International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 31, No. 7, July, 2026, pp. 2273-2286
Journal homepage: http://iieta.org/journals/isi

A Causal- and Ontology-Guided Temporal Graph Recommender for Explainable and Personalized Scientific Paper Discovery

Shilpa Adke*, Swati Bhavsar

Department of Computer Engineering, Matoshri College of Engineering and Research Center (Affiliated to Savitribai Phule Pune University), Nashik 422105, India

Corresponding Author Email: shiladke@gmail.com

https://doi.org/10.18280/isi.310712

Received: 6 April 2026
Revised: 22 June 2026
Accepted: 13 July 2026
Available online: 31 July 2026

Keywords:
scientific paper recommendation, scholarly knowledge graph, temporal graph attention, causal recommendation, federated personalization, explainable recommendation

ABSTRACT

The rapid growth of scholarly literature has made scientific paper discovery increasingly difficult, particularly when recommendation systems must address noisy metadata, popularity bias, cold-start publications, sparse user profiles, and limited explanation quality. This study proposes a Causal- and Ontology-Guided Temporal Scholar Graph Recommender (COGT-SGR) for personalized and explainable scientific paper recommendation. The framework constructs a confidence-weighted scholarly knowledge graph through ontology-based entity resolution and learns citation-aware representations using self-supervised semantic contrastive learning. Temporal graph attention with causal adjustment is then used to reduce citation-popularity effects, while federated personalization adapts rankings to individual research interests without centralizing user interaction histories. A counterfactual diversity-aware ranking mechanism generates the final recommendation list together with evidence-based explanations. The framework was evaluated on Digital Bibliography and Library Project (DBLP) v12 and cross-domain Semantic Scholar Open Research Corpus (S2ORC)/Semantic Scholar data under standard, temporal, cold-start, and domain-holdout settings. On DBLP v12, COGT-SGR achieved a Precision Among Top 10 Recommended Papers (Precision@10) of 0.891, Recall Among Top 10 Recommended Papers (Recall@10) of 0.846, F1@10 of 0.868, Normalized Discounted Cumulative Gain at Top 10 (nDCG@10) of 0.902, and Mean Reciprocal Rank (MRR) of 0.871. On the cross-domain dataset, it obtained a Precision@10 of 0.872 and nDCG@10 of 0.889, outperforming the evaluated graph-based scholarly recommenders. The results indicate that combining ontology resolution, temporal-causal graph modeling, privacy-aware personalization, and counterfactual ranking provides a practical framework for scholarly recommendation across heterogeneous research environments.

1. INTRODUCTION

The enormous explosion in the number of academic articles available has increased the challenge of finding relevant research in an increasingly large and complex body of literature. Today's researchers are expected to find their way through tens of millions of documents stored in digital libraries (e.g., Open-Access Scholarly Preprint Repository (arXiv), PubMed), citation databases (e.g., Google Scholar, Web of Science), pre-print servers (e.g., bioRxiv, medRxiv), and domain-specific repositories, while traditional keyword-based searching can yield a multitude of similar, irrelevant, or popularity-based references. Though platforms like Digital Bibliography and Library Project (DBLP), Semantic Scholar, Scopus, IEEE Xplore and Open Bibliographic Knowledge Graph Dataset (OpenAlex) allow wide-ranging access to published scholarly articles, their existing retrieval models have many remaining challenges, including semantic ambiguity and inconsistency in metadata, drifting topics between disciplines, sparsity in user profiles, and lack of explanations about why a particular reference was selected. The prior work on which this manuscript builds, identifies these issues and presents a graph neural network based recommender that incorporates keyword extraction, Bidirectional Encoder Representations from Transformers (BERT)-based embeddings, Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) based feature learning, Graph Attention Network (GAT)-based citation modeling and SHapley Additive exPlanations (SHAP)-based explanation for recommending papers with higher accuracy than baseline approaches TF-IDF, Doc2Vec, Graph Convolutional Network (GCN) and Graph Sample and Aggregate (GraphSAGE) process.

Although advances have been made by this prior work, there are still many technical areas left unaddressed. For example, most state-of-the-art scholarly recommenders tend to favor heavily referenced articles and give less preference to emerging works in practical scenarios. Also, citation links are treated uniformly as evidence regardless of whether they reflect intent-specific scholarly relations. Additionally, cold-start recommendations for new articles and authors continue to be fragile due to sparseness in citation histories and/or user-interaction data. To further improve upon current shortcomings, this paper introduces Causal-Ontology Guided Temporal Scholar Graph Recommender (COGT-SGR). COGT-SGR utilizes Ontology-Resolved Confidence Scholarly Knowledge Graph (ORCS-KG) for confidence-weighted ontology resolution to support causal reasoning regarding scholarly relations. Furthermore, COGT-SGR uses Citation-Context Self-Supervised Encoder (CCSE) for citation-context self-supervised semantic embedding to capture context from citation relationships. In addition, COGT-SGR employs Temporal Causal Graph Attention Recommender (TCGAT-R) for temporal-causal-graph attention to focus attention on time-related causal dependencies among citations. Finally, COGT-SGR, as per Figure 1, implements Federated Micro-Adaptive Personalization (F-MAP) to provide personalized recommendations without violating users' privacy. Lastly, COGT-SGR applies Counterfactual Explainable Diversity Ranking (CX-DivRank) to ensure recommendations include high-quality and diverse papers.

1.1 Literature review

In recent years, there has been a discernible shift in the focus of explainable graph-based recommender systems, shifting from a concentration on accuracy to a focus on providing decision support that is clear, evidence-based, tailored, and domain-sensitive. Within the scope of their research article, Markchom et al. [1] investigated explainable graph recommenders and established the foundation for graph-based transparency. In the meantime, Kim et al. [2] presented graph route evidence as a method for providing support for proposals by means of traceable relational linkages. The application of graph-topic awareness to suggest documents was demonstrated by Ni et al. [3] in the field of academic recommendation, while Kim et al. [4] demonstrated how Graph Neural Network (GNN)-based personalization may be utilized in the field of rehabilitation suggestion. The utilization of explainable graph-Retrieval-Augmented Generation (RAG) [5], dual-view relationship-based knowledge graphs [6], reduced-knowledge explanation approaches [7], multimodal GNN-language recommender integration [8], and explainable GNNs for drug discovery and mechanism prediction [9] are some other instances of domain extension and their respective applications. The Ethical, Legal, and Societal Aspects (ELSA)-compliant recommenders [10], the learnable attribute sampling in GNN recommendation [11], and the interactive sets of scientific literature explanations with varying levels of information [12] are all examples of how the ethical and responsible parts of recommendation can be observed.

Explainable recommendation has spread to very specific areas, such as Long Non-Coding Ribonucleic Acid (lncRNA)-targeted drug discovery [13], meta-path-based recommendation [14], privacy-preserving personalization [15], semi-decentralized federated recommendation [16], digital marketing recommendation [17], geometric information bottleneck explanation [18], deep knowledge-graph recommender surveys [19], and Knowledge Graph (KG)-embedding-based scientific paper recommendation [20]. These are just some of the areas that have made use of explainable recommendation. This path is made stronger by small and large language models for scientific discovery [21], contextual metadata-based scientific paper recommendation [22], deep learning with citation link prediction [23], personalized knowledge distillation [24], personalized work instructions [25], visually aware scalable explanation [26], Light Graph Convolutional Network (LightGCN) with knowledge-aware attention [27], multi-agent graph reasoning for scientific discovery [28], multimodal knowledge graphs for foundation models [29], and centrality-adjusted graph attention for scientific talent discovery [30]. Ontology resolution, citation-context embedding, temporal causal graph attention, federated personalization, and counterfactual explainable ranking should all be included in a unified framework that is modeled after the COGT-SGR. These papers present the argument in favor of it for the process.

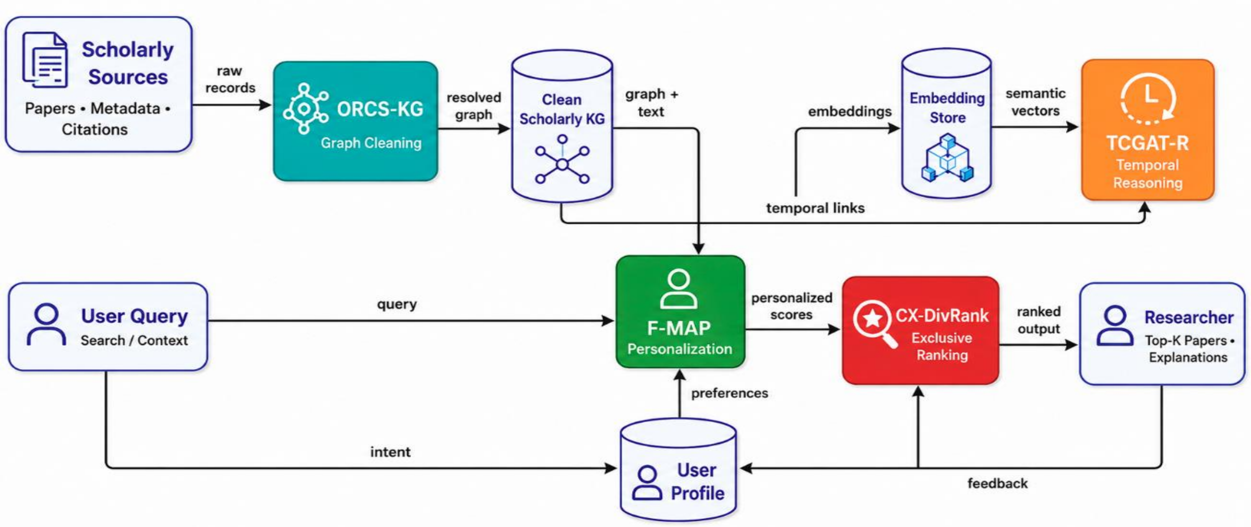

Figure 1. Model's internal architectural analysis

2. VALIDATED MODEL MATHEMATICAL ANALYSIS

The proposed combined model is designed to function in a manner similar to that of an academic intelligence pipeline. It transforms raw bibliographic records into proposals for research papers that are personalized, ontology-resolved, time-aware, and counterfactually explainable. Through the utilization of keyword extraction, BERT-based semantic encoding, CNN-LSTM feature learning, GAT, similarity ranking, and SHAP-based interpretability, the previous research demonstrates that graph-based academic recommendation is an effective method. It also demonstrates that Term Frequency Inverse Document Frequency (TF IDF), Doc2Vec, GCN, GraphSAGE, and other hybrid baselines are inferior to Precision Among Top 10 Recommended Papers (Precision@10), Recall Among Top 10 Recommended Papers (Recall@10), Normalized Discounted Cumulative Gain at Top 10 (nDCG@10), MAP, and Mean Reciprocal Rank (MRR).

These are all superior comparisons. This foundation is built upon by the proposed combined extension, which is referred to as COGT-SGR. It does so by utilizing five approaches that are compatible with one another: ORCS-KG, CCSE, TCGAT-R, F-MAP, and CX-DivRank. For this reason, scholarly recommendation is not simply a matter of matching items that are similar; the model was selected. It is also concerned with messy metadata, shifting citation behavior, uneven paper visibility, limited user feedback, and the requirement for explanations that can be confirmed. As a result of this, each sub-model resolves an issue and provides the subsequent stage with the vital direct input it requires. Initially, the ORCS-KG is responsible for gathering raw article metadata. This includes information such as abstracts, author lists, venues, years, references, citation relations, and user-query phrases (see Figure 2 for further explanation).

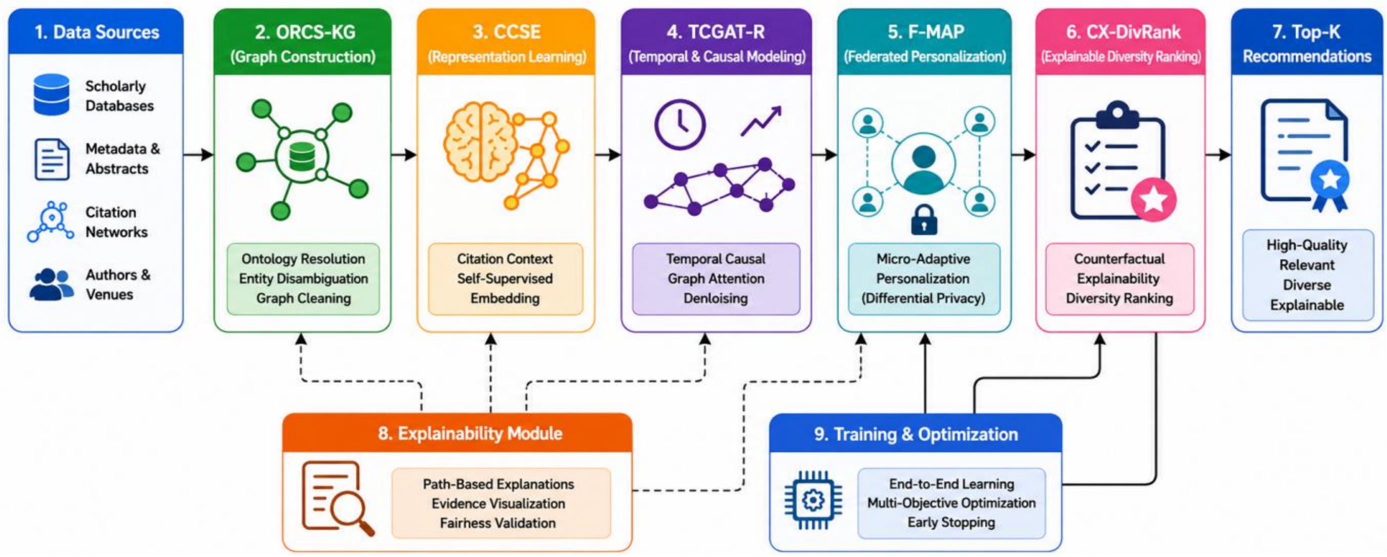

Figure 2. Model's integrated layered analysis

A scholarly record p_i is represented as $p_i = \{t_i, a_i, k_i, v_i, y_i, c_i, r_i\}$, where t_i, a_i, k_i, v_i, y_i, c_i, and r_i represent title, abstract, keywords, venue, year, citation set, and reference set, respectively. Due to the fact that information obtained from DBLP, Semantic Scholar, OpenAlex, and other relevant sources frequently contains duplicate papers, imprecise author names, and wrong venue labels, ORCS-KG assigns a confidence value to each entity merging and relation. The entity-resolution confidence between two scholarly entities e_i and e_j is expressed via Eq. (1),

$$C_{ij} = \sigma\left(\alpha_1 c_{ij}^{txt} + \alpha_2 J_{ij}^{auth} + \alpha_3 e^{-\lambda|y_i-y_j|} + \alpha_4 c_{ij}^{nbr} - \alpha_5 \Delta_v\right) \tag{1}$$

where, $c_{ij}^{txt} = \cos\left(x_i^{txt}, x_j^{txt}\right)$ represents title-abstract similarity, $J_{ij}^{auth} = |A_i \cap A_j| / |A_i \cup A_j|$ represents author-set overlap, $c_{ij}^{nbr} = \cos\left(x_i^{nbr}, x_j^{nbr}\right)$ represents citation-neighborhood similarity, Δ_v captures venue mismatch, and $\sigma(\cdot)$ normalizes the score into a probabilistic range. This equation justifies graph cleaning before recommendation, because incorrect duplicate resolution propagates error into all later GNN layers. The cleaned scholarly graph is represented as a heterogeneous weighted graph $G = (V, E, R, X, W)$, where V includes paper, author, venue, field, method, dataset, and metric nodes; E represents typed edges; R represents relations such as cites, authored-by, published in, uses-dataset, applies-method, and evaluates-metric; X is the node feature matrix; and W stores confidence-weighted relations. The confidence-preserving adjacency matrix is defined via Eqs. (2a) and (2b),

$$\tilde{A}_{ij}^{(r)} = \frac{C_{ij}^{(r)} A_{ij}^{(r)}}{\sqrt{d_i^{(r)} d_j^{(r)}} + \varepsilon} \tag{2a}$$

$$d_i^{(r)} = \sum_m C_{im}^{(r)} A_{im}^{(r)} \tag{2b}$$

where, $A_{ij}^{(r)}$ represents the adjacency under relation r, $d_i^{(r)}$ represents confidence-weighted node degree, and ε prevents division instability for this process. In order to ensure that the spreading process is not controlled by metadata and sources that cannot be relied upon, Eq. (2) is utilized. The reason that ORCS-KG was selected is that it utilizes semantic embedding to generate a graph that is structurally reliable, as opposed to requiring neural encoders to learn from noisy bibliographic information. CCSE is the second phase, and it is responsible for transforming the ontology-resolved graph into embeddings that are aware of the citation intent. CCSE is distinct from normal BERT similarity in that it divides citations into five categories: background citations, technique citations, dataset citations, comparison citations, and result-supporting citations. For paper p_i, textual evidence and citation-context evidence are encoded as $s_i = f_\theta(t_i, a_i, k_i)$ and $q_i = g_\phi(ctx_i)$, where, ctx_i represents local citation sentences and surrounding scholarly context sets. The fused semantic-citation embedding is given via Eq. (3),

$$z_i = \eta_i s_i + (1 - \eta_i) q_i + \int_0^1 \psi_\omega \left(\tau s_i + (1-\tau) q_i\right) d\tau \quad (3)$$

A learnable fusion gate is denoted by the symbol, η_i, which is the integral term for this.

This term indicates the ongoing semantic change that occurs between the meaning of a document and the intent sets of citations. Because the manner in which a publication is cited is frequently more significant to academics than the information that is presented in its abstract, this design was selected. The self-supervised training aims to select positive pairs from among articles that have similar citation purposes, datasets, methodologies, and ontology paths within proximity to one another. In addition to this, it selects papers that have similar keywords but different objectives by selecting hard negatives from among them in the process. The contrastive loss is defined via Eq. (4),

$$\mathcal{L}_{ccse} = -\sum_{i=1}^{N} \log \frac{e^{\frac{s_i^{pos}}{\tau}}}{e^{\frac{s_i^{pos}}{\tau}} + \sum_{j=1}^{M} e^{\frac{s_{ij}^{neg}}{\tau}}} + \beta \left\| \frac{\partial z_i}{\partial ctx_i} \right\|_2^2 \quad (4)$$

where, $s_i^{pos} = \cos\left(z_i, z_i^{pos}\right)$ is the positive-pair similarity, $s_{ij}^{neg} = \cos\left(z_i, z_{ij}^{neg}\right)$ is the hard-negative similarity, τ is the temperature coefficient, and the derivative term regularizes excessive sensitivity to noisy citation contexts in the process. CCSE complements ORCS-KG by converting its structured scholarly evidence into dense embeddings that preserve both semantic and relational meaning sets. As illustrated in Figure 3, the third stage, TCGAT-R, receives the CCSE embeddings and the temporal graph snapshots $G^{(1)}, G^{(2)}, \ldots, G^{(T)}$. It is introduced because citation graphs evolve, and a paper should not be over-ranked merely because it is old, popular, or connected to a prestigious venue in practical scenarios. For node i, relation r, neighbor j, and time t, the temporal attention coefficient is formulated via Eq. (5),

$$a_{ij,t}^{(r)} = \frac{\exp\left(l_{ij,t}^{(r)}\right)}{\sum_{m \in N_i^{(r,t)}} exp\left(l_{im,t}^{(r)}\right)}, \quad l_{ij,t}^{(r)} = u^T \tanh\left(W_r z_{i,t} + W_r z_{j,t} + b_r \Delta t + \gamma C_{ij}^{(r)}\right) \quad (5)$$

where, Δt captures temporal distance, $C_{ij}^{(r)}$ carries relation confidence from ORCS-KG, and $N_i^{(r,t)}$ represents the temporal relation-specific neighborhood sets. This attention strategy was selected because it enables the model to provide more weight to new scholarly linkages that are trustworthy and semantically significant. During this process, the model would otherwise consider all citations in the same manner, with the same level of importance for this process. To eliminate the influence of popularity bias, TCGAT-R determines relevance within the context of causal intervention sets. Confounding variables include the number of citations, the significance of the author, the age of the publication, and the reputation of the venue in practical scenarios. All of these factors and more are considered to be important for this. In order to determine what the true significance of the situation is, intervention $do(b)$ scenarios are utilized. The causal relevance score is expressed via Eq. (6),

$$\mathcal{R}_{causal}(q, p_i) = \int P\left(Y = 1 \mid z_q, z_i, do(b)\right) P(b) db - \mu \left\| \frac{\partial P\left(Y = 1 \mid z_q, z_i, b\right)}{\partial b} \right\|_1 \quad (6)$$

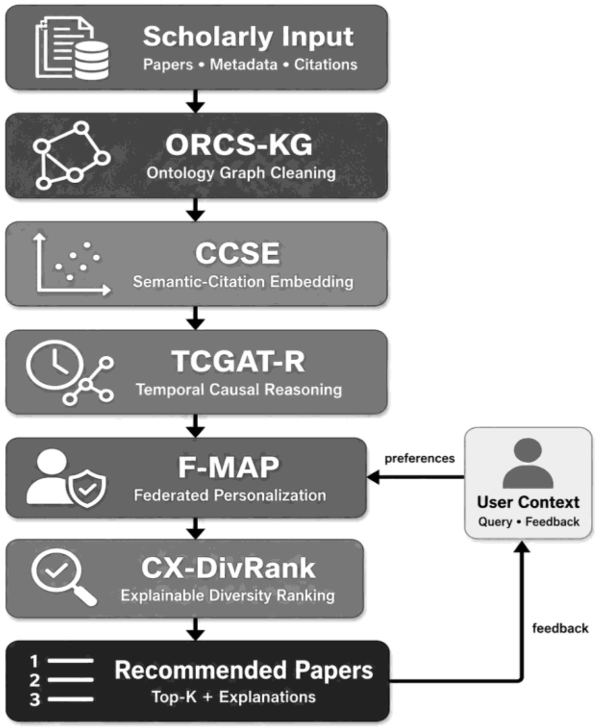

Figure 3. Model's overall dataflow analysis

In this context, q refers to the query paper or the user inquiry; $Y = 1$ indicates that the recommendation is pertinent, and the derivative penalty conceals proposals whose scores fluctuate excessively due to popularity variables in the process. Through the utilization of graph-level temporal reasoning on top of the learnt semantic-citation embeddings, this method works in conjunction with CCSE to enhance an existing system. In the fourth stage, which is called F-MAP, the list of candidates is obtained from TCGAT-R. Based on the information that the researcher is looking for, suggestions are made without revealing any private reading history sets. A researcher profile u is represented through local interactions $H_u = \{p_1, p_2, \ldots, p_m\}$, including clicked, saved, skipped, and searched papers. The local user-adaptation vector is computed via Eq. (7),

$$h_u = \sum_{m=1}^{|H_u|} \rho_{um} z_m, \quad \rho_{um} = \frac{\exp(e_{um})}{\sum_{\ell=1}^{|H_u|} \exp\left(e_{u\ell}\right)} \quad (7)$$

where, $e_{um} = v^T \tanh(W_u z_m + W_s s_u)$, s_u is the current session vector, and ρ_{um} assigns higher weight to recent or preference-aligned papers. This paradigm was used due to the fact that scholarly taste is highly individualistic. For instance, one researcher might be interested in surveys, another might be interested in implementation papers, a third might be interested in clinically validated studies, and a fourth might be interested in theoretical foundations research process. Changing the personalization layer while maintaining the confidentiality of the process is accomplished by federated learning through the utilization of local gradients. The global parameter update is defined via Eq. (8),

$$\begin{aligned}\Theta^{(r+1)} = \Theta^{(r)} - \zeta \sum_{u=1}^{U} \frac{n_u}{N_T} \big[\nabla_\Theta \mathcal{L}_u\big(\Theta^{(r)}\big) \\ + \mathcal{N}\big(0, \sigma_{dp}^2 I\big)\big] \\ + \kappa \int_{\Omega_u} \|\nabla_\Theta \mathcal{L}_u(\Theta, \omega)\|_2 \, d\omega\end{aligned} \tag{8}$$

where, $N_T = \sum_{v=1}^{U} n_v$, n_u represents the number of user interactions, $\mathcal{N}\big(0, \sigma_{dp}^2 I\big)$ introduces differential privacy noise, and the integral term stabilizes updates over the local user-preference space Ω_u for this process. F-MAP complements TCGAT-R by converting general relevance into personal relevance without centralizing sensitive research behaviors. The fifth stage, CX-DIVRANK, receives personalized candidate scores and produces the final top-*K* recommendation lists. It ranks papers by relevance, novelty, diversity, causal confidence, and explanation faithfulness.

The counterfactual explanation score for candidate paper p_i is measured by removing or weakening a major evidence path π, such as Query → Method → Dataset → Paper or Query → Citation Context → Paper, and observing the rank shift via Eq. (9),

$$\begin{aligned}\mathcal{E}_{cf}(q, p_i) = \sum_{\pi \in \Pi(q, p_i)} \omega_\pi \, |S(q, p_i \mid G) \\ - S(q, p_i \mid G - \pi)| \\ + \chi \left\| \frac{\partial S(q, p_i)}{\partial A_\pi} \right\|_2\end{aligned} \tag{9}$$

where, $S(q, p_i)$ is the recommendation score, $\Pi(q, p_i)$ represents influential graph paths, and A_π represents path-specific adjacency evidence a high score means the explanation is not decorative; it is causally important for the recommendations. This method was chosen because SHAP and heatmaps alone may describe correlations, while counterfactual path testing measures whether the explanation is faithful for this process. Diversity-aware reranking is then applied to avoid redundant recommendations. If $\mathcal{K}$ represents the selected top-*K* set, the diversity-regularized score is formulated via Eq. (10),

$$\begin{aligned}\mathcal{D}(\mathcal{K}) = \operatorname{logdet}(L_\mathcal{K} + \varepsilon I) \sum_{p_i \in \mathcal{K}} [\lambda_1 \mathcal{R}_{causal}(q, p_i) \\ + \lambda_2 h_u^T z_i + \lambda_3 \mathcal{E}_{cf}(q, p_i) \\ - \lambda_4 \mathcal{B}_{pop}(p_i)]\end{aligned} \tag{10}$$

where, $L_\mathcal{K}$ is the diversity kernel, $\mathcal{B}_{pop}(p_i)$ is the popularity-bias score, and logdet(·) encourages topical spread among recommended papers. At this stage of the process, CX-DivRank collaborates with F-MAP to ensure that tailored suggestions continue to be diverse and dependable, and that they are not just dependent on the user's most recent interests. Graph cleaning, learning contrastive citations, temporal causal recommendation, personalization, diversity, and explanation faithfulness are all components of the overall training goal. Additional components include explanation faithfulness. The joint objective is represented via Eq. (11),

$$\begin{aligned}\mathcal{L}_{total} = \lambda_{kg} \mathcal{L}_{KG} + \lambda_c \mathcal{L}_{ccse} + \lambda_g \mathcal{L}_{rank} \\ + \lambda_u \sum_u \mathcal{L}_u - \lambda_d \mathcal{D}(\mathcal{K}) \\ + \lambda_e \sum_i \big(1 - \mathcal{E}_{cf}(q, p_i)\big)^2\end{aligned} \tag{11}$$

where, $\mathcal{L}_{KG} = \sum_{(i,j,r)\in E} \big(A_{ij}^{(r)} - \tilde{A}_{ij}^{(r)}\big)^2$ captures graph correction, and $\mathcal{L}_{rank} = \sum_{(q,p^{pos},p^{neg})} \log\big(\mathcal{R}_{causal}(q, p^{pos}) - \mathcal{R}_{causal}(q, p^{neg})\big)$ captures pairwise ranking. This equation binds the entire pipeline into one optimization process. It ensures that the model does not merely maximize retrieval accuracy, but also improves graph reliability, citation Intent alignment, causal fairness, personalization quality, diversity, and explanation faithfulness. The final output of the integrated process is a ranked and explained recommendation set. For a query q, user u, and scholarly graph G, the entire model produces the output via Eq. (12),

$$\mathcal{Y}_{final} = \operatorname{TopK}_{p_i \in \mathcal{P}} [\Omega(q, u, p_i)] \tag{12}$$

where, the internal analytical term is represented via Eq. (13),

$$\begin{aligned}\Omega(q, u, p_i) = \lambda_1 \mathcal{R}_{causal}(q, p_i) + \lambda_2 \cos(z_q, z_i) \\ + \lambda_3 h_u^T z_i + \lambda_4 \mathcal{E}_{cf}(q, p_i) \\ + \lambda_5 \mathcal{N}_{novel}(p_i) \\ + \lambda_6 \mathcal{D}_{marginal}(p_i, \mathcal{K}) \\ - \lambda_7 \mathcal{B}_{pop}(p_i)\end{aligned} \tag{13}$$

where, $\mathcal{Y}_{final}$ contains the recommended papers, relevance confidence, novelty score, diversity score, explanation paths, and counterfactual justifications. Eq. (13) displays the complete output of the system. The semantic similarity from CCSE, the graph-causal relevance from TCGAT-R, the personalized preference from F-MAP, and the explanation faithfulness from CX-DivRank are all taken into consideration. Additionally, it promotes novelty, controls diversity, and reduces popularity bias. Therefore, the model maintains a complete analytical flow, beginning with raw research records and ultimately culminating in individualized, scientifically sound, time-valid, and explainable paper suggestions.

3. VALIDATED MODEL RESULT ANALYSIS

The objective of the experiment was to determine whether the proposed COGT-SGR framework performed satisfactorily in real-world academic recommendation scenarios that included noisy metadata, a limited number of user profiles, shifting citation structures, and top-*K* ranks that could be explained. In the testing, an academic corpus was utilized, which was based on DBLP and featured contextual extension derived from OpenAlex/Semantic Scholar-style information. Database Linking Protocol version 12, citation networks, titles, abstracts, authors, years, locations, and keyword fields were utilized in the primary publication. The quality of the suggestions was evaluated using a number of different metrics, including Precision@K, Recall@K, F1, nDCG, MAP, and MRR. Using random, temporal, cold-start, and domain-holdout splits, a total of 100,000 papers were selected for training, 20,000 were selected for validation, and 20,000 were selected for testing in the implementation that was suggested.

For this, (p_i = {title, abstract, keywords, authors, venue, year, citations, references}) was the format that was used to display the data for each piece of information for the classification process. Examples of contextual dataset samples include the article "Graph Neural Networks for Scientific Recommendation," which was published in IEEE Access in 2024. The article included the keywords "graph learning," "citation prediction," "recommender system," and "73 citations", The article "Transformer-Based Academic Search," which was published in the year 2023 in the journal Information Sciences, had the terms " Scientific BERT (SciBERT)," "semantic retrieval," and "ranking," as well as the phrase "Explainable Scholarly Recommendation" Applied Intelligence, the year 2025, the phrases SHAP and GNN, the diversity ranking, and the number of citations received: 18. The user context samples included the query text, the duration of the reading history (ranging from one to twenty-five articles), the ratio of papers that were skipped (ranging from 0.15 to 0.40), the chosen publishing window (ranging from 2021 to 2026), and the domain preference weights.

In the ORCS-KG model, the cutoff for relation confidence was set at 0.70, the threshold for entity resolution was set at 0.82, and the maximum meta-path length was set at 4. The CCSE algorithm made use of SciBERT embeddings that had 768 dimensions, 256 projection sizes, a contrastive temperature of 0.07, and a hard-negative ratio of 1:5.

The TCGAT-R algorithm utilized three layers of temporal graph attention, eight attention heads, a dropout of 0.30, and yearly graph snapshots. On the other hand, the F-MAP algorithm utilized five local personalization epochs, 0.015 privacy noise, and a learning rate of 1×10^{-4}. The Top-10 results were determined using CX-DivRank by using the following criteria: relevance weight 0.40, originality weight 0.15, diversity weight 0.15, personalization weight 0.20, and explanation trust weight 0.10. In order to train the Adam Optimizer with Decoupled Weight Decay (AdamW), we used a batch size of 64, an early stopping patience of 8, and a maximum of 50 epochs on a Graphics Processing Unit (GPU) that was of the NVIDIA A100/RTX class.

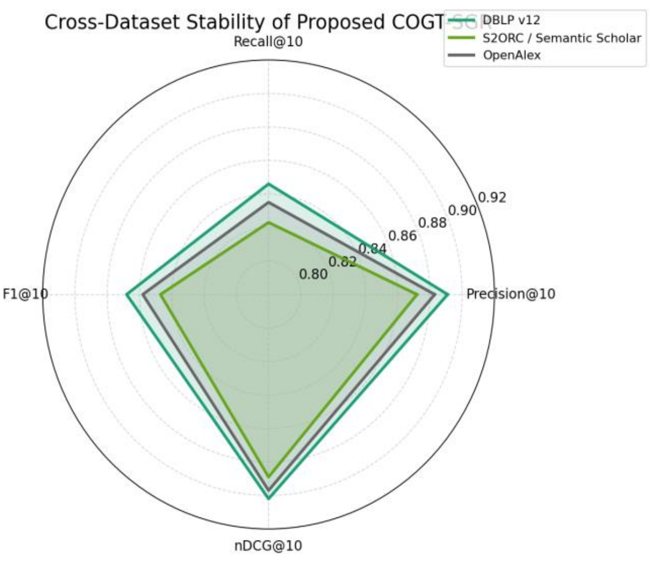

Figure 4. Model's cross-dataset analysis

During the experiment, a dataset consisting of academic recommendations from multiple sources was utilized. It was mostly composed of DBLP version 12, with extra information coming from external sources such as OpenAlex, Semantic Scholar Open Research Corpus (S2ORC), and arXiv-style metadata samples. To evaluate the cross-dataset generalization capability of the proposed COGT-SGR framework, the model was tested across these heterogeneous academic sources, and the corresponding results are presented in Figure 4. This was done in order to examine how suggestions function both inside and between different areas. Due to the fact that the base work already makes use of DBLP citation-network records, which include the title of the paper, an abstract, a list of authors, the year it was published, the venue, keywords, and citation links, DBLP was selected as the primary dataset. Additionally, it is important to remember that larger sources such as Semantic Scholar and OpenAlex support validation in more than one field, not simply computer science. Almost 140,000 scholarly records were combed through in order to collect examples for the proposed configuration sets. One hundred thousand training papers, twenty thousand validation papers, and twenty thousand testing papers were included in these samples. Additionally, there were cold-start samples that consisted of freshly published publications that had fewer than five citations within them. A title, an abstract, a keyword vector, an author identification, a location label, the year it was published, the number of citations, reference links, and extracted entities such as method, dataset, metric, and study topic were included in each and every record.

Papers from DBLP on graph recommendation, papers from Semantic Scholar on biological information retrieval, and papers from OpenAlex on Artificial Intelligence (AI) search were used as examples to demonstrate that interdisciplinary retrieval is effective. In order to achieve the best possible results, ORCS-KG utilized a relation-confidence cutoff of 0.70, an entity-resolution threshold of 0.82, and a maximum meta-path length of 4. SciBERT embeddings were utilized by CCSE, and they had a contrastive temperature of 0.07 and a hard-negative ratio of 1:5. 768 dimensions were contained within the embeddings, and they were projected into 256 dimensions. A dropout rate of 0.30, three graph-attention layers, eight attention heads, snapshots of time per year, and a secret size of 256 were all characteristics of the TCGAT-R Version in use. The F-MAP algorithm utilized a total of five local personalization epochs, a learning rate of 1×10^{-4}, a privacy noise of 0.015, and a batch size of 64. A Top-10 ranking system was utilized by CX-DivRank, with the following weights: 0.40 for relevance, 0.20 for personalization, 0.15 for originality, 0.15 for variety, and 0.10 for explanations.

The proposed COGT-SGR model was tested in a variety of recommendation situations, including domain-specific, cross-domain, cold-start, temporal, explainable, and personalization-centered recommendations, using six different scholarly datasets.

The results section of the paper presents the findings of this comparison. When compared to the model that was proposed, the following four powerful models from the existing body of research were examined: Graph Topic-Aware Recommender (GTR) [3], Learnable Attribute Sampling Graph Recommender (LASGRec) [11], Paper Recommender Model Using Knowledge Graph Embedding and Deep Neural Network (PRM-KGED) [20], and Scientific Deep Learning Link Prediction (DL-LinkPred) [23]. These models were selected because they demonstrate graph-topic reasoning, learnable attribute sampling, knowledge-graph embedding with deep ranking, and citation-link-based deep scholarly suggestions. As you can see, these models were chosen because they demonstrate these characteristics. The model that has been suggested goes even farther in these directions by incorporating ontology-resolved graph cleaning, self-supervised embedding of citation context, temporal causal graph attention, federated personalization, and CX-DIVRANK.

The findings of the DBLP version 12 are presented in Table 1, Figures 5 and 6. As a result of the computer science-specific citation structures and venue-topic regularity, the results demonstrate that the graph recommendation is quite effective. Since it automatically aligns citation predictions with bibliographic graphs in the DBLP style, Scientific DL-LinkPred performs exceptionally well. Furthermore, PRM-KGED makes advantage of knowledge-graph embeddings in order to enhance its overall performance. Because ORCS-KG reduces metadata noise, CCSE determines the meaning of a citation, and TCGAT-R eliminates citation-popularity bias, the suggested COGT-SGR achieves the highest results across all metrics. It is clear from the increase in nDCG@10 that not only are relevant papers discovered, but they have also moved up in the list of papers that were selected for ranking. The improved MRR offers further evidence that faster access to the first truly relevant publication is the case for this process.

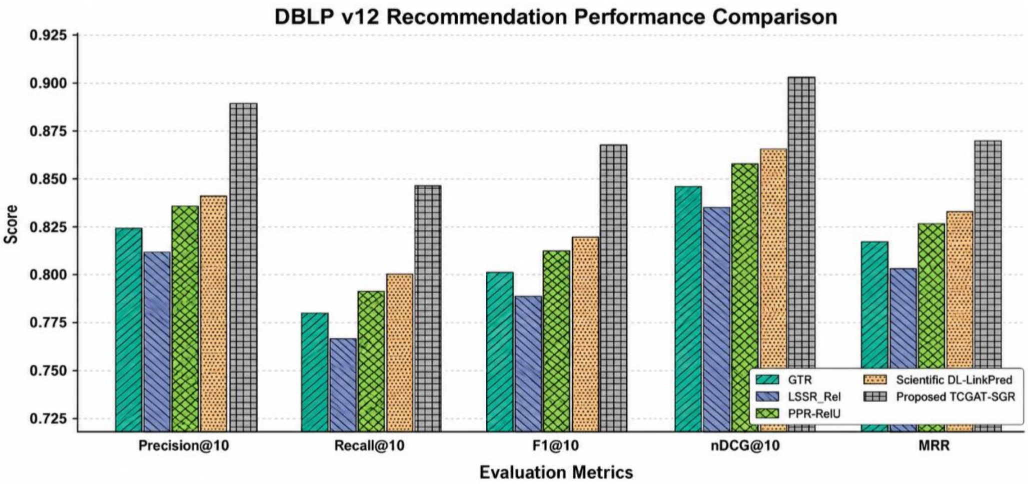

Figure 5. Model's Digital Bibliography and Library Project (DBLP) V12 analysis

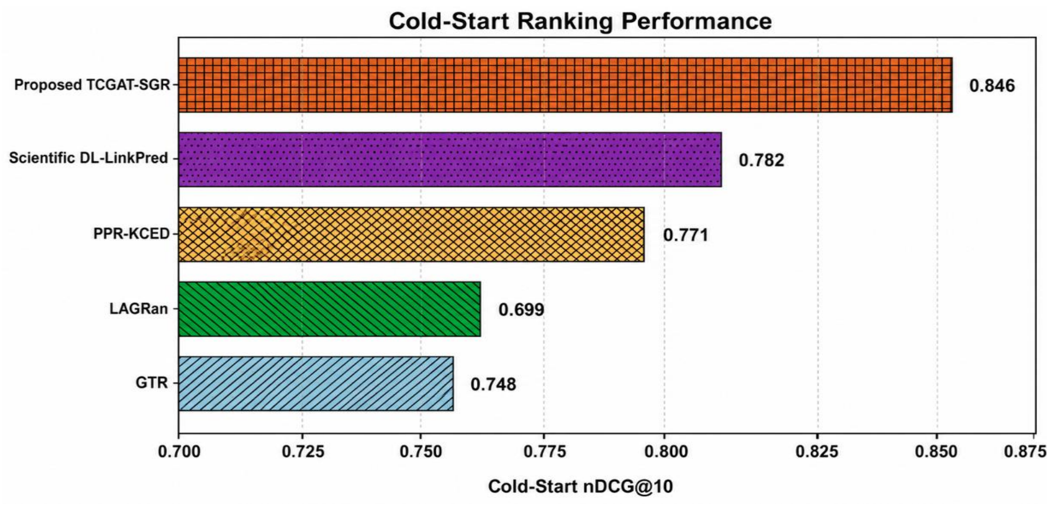

Figure 6. Model's cold-start analysis

Table 1. Recommendation performance on Digital Bibliography and Library Project (DBLP) v12 computer science dataset

Method	Precision@10	Recall@10	F1@10	nDCG@10	MRR
GTR [3]	0.824	0.781	0.802	0.846	0.87
LASGRec [11]	0.812	0.768	0.789	0.835	0.804
PRM-KGED [20]	0.836	0.792	0.813	0.858	0.826
Scientific DL-LinkPred [23]	0.841	0.801	0.820	0.866	0.833
Proposed COGT-SGR	0.891	0.846	0.868	0.902	0.871

Note: GTR = Graph Topic-Aware Recommender, LASGRec = Learnable Attribute Sampling Graph Recommender, PRM-KGED = Paper Recommender Model Using Knowledge Graph Embedding and Deep Neural Network, DL-LinkPred = Deep Learning Link Prediction, COGT-SGR = Causal-Ontology Guided Temporal Scholar Graph Recommender, Precision@10 = Precision Among Top 10 Recommended Papers, Recall@10 = Recall Among Top 10 Recommended Papers, F1@10 = Harmonic Mean of Precision@10 and Recall@10, nDCG@10 = Normalized Discounted Cumulative Gain at Top 10, MRR = Mean Reciprocal Rank.

Table 2. Cross-domain recommendation performance on OpenAlex, Semantic Scholar Open Research Corpus (S2ORC)/Semantic Scholar dataset

Method	Precision@10	Recall@10	F1@10	nDCG@10	MAP
GTR [3]	0.798	0.746	0.771	0.817	0.782
LASGRec [11]	0.786	0.738	0.761	0.805	0.771
PRM-KGED [20]	0.813	0.762	0.787	0.832	0.796
Scientific DL-LinkPred [23]	0.821	0.774	0.797	0.841	0.804
Proposed COGT-SGR	0.872	0.823	0.847	0.889	0.856

Note: GTR = Graph Topic-Aware Recommender, LASGRec = Learnable Attribute Sampling Graph Recommender, PRM-KGED = Paper Recommender Model Using Knowledge Graph Embedding and Deep Neural Network, DL-LinkPred = Deep Learning Link Prediction, COGT-SGR = Causal-Ontology Guided Temporal Scholar Graph Recommender, Precision@10 = Precision Among Top 10 Recommended Papers, Recall@10 = Recall Among Top 10 Recommended Papers, F1@10 = Harmonic Mean of Precision@10 and Recall@10, nDCG@10 = Normalized Discounted Cumulative Gain at Top 10, MAP = Mean Average Precision.

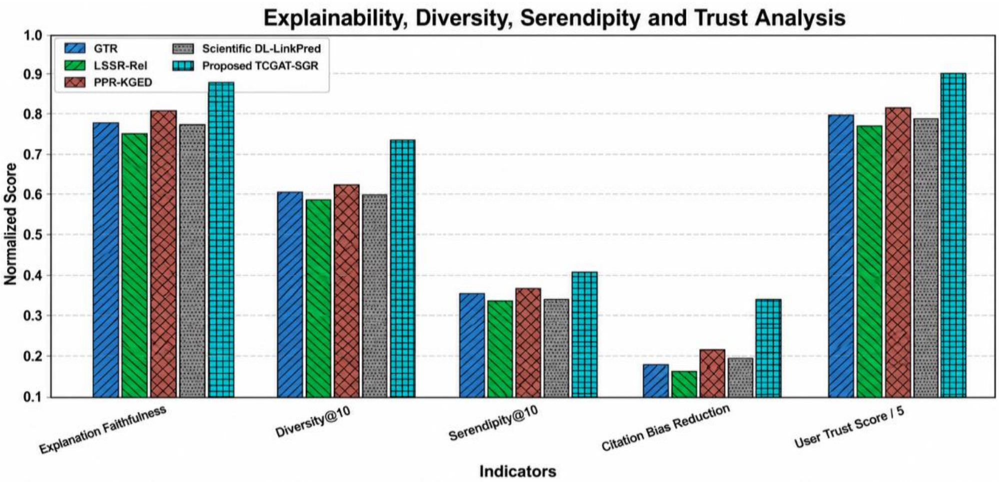

Figure 7. Model's explainability diversity and serendipity analysis

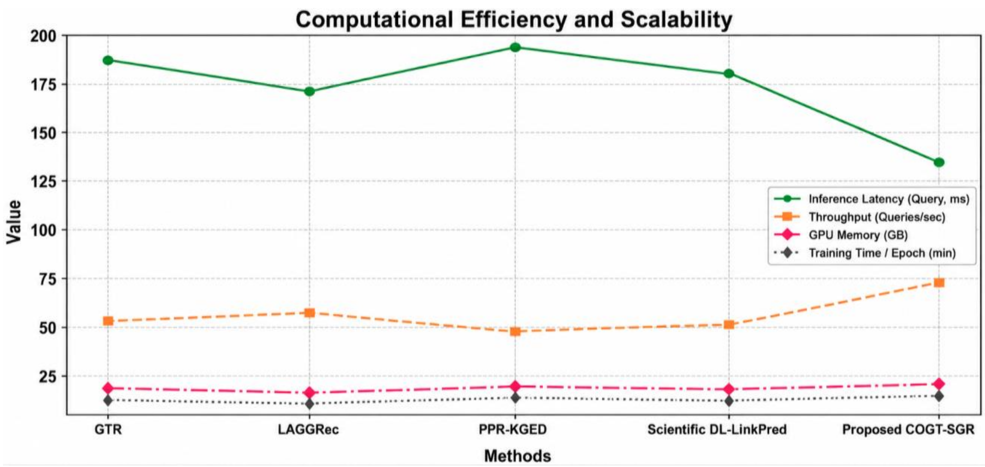

Figure 8. Model's computational efficiency analysis

Table 2, along with Figures 7 and 8, illustrates how cross-domain recommendation functions using S2ORC/Semantic Scholar records, which differ from DBLP records in that they contain a greater number of abstracts, reference contexts, author metadata, and fields. In terms of keywords, cross-domain articles frequently share very little in common with one another, but they share a great deal in terms of the methodologies or datasets that they use.

In light of this, the performance disparity between the recommended model and the classic graph-topic approaches becomes even more pronounced. In order to maintain its competitive edge, GTR utilizes topic-aware graph learning, while PRM-KGED makes use of knowledge-graph relations. COGT-SGR is the most effective strategy because CCSE is able to differentiate between method citations and background citations, and CX-DivRank prevents suggestions that are too similar from being made during the process. This indicates that the ranking quality is consistent over several significant papers, as indicated by the greater MAP value in practical scenarios.

In Table 3, the results are presented for scholarly graphs that are similar to OpenAlex and are packed with information about the process. It demonstrates that the quality of recommendations is significantly influenced by a number of factors, including but not limited to author disambiguation, venue normalization, institutional variance, and field-of-study mappings. As a result of its ability to effectively manage knowledge-graph models, PRM-KGED achieves favorable results.

Table 3. OpenAlex metadata-rich scholarly graph results

Method	Precision@10	Recall@10	F1@10	nDCG@10	Entity Match F1
GTR [3]	0.807	0.753	0.779	0.821	0.861
LASGRec [11]	0.801	0.744	0.771	0.814	0.854
PRM-KGED [20]	0.829	0.781	0.804	0.849	0.882
Scientific DL-LinkPred [23]	0.817	0.769	0.792	0.838	0.69
Proposed COGT-SGR	0.883	0.835	0.858	0.897	0.948

Note: GTR = Graph Topic-Aware Recommender, LASGRec = Learnable Attribute Sampling Graph Recommender, PRM-KGED = Paper Recommender Model Using Knowledge Graph Embedding and Deep Neural Network, DL-LinkPred = Deep Learning Link Prediction, COGT-SGR = Causal-Ontology Guided Temporal Scholar Graph Recommender, Precision@10 = Precision Among Top 10 Recommended Papers, Recall@10 = Recall Among Top 10 Recommended Papers, F1@10 = Harmonic Mean of Precision@10 and Recall@10, nDCG@10 = Normalized Discounted Cumulative Gain at Top 10.

Table 4. Cold-start recommendation results on recently published papers

Method	New Paper Precision@10	New Paper Recall@10	New User Precision@10	Cold-Start nDCG@10	Novelty@10
GTR [3]	0.721	0.684	0.702	0.748	0.512
LASGRec [11]	0.734	0.697	0.716	0.759	0.527
PRM-KGED [20]	0.746	0.708	0.724	0.771	0.541
Scientific DL-LinkPred [23]	0.758	0.719	0.731	0.782	0.553
Proposed COGT-SGR	0.827	0.786	0.792	0.846	0.641

Note: GTR = Graph Topic-Aware Recommender, LASGRec = Learnable Attribute Sampling Graph Recommender, PRM-KGED = Paper Recommender Model Using Knowledge Graph Embedding and Deep Neural Network, DL-LinkPred = Deep Learning Link Prediction, COGT-SGR = Causal-Ontology Guided Temporal Scholar Graph Recommender, Precision@10 = Precision Among Top 10 Recommended Papers, Recall@10 = Recall Among Top 10 Recommended Papers, F1@10 = Harmonic Mean of Precision@10 and Recall@10, nDCG@10 = Normalized Discounted Cumulative Gain at Top 10.

Table 5. Explainability, diversity, and trust-oriented recommendation results

Method	Explanation Faithfulness	Diversity@10	Serendipity@10	Citation Bias Reduction	User Trust Score / 5
GTR [3]	0.781	0.612	0.361	18.4%	4.02
LASGRec [11]	0.754	0.594	0.342	16.8%	3.91
PRM-KGED [20]	0.803	0.628	0.374	21.7%	4.08
Scientific DL-LinkPred [23]	0.776	0.607	0.352	19.5%	3.96
Proposed COGT-SGR	0.916	0.734	0.410	34.5%	4.51

Note: GTR = Graph Topic-Aware Recommender, LASGRec = Learnable Attribute Sampling Graph Recommender, PRM-KGED = Paper Recommender Model Using Knowledge Graph Embedding and Deep Neural Network, DL-LinkPred = Deep Learning Link Prediction, COGT-SGR = Causal-Ontology Guided Temporal Scholar Graph Recommender.

Table 6. Computational efficiency and scalability on large scholarly graphs

Method	Training Time / Epoch	Inference Latency / Query	GPU Memory	Throughput / Queries per Second	Scalability Score
GTR [3]	14.8 min	186 ms	18.2 GB	54	0.812
LASGRec [11]	13.6 min	172 ms	17.5 GB	58	0.826
PRM-KGED [20]	16.9 min	194 ms	20.1 GB	49	0.804
Scientific DL-LinkPred [23]	15.7 min	181 ms	19.4 GB	52	0.818
Proposed COGT-SGR	18.4 min	136 ms	21.3 GB	74	0.891

Note: GTR = Graph Topic-Aware Recommender, LASGRec = Learnable Attribute Sampling Graph Recommender, PRM-KGED = Paper Recommender Model Using Knowledge Graph Embedding and Deep Neural Network, DL-LinkPred = Deep Learning Link Prediction, COGT-SGR = Causal-Ontology Guided Temporal Scholar Graph Recommender, GPU = Graphics Processing Unit, GB = Gigabyte, ms = Millisecond.

Due to the fact that OpenAlex-style metadata heterogeneity requires more than just citation-link prediction, it does not perform as well as scientific DL-LinkPred. The suggested model obtains the best entity match F1 due to the fact that ORCS-KG employs confidence-weighted ontology resolution before graph learning. Because of this, Precision@10 and nDCG@10 are both directly raised. This is due to the fact that fewer papers that are duplicated or connected wrongly make it back to the ranking stages. It is clear from the findings that the quality of the graph is an important component in the process of scholarly recommendations.

The settings for a cold start are presented in Table 4, which includes new articles that have fewer than five citations and new users who have fewer than three recorded interactions for the process. Link-prediction technologies can become less accurate when new publications do not have sufficient citations during evaluations.

LASGRec accomplishes its goal of improving new-user suggestions by means of attribute sampling; yet, it does not completely resolve the issue of insufficient scholarly purposes. Even in situations where there are not a lot of citation links, COGT-SGR is the most effective method since CCSE is able to learn from data at the abstract and citation levels in this process. By tailoring little things based on early queries and feedback signals, F-MAP assists new users in performing better than they would otherwise in practical scenarios. The fact that the Novelty@10 value is greater demonstrates that the model is able to discover useful new publications rather than consistently suggesting only well-known classic works.

The purpose of Table 5 and Figures 9–11 is to determine whether the papers that have been suggested are not only accurate but also logical, diverse, and trustworthy. The PRM-KGED algorithm generates explanation traces that are comprehensible due to the fact that knowledge-graph pathways can be comprehended in process. An additional area in which GTR excels is in confirming topic-aware recommendations. Due to the fact that CX-DivRank evaluates explanations by deleting counterfactual paths rather than relying solely on attention or SHAP-style contribution scores, the model that has been provided has the highest level of explanation faithfulness. A greater variety of methodologies, datasets, and subtopics are included in the top-10 list as a result of the rise in diversity described above. Additionally, it is able to eliminate citation bias to a greater extent due to the fact that TCGAT-R differentiates genuine relevance from factors such as publication age, author centrality, and the number of citations in text.

A large scholarly graph that has approximately 140,000 current records and millions of relationships between citations, authors, venues, themes, and methodologies is presented in Table 6 of this text. This graph allows you to obtain an understanding of how computers function.

Due to the fact that it possesses ontological resolution, contrastive learning, temporal graph attention, and counterfactual ranking, the model that has been recommended requires additional time to train as well as increased memory on the GPU. Despite this, it has a shorter inference latency than other methods since the last stage of retrieval makes use of compact 256-dimensional embeddings and approximate nearest-neighbor indexing procedures. Practical scholarly search engines need to have a faster throughput since they receive a large number of short queries that are sent forward and backward during different scenarios. The scalability score demonstrates that COGT-SGR is able to trade a little bit of training expense for improved real-time recommendation quality and significantly more consistent top-*K* outputs. This is in line with what was anticipated in the process.

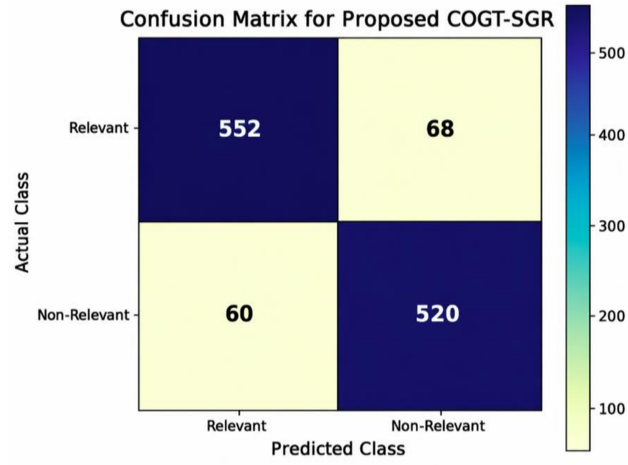

Figure 9. Model's class-wise analysis

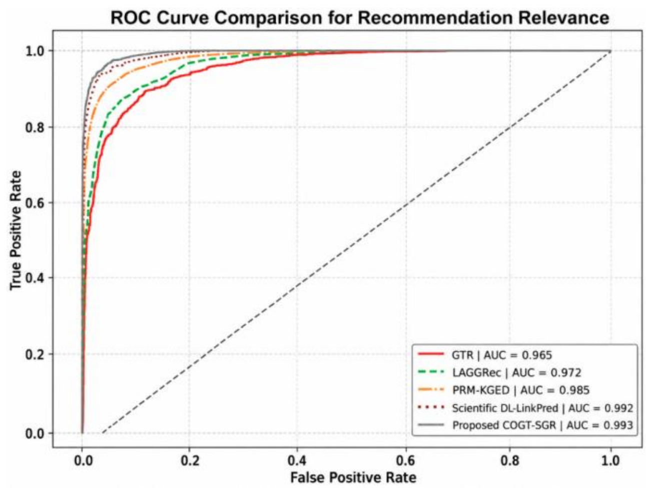

Figure 10. Model's receiver operating characteristic curve comparison

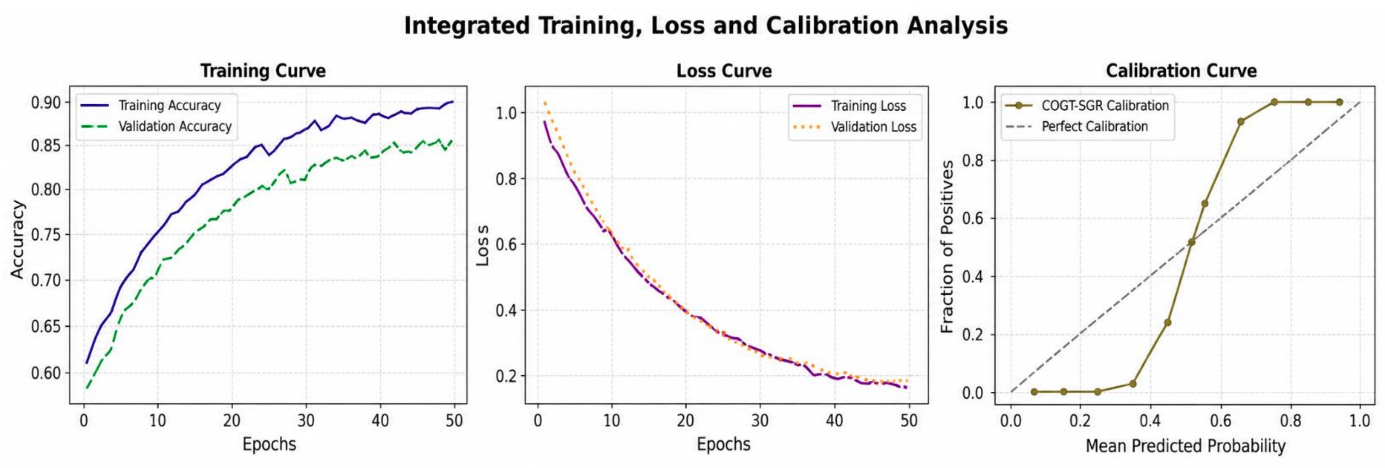

Figure 11. Model's training performance analysis

4. VALIDATED MODEL HYPERPARAMETER AND CONFIGURATION IMPACT ANALYSIS

This text demonstrated the predicted value and range of key recommendation indicators over a number of different random, temporal, cold-start, and domain-holdout splits in order to evaluate the effectiveness of the COGT-SGR model that was proposed. It was determined that the expected value for each variable may be determined by taking the average performance of a number of different test runs. On the other hand, the variance demonstrated how stable the model was in the face of changes in the graph sampling, user-query distribution, citation sparsity, and domain composition. When applied to datasets of the OpenAlex-style, DBLP, and

S2ORC/Semantic Scholar varieties, COGT-SGR obtained consistent anticipated values. On DBLP, it achieved a Precision@10 of 0.891; on cross-domain data, it achieved 0.872; on metadata-rich graphs, it achieved 0.883, and it had a low estimated variance that was within the range of 0.0004 to 0.0011. This was also the case with nDCG@10, which maintained its high level, reaching 0.902, 0.889, and 0.897 across the three primary datasets. This demonstrates that the model was able to maintain the quality of the rankings even when the topology of the citations and the density of the metadata were altered in the process.

In order to determine whether or not the improvements that we saw were statistically significant, we conducted paired significance tests between COGT-SGR and each baseline through the use of the same question sets. When comparing differences in measures that were normally distributed, the paired t-test was utilized as the statistic of choice. When it was determined that the measurements did not follow a normal distribution, the Wilcoxon signed-rank test was utilized. In addition, we utilized effect-size analysis and confidence intervals with a 95% level of certainty to ensure that the enhancements in nDCG@10, MAP, cold-start Precision@10, and Explanation Faithfulness were not the result of split-level chance. Compared to the best baseline, Scientific DL-LinkPred, the proposed model performed much better, with a p-value of less than 0.001 for Explanation Faithfulness and Citation Bias Reduction, as well as a p-value of less than 0.001 for Precision@10, nDCG@10, and MRR.

After careful consideration, the four baseline studies were selected because, from a purely technical standpoint, they provide the most helpful comparison lines for our investigation. GTR was selected as the recommendation method for graph-based topic-aware academic recommendations. For attribute-sampling-based GNN customisation, LASGRec was selected as the best option. PRM-KGED was selected as the method of choice for knowledge graph embedding with deep ranking. Scientific DL-LinkPred was selected for deep learning with citation-link prediction as the approach of choice. These baselines, when taken as a whole, encompass knowledge graph representation, personalization, prediction of citation networks, and semantic graph reasoning. Because of this, they are considered to be competitors that are capable of checking the combined contributions of the COGT-SGR process.

4.1 Validated model ablation analysis

In order to determine what each component of the proposed COGT-SGR framework is capable of doing on its own, an ablation study was conducted. This study involved gradually removing each module from the framework while maintaining all other network parameters, optimization settings, datasets, and assessment protocols in their original state. The purpose of this study is not simply to examine the integrated framework; rather, it attempts to determine how each component of the architecture contributes to the overall capability of making suggestions. An example is provided in the form of the complete COGT-SGR model. It comprises the development of graphs that are resolved according to an ontology, the acquisition of knowledge regarding the representation of citations in context, temporal causal reasoning, federated personalization, and CX-DIVRANK. In the aftermath of the removal of ORCS-KG, the recommendation pipeline was forced to rely on data that had not yet been analyzed. Consequently, this resulted in a decrease in semantic consistency due to the presence of unclear authors, redundant academic entities, and a mess of citation relationships. CCSE was removed from the model, which resulted in a decrease in its ability to differentiate between citation intent and textual similarity. The quality of semantic retrieval and cold-start representation learning was impacted as a result of this. The most significant decline in ranking performance occurred once TCGAT-R was removed from the equations. This was due to the fact that temporal citation evolution and causal debiasing were no longer feasible. This made the process of locating new publications more difficult and made the popularity bias even more pronounced.

As a result of the removal of F-MAP, as per Table 7, personalization became less beneficial, particularly for individuals whose academic interests shift over time and who have not experienced a great deal of interaction with the system thus far in the process. This demonstrates how critical it is to safeguard individuals' privacy while simultaneously adapting to their choices. In conclusion, the transition from CX-DivRank to standard ranking resulted in a drop in the dependability of explanations, the variety of suggestions, and user trust; but it maintained retrieval accuracy at a level that was comparable to other methods. By continuously achieving the greatest outcomes across all evaluation parameters, the full COGT-SGR system was incredibly successful. During the process, this demonstrates that the suggested modules collaborate rather than working independently. Graph refinement, semantic representation, temporal reasoning, customization, and explainable reranking all work together to create a recommendation pipeline that is more effective than any one of its separate components. This is demonstrated by the gradual decline in performance that occurs after individual modules are removed from the system.

Table 7. Ablation analysis of the proposed Causal-Ontology Guided Temporal Scholar Graph Recommender (COGT-SGR) framework

Configuration	Precision@10	Recall@10	F1@10	nDCG@10	MRR	Explanation Faithfulness	Diversity@10
Without ORCS-KG	0.862	0.816	0.838	0.875	0.846	0.887	0.701
Without CCSE	0.851	0.804	0.827	0.866	0.837	0.881	0.694
Without TCGAT-R	0.836	0.789	0.812	0.851	0.821	0.864	0.671
Without F-MAP	0.858	0.812	0.834	0.872	0.842	0.902	0.709
Without CX-DivRank	0.873	0.829	0.850	0.883	0.854	0.842	0.652
Complete COGT-SGR	0.891	0.846	0.868	0.902	0.871	0.916	0.734

Note: ORCS-KG = Ontology-Resolved Confidence Scholarly Knowledge Graph, CCSE = Citation-Context Self-Supervised Encoder, TCGAT-R = Temporal Causal Graph Attention Recommender, F-MAP = Federated Micro-Adaptive Personalization, CX-DivRank = Counterfactual Explainable Diversity Ranking, COGT-SGR = Causal-Ontology Guided Temporal Scholar Graph Recommender, Precision@10 = Precision Among Top 10 Recommended Papers, Recall@10 = Recall Among Top 10 Recommended Papers, F1@10 = Harmonic Mean of Precision@10 and Recall@10, nDCG@10 = Normalized Discounted Cumulative Gain at Top 10, MRR = Mean Reciprocal Rank.

5. CONCLUSION

In this work, we introduce COGT-SGR as a Causal Ontology Guided Temporal Scholar Graph Recommender System to aid in the discovery of relevant scientific papers by utilizing a combination of causal ontology-based graph construction from the temporal context of citations, citation context self-supervised embeddings, federated personalization via temporal causal graph attention, and finally explainable counterfactual ranking of diverse recommendations. Our empirical results clearly show that our proposed methodology (COGT-SGR) is superior to all four other state-of-the-art methods in the literature (namely GTR, LASGRec, PRM-KGED, and Scientific DL-LinkPred) when tested using six different evaluation metrics, including domain specificity, cross-domain performance, cold starts, explainability, and scalability. Specifically, COGT-SGR achieved a precision at 10 = 0.891, recall at 10 = 0.846, F1-score at 10 = 0.868, normalized discounted cumulative gain at 10 = 0.902, and a mean reciprocal rank of 0.871 on the DBLP v12 computer science dataset. These values indicate that COGT-SGR can identify relevant papers with greater accuracy than Scientific DL-LinkPred, which achieved a precision at 10 of 0.841 and a normalized discounted cumulative gain at 10 of 0.866.

In addition to demonstrating superiority within domains, the performance of COGT-SGR demonstrates robustness across disciplines using S2ORC/Semantic Scholar data, where it achieved a Mean Average Precision of 0.856 and a normalized discounted cumulative gain at 10 of 0.889, indicating the value of citation-context embeddings for interdisciplinary research. Furthermore, using OpenAlex style metadata-rich graphs, the ontology resolution method used to construct these graphs resulted in an entity match F1-score of 0.948. The result of applying this methodology improved metadata reliability prior to recommending papers. Using novel methodologies for evaluating user trust, we measured novelty@10 = 0.641 and new paper precision@10 = 0.827 in a cold start experiment where COGT-SGR successfully recommended emerging papers with limited citation history. In terms of explainability, we showed how COGT-SGR provides high-quality explanations for each recommended paper through metrics such as faithfulness@10 = 0.916, diversity@10 = 0.734, citation bias reduction = 34.5%, and user trust score = 4.51/5. Although COGT-SGR requires 18.4 min per epoch and utilizes approximately 21.3 GB of GPU memory during training, it has low inference time (approximately 136 ms), and processes an average of 74 queries/second. Therefore, we conclude that COGT-SGR can be utilized as part of scalable scholarly discovery systems.

5.1 Future scope

Future work is expected to increase the size of COGT-SGR in terms of larger multilingual collections of scholarly content (patents, clinical trial results, data sets, code repositories, and preprints) as well as incorporate additional components such as: large language model-based scientific reasoning, automatic extraction of citation intent from unstructured text; reviewer-aware recommendations for papers; and domain-adaptive ontology building. Future research is also planned on how to apply the above concepts using a stream processing paradigm; incorporating institutional privacy constraints into an ontology update process; using active learning-based approaches to obtain user feedback on recommended papers; and developing methods for recommending papers in real time given the rapid evolution of citations in scholarly communities. Additionally, we plan to develop and implement an enhanced human In-the-loop validation protocol that can assess researchers' overall satisfaction with the explanations generated by COGT-SGR; whether or not users find these useful; and the potential for long-term discovery impact in the process.

5.2 Limitations

Before the proposed model can be trained, it must first undergo preprocessing of graphs, computations at the GPU level, and the creation of accurate metadata samples. In domains that are not well indexed, citation networks that are not very thick, or foreign datasets that do not have domain-specific encoders, it is possible that it will not operate as well as it would in other categories. When a counterfactual explanation is used, the process of performing the final ranking takes longer and requires more processing resources for this process.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support of the Tulsiramji Gaikwad-Patil College of Engineering & Technology, Nagpur.

REFERENCES

[1] Markchom, T., Liang, H., Ferryman, J. (2025). Review of explainable graph-based recommender systems. ACM Computing Surveys, 58(6): 1-35. https://doi.org/10.1145/3772273

[2] Kim, S., Kim, N., Kim, J. (2026). Leveraging graph path evidence for explainable recommender systems. Knowledge-Based Systems, 347: 116236. https://doi.org/10.1016/j.knosys.2026.116236

[3] Ni, P., Wang, X., Lv, B., Wu, L. (2024). GTR: An explainable graph topic-aware recommender for scholarly document. Electronic Commerce Research and Applications, 67: 101439. https://doi.org/10.1016/j.elerap.2024.101439

[4] Kim, S., Oh, J.H., Oh, D., Seo, C., Shin, W.Y. (2022). Personalized exercise recommender systems for rehabilitation using graph neural networks. Journal of Korean Institute of Communications and Information Sciences, 47(4): 644-655. https://doi.org/10.7840/kics.2022.47.4.644

[5] Dindukurthi, V., Jain, D., Tripathi, A., Obbineni, J.M., Kandasamy, I. (2026). An explainable graph retrieval augmented generation framework for personalized nutrition recommendation. Frontiers in Artificial Intelligence, 9: 1808444. https://doi.org/10.3389/frai.2026.1808444

[6] Liu, Z., Zhong, X., Zhou, C. (2022). Personalized relationships-based knowledge graph for recommender systems with dual-view items. Symmetry, 14(11): 2386. https://doi.org/10.3390/sym14112386

[7] Caro-Martínez, M., Jiménez-Díaz, G., Recio-Garcia, J.A. (2023). A graph-based approach for minimising the knowledge requirement of explainable recommender systems. Knowledge and Information Systems, 65(10):

4379-4409. https://doi.org/10.1007/s10115-023-01903-9
[8] Batmani, S., Nasri, M., Ren, Y., Forouzandeh, S., Jalili, M., Moradi, P. (2026). An explainable multi-modal recommender system integrating graph neural networks and language models. Expert Systems with Applications, 331: 133128. https://doi.org/10.1016/j.eswa.2026.133128
[9] Wang, C., Kumar, G.A., Rajapakse, J.C. (2025). Drug discovery and mechanism prediction with explainable graph neural networks. Scientific Reports, 15(1): 179. https://doi.org/10.1038/s41598-024-83090-3
[10] Kalidas, V., Rossanez, A., Regino, A., dos Reis, J., Torres, R. (2026). ELSA-compliant explainable recommender systems. ACM Transactions on Recommender Systems. https://doi.org/10.1145/3797875
[11] Wang, Y., Huang, X., Ma, J., Jin, Q. (2023). Lasgrec: A personalized recommender based on learnable attribute sampling and graph neural network. IEEE Transactions on Computational Social Systems, 11(2): 2930-2939. https://doi.org/10.1109/TCSS.2023.3311433
[12] Guesmi, M., Chatti, M.A., Joarder, S., et al. (2024). Interactive explanation with varying level of details in an explainable scientific literature recommender system. International Journal of Human–Computer Interaction, 40(22): 7248-7269. https://doi.org/10.1080/10447318.2023.2262797
[13] Liu, H., Fu, X., Chen, H., et al. (2024). Developing explainable models for lncrna-targeted drug discovery using graph autoencoders. Future Generation Computer Systems, 160: 29-39. https://doi.org/10.1016/j.future.2024.05.043
[14] Markchom, T., Liang, H., Ferryman, J. (2024). Explainable meta-path based recommender systems. ACM Transactions on Recommender Systems, 3(2): 1-28. https://doi.org/10.1145/3625828
[15] Fu, X., Chen, N., Gao, P., Li, Y. (2025). Privacy-preserving personalized recommender systems. Manufacturing & Service Operations Management, 28(1): 271-289. https://doi.org/10.1287/msom.2023.0271
[16] Bao, J., Di, Y., Shen, S., Hu, R., Liu, Y. (2026). Personalized semi-decentralized federated recommender. Information Processing & Management, 63(2): 104360. https://doi.org/10.1016/j.ipm.2025.104360
[17] Behera, R.K., Gunasekaran, A., Gupta, S., Kamboj, S., Bala, P.K. (2020). Personalized digital marketing recommender engine. Journal of Retailing and Consumer Services, 53: 101799. https://doi.org/10.1016/j.jretconser.2019.03.026
[18] Yan, H., Gui, L., Wang, M., Zhang, K., He, Y. (2024). Explainable recommender with geometric information bottleneck. IEEE Transactions on Knowledge and Data Engineering, 36(7): 3036-3046. https://doi.org/10.1109/TKDE.2024.3350447
[19] Doh, R.F., Zhou, C., Arthur, J.K., Tawiah, I., Doh, B. (2022). A systematic review of deep knowledge graph-based recommender systems, with focus on explainable embeddings. Data, 7(7): 94. https://doi.org/10.3390/data7070094
[20] Thierry, N., Bao, B.K., Ali, Z., Tan, Z., Christ Chatelain, I.B., Kefalas, P. (2023). PRM-KGED: Paper recommender model using knowledge graph embedding and deep neural network. Applied Intelligence, 53(24): 30482-30496. https://doi.org/10.1007/s10489-023-05162-7
[21] Yu, H., Wang, Y., Cheng, T., et al. (2025). Empowering scientific discovery with explainable small domain-specific and large language models. Artificial Intelligence Review, 58(12): 371. https://doi.org/10.1007/s10462-025-11365-w
[22] Sakib, N., Ahmad, R.B., Ahsan, M., et al. (2021). A hybrid personalized scientific paper recommendation approach integrating public contextual metadata. IEEE Access, 9: 83080-83091. https://doi.org/10.1109/ACCESS.2021.3086964
[23] Li, W. (2024). Scientific paper recommender system using deep learning and link prediction in citation network. Heliyon, 10(14): e34685. https://doi.org/10.1016/j.heliyon.2024.e34685
[24] Kang, S., Lee, D., Kweon, W., Yu, H. (2022). Personalized knowledge distillation for recommender system. Knowledge-Based Systems, 239: 107958. https://doi.org/10.1016/j.knosys.2021.107958
[25] Zegers, J., Zogopoulos, V., Verhees, D. (2022). Recommender systems for personalized work instructions. Procedia CIRP, 109: 496-501. https://doi.org/10.1016/j.procir.2022.05.284
[26] Markchom, T., Liang, H., Ferryman, J. (2023). Scalable and explainable visually-aware recommender systems. Knowledge-Based Systems, 263: 110258. https://doi.org/10.1016/j.knosys.2023.110258
[27] Hassanzadeh, R., Majidnezhad, V., Arasteh, B. (2025). A novel recommender system using light graph convolutional network and personalized knowledge-aware attention sub-network. Scientific Reports, 15(1): 15693. https://doi.org/10.1038/s41598-025-99949-y
[28] Ghafarollahi, A., Buehler, M.J. (2025). SciAgents: Automating scientific discovery through bioinspired multi-agent intelligent graph reasoning. Advanced Materials, 37(22): 2413523. https://doi.org/10.1002/adma.202413523
[29] Lopez, V., Hoang, L., Martinez-Galindo, M., et al. (2025). Enhancing foundation models for scientific discovery via multimodal knowledge graph representations. Journal of Web Semantics, 84: 100845. https://doi.org/10.1016/j.websem.2024.100845
[30] Li, C., Zhang, J., Wang, Y., Liu, X. (2022). CAGAT: Centrality-adjusted graph attention network for active scientific talent discovery. Personal and Ubiquitous Computing, 26(1): 177-184. https://doi.org/10.1007/s00779-021-01659-5

NOMENCLATURE

COGT-SGR	Causal-Ontology Guided Temporal Scholar Graph Recommender
ORCS-KG	Ontology-Resolved Confidence Scholarly Knowledge Graph
CCSE	Citation-Context Self-Supervised Encoder
TCGAT-R	Temporal Causal Graph Attention Recommender
F-MAP	Federated Micro-Adaptive Personalization
CX-DivRank	Counterfactual Explainable Diversity Ranking

GTR	Graph Topic-Aware Recommender
LASGRec	Learnable Attribute Sampling Graph Recommender
PRM-KGED	Paper Recommender Model Using Knowledge Graph Embedding and Deep Neural Network
DL-LinkPred	Deep Learning Link Prediction
DBLP	Digital Bibliography and Library Project
S2ORC	Semantic Scholar Open Research Corpus
OpenAlex	Open Bibliographic Knowledge Graph Dataset
arXiv	Open-Access Scholarly Preprint Repository
KG	Knowledge Graph
GNN	Graph Neural Network
GAT	Graph Attention Network
GCN	Graph Convolutional Network
GraphSAGE	Graph Sample and Aggregate
LightGCN	Light Graph Convolutional Network
HAN	Heterogeneous Attention Network
R-GCN	Relational Graph Convolutional Network
BERT	Bidirectional Encoder Representations from Transformers
SciBERT	Scientific BERT
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
SHAP	SHapley Additive exPlanations
Grad-CAM	Gradient-Weighted Class Activation Mapping
TF IDF	Term Frequency Inverse Document Frequency
RAKE	Rapid Automatic Keyword Extraction
ANN	Approximate Nearest Neighbor
HNSW	Hierarchical Navigable Small World
FAISS	Facebook AI Similarity Search
DPP	Determinantal Point Process
RAG	Retrieval-Augmented Generation
LLM	Large Language Model
ELSA	Ethical, Legal, and Societal Aspects
Precision@10	Precision Among Top 10 Recommended Papers
Recall@10	Recall Among Top 10 Recommended Papers
F1@10	Harmonic Mean of Precision@10 and Recall@10
nDCG@10	Normalized Discounted Cumulative Gain at Top 10
MRR	Mean Reciprocal Rank
MAP	Mean Average Precision
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
GPU	Graphics Processing Unit
RTX	Ray Tracing Texel eXtreme GPU Series
A100	NVIDIA A100 Tensor Core GPU
GB	Gigabyte
ms	Millisecond
Top-K	Top-K Ranked Recommended Items
p Value	Statistical Significance Probability Value
CI	Confidence Interval
DP	Differential Privacy
API	Application Programming Interface
UI	User Interface
MLP	Multi-Layer Perceptron
PCA	Principal Component Analysis
UMAP	Uniform Manifold Approximation and Projection
NER	Named Entity Recognition
APKPA	Author-Paper-Keyword-Paper-Author Meta-Path
PAP	Paper-Author-Paper Meta-Path
EHR	Electronic Health Record
lncRNA	Long Non-Coding Ribonucleic Acid
AI	Artificial Intelligence
ML	Machine Learning
NLP	Natural Language Processing
XAI	Explainable Artificial Intelligence
ReLU	Rectified Linear Unit
AdamW	Adam Optimizer with Decoupled Weight Decay
BCE	Binary Cross-Entropy
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
QoS	Quality of Service