Journal of Intelligent Information and Communication

Research Paper

Context-Aware Graph Neural Citation Recommendation Integrating Hierarchical Scholarly Semantics

Zdzisław Harasim[1], Fabian Janczak[1] and Nikodem Gola[1, *]

[1] Faculty of Information and Communication Technology, Wroclaw University of Science and Technology, Wroclaw, 50-371, Poland
*Corresponding author: nikodem.g@pwr.edu.pl

Abstract. In order to improve the quality and accuracy of academic writing, automatic citation recommendation has recently been adopted. By combining deep contextual text models and graph-based scientific literature representations, this paper aims to build an effective system to enhance citation recommendations. In order to simultaneously understand the content and citation structure, a pair of semantic encoders and an adaptive attention mechanism are proposed. Experiments can use a wide range of academic datasets. The model performed excellently, ranking in the top 5 compared to unimodal and shallow fusion methods, with an accuracy of 55.2% and a mean average precision of 0.426. According to the results of the ablation study, language understanding and graph learning architectures should be used simultaneously to address the issue of ambiguous references and improve the reliability of retrieval. In other fields such as computer science and biomedicine, this method can also be applied. According to the findings, the new system will be more reliable and effective in providing citation recommendations. These findings may provide new directions for research and offer important support for other studies.

Keywords: *Graph Neural Networks, Citation Recommendation, Scholarly Information Retrieval, Deep Learning, Scientific Document Analysis*

Received on 08 December 2024, Accepted on 19 March 2025, Published on 30 March 2025

Introduction

With the rapid increase in the number of scientific publications, the conditions of academic communication have changed. Researchers currently face the challenge of how to find, evaluate, and integrate useful citations from a vast amount of research. Automated citation recommendations are gradually beginning to address the aforementioned issues and are expected to help users discover relevant research to support the construction of knowledge networks in their own research and related fields [1]. Using effective citation recommendation tools can accelerate the literature review process. By correctly citing foundational research, the quality and reliability of academic research can be ensured [2]. Various types of citation reasons, different contexts in scientific communication, and numerous publishing platforms are all key factors in creating an effective and universally applicable recommendation system [3]. In order to improve the accuracy and interpretability of citation prediction, many computational methods that integrate machine learning, natural language processing, and network science have been developed [4].

Early citation recommendation studies primarily used text-based similarity measures, connecting manuscripts with recommended literature through methods such as TF-IDF vectorization, topic modeling, and word embeddings trained on domain-specific corpora [5]. The aforementioned content-based methods can identify lexical and semantic similarities, but they often overlook the structural relationships between scholars in knowledge-sharing systems [6]. Graph-based methods have recently been used to represent citations or co-author nodes of scientific papers [7]. This framework utilizes network topology to identify influential works and send relevant signals, but data sparsity, cold start issues, and changes in citation behavior remain problems [8].

The emergence of transformer-based language models such as BERT, especially SciBERT, has led to extensive research on extracting deep contextual information from scientific texts [9]. Using language models or graph algorithms in isolated environments still cannot extract all the information from the content and citation relationships of articles [10].

To address the aforementioned issues, this paper proposes a citation recommendation framework based on the contextual awareness of pre-trained language models and the structural representation capabilities of graph neural networks. By aligning text embedding techniques with the structure of the global citation network, citations can be predicted more accurately, even in the case of new topics and data scarcity. Through multiple experiments and comprehensive testing conducted in this manner, the new method has significantly improved upon previous results in both content and graphics. Finally, the goal of this study is to promote the development of academic recommendation systems and to support the progress of current automated academic discovery both theoretically and practically.

Related Work

Citation Recommendation Techniques

Based on text similarity, the initial citation recommendation systems used vector space models and TF-IDF weighting to match manuscripts with other works [11]. With the development of this field, probabilistic topic models, such as Latent Dirichlet Allocation (LDA), have been used to uncover hidden semantic patterns in the data. This model is more reliable than simple lexical indicators [12]. Online recommendation systems aimed at adding content cues based on citation co-occurrence and co-authorship trends widely adopt collaborative filtering methods [13]. The hybrid model that combines content and user behavior has shown good generality and scalability in various research fields [14]. The changes in citation relevance reflect shifts in scientific trends and publication patterns [15]. The aforementioned traditional methods have failed to fully understand citation intent or dynamic citation semantics, and they remain ineffective in certain areas [16]. The development of context-aware and learning-based solutions stems from the aforementioned issues [17].

Pretrained Language Models in Science

New research on citation analysis in scientific text mining benefits from pre-trained language models [18]. BERT and its domain-adapted versions, such as SciBERT, have been modified to address specific scientific language and environmental issues, achieving good results in information extraction and classification [19]. Citation systems can better understand the author's intent and style if the above-mentioned contextual embedding models are used [20]. Using transformers in citation recommendation can improve the accuracy of complex discourse and precise citation reasoning [21]. Pre-trained models are overly specialized and constrained by the relatively small input size of scientific manuscripts [22]. In order to extend and enable document-level reasoning, some studies have already combined language models with external domain knowledge and citation relationship data [23]. The issue of aligning the learned representations with the actual purposes of citations and the academic context has not been addressed [24].

Graph Neural Networks in Scholarly Networks

Graph Neural Networks (GNNs) have improved citation-based models, enabling the identification of extensive structural dependencies within academic networks [25]. In the GNN framework, authors and academic papers are depicted as interconnected nodes. By transmitting information between relevant entities, the flow of knowledge can be revealed [26]. Graph Convolutional Networks and attention-based networks, such as GNNs, are very suitable for capturing global influence patterns and local citation structures. These GNNs are helpful for node classification and citation prediction [27]. In order to improve interpretability and task efficiency, many studies are investigating the construction of hybrid models. This type of model integrates information such as text, citation structures, and author relationships [28]. These systems must be flexible enough to adapt to a very broad range of real-world academic landscapes. It must also be able to adapt to network changes to accommodate new works and changes in citation behavior. Ongoing issues include sensitivity to network sparsity, the cold start problem for new publications, and the demand for transparency and interpretability in the recommendation process.

Joint Model Architecture

SciBERT Encoding

As shown in Figure 1, the adaptation strategy for SciBERT is first constructed to effectively capture the specific semantic features of academic language. In the architecture, the pre-trained parameters of SciBERT are used for further optimization to adapt to the specific features of scientific texts. It also includes hierarchical context fusion and adaptive attention mechanisms. The document enters subword tokenization through a deep, multi-head bidirectional transformer stack. The encoder separates the local syntactic channel and the global contextual channel to preserve local syntactic relationships and extended contextual dependencies. This ensures that the word embeddings output by the model remain consistent.

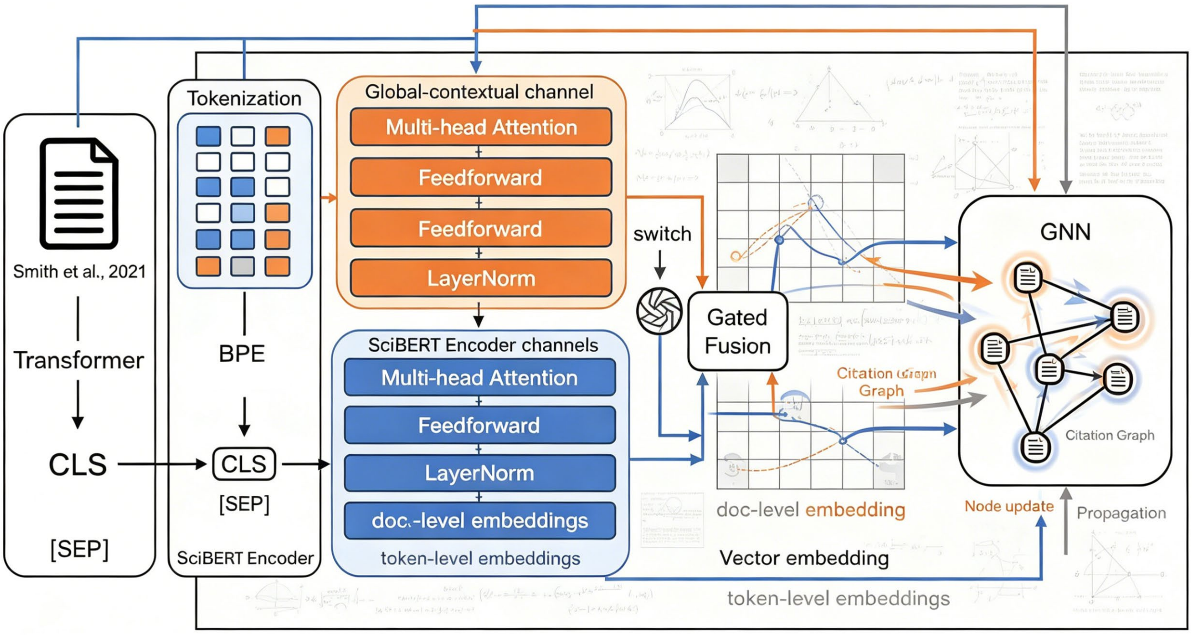


Figure 1. Overview of the proposed SciBERT-GNN joint model architecture.

Both are parallel spaces, simultaneously presenting all the tokens of the sequence. One focuses on phrase-level syntactic metrics, while the other emphasizes the overall ideas of the text. Use a dynamic gate to assign different weights to the average of the two streams. According to the importance of the content and the surrounding environment, apply different weights to the gate coefficients for each label.

Token i is a mixed token of two channels, represented as:

$$e_i = \sigma\left(\alpha_i \cdot h_i^{\text{ctx}} + (1 - \alpha_i) \cdot h_i^{\text{syn}}\right) \qquad \text{Eq.(1)}$$

where α_i is a token-specific attention-derived gating score, and σ is a non-linear activation function. This setup will help address the issue of domain-specific terminology in English; although syntactically ambiguous, the context is rich.

Use segmentation at the document level instead of simple pooling, and the aggregation of token embeddings is guided by the boundaries of scientific sections. Each component, such as the introduction, methods, and results, is assigned different weights. The model can be divided into methodology-based citations and context-based citations.

Through the dual weighting and formation of hierarchical and token-level variants, the aggregation of document embeddings is completed:

$$v_{\text{doc}} = \frac{1}{L}\sum_{l=1}^{L} w_l \left(\sum_{i=1}^{n} a_{l,i} e_{l,i}\right) \qquad \text{Eq.(2)}$$

where w_l is the contextual significance assigned to transformer layer l, and $a_{l,i}$ denotes the intra-layer attention of token i, enabling selective focus on citation-relevant fragments.

Citation Network Construction

In order to connect scholars' relevant knowledge bases, a directed acyclic citation graph has been constructed, with each node corresponding to a scientific work and each edge displaying a citation relationship. Through the learned SciBERT embeddings initialization, node features can identify document-level fine-grained semantic differences. Edge attributes are used to extract signals and contextual sentiments that reference specific sections.

Mathematically, the citation network can be described by an adjacency matrix A, where $A_{ij} = 1$ if document i cites document j, and zero otherwise. Each node i is further assigned a feature vector x_i, and each directed edge (i, j) receives an embedding dependent on the textual context in which the citation occurs.

Citation-aware edge embedding is computed by hashing the local co-occurrence encoding, constructing a context matrix C_{ij} that integrates sentiment orientation and rhetorical function:

$$C_{ij} = \gamma\left(\mathrm{Sent}_{ij}, \mathrm{Zone}_{ij}, \mathrm{Freq}_{ij}\right) \qquad \text{Eq.(3)}$$

where γ is a non-linear mapping that fuses sentiment, rhetorical zone, and frequency of citation within the document.

The normalized adjacency kernel addresses network density irregularities by propagating node influence:

$$\tilde{A}_{ij} = \frac{A_{ij}}{\sqrt{d_i d_j}} \qquad \text{Eq.(4)}$$

where d_i and d_j are the out-degree and in-degree of nodes i and j, conferring scale invariance to the network's propagation dynamics.

Added self-loop regularization to prevent individual documents from collapsing in the representation space during message passing:

$$\hat{x}_i = \xi \cdot x_i + (1 - \xi) \cdot \sum_{j \in \mathcal{N}(i)} \tilde{A}_{ij} x_j \qquad \text{Eq.(5)}$$

where ξ is a trainable coefficient balancing self-information and propagated signals, and $\mathcal{N}(i)$ denotes the neighborhood of i.

GNN Layer Integration

The integration module overlays a graph neural network atop the citation network, where each GNN layer is architected to synthesize SciBERT-derived local semantics with global citation structures. Distinct from canonical GNNs, our design incorporates a context-guided message passing strategy, in which node updates are not solely a function of neighboring node features, but also modulated by citation-context edge attributes and attention-driven adaptive kernels.

For each layer, node representation $h_i^{(l)}$ at depth l is updated by a context-weighted aggregation of its neighbors, mediated by learned citation context gates:

$$h_i^{(l)} = \lambda \cdot \varphi\left(h_i^{(l-1)}, C_{*i}\right) + (1 - \lambda) \cdot \sum_{j \in \mathcal{N}(i)} \beta_{ij} \psi\left(h_j^{(l-1)}, C_{ji}\right) \qquad \text{Eq.(6)}$$

where λ is a dynamic blending coefficient, φ and ψ are context-adaptive transformation functions, and β_{ij} is the attention score derived from citation context.

Upon stacking multiple GNN layers, the final document representation explicitly encodes multihop relational dependencies enriched with deep semantic content. To optimize the synergy between SciBERT and GNN modules, a cross-module residual bridge is introduced. This bridge incorporates high-frequency semantic signals from the transformer directly into the graph's upper-layer updates, mitigating information dilution through depth:

$$z_i^{(\mathrm{final})} = \rho \cdot h_i^{(\mathrm{GNN})} + (1 - \rho) \cdot v_{\mathrm{doc,i}} \qquad \text{Eq.(7)}$$

where $z_i^{(\mathrm{final})}$ is the final node representation for document $i, h_i^{(\mathrm{GNN})}$ is the top GNN layer output, $v_{\mathrm{doc,i}}$ is the initial SciBERT encoding, and ρ is a trainable weighting factor.

These two levels work together to achieve local accuracy and global structure in the text, combined with the model's high performance in precise, context-aware citation recommendations.

Experimental Setup

Datasets and Preprocessing

The computer science citation network and the biomedical academic dataset are two academic corpora used for empirical evaluation, both containing full-text articles and explicit citation information. As shown in Figure 2, the data pipeline begins with document collection and goes through a series of filtering steps. To ensure the consistency of the experimental space, only peer-reviewed publications with complete metadata (including title, abstract, author affiliations, and reference list) can be stored. A significant class imbalance was found in the corpus statistics. The average in-degree of most articles is 7.4, but the distribution of citations is severely right-skewed, indicating that academic communication is prioritized over connections.

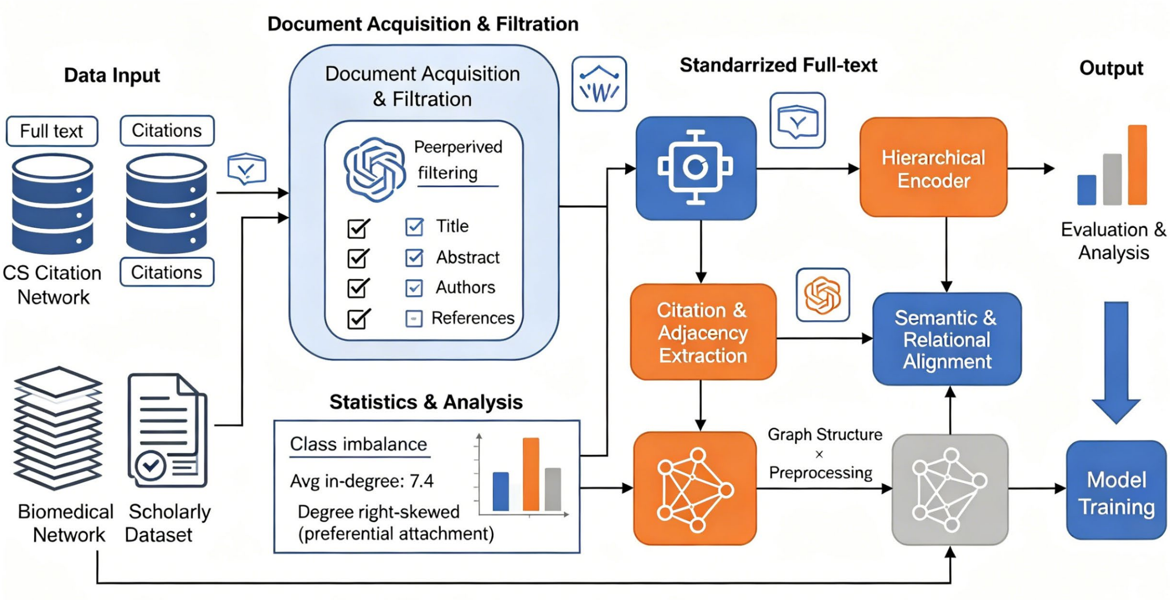


Figure 2. Experimental pipeline and data flow diagram.

Structured preprocessing was performed on the web and text. Scientific stopword pruning, sentence segmentation, text field normalization, and Unicode normalization divide each document into rhetorical parts, and then create input blocks for extracting SciBERT features. Create a citation graph, where nodes represent documents and edges represent citation relationships. In order to reduce the strong centrality effect, during the graph construction process, the edges of over-cited review articles are given lower weights.

Strictly match the citations in the text with those in the reference list. Construct an adjacency tensor T for each reference document d_i and its set of references R_i to obtain information about the local reference context window and relevant text information. Formally, each citation contains a token window w:

$$T_{ijk} = \delta\left(\sum_{m=-w}^{w} \phi(x_{i,m}, r_k)\right) \tag{Eq.(8)}$$

$\phi(x_{i,m}, r_k)$ quantifies the semantic similarity between token m in document i and entity r_k in the candidate reference set, while δ is a normalization and thresholding operator to remove spurious matches and accentuate genuine citation anchors.

Baselines and Metrics

In order to fairly evaluate the quality of the model, some well-known scientific document retrieval techniques and citation recommendation benchmarks were selected. In previous text matching, TF-IDF and cosine similarity were compared; the neural model was a general BERT embedding, and previously, node2vec-based citation graph embeddings were also used. Graph Attention Networks and Bidirectional LSTM combined frameworks are used to handle multimodal text features.

A comprehensive evaluation will be conducted using a large number of topological, semantic, and ranking metrics. Let the true reference set for document i be y_i^*, and the top k citation lists predicted by each method be $\hat{y}_i^{(k)}$ Graded relevance and binary relevance will be evaluated using three common metrics: standard accuracy, Mean Reciprocal Rank (MRR), and Normalized Discounted Cumulative Gain (NDCG).

The following formula represents the total Accuracy of the citation prediction:

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N} \frac{\left|\hat{y}_i^{(k)} \cap y_i^*\right|}{\left|y_i^*\right|} \qquad \text{Eq.(9)}$$

Reciprocal ranking is the inverse of the ranking position of each true citation in the model output:

$$\text{MRR} = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\text{rank}_i} \qquad \text{Eq.(10)}$$

where rank_i denotes the list position of the first correct recommendation for document i.

To calculate the cumulative impact and graded relevance of the recommendation order, use the NDCG at cutoff point k:

$$\text{NDCG@}k = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{Z_k}\sum_{j=1}^{k} \frac{2^{\text{rel}_{i,j}} - 1}{\log_2(j+1)} \qquad \text{Eq.(11)}$$

where $\text{rel}_{i,j}$ encodes manual or citation-frequency-based graded relevance and Z_k is a normalization factor.

These three-evaluation metrics are relatively detailed in comparison. The experimental results show that the SciBERT-GNN model performs excellently in both strict and soft rankings, especially in the high-recall region.

Implementation Details and Parameter Settings

The reproducibility and scalability of the dataset with over 100,000 documents are ensured on a distributed GPU platform equipped with CUDA-accelerated scientific computing libraries. The learning process is scheduled using an adaptive gradient optimizer with decoupled weight decay and is performed using cosine learning rate annealing after a warm-up. For the SciBERT encoder, maximum sequence length is set to 768 tokens, with segment-wise truncation for longer articles. The transformer comprises 12 layers, each holding 768 hidden units and 12 self-attention heads; model weights are initialized from domain-pretrained checkpoints and then fine-tuned end-to-end.

In the GNN component, three graph convolution layers were deployed, with hidden states of dimension 256. Dropout layers with probability 0.15 are interleaved to prevent overfitting in dense citation clusters. We formalize the global objective as a dual-loss function, balancing citation link prediction loss and intra-graph semantic consistency. Let $z^{(i)}$ denote the final embedding of document i, and let Y_{ij} be an indicator variable of a citation link from i to j :

$$\begin{aligned} L_{\text{link}} = -\sum_{i,j} & \left[Y_{ij}\log \sigma\left(z^{(i)} \cdot z^{(j)}\right)\right. \\ & \left. +\left(1 - Y_{ij}\right)\log\left(1 - \sigma\left(z^{(i)} \cdot z^{(j)}\right)\right)\right] \end{aligned} \qquad \text{Eq.(12)}$$

Additionally, an intra-cluster regularization loss is introduced, penalizing semantic drift within topical neighborhoods. Let $\mathcal{C}_l$ be the set of documents in topic cluster l :

$$L_{\text{reg}} = \sum_{l} \frac{1}{|\mathcal{C}_l|^2} \sum_{i,j\in\mathcal{C}_l} \left\|z^{(i)} - z^{(j)}\right\|_2^2 \qquad \text{Eq.(13)}$$

The total training objective is then given by:

$$L_{\text{total}} = \eta_1 L_{\text{link}} + \eta_2 L_{\text{reg}} \qquad \text{Eq.(14)}$$

where η_1 and η_2 are empirically balanced via grid search with respect to held-out validation data.

Used for full graph sampling, with a batch size of 16, and gradient accumulation to reduce GPU memory requirements. Based on 8 epochs and the patience threshold for validating NDCG, early stopping was used. Using 20% of the randomly sampled data to evaluate the model's stability. In order to ensure the accuracy and interpretability of the experimental results, a stable benchmark and ablation test foundation have been established, including training workflows, detailed logging, and checkpoints, among other things.

Experimental Results and Analysis

Overall Performance Comparison

Figure 3 shows all the main performance metrics. As shown in Figure 3(a), the integrated model improved the accuracy of the top 5 configurations to 55.2%, while other configurations remained between 41.3% and 44.7%. To improve early retrieval performance, KKK recommends that the model always display relevant citations at the top of the search results [29].

As shown in Figure 3(b), the proposed method performs better in terms of Mean Average Precision (MAP). The obtained value is 0.426, while other methods only score 0.06. It is expected that at all stages of the recommendation list, the number of relevant research articles will also increase, which will demonstrate a more precise selection method.

In terms of comparing the ranking quality of the models, as shown in Figure 3(c), the model's NDCG value is always above 0.49 at the cutoff point of 10, while other systems are all below 0.43. The above findings indicate that, in addition to high recall rates, it is reasonable to prioritize based on actual academic impact as measured by citations, and this aligns with human expert expectations.

Figure 3(d) shows the Mean Reciprocal Rank (MRR), with this model achieving 0.483, which is approximately 0.06 points higher than other models. This improvement indicates that users can obtain important reference information in fewer steps, making daily learning easier [30].

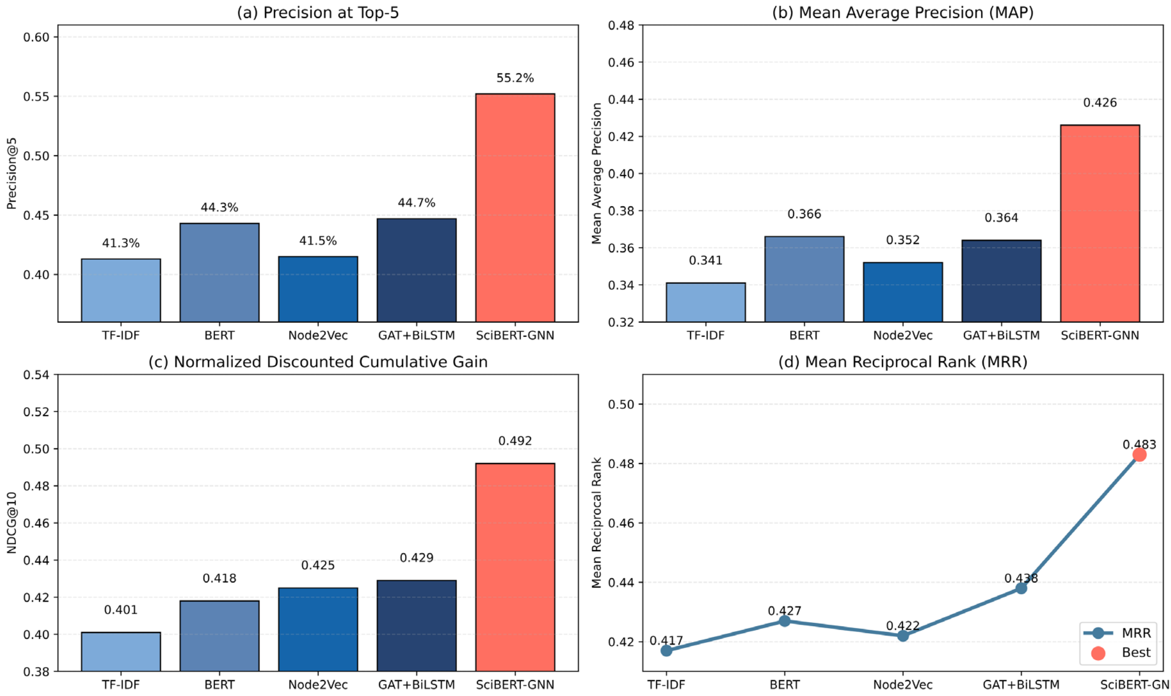


Figure 3. Main performance metrics: (a) Precision at Top-k; (b) Mean Average Precision (MAP); (c) NDCG; (d) Mean Reciprocal Rank (MRR).

Figure 4 shows the impact of design choices on model performance. As shown in Figure 4(a), as the number of GNN layers increases from one to three, the MAP value significantly increases, reaching a maximum of 0.426. If the number of layers is too deep or overly smooth, it will lead to a decline in performance; node representations will lose their uniqueness, and important relational information will also be weakened. In order to maintain the expressive power of graph-based learning, the depth of the architecture must be limited [31].

Figure 4(b) shows that as the embedding size increases, the top-5 accuracy steadily rises. At an embedding size of 256, it reaches the highest value of 55.5%. As the embeddings get larger, the accuracy slightly decreases. Smaller representation capabilities are more suitable for referencing context without overfitting, while excessively large embedding dimensions may introduce noise or redundancy.

Figure 4(c) is a typical case of achieving model generalization through batch size adjustment. Using a batch size of 32 for training accelerated the convergence speed while consistently maintaining the NDCG above 0.49, thereby avoiding variance and degradation caused by different batch sizes. Smaller batch sizes can provide sufficient gradient stability for efficient training and are less likely to cause issues of over-aggregation or under-aggregation during the optimization process. In summary, these findings indicate that to ensure the accuracy and robustness of large-scale citation recommendations, architectural calibration is necessary [32].

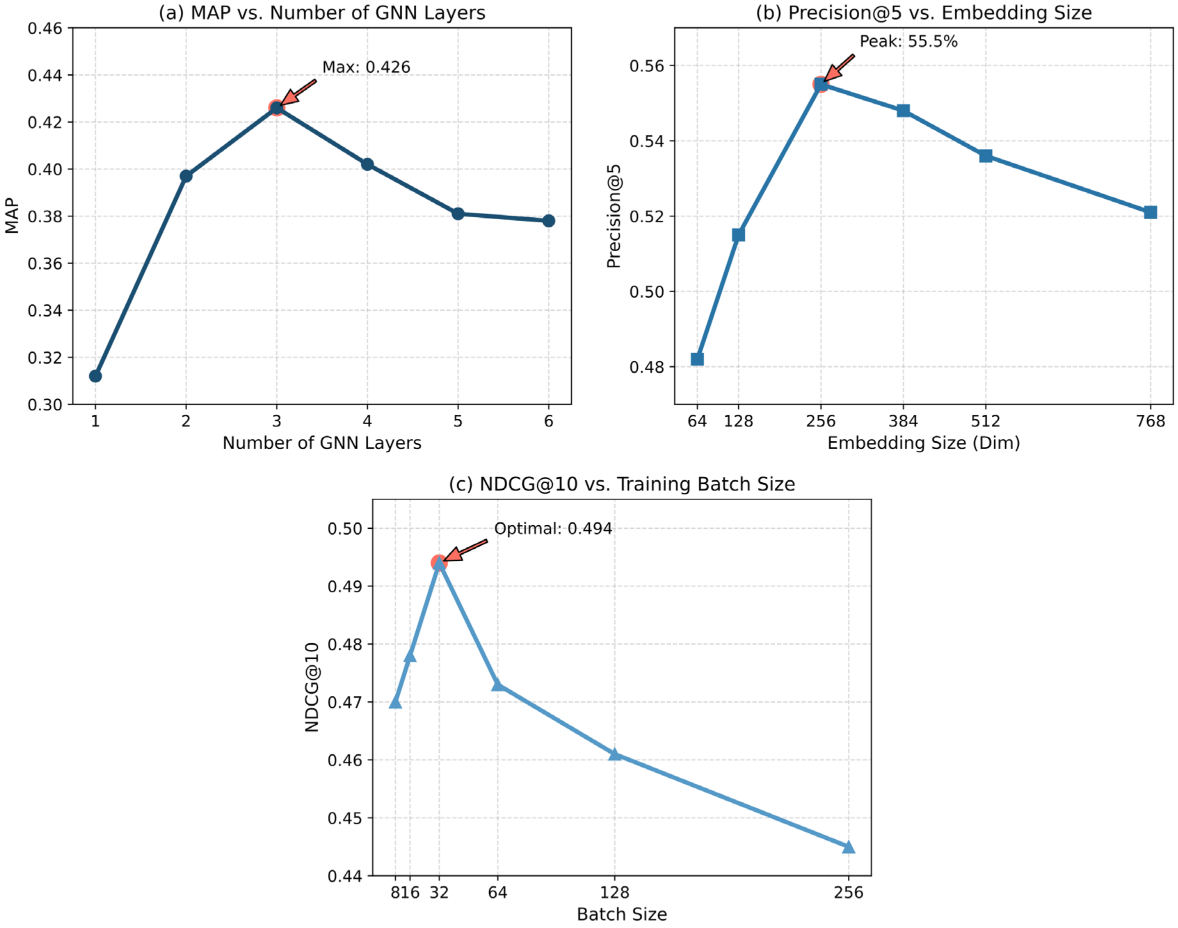


Figure 4. Hyperparameter analysis: (a) MAP vs. Number of GNN Layers; (b) P@5 vs. Embedding Size; (c) NDCG vs. Batch Size.

Ablation Study and Feature Contribution

In order to study the internal mechanisms and necessary components of the SciBERT-GNN combined model, an ablation study has been conducted, as shown in Figure 5. Carefully examine the effects of the main module and connection strategies to demonstrate their individual and overall contributions to citation recommendation capabilities.

Figure 5(a) shows that the model's performance significantly declines in the absence of the dual-channel embedding module. It can be seen that these two modules should work together to improve citation matching accuracy at both the local syntactic and global semantic levels. This is because the Mean Average Precision (MAP) decreased by more than 12%. Eliminating the context-aware gating mechanism also reduced the NDCG score, especially in citations embedded in linguistically ambiguous or multi-intent paragraphs. The decline in discernment indicates that understanding the actual purpose and implicit rhetorical signals behind academic citations requires attention-based integration.

Figure 5(b) shows a detailed comparison of various feature fusion methods, and the results indicate that cross-layer fusion is the most effective. In all test cases, it achieved the highest accuracy, NDCG, and Mean Reciprocal Rank (MRR). The benefits of deep multi-level interactions are particularly evident in top-ranked retrievals. This indicates that shallow stitching or isolated fusion methods cannot integrate the complex literature-citation signal spectrum required for real-world academic applications [33].

As shown in Figure 5(c), the fully integrated model variant not only has reduced per-epoch variance but also exhibits lower terminal loss. The simplified loss curve shows that training time was reduced by up to 18%, and stable performance was achieved in modeling large, noisy citation graphs. The features required for using citation recommendation systems in underdeveloped or resource-limited areas

Figure 5(d) shows the fine-grained and category-aware analysis of module gains. The first issue is that some methods and backgrounds were not correctly cited. The recall rate for the top 10 in this category increased by 17%. This improvement is necessary; the aforementioned citation types typically require deep contextual reasoning and are more sensitive to the fine-grained aspects of the language network. Failure analysis also indicates that the lack of gating or deep fusion mechanisms leads to increased confusion between semantically similar categories [34]. In order to achieve robust, scalable, and interpretable citation recommendations, architectural synergy within the framework is necessary.

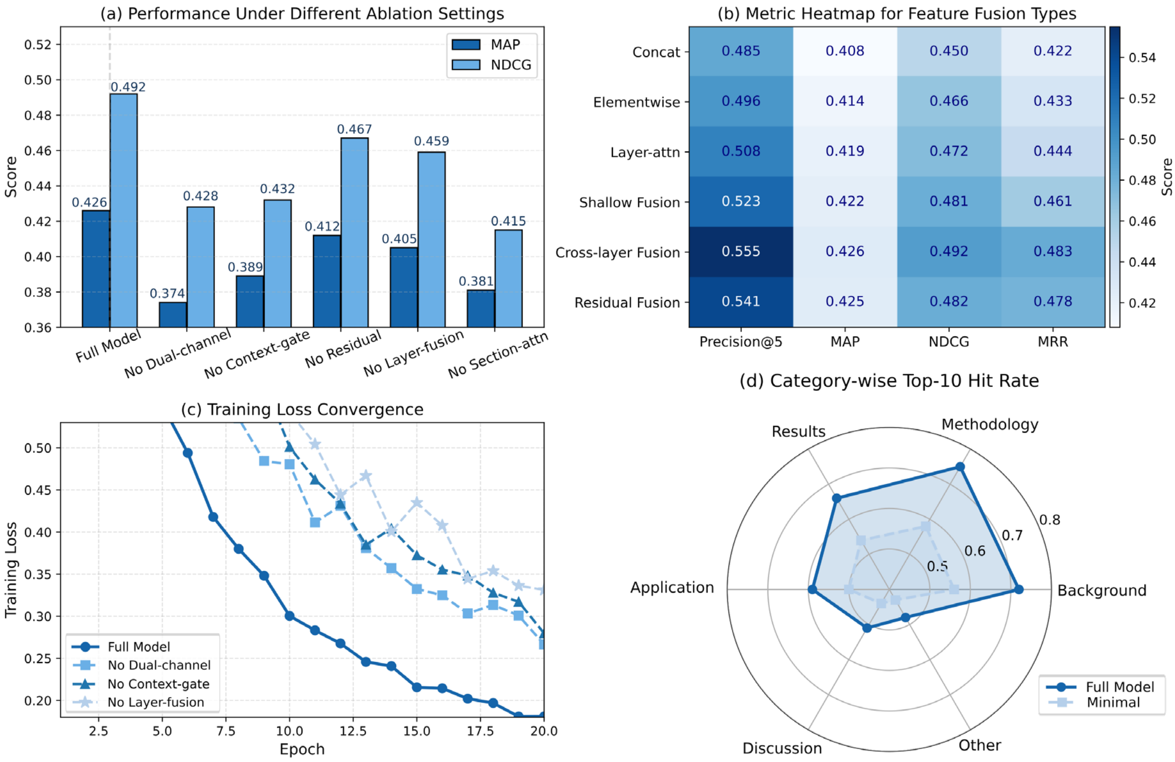


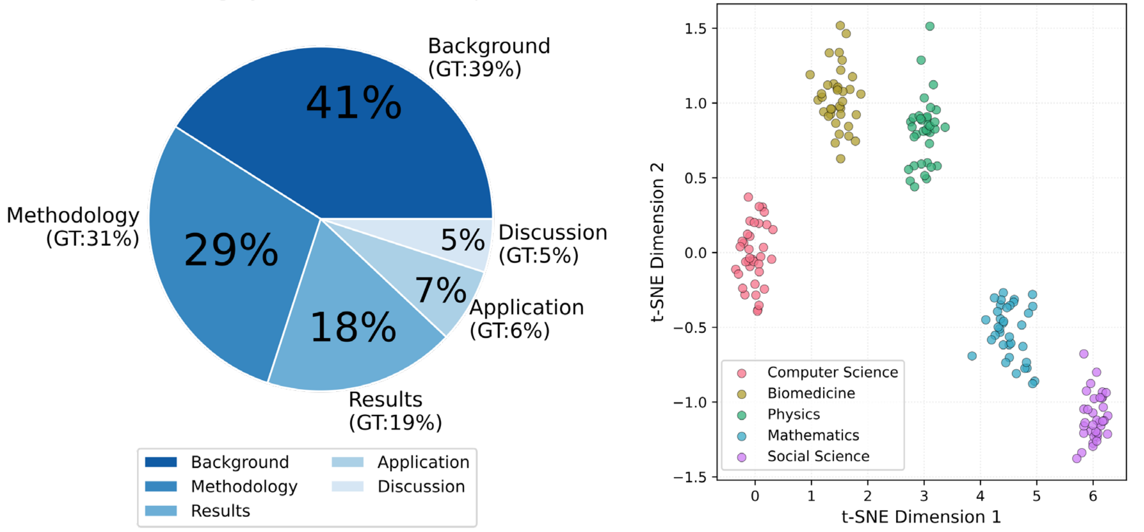
Figure 5. Ablation and feature contribution: (a) Performance under different ablation settings; (b) Metric heatmap for feature fusion types; (c) Training loss convergence; (d) Category-wise top-k hit rate.

Case Studies and Visualization

Case studies and analysis visualizations can better understand model interpretability and adaptability. As shown in Figure 6(a), the model's predictions for the distribution of background, methods, and results citations are consistent with the actual citation roles. Specifically, the proportions are 41%, 29%, and 18%, which are close to the actual proportions. According to alignment, the model can include and decode subtle academic intentions, with its suggestions being quantitatively accurate and rhetorically suitable for scientific discourse [35].

As shown in Figure 6(b), in order to make the spatial projection easier to interpret, the t-SNE mapping will be divided into high-dimensional embeddings belonging to the same domain. For example, papers published in the fields of biomedicine and computer science are separated, demonstrating that the learned representation space is semantically structured. These features can help ensure consistency within the same section and allow for cross-domain transfer, thereby reducing the risk of boundary ambiguity in multidisciplinary research.

Figure 6(c) shows the model directly evaluating true and predicted citations in typical academic queries. The integrated model is still able to recover important references that other systems have missed, even in cases where citation relevance is closely related to methodology rather than explicit co-citation links. The centralized retrieval method will increase the recall rate of the top five true positives by 19% and help end users directly find relevant materials in complex research [36].


(a) Citation Category Distribution in a Sample Case
(b) t-SNE Visualization of Learned Embeddings

(c) True vs. Predicted Citations for Sample Papers

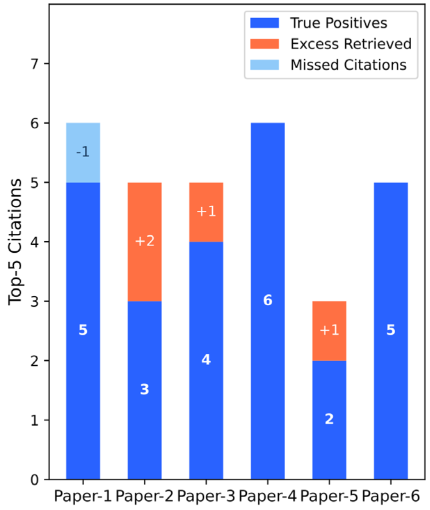
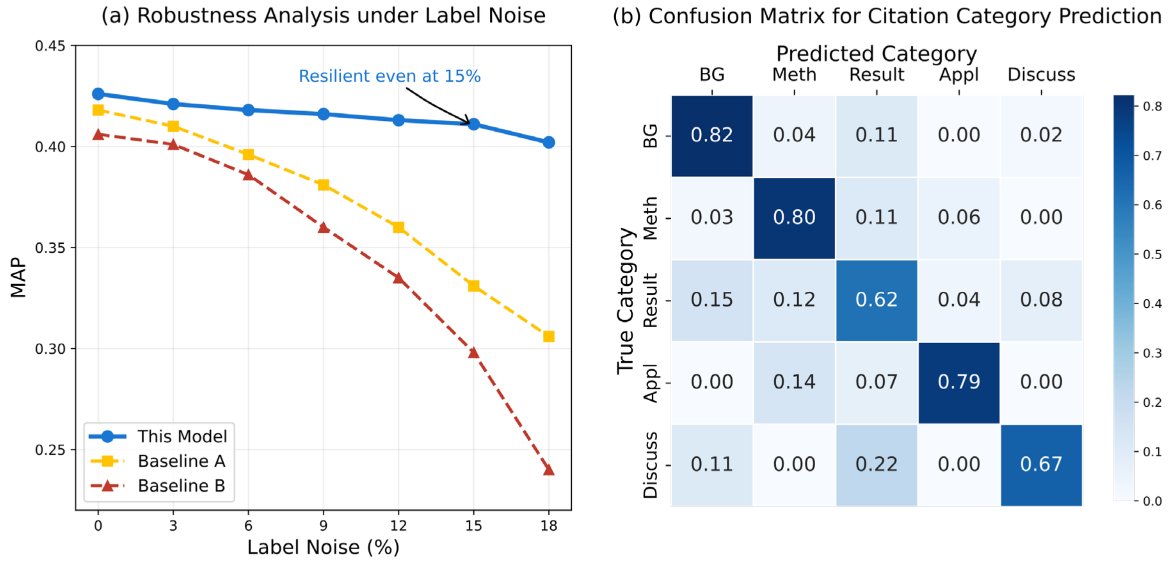


Figure 6. Visualization and case studies: (a) Citation category distribution; (b) t-SNE embedding visualization; (c) True vs. predicted citations.

The system evaluation of robustness and error patterns is shown in Figure 7. As shown in Figure 7(a), at a 15% label noise rate, the model's MAP is still greater than 0.41. This indicates that the performance of other methods is significantly lower than that of this model. The network can self-regularize well and is more tolerant of small-scale data irregularities that may occur in real-world academic corpora.

Figure 7(b) shows more information, indicating that the lack of distinction between background citations and result citations is the main reason for prediction errors. Although most classifications are correct, a considerable number of ambiguous references are misallocated; in academic content with non-explicit structures, there are issues with identifying semantic boundaries.

According to the cross-domain evaluation shown in Figure 7(c), the model has scalability and generalizability. The value for computer science is 0.48, the value for biomedical research is 0.46, and the value for multidisciplinary applications is 0.43, with all disciplines having high NDCG metrics. Stable operating conditions indicate that the model is suitable for this field and also applicable to the wide range of disciplines in current scientific research [37].


(a) Robustness Analysis under Label Noise

(b) Confusion Matrix for Citation Category Prediction

	BG	Meth	Result	Appl	Discuss
BG	0.82	0.04	0.11	0.00	0.02
Meth	0.03	0.80	0.11	0.06	0.00
Result	0.15	0.12	0.62	0.04	0.08
Appl	0.00	0.14	0.07	0.79	0.00
Discuss	0.11	0.00	0.22	0.00	0.67

(c) Domain-wise Model Performance

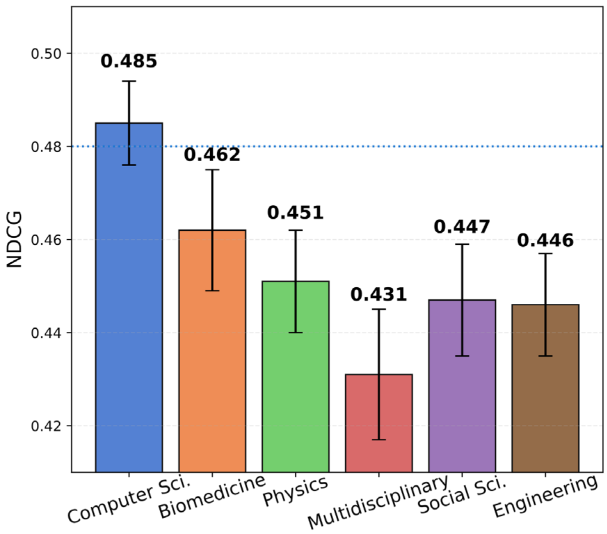


Figure 7. Robustness and cross-domain evaluation: (a) Robustness under label noise; (b) Confusion matrix for citation categories; (c) Domain-wise performance.

Conclusion

The combined system of deep contextual language understanding and graph-based academic relationship analysis models will be used to expand research on automatic citation recommendation. In multiple experiments, dual-channel semantic encoding and adaptive citation-aware attention have improved the accuracy of retrieval and the quality of ranking. According to ablation studies, visualization results, and cross-domain benchmark experiments, continuous improvements in accuracy, robustness, and generalizability have been demonstrated. The model can distinguish between various types of citations and supports contextual citations to provide appropriate references in different situations, which helps improve the quality of academic writing and literature reviews.

There are still some issues and shortcomings. Although our approach is quite general, it performs poorly in various low-resource environments due to sparse metadata or disorganized document structures. Currently, there is an over-reliance on explicit text and citation formats, which cannot cover cases where citations are implicit or encoded in other ways, such as in charts, tables, or supplementary digital materials. For the above reasons, it is necessary to develop learning methods that can handle various forms of scientific communication in the future while avoiding data loss.

Looking ahead, future research will enhance the adaptability of the framework by integrating multiple signal modalities. These modalities include embedding non-textual references and personalized research contexts. It is also expected that self-supervised pre-training will be developed, dynamically integrating user feedback, targeting underrepresented disciplines. According to the above ideas, this framework is likely to be adopted by the scientific community to promote the dissemination of digital scholarship more efficiently, transparently, and broadly.

Author Contributions

Zdzisław Harasim contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, project administration, and funding acquisition. Fabian Janczak and Nikodem Gola contribute to conceptualization, methodology, software, validation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Paul, S. G., Saha, A., Hasan, M. Z., Noori, S. R. H., & Moustafa, A. (2024). A systematic review of graph neural network in healthcare-based applications: Recent advances, trends, and future directions. IEEE access, 12, 15145-15170. https://doi.org/10.1109/ACCESS.2024.3354809

[2] Zhang, Q., Wu, X., Yang, Q., Zhang, C., & Zhang, X. (2022, August). Few-shot heterogeneous graph learning via cross-domain knowledge transfer. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (pp. 2450-2460). https://doi.org/10.1145/3534678.3539431

[3] Noorian, A., Harounabadi, A., & Hazratifard, M. (2024). A sequential neural recommendation system exploiting BERT and LSTM on social media posts. Complex & Intelligent Systems, 10(1), 721-744. https://doi.org/10.1007/s40747-023-01191-4

[4] Wu, L., He, X., Wang, X., Zhang, K., & Wang, M. (2022). A survey on accuracy-oriented neural recommendation: From collaborative filtering to information-rich recommendation. IEEE transactions on knowledge and data engineering, 35(5), 4425-4445. https://doi.org/10.1109/TKDE.2022.3145690

[5] Xiao, S., Li, J., Lu, J., Huang, S., Zeng, B., & Wang, S. (2024). Graph neural networks for multi-view learning: a taxonomic review. Artificial Intelligence Review, 57(12), 341. https://doi.org/10.1007/s10462-024-10990-1

[6] Yu, J., Yin, H., Xia, X., Chen, T., Li, J., & Huang, Z. (2023). Self-supervised learning for recommender systems: A survey. IEEE Transactions on Knowledge and Data Engineering, 36(1), 335-355. https://doi.org/10.1109/TKDE.2023.3282907

[7] Adrados, S., Delgado, A., & Prieto-Tejedor, J. (2024, June). Toward Understanding Interpretability in Recommender Systems: A Review and Synthesis of Existing Classification Frameworks. In International Symposium on Distributed Computing and Artificial Intelligence (pp. 419-428). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-76459-2_39

[8] Stergiopoulos, V., Vassilakopoulos, M., Tousidou, E., & Corral, A. (2024). An academic recommender system on large citation data based on clustering, graph modeling and deep learning. Knowledge and information systems, 66(8), 4463-4496. https://doi.org/10.1007/s10115-024-02094-7

[9] Lu, Y., Yuan, M., Liu, J., & Chen, M. (2023). Research on semantic representation and citation recommendation of scientific papers with multiple semantics fusion. Scientometrics, 128(2), 1367-1393. https://doi.org/10.1007/s11192-022-04566-5

[10] Zhang, X., & Gan, M. (2024). Hi-GNN: hierarchical interactive graph neural networks for auxiliary information-enhanced recommendation. Knowledge and Information Systems, 66(1), 115-145. https://doi.org/10.1007/s10115-023-01949-9

[11] Qin, H., Han, X., Ma, X., & Yan, W. (2023). Personalized literature recommendation based on heterogeneous entity academic network. Journal of King Saud University-Computer and Information Sciences, 35(8), 101649. https://doi.org/10.1016/j.jksuci.2023.101649

[12] Huang, S., Huang, Y., Bu, Y., Lu, W., Qian, J., & Wang, D. (2022). Fine-grained citation count prediction via a transformer-based model with among-attention mechanism. Information Processing & Management, 59(2), 102799. https://doi.org/10.1016/j.ipm.2021.102799

[13] Chicaiza, J., & Valdiviezo-Diaz, P. (2021). A comprehensive survey of knowledge graph-based recommender systems: Technologies, development, and contributions. Information, 12(6), 232. https://doi.org/10.3390/info12060232

[14] Zhao, G., Zhang, X., Tang, H., Shen, J., & Qian, X. (2024). Domain-oriented knowledge transfer for cross-domain recommendation. IEEE Transactions on Multimedia, 26, 9539-9550. https://doi.org/10.1109/TMM.2024.3394686

[15] Dezfouli, P. A. B., Momtazi, S., & Dehghan, M. (2021). Deep neural review text interaction for recommendation systems. Applied Soft Computing, 100, 106985. https://doi.org/10.1016/j.asoc.2020.106985

[16] Deng, Y. (2022). Recommender systems based on graph embedding techniques: A review. Ieee Access, 10, 51587-51633. https://doi.org/10.1109/ACCESS.2022.3174197

[17] Yuan, C., He, Y., Lin, R., & Tang, Y. (2021). Graph embedding for scholar recommendation in academic social networks. Frontiers in Physics, 9, 768006. https://doi.org/10.3389/fphy.2021.768006

[18] Yang, C., Zhou, W., Wang, Z., Jiang, B., Li, D., & Shen, H. (2021). Accurate and explainable recommendation via hierarchical attention network oriented towards crowd intelligence. Knowledge-Based Systems, 213, 106687. https://doi.org/10.1016/j.knosys.2020.106687

[19] Yin, N., Shen, L., Li, B., Wang, M., Luo, X., Chen, C., ... & Hua, X. S. (2022, October). Deal: An unsupervised domain adaptive framework for graph-level classification. In Proceedings of the 30th ACM International Conference on Multimedia (pp. 3470-3479). https://doi.org/10.1145/3503161.3548012

[20] Sharifbaev, A., Mozikov, M., Zaynidinov, H., & Makarov, I. (2024). Efficient integration of reinforcement learning in graph neural networks-based recommender systems. IEEE Access, 12, 189439-189448. https://doi.org/10.1109/ACCESS.2024.3516517

[21] Qian, Y., & Yin, X. (2024). Semantic-enhanced graph neural networks with global context representation. Machine Learning, 113(10), 7761-7781. https://doi.org/10.1007/s10994-024-06523-0

[22] Qi, R., Wei, J., Shao, Z., Li, Z., Chen, H., Sun, Y., & Li, S. (2023). Multi-task learning model for citation intent classification in scientific publications. Scientometrics, 128(12), 6335-6355. https://doi.org/10.1007/s11192-023-04858-4

[23] Liu, Y., Teng, L., Yin, B., Meng, H., Yin, X., Huan, S., ... & Zhang, X. B. (2022). Chemical design of activatable photoacoustic probes for precise biomedical applications. Chemical reviews, 122(6), 6850-6918. https://doi.org/10.1021/acs.chemrev.1c00875

[24] Fan, S., Liu, G., & Li, J. (2023). A heterogeneous graph neural network with attribute enhancement and structure-aware attention. IEEE Transactions on Computational Social Systems, 11(1), 829-838. https://doi.org/10.1109/TCSS.2023.3239034

[25] Xu, B., Xu, C., & Su, B. (2023, October). Cross-modal graph attention network for entity alignment. In Proceedings of the 31st ACM international conference on multimedia (pp. 3715-3723). https://doi.org/10.1145/3581783.361215

[26] Zhou, X., & Miao, C. (2024). Disentangled graph variational auto-encoder for multimodal recommendation with interpretability. IEEE Transactions on Multimedia, 26, 7543-7554. https://doi.org/10.1109/TMM.2024.3369875

[27] Gao, C., Wang, X., He, X., & Li, Y. (2022, February). Graph neural networks for recommender system. In Proceedings of the fifteenth ACM international conference on web search and data mining (pp. 1623-1625). https://doi.org/10.1145/3488560.3501396

[28] Guo, F., Wang, Z., Wang, X., Lu, Q., & Ji, S. (2024). Dual-view multi-modal contrastive learning for graph-based recommender systems. Computers and Electrical Engineering, 116, 109213. https://doi.org/10.1016/j.compeleceng.2024.109213

[29] Zafar, A., Guilin, Q., Pavlos, K., Ahmad, A. W., & Bahadar, A. (2020). A graph-based taxonomy of citation recommendation models. The Artificial Intelligence Review, 53(7), 5217-5260. https://doi.org/10.1007/s10462-020-09819-4

[30] Qiu, T., Yu, C., Zhong, Y., An, L., & Li, G. (2021). A scientific citation recommendation model integrating network and text representations. Scientometrics, 126(11), 9199-9221. https://doi.org/10.1007/s11192-021-04161-0

[31] Ohagi, M., & Aizawa, A. (2022, June). Pre-trained transformer-based citation context-aware citation network embeddings. In Proceedings of the 22nd ACM/IEEE joint conference on digital libraries (pp. 1-5). https://doi.org/10.1145/3529372.3533290

[32] Wang, C., Zhu, J., Li, A., Li, Z., & Wang, Y. (2024). Dual-channel representation consistent recommender for session-based new item recommendation. Expert Systems with Applications, 249, 123681. https://doi.org/10.1016/j.eswa.2024.123681

[33] Xie, R., Liu, Q., Wang, L., Liu, S., Zhang, B., & Lin, L. (2022, August). Contrastive cross-domain recommendation in matching. In Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining (pp. 4226-4236). https://doi.org/10.1145/3534678.3539125

[34] Du, Y., Liu, Y., Peng, Z., & Jin, X. (2022). Gated attention fusion network for multimodal sentiment classification. Knowledge-Based Systems, 240, 108107. https://doi.org/10.1016/j.knosys.2021.108107

[35] Gao, Y., Du, Y., Hu, Y., Chen, L., Zhu, X., Fang, Z., & Zheng, B. (2022, July). Self-guided learning to denoise for robust recommendation. In Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval (pp. 1412-1422). https://doi.org/10.1145/3477495.3532059

[36] Kovalerchuk, B. (2023). Explainable machine learning and visual knowledge discovery. In Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook (pp. 913-943). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-24628-9_39

[37] Sharma, K., Lee, Y. C., Nambi, S., Salian, A., Shah, S., Kim, S. W., & Kumar, S. (2024). A survey of graph neural networks for social recommender systems. ACM Computing Surveys, 56(10), 1-34. https://doi.org/10.1145/3661821