities influence sustainable performance, emphasizing the mediating role of green entrepreneurial orientation.
8	Collaboration	E-Government	Pan and Fan (2024) assess how policy attention affects e-government performance, highlighting the importance of public–private collaboration and strategic resource allocation.

9	Public administration	Resilience	Osei-Kyei et al. (2024) review strategies for enhancing resilience in urban governance, underscoring the role of public–private partnerships in disaster preparedness.
10	Data analytics	Innovation performance	Abera et al. (2024) explore the link between digital strategy and innovation performance, identifying innovation capability as a key mediator in microfinance institutions.

As shown in Table 7, the representative predicted edges are clearly reflected in the literature published in 2024， demonstrating the effectiveness and accuracy of our method.

Expert verification method

To enhance the reliability and practical relevance of our link prediction results, we incorporate an Expert knowledge-based validation. Since co-word network link prediction focuses on identifying emerging themes, the generated keyword pairs should possess genuine conceptual relevance. Such relevances cannot be determined solely through simple co-occurrence in literature but require evaluation by experienced domain experts. We invited five experts with extensive knowledge in the ISLS field to evaluate and score our predicted edges. We used a five-point Likert scale to score the predicted edges, where 1 to 5 respectively indicate increasing agreement that the predicted links align with the development trends in the ISLS field. Scores closer to 5 represent stronger consensus supporting the predicted associations, while scores near 1 suggest disagreement. Table 8 shows the result of Expert Evaluation.

Table 8 Expert Assessment-based validation of keyword pairs

No	Keyword 1	Keyword 2	Exp.1	Exp.2	Exp.3	Exp.4	Exp.5
1	Policy making	Open data	4	5	5	4	3
2	Firm performance	Artificial intelligence	5	4	5	5	4
3	Blockchain	Records Management	3	4	3	3	4
4	ChatGPT	Academic Libraries	4	5	4	4	5
5	Diffusion of innovation	Social influence	5	5	4	3	4
6	Technology	Perception	4	4	3	5	5
7	Digitalization	Sustainable performance	5	5	4	4	5
8	Collaboration	E-Government	5	4	4	4	4
9	Public administration	Resilience	5	4	5	4	5
10	Data analytics	Innovation performance	5	4	5	5	4

Based on the expert scores in Table 8, we calculated an average rating of 4.4, indicating strong consensus among experts and validating the effectiveness of our link prediction method.

Visualizing Network Evolution: A Macro-to-Micro Analysis of Link Prediction

This study further employs a visualization-based approach at both macro and micro levels to conduct a comparative analysis of the performance of the proposed methodology.

Macro-level visual analysis. At the macro level, the comparison centers on the overall structure of the observed future network versus the structure forecasted by the method. The macro-level visualization of the prediction results is provided in Fig. 11: G_HTST and G_LPSI represent the the visualization of the edges in the future network as predicted by HTST and LPSI, respectively; G_actual displays the visualization of the true edges in the future network, serving as the ground truth for our validation. All networks are constructed from links emerging in 2024, which are excluded from the training dataset (2014–2023). To avoid a cluttered network where excessive nodes and connections obscure the overall structure, this study constructs the graph using only keywords that appear more than three times in the testing set. It should be noted that all relevant areas across the three networks have been annotated with red boxes for the purposes of comparison.

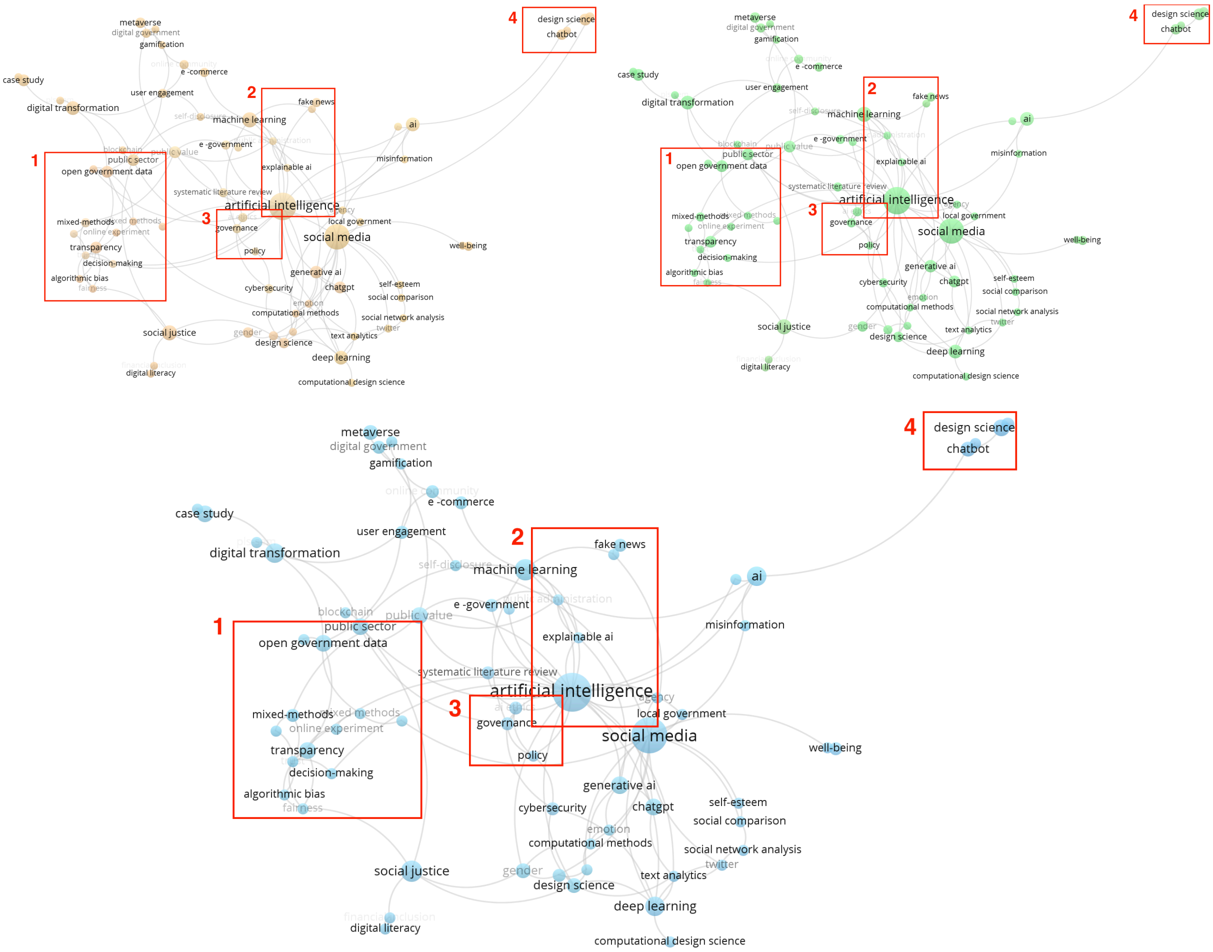


Fig.11 The predictive network G_LPSI (orange), the predictive network G_HTST (green), and the actual network G_actual (blue)

From the visual results in Fig.11, it can be observed that, compared to the network G_LPSI, the G_HTST network is more structurally consistent with the actual network G_actual. Through careful observation, we can identify the following phenomena. First, the G_actual is more strongly centered on the keywords "artificial intelligence" and "social media," a pattern well captured by G_HTST but less so by G_LPSI. This observation confirms that HTST effectively models future relationships, as evidenced by its accurate reflection of the rising focus on AI and social media in 2024. In addition, HTST captures emerging trends more effectively, such as "chatgpt" and "text analysis", which do not exist in G_LPSI. Since the connection between "chatgpt" and "text analysis" reflects the recent surge of research integrating large language models into text-based analytical tasks, its accurate prediction is noteworthy. Overall, these visual comparisons provide evidence that HTST is more effective than the baseline in modeling newly emerging relationships in future networks.

At a finer level of inspection, we can see specific mispredictions of LPSI. For example, in red box 4, LPSI incorrectly links "chatbot" with "design science," two concepts with minimal conceptual overlap. Similarly, it establishes an unwarranted connection between "fake news" and "artificial intelligence," despite their weak association, which is highlighted in red box 2. As presented in red box 1, LPSI incorrectly predicts a link between "mixed methods" and "transparency," two concepts that share minimal semantic relevance. In addition, LPSI fails to capture the strong relationship between "AI ethics" and "policy" , which is correctly reflected in HTST, as shown in red box 3. We can observe "AI ethics" and "policy" have strong semantic relationship, since we need policy to improve AI ethics, which demonstrates the model's capability of capturing latent semantic associations between conceptually related topics, rather than relying on mere co-occurrence frequency. Taken together, Fig.11 provides clear evidence that HTST achieves higher accuracy and greater semantic coherence than LPSI in predicting future links.

Micro-level visual analysis. To complement the macroscopic validation, a micro-level analysis is necessary for a more fine-grained comparison between the two methods. At the micro level, it examines a specific node, comparing its actual edge connections in a future time period against the connections predicted for it by the applied method.Consistent with the macro-level visualization, only keywords that appear more than three times in the test set are retained to ensure clarity and comparability of the network structure. By tracing the evolution of a single, highly central keyword, we can illustrate the effectiveness of our approach from a micro perspective. For this purpose, we selected the “Artificial Intelligence” node—a pivotal concept within the ISLS discipline—as a case study to shed light on the model’s inner mechanisms. This node is chosen because it occupies a central position in the network, exhibits a relatively high degree of connectivity, and frequently interacts with multiple thematic clusters, making it an ideal candidate for detailed analysis.

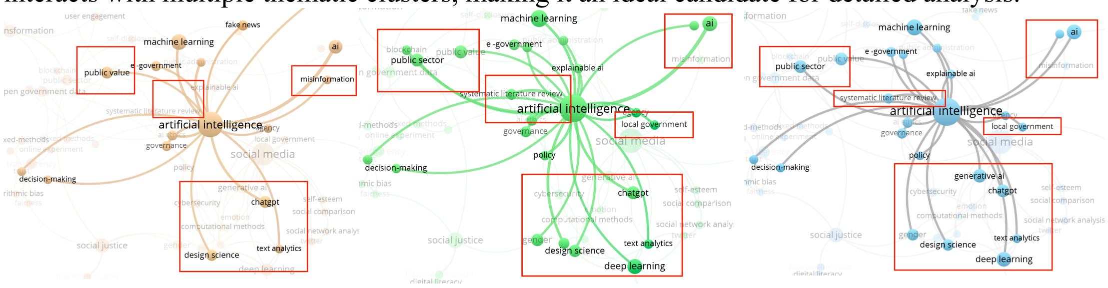


Fig.12 Focal node analysis of ‘Artificial Intelligence’ in the predictive network G_LPSI (orange), the predictive network G_HTST (green), and the actual network G_actual (blue)

As shown in the focal analysis in Figure 12, we compare the predictive edges related to the “Artificial Intelligence” node generated by the two methods with the actual edges in the G_actual network. The results indicate that the HTST model attains substantially higher precision than the LPSI model in forecasting the future relational pathways of the “Artificial Intelligence” node. For instance, as highlighted in the red boxes, HTST correctly predicts more actual links between “Artificial Intelligence” and related concepts such as “deep learning”, “policy”, and “explainable AI”, all of which are strongly associated with “Artificial Intelligence” in contemporary research and practice. This indicates that our HTST model effectively captures the deep semantic relationships among links, as “deep learning” and “explainable AI” are closely associated with “Artificial Intelligence.” Furthermore, given the central role of deep learning in the development of Artificial Intelligence, the strong linkage between “deep learning” and “Artificial Intelligence” further demonstrates the effectiveness of HTST in representing temporal features. In contrast, LPSI produces several erroneous predictions, such as links between “Artificial Intelligence” and “fake news” or “misinformation”, which share only marginal relevance. In summary, HTST demonstrates higher accuracy than LPSI in both the macro-level and focal node analyses.

Conclusions

This study proposes HTST (Hybrid Topology Representation and deep Semantics with Temporal Weighting for Link Prediction in Co-word Network), a novel framework that effectively integrates topology and semantic information for co-word network analysis. Unlike previous approaches that focused on either topological similarity or shallow semantic features, HTST simultaneously captures global and local topological characteristics through the combination of Node2Vec and GAT, while employing a deep semantic modeling to incorporate syntactic and contextual cues into node embeddings. Furthermore, the introduction of a time-decay model allows the framework to dynamically weight temporal signals, thereby improving its ability to model the evolving nature of scientific domains. Experimental results across four disciplines demonstrate that HTST consistently outperforms baseline models in both predictive accuracy and robustness, providing a powerful tool for uncovering emerging research trends and the dynamic evolution of knowledge structures.

Nevertheless, several promising directions remain for future exploration. First, while the exponential time-decay model effectively captures temporal trends, developing adaptive or self-learned temporal dynamics, could provide more flexible temporal modeling across disciplines with heterogeneous growth rates. Second, incorporating large language models (LLMs) or graph foundation models into semantic processing may further enhance the contextual representation of keywords and relations. Finally, integrating multimodal information (e.g., citation networks, author affiliations, or full-text embeddings) could enrich the interpretability of predicted links and offer a more comprehensive view of scientific knowledge evolution.

Reference

Abera, M., Marvadi, C., & Suthar, D. (2024). Digitization strategy and innovation performance of microfinance institutions: mediating role of innovation capability. *Journal of Accounting & Organizational Change*, 21 (2), 250–277. https://doi.org/10.1108/JAOC-02-2024-0026

Altendeitering, M., Guggenberger, T. M., & Möller, F. (2024). A design theory for data quality tools in data ecosystems: Findings from three industry cases. *Data & Knowledge Engineering*, 102333. https://doi.org/10.1016/j.datak.2024.102333

Austin, E., Makwana, S., Trabelsi, A., Largeron, C., & Zaïane, O. R. (2024). Uncovering flat and hierarchical topics by community discovery on word co-occurrence network. *Data Science and Engineering*, 9(1), 41-61. https://doi.org/10.1007/s41019-023-00244-y

Bedi, P., Gupta, N., & Jindal, V. (2021). I-SiamIDS: an improved Siam-IDS for handling class imbalance in network-based intrusion detection systems. *Applied Intelligence,* 51(2), 1133-1151. https://doi.org/10.1007/s10489-020-01886-y

Behrouzi, S., Sarmoor, Z. S., Hajsadeghi, K., & Kavousi, K. (2020). Predicting scientific research trends based on link prediction in keyword networks. *Journal of Informetrics*, 14(4), 101079. https://doi.org/10.1016/j.joi.2020.101079

Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A pretrained language model for scientific text. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3687–3691. https://doi.org/10.18653/v1/D19-1383

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. J*ournal of Machine Learning Research*, 3(Jan), 993–1022. https://doi.org/10.1162/jmlr.2003.3.1.993

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech.*, 2008(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Cai, H., Zheng, V. W., & Chang, K. C. C. (2018). A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE transactions on knowledge and data engineering*, 30(9):1616–1637. https://doi.org/10.1109/TKDE.2017.2766838

Cai, X., Liu, S., Yang, L., Lu, Y., Zhao, J., Shen, D., & Liu, T. (2022). COVIDSum: A linguistically enriched SciBERT-based summarization model for COVID-19 scientific papers. *Journal of Biomedical Informatics*, 127, 103999. https://doi.org/10.1016/j.jbi.2022.103999

Cao, X., Chen, X., Huang, L., Deng, L., Cai, Y., & Ren, H. (2024). Detecting technological recombination using semantic analysis and dynamic network analysis. *Scientometrics*, 129(11), 7385–7416. https://doi.org/10.1007/s11192-023-04812-4

Chaoguang, H., Yueji, H., Fanfan, H., & Chenwei, Z. (2025). An approach for interdisciplinary knowledge discovery: Link prediction between topics. *Physica A: Statistical Mechanics and its Applications*, 665, 130517. https://doi.org/10.1016/j.physa.2025.130517

Chatterjee, A., Ikica, B., Ravandi, B., & Palowitch, J. (2025). Transfer learning for temporal link prediction. arXiv preprint arXiv:2504.10925. https://doi.org/10.48550/arXiv.2504.10925

Chen, G., Hong, S., Du, C., Wang, P., Yang, Z., & Xiao, L. (2024). Comparing semantic representation methods for keyword analysis in bibliometric research. *Journal of Informetrics*, 18(3), 101529. https://doi.org/10.1016/j.joi.2024.101529

Chen, H., & Li, J. (2018). Exploiting structural and temporal evolution in dynamic link prediction. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 427-436. https://doi.org/10.1145/3269206.3271738

Chiu, S.-H., Teng, Y.-W., Yang, D.-N., & Chen, M.-S. (2025). Human-Centric Community Detection in Hybrid Metaverse Networks with Integrated AI Entities. In *Proceedings of the ACM on Web Conference 2025*, 3794–3808.

Choudhary, S., & Kumar, G. (2025). Enhancing link prediction in dynamic social networks through hybrid GCN-LSTM models. Knowledge and Information Systems, 67, 6717–6751. https://doi.org/10.1007/s10115-025-02430-5

Choudhury, N., & Uddin, S. (2016). Time-aware link prediction to explore network effects on temporal knowledge evolution. *Scientometrics*, 108(2), 745–776. https://doi.org/10.1007/s11192-016-2003-5

Cole, N. L., Kormann, E., Klebel, T., & Lindl, S. (2024). The societal impact of Open Science: A scoping review. *Royal Society Open Science*, 11(6), 240286. https://doi.org/10.1098/rsos.240286

Curiac, C.-D., Doboli, A., & Curiac, D.-I. (2022). Co-occurrence-based double thresholding method for research topic identification. *Mathematics*, 10(17), 3115. https://doi.org/10.3390/math10173115 de Bruin, G. J., Veenman, C. J., van den Herik, H. J., & Takes, F. W. (2021). Supervised temporal link prediction in large-scale real-world networks. *Social Network Analysis and Mining*, 11(1), 80. https://doi.org/10.1007/s13278-021-00806-y

Feng, J., Zhang, Y. Q., & Zhang, H. (2017). Improving the co-word analysis method based on semantic distance. *Scientometrics*, 111(3), 1521–1531. https://doi.org/10.1007/s11192-017-2286-1

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3–5), 75–174. https://doi.org/10.1016/j.physrep.2009.11.002

Gao, H., & Huang, H. (2018). Deep Attributed Network Embedding. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI)*, 3364–3370. https://doi.org/10.24963/ijcai.2018/467

Grover, A., & Leskovec, J. (2016). node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 855–864.

Gu, W., Gao, F., Lou, X., & Zhang, J. (2019). Link prediction via graph attention network. arXiv preprint arXiv:1910.04807. https://doi.org/10.48550/arXiv.1910.04807

Gu, W., Hou, J., & Gu, W. (2023). Improving link prediction accuracy of network embedding algorithms via rich node attribute information. *J. Soc. Comput.*, 4(4), 326–336. https://doi.org/10.23919/JSC.2023.0018

Guo, J., Wang, X., Li, Q., & Zhu, D. (2016). Subject–action–object-based morphology analysis for determining the direction of technological change. *Technological Forecasting and Social Change*, 105, 27–40. https://doi.org/10.1016/j.techfore.2016.01.028

Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30. https://doi.org/10.55948/cy2t3946

Hua, S., Chen, W., Li, Z., Zhao, P., & Zhao, L. (2020). Path-based academic paper recommendation. In: *International Conference on Web Information Systems Engineering*, 343–356. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-62584-0_26

Huang, A. (2008). Similarity measures for text document clustering. In *Proceedings of the Sixth New Zealand Computer Science Research Student Conference (NZCSRSC2008)*, 4, 9–56.

Huang, D., & Lei, F. (2023). Temporal group-aware graph diffusion networks for dynamic link prediction. *Information Processing & Management*, 60(3), 103292. https://doi.org/10.1016/j.ipm.2023.103292

Huang, L., Chen, X., Ni, X., Liu, J., Cao, X., & Wang, C. (2021). Tracking the dynamics of co-word networks for emerging topic identification. *Technological Forecasting and Social Change*, 170, 120944. https://doi.org/10.1016/j.techfore.2021.120944

Hussain, A. (2024). Unlocking the potential of ChatGPT in academic libraries. IP Indian J. Libr. Sci. Inf. *Technol.*, 9(2), 88–97. https://doi.org/10.18231/j.ijlsit.2024.015

Katsurai, M., & Ono, S. (2019). TrendNets: mapping emerging research trends from dynamic co-word networks via sparse representation. *Scientometrics*, 121(3), 1583–1598. https://doi.org/10.1007/s11192-019-03241-6

Khanam, K. Z., Singhal, A., & Mago, V. (2021). Noddle: Node2vec based deep learning model for link prediction. In: *National Conference on Big Data Technology and Applications (pp. 196–212)*. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-33614-0_14

Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR 2017).* https://arxiv.org/abs/1609.02907

Kumari, A., Behera, R. K., Sahoo, K. S., Nayyar, A., Kumar Luhach, A., & Prakash Sahoo, S. (2022). Supervised link prediction using structured-based feature extraction in social network. *Concurrency and Computation*, 34(13), e5839. https://doi.org/10.1002/cpe.5839

Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2010). Governance in social media: A case study of the Wikipedia promotion process. *ICWSM*, 4(1), 98–105. https://doi.org/10.1609/icwsm.v4i1.14013

Levy, O., & Goldberg, Y. (2014). Dependency-based word embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Association for Computational Linguistics, Baltimore, Maryland, 302–308. https://doi.org/10.3115/v1/p14-2050

Liang, Y., Koo, J. M., & Lee, M. J. (2024). The interplay of environmental dynamism, digitalization capability, green entrepreneurial orientation, and sustainable performance. *Sustainability*, 16(17), 7674. https://doi.org/10.3390/su16177674

Liben-Nowell, D., & Kleinberg, J. (2003). The link prediction problem for social networks. In *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, 556–559. https://doi.org/10.1145/956863.956972

Lin, J. C.-W., Tomasiello, S., & Srivastava, G. (2023). Integrated artificial intelligence in data science. *Applied Sciences*, 13(21), 11612. https://doi.org/10.3390/app132111612

Ling, C., Li, Z., Hu, Y., Zhang, Z., Liu, Z., Zheng, S., & Zhao, L. (2024). Link prediction on textual edge graphs. In arXiv preprint arXiv:2405.16606.

Liu, C., Han, Y., Xu, H., Yang, S., Wang, K., & Su, Y. (2024). A community detection and graph-neural-network-based link prediction approach for scientific literature. *Mathematics*, 12(3), 369. https://doi.org/10.3390/math12030369

Liu, J., Chen, J., Fan, C., & Zhou, F. (2023). Joint embedding in hierarchical distance and semantic representation learning for link prediction. In arXiv preprint arXiv:2303.15655.

Liu, J., Hua, Z., Xie, Y., Li, B., Shomer, H., Song, Y., Hassani, K., & Tang, J. (2025). Higher-order structure boosts link prediction on temporal graphs. In arXiv preprint arXiv:2501.03456.

Liu, M., Zhang, X., & Yan, Y. (2020). Research on method of determining scope of word set in co-word analysis based on word frequency, number of words, cumulative word frequency in proportion. *Journal of Information Science*, 29(4), 112–118. https://doi.org/10.13266/j.issn.0252-3116.2016.23.017

Liu, S., Xia, Y., & Wang, D. (2024). A human-in-the-loop anomaly detection architecture for big traffic data of cellular network. *IEEE Access*, 12, 41787–41797. https://doi.org/10.1109/ACCESS.2024.3394629

Liu, Y., Jin, M., Pan, S., Zhou, C., Zheng, Y., Xia, F., & Yu, P. S. (2022). Graph self-supervised learning: A survey. *IEEE transactions on knowledge and data engineering*, 35(6):5879–5900. https://doi.org/10.1109/TKDE.2022.3155798

Lü, L., & Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications*, 390(6), 1150–1170. https://doi.org/10.1016/j.physa.2010.11.027

Ma, M., Mao, J., & Li, G. (2024). Discovering weak signals of emerging topics with a triple-dimensional framework. *Information Processing & Management*, 61(5), 103793. https://doi.org/10.1016/j.ipm.2024.103793

Mane, K. K., & Börner, K. (2004). Mapping topics and topic bursts in PNAS. *Proceedings of the National Academy of Sciences*, 101(suppl_1), 5287–5290. https://doi.org/10.1073/pnas.0307626100

Martínez, V., Berzal, F., & Cubero, J. C. (2016). A survey of link prediction in complex networks. *ACM Computing Surveys (CSUR)*, 49(4), 1–33. https://doi.org/10.1145/2993469

Matthews, J. A., Starr, J. R., & van Schijndel, M. (2024). Semantics or spelling? Probing contextual word embeddings with orthographic noise. arXiv preprint arXiv:2408.04162. https://arxiv.org/abs/2408.04162

McAllister, J. T., Lennertz, L., & Atencio Mojica, Z. (2022). Mapping a discipline: a guide to using VOSviewer for bibliometric and visual analysis. *Science & Technology Libraries*, 41(3), 319-348. https://doi.org/10.1080/0194262X.2021.1991547

Miranda, M., Pereda, M., Sánchez, A., & Calvo, M. G. (2024). Indirect social influence and diffusion of innovations: An experimental approach. *PNAS nexus*, 3(10), pgae409. https://doi.org/10.1093/pnasnexus/pgae409

Pan, T., & Fan, B. (2024). How does policy attention affect e-government performance? The role of resource allocation and public–private collaboration. *International Review of Administrative Sciences*, 90(2), 369–384. https://doi.org/10.1177/00208523241228229

Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2020). Pre-trained models for natural language processing: A survey. *Science China technological sciences*, 63(10):1872–1897. https://doi.org/10.1007/s11431-019-1504-2

Osei-Kyei, R., Ampratwum, G., Tam, V. W. Y., & Torgbor, E. (2024). Building urban community resilience against hazards through public-private partnerships: A review of critical resilience strategies. *Buildings*, 14(7), 1947. https://doi.org/10.3390/buildings14071947

Quispe, L. V. C., Tohalino, J. A. V., & Amancio, D. R. (2021). Using word embeddings to improve the discriminability of co-occurrence text networks. *Physica A: Statistical Mechanics and its Applications*, 562, 125344. https://doi.org/10.1016/j.physa.2020.125344

Schwarz, A., & Chin, W. W. (2024). Information technology acceptance: Construct development and empirical validation. *International Journal of Information Management*, 78, 102810. https://doi.org/10.1016/j.ijinfomgt.2024.102810

Shang, K. K., Yi, J., Small, M., & Zhou, Y. (2025). Triadic Closure-Heterogeneity-Harmony GCN for Link Prediction. *arXiv preprint arXiv:2504.20492*.

Shi, F., & Evans, J. (2023). Surprising combinations of research contents and contexts are related to impact and emerge with scientific outsiders from distant disciplines. *Nature Communications*, 14(1):1641. https://doi.org/10.1038/s41467-023-37330-8

Si, S., Wang, J., Zhang, R., Su, Q., & Xiao, J. (2022). Federated non-negative matrix factorization for short texts topic modeling with mutual information. *2022 International Joint Conference on Neural Networks (IJCNN)*, 1–7. https://doi.org/10.1109/IJCNN55064.2022.9892305

Sugiyama, K., & Kan, M. Y. (2010). Scholarly paper recommendation via user's recent research interests. In: *Proceedings of the 10th annual joint conference on Digital libraries*, 29–38. https://doi.org/10.1145/1816999.1817006

Small, H., Boyack, K. W., & Klavans, R. (2014). Identifying emerging topics in science and technology. *Research Policy*, 43(8), 1450–1467. https://doi.org/10.1016/j.respol.2014.02.005

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Tahir, N. U. A., Rashid, U., Hadi, H. J., Ahmad, N., Cao, Y., Alshara, M. A., & Javed, Y. (2024). Blockchain-based healthcare records management framework: Enhancing security, privacy, and interoperability. *Technologies*, 12(9), 168. https://doi.org/10.3390/technologies12090168

Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., & Mei, Q. (2015). LINE: Large-scale Information Network Embedding. In *Proceedings of the 24th International Conference on World Wide Web*, 1067–1077. https://doi.org/10.1145/2736277.2741093

Tylenda, T., Angelova, R., & Bedathur, S. (2009). Towards time-aware link prediction in evolving social networks. In *Proceedings of the 3rd workshop on social network mining and analysis*, 1–10. https://doi.org/10.1145/1731011.173102

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018). Graph attention networks. In *International Conference on Learning Representations (ICLR 2018)*. https://doi.org/10.48550/arXiv.1710.10903

Wang, A., Luo, K., & Nie, Y. (2024). Can artificial intelligence improve enterprise environmental performance: Evidence from China. *Journal of Environmental Management*, 370, 123079. https://doi.org/10.1016/j.jenvman.2024.123079

Wang, D., Cui, P., & Zhu, W. (2016). Structural Deep Network Embedding. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, San Francisco, California, USA, 1225–1234. https://doi.org/10.1145/2939672.2939753

Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics,* 10(2), 347–364. https://doi.org/10.1016/j.joi.2016.02.003

Wang, X., Cheng, Q., & Lu, W. (2014). Analyzing evolution of research topics with NEViewer: a new method based on dynamic co-word networks. *Scientometrics*, 101(2), 1253–1271. https://doi.org/10.1007/s11192-014-1347-y

Wang, Y., Li, X., Zhou, X., & Ouyang, J. (2021). Extracting topics with simultaneous word co-occurrence and semantic correlation graphs: Neural topic modeling for short texts. In *Findings of the Association for Computational Linguistics: EMNLP 2021*. Association for Computational Linguistics, Punta Cana, Dominican Republic, 18–27. https://doi.org/10.18653/v1/2021.findings-emnlp.2

Wang, Z.-Y., Li, G., Li, C.-Y., & Li, A. (2011). Research on the semantic-based co-word analysis. *Scientometrics*, 90(3), 855–875. https://doi.org/10.1007/s11192-011-0563-y

Xenopoulos, P. (2017). Introducing DeepBalance: Random deep belief network ensembles to address class imbalance. In *2017 IEEE International Conference on Big Data (Big Data)*, 3684-3689. https://doi.org/10.1109/BigData.2017.8258288

Xiong, T., Zhou, L., Zhao, Y., & Zhang, X. (2022). Mining semantic information of co-word network to improve link prediction performance. *Scientometrics*, 127(6), 2981–3004. https://doi.org/10.1007/s11192-021-04247-9

Yang, C., Zhu, D., & Wang, X. (2017). SAO Semantic Information Identification for Text Mining. *IJCIS*, 10(1), 593. https://doi.org/10.2991/ijcis.2017.10.1.40

Yang, J., & Wu, Y. (2024). Link prediction based on depth structure in social networks. *International Journal of Machine Learning and Cybernetics*, 15(10), 4639-4657. https://doi.org/10.1007/s13042-024-02178-4

Yang, Y. Y., Chou, C. N., & Chaudhuri, K. (2022). Understanding rare spurious correlations in neural networks. *arXiv preprint arXiv:2202.05189*. https://arxiv.org/abs/2202.05189

Yang, Y., Xiang, Y., & Zhang, C. (2024). Identifying Emerging Topics in Specific Domains via Novelty Analysis of Entities in Future Work Sentences from Academic Articles. In: *Joint Workshop of the 2th Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data (IMSC2024)*. Hong Kong, China. CEUR-WS.org/Vol-3982/paper2.pdf

Yao, L., Mao, C., & Luo, Y. (2019). Graph convolutional networks for text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01), 7370-7377. https://doi.org/10.1609/aaai.v33i01.33017370

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

Zhang, Y.-J., Yang, K.-C., & Radicchi, F. (2021). Systematic comparison of graph embedding methods in practical tasks. *Phys. Rev. E*, 104(4), 044315. https://doi.org/10.1103/PhysRevE.104.044315

Zhang, H., Huating, S., & Wu, X. (2020). Topic model for graph mining based on hierarchical Dirichlet process. *Statistical Theory and Related Fields*, 4(1):66–77. https://doi.org/10.1080/24754269.2020.1770966

Zhao, Y., Fesharaki, N. J., Li, X., Patrick, T. B., & Luo, J. (2018). Semantic-Enhanced Query Expansion System for Retrieving Medical Image Notes. *J. Med. Syst.*, 42(6), 105. https://doi.org/10.1007/s10916-018-0954-1

Zhou, W., Huang, H., Shi, R., Song, X., Lin, X., Wang, X., & Jin, H. (2023). Temporal Heterogeneous Information Network Embedding via Semantic Evolution. *IEEE Trans. Knowl. Data Eng.*, 35(12), 13031–13042. https://doi.org/10.1109/TKDE.2023.3287260

Zhu, X., & Zhang, Y. (2020). Co-word analysis method based on meta-path of subject knowledge network. *Scientometrics*, 123(2), 753–766. https://doi.org/10.1007/s11192-020-03400-0

Zhao, Y., Yin, J., Zhang, J., et al. (2023). Identifying the driving factors of word co-occurrence: a perspective of semantic relations. *Scientometrics*, 128:6471–6494. https://doi.org/10.1007/s11192-023-04740-4

Zuckerman, M., & Last, M. (2019). Using Graphs for Word Embedding with Enhanced Semantic Relations. In: *Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing* (TextGraphs-13), 32–41. Hong Kong: Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-3304