Proceedings of the 4th International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2026.32917

Dual-Pathway Heterogeneous Graph Neural Networks for Academic Impact Prediction

Chunping Cao[1], Xinqi Li[1*]

[1]*School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, China*

**Corresponding Author. Email: 335895802@qq.com*

Abstract. This study aims to improve the early prediction of academic paper impact. To this end, a dual-pathway dynamic heterogeneous graph neural network model is proposed. The model constructs a temporal heterogeneous graph consisting of multiple types of nodes, including papers, authors, institutions, and journals. Citation and non-citation pathways are designed to model knowledge diffusion relations and social-semantic associations, respectively. In addition, a life-cycle-aware mechanism is introduced to capture feature variations across the emerging, growing, and mature stages of papers. On this basis, the prediction outcome is further disentangled into three independent components, namely diffusion effect, social bias effect, and intrinsic value effect, thereby enhancing the interpretability of the model. Experimental results on datasets from multiple disciplines, including computer science, chemistry, and psychology, demonstrate that the proposed method outperforms existing mainstream models in terms of mean absolute log error and log-transformed coefficient of determination. It is especially more accurate and stable in cold-start scenarios. The results indicate that the proposed method can provide effective support for the early identification of high-potential papers.

Keywords: Citation prediction, Heterogeneous graph neural network, Impact disentanglement, Dynamic modeling, Attention mechanism, Cold start.

1. Introduction

Academic papers serve as the primary carrier of scientific research outcomes. With the exponential growth in the number of academic papers published globally, early identification of high-impact papers from the massive body of literature is of great significance for optimizing the allocation of research resources and grasping the frontiers of disciplinary development. However, current academic impact evaluation typically relies on cumulative citation counts, which cannot fairly compare papers of different ages [1,2]. Therefore, future-oriented prediction models are urgently needed to assess the potential impact of papers [2,3].

Early citation prediction relied on time-series fitting, such as the WSB model [4] and reinforced Poisson process [5], later extended by deep sequence models based on LSTM [6] and Transformer [7]; these methods depend on historical citations and ignore network topology. PageRank-based approaches [8-10] introduced network coupling but remain limited to single-relation graphs.

DOI: 10.54254/2755-2721/2026.32917

Heterogeneous GNNs have since become mainstream: HGT [11] models heterogeneous propagation via meta-relation attention but applies uniform aggregation regardless of paper age; HINTS [12] addresses cold-start via dynamic heterogeneous embeddings, yet is evaluated only on single-domain networks; H²CGL [13] and ROLAND [14] enhance temporal sensitivity through dynamic modeling but lack explicit impact disentanglement. For disentanglement, DisenGCN [15] and SILD [16] separate latent factors on graphs, while DPPDCC [17] decomposes citations into diffusivity, conformity, and intrinsic contribution via orthogonality constraints.

However, existing models still have the following limitations: First, newly published papers have extremely sparse information, making it difficult for models to accurately capture early potential impact signals, resulting in underestimated predictions for cold-start papers. Second, papers at different life-cycle stages exhibit significant differences in feature distribution and growth patterns, yet most models still adopt a unified aggregation strategy without stage-specific adaptation. Third, the coupled modeling of academic contribution and social diffusion effects causes social conformity to mask the intrinsic academic value of papers. In particular, the hard orthogonal constraint of DPPDCC ignores the positive correlation between the intrinsic contribution and diffusive citations of high-quality papers, and for new papers lacking diffusion and conformity signals, the model effectively degenerates into purely content-driven prediction.

To address these issues, we propose a Dual-Pathway Dynamic Heterogeneous Graph Neural Network (DP-DHGNN) with three key innovations: (1) a dual-pathway encoding mechanism that separately models knowledge propagation via a citation pathway and social-semantic relations (authorship, affiliation) via a non-citation pathway, alleviating information sparsity under cold-start conditions; (2) a life-cycle-aware mechanism that dynamically adjusts aggregation strategies across evolutionary stages to capture non-stationary early-stage growth patterns; and (3) a multi-factor disentanglement framework with soft orthogonal constraints that separates propagation, conformity, and intrinsic value effects, improving interpretability and early identification of high-potential papers.

2. Dual path dynamic heterogeneous graph neural network model

The citation prediction task aims to predict the citation increment within the future time window of a paper based on its historical characteristics and network structure. Formally, given the set of papers $\mathrm{P} = \{\mathrm{p}_1, \mathrm{p}_2, \cdots, \mathrm{p}_\mathrm{n}\}$, observing the time point tobs and the prediction time window $\Delta \mathrm{t}$, the goal is to learn the mapping function f such that:

$$= f(p_1, G_{\{t_{obs}\}}) \tag{1}$$

In the formula, $\hat{\mathrm{y}}_\mathrm{i}$ is the predicted quoted increment of the paper p_i in the time window $[\mathrm{t}_\mathrm{obs}, \mathrm{t}_\mathrm{obs} + \Delta \mathrm{t}]$; $G_{\{t_{obs}\}}$ is the dynamic heterogeneous citation network at the observation time point. The task is modeled as a regression problem, optimized with log-transformed mean squared error to accommodate the long-tail distribution of citation data.

2.1. Overall framework of the model

The overall methodological framework is shown in Figure 1.

DOI: 10.54254/2755-2721/2026.32917

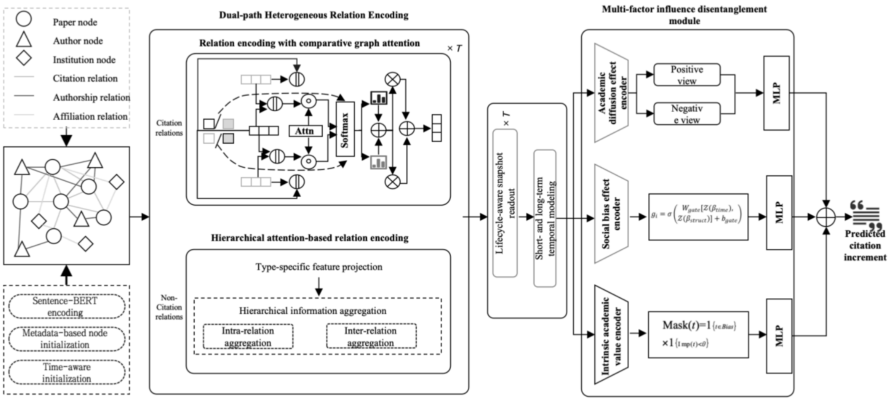

Figure 1. Overall architecture

The model primarily consists of four modules: the Dynamic Heterogeneous Network Construction module, the Dual-Pathway Heterogeneous Relation Encoding module, the Lifecycle-Aware Snapshot Readout module, and the Multi-Factor Influence Disentanglement module. The Dual-Pathway Encoding module splits information flow into a citation pathway and a non-citation pathway, separately capturing knowledge propagation structure and social-semantic associations to alleviate information sparsity under cold-start conditions. The Lifecycle-Aware Readout module adaptively adjusts feature aggregation weights according to the evolutionary stage of each paper. The Disentanglement module decomposes the citation increment prediction into propagation effect, social bias effect, and intrinsic value effect, enhancing both interpretability and early-stage prediction accuracy.

2.2. Dynamic heterogeneous citation network construction

The academic network is represented as a sequence of snapshots $\{G_\tau\}_{\tau=1}^{t}$, where each snapshot $G_\tau = (V_\tau, E_\tau, \phi, \psi)$ contains multi-type nodes (ϕ) and multi-type edges (ψ). Co-citation connections are constructed based on normalized Jaccard coefficients to capture latent semantic associations between papers. Node representations are initialized by combining Sentence-BERT semantic embeddings for paper nodes, weighted-average semantic vectors for auxiliary nodes (authors, institutions, journals), structural features, and sinusoidal time encodings fused via a gating mechanism. For each target paper, a k-hop (k=2) local subgraph is extracted with hierarchical sampling to form the model input.

2.3. Dual-path heterogeneous relationship coding mechanism

In the early stage of publication, citation relations have not fully formed, making it difficult to capture a paper's potential impact from citation structure alone. Meanwhile, simply mixing different relation types introduces noise. We therefore split the information flow into two complementary pathways.

2.3.1. Citation relationship coding based on co-citation structure

The hierarchical attention mechanism is shown in Figure 2.

DOI: 10.54254/2755-2721/2026.32917

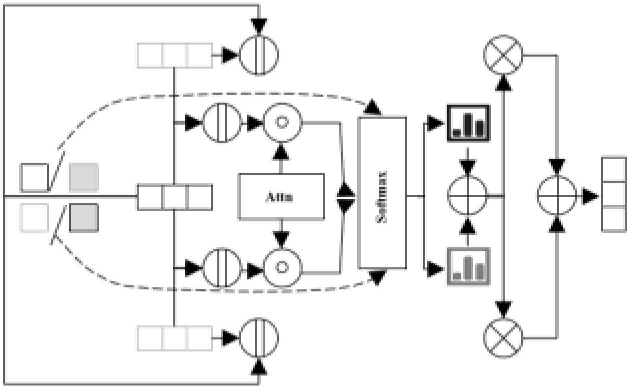

Figure 2. CiteLink-GAT block diagram

To balance structural and semantic information under varying local graph densities, the aggregation weight between nodes is defined as a convex combination:

$$w_{\nu u} = \rho \cdot S_{cocite}(\nu, u) + (1 - \rho) \cdot S_{semantic}(\nu, u) \tag{2}$$

where $\rho = |E_local|/(|V_local| \cdot (|V_local| - 1))$ is the local graph density computed at the subgraph level for each target paper within each snapshot, S_{cocite} is the co-citation strength based on normalized Jaccard coefficients, and $S_{semantic}$ is the node embedding similarity. Node representations are updated by weighted neighborhood aggregation using the above weights. Local attention is further applied to suppress noise in high-density regions and strengthen semantic propagation in low-density regions.

2.3.2. Non-citation relational coding for hierarchical attention

The hierarchical attention mechanism is shown in Figure 3.

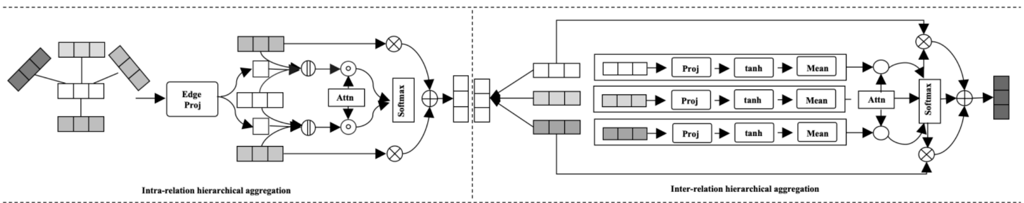

Figure 3. The hierarchical attention mechanism

For papers that have not yet been cited, social attributes such as authorship, institutional affiliation, and journal venue are the primary sources affecting visibility. The non-citation pathway first maps heterogeneous nodes into a unified semantic space via type-specific projections [18]:

$$h^{r}_{0,u} = \pi_u(W_{\pi(u)} x_u), h^{r}_{0,(u,\nu)} = \pi_e(W_{\pi(u,\nu)} x_{(u,\nu)}) \tag{3}$$

Then, a hierarchical attention mechanism is applied: intra-relation attention aggregates same-type neighbors to highlight key nodes, and inter-relation attention dynamically fuses information across different relational channels:

DOI: 10.54254/2755-2721/2026.32917

$$h_u^r = LayerNorm(\Psi^r_{\pi(u)} \cdot h_u^r + (1 - \Psi^r_{\pi(u)}) \cdot h_{u,r-1}^r) \tag{4}$$

The dual-pathway mechanism separately encodes structural propagation and social-semantic information, then merges them during snapshot fusion, enabling robust node representations even with extremely sparse early citations

2.4. Lifecycle-aware snapshot readout mechanism

The lifecycle aware snapshot readout mechanism is shown in Figure 4.

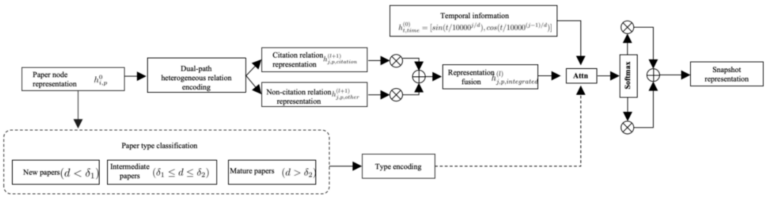

Figure 4. Lifecycle-aware snapshot readout mechanism

Citation evolution typically follows a "latent–burst–stability" pattern with distinct nonlinear characteristics. Traditional models assume a consistent growth pattern, ignoring lifecycle differences, making it difficult to capture early potential signals. We propose a lifecycle-aware mechanism that introduces stage division and time gating.

Papers are classified into emerging, growing, and mature stages based on logistic citation growth curve fitting. It should be noted that logistic curve fitting is only applied to papers with sufficient citation history (typically ≥3 years). For newly published papers that lack citation records, the model assigns them to the emerging stage by default. In this case, the time-gating function automatically assigns higher weight to the non-citation pathway, effectively relying on social-semantic signals rather than citation structure.A time-gating function dynamically fuses the citation and non-citation representations for each paper node:

$$w_i^{cite} = \sigma(W_{gate}[h_i^{cite}; h_i^{non}; e_{type}(i)]) \tag{5}$$

The fused representation $h_i^{fused} = w_i^{cite} \cdot h_i^{cite} + w_i^{non} \cdot h_i^{non}$ ensures smooth transition from social-semantic dominance in early stages to citation-propagation dominance in mature stages. Snapshot representations are obtained via attention aggregation and processed by a Transformer encoder to capture cross-stage temporal dependencies.

2.5. Multi-factor influence decoupling module

Citation behavior is driven by multiple factors: diffusion through citation networks, social conformity from author reputation and institutional prestige, and intrinsic academic quality. Mixing these effects reduces interpretability and prediction stability [19]. We decompose the predicted citation increment as:

$$\Delta\widehat{C}_i = \lambda_P \cdot \Delta\widehat{C}_{i,P} + \lambda_S \cdot \Delta\widehat{C}_{i,S} + \lambda_I \cdot \Delta\widehat{C}_{i,I} \tag{6}$$

Where $\lambda_P, \lambda_S, \lambda_I$ are learnable weights for propagation, social bias, and intrinsic value effects respectively [20]. To encourage decoupled representations while retaining reasonable inter-factor

DOI: 10.54254/2755-2721/2026.32917

correlations, soft orthogonal constraints are adopted [21,22],:

$$\mathscr{L}_{orth} = \sum_{i \neq j} \|Z_i Z_j\|_F^2 \quad (7)$$

Unlike the hard orthogonal constraint, this soft regularization only encourages rather than forces representation independence, allowing the model to adaptively balance between decoupling and factor correlation.

Propagation Effect Encoder .A multi-view contrastive learning framework derives three views from the citation graph. The positive view G^+ is constructed by adding edges between high-cited node pairs with probability proportional to the product of their citation counts. The negative view G^- is generated by randomly dropping edges with probability p_drop = 0.2. The propagation representation is optimized using InfoNCE loss.

Social Bias Effect Encoder .Temporal bias (early-stage abnormal citation attention) and structural bias (PageRank centrality) are adaptively fused through a gating mechanism and mapped to the social bias representation via MLP.

Intrinsic Value Effect Encoder.Intrinsic academic value is evaluated from three dimensions: content richness (information entropy of paper embeddings), terminological novelty (TF-IDF weighted novel term scores), and structural importance (degree and eigenvector centrality). These features are concatenated and linearly projected to obtain the intrinsic value representation.

Fusion and Training. An attention-gated mechanism dynamically assigns weights to the three effects based on lifecycle embeddings:

$$[\alpha_P, \alpha_S, \alpha_I] = Softmax(W_f[h_i^P; h_i^S; h_i^I; e_i]) \quad (8)$$

where e_i is the lifecycle embedding of paper i. The fused representation $h_i^{fused} = \alpha_P \cdot h_i^P + \alpha_S \cdot h_i^S + \alpha_I \cdot h_i^I$ is passed through a prediction layer to output the citation increment. This mechanism forms a dynamic progression from "social exposure → diffusion propagation → intrinsic accumulation" across lifecycle stages. The total loss combines main regression loss, soft orthogonal regularization, and parameter regularization for end-to-end training.

3. Experimental results and discussion

3.1. Experimental setup

Datasets. Experiments are conducted on three disciplinary subsets of the Semantic Scholar Academic Graph (S2AG): Computer Science (CS), Chemistry (CHM), and Psychology (PSY), containing 1.6M, 1.4M, and 1.3M papers respectively. Data is split by observation time rather than random sampling: training at 2012, validation at 2014, and testing at 2017. The test set covers three paper types — seen (existing at training time), recent (newly added), and instant (just published at test time) — to evaluate cold-start performance.

Baselines. We compare against eight methods: content-based methods (SciBERT, SPECTER2), heterogeneous graph methods (HINTS, H^2CGL, ROLAND), disentanglement-based methods (DisenGCN, SILD), and the state-of-the-art DPPDCC (hard and soft variants).

Metrics. MALE (Mean Absolute Log Error, ↓) and LogR^2 (log-space coefficient of determination, ↑) are used to evaluate prediction accuracy under the long-tail citation distribution.

Implementation. All models are implemented in PyTorch 1.10 and DGL 0.8.0 on NVIDIA V100 GPUs. Key hyperparameters: k=2 hops, L=4 layers, d=128 (CS/PSY) / 192 (CHM), learning rate 1e-

DOI: 10.54254/2755-2721/2026.32917

4, Adam optimizer.

3.2. Overall performance

Table 1 reports MALE results on Overall and Instant (cold-start) settings across three datasets. DP-DHGNN consistently achieves the best performance, outperforming the strongest baseline DPPDCC(soft) by 5.2%/3.8%/2.8% on CS/CHM/PSY overall (all $p<0.0042$, Bonferroni correction). Gains are most pronounced on cold-start Instant papers, reaching 8.6% on CHM and 4.1% on PSY, confirming the effectiveness of dual-pathway encoding and lifecycle-aware mechanisms for early-stage prediction.

Among graph-based methods, H²CGL and ROLAND improve over static baselines via dynamic modeling but lack explicit disentanglement, while content-based methods (SciBERT, SPECTER2) and meta-path methods (HINTS) perform substantially worse (MALE > 0.50 on CS) and are omitted from the table. DPPDCC(soft) validates that soft orthogonal constraints outperform hard ones (+5.5% average), consistent with the observed moderate correlation ($r=0.673$) between propagation and intrinsic value effects. DP-DHGNN further improves upon this by jointly optimizing dual-pathway encoding (+4.5% when removed), lifecycle awareness (+12.5% on Instant papers), and multi-factor disentanglement. LogR^2 results show consistent trends, with the largest gains again on cold-start scenarios (CHM: +3.17%, PSY: +2.12%); full LogR^2 results are omitted for space.

Table 1. Model performance comparison (MALE ↓, mean ± standard deviation over five repeated experiments)

Model	CS (Overall)	CS (Instant)	CHM (Overall)	CHM (Instant)	PSY (Overall)	PSY (Instant)
H²CGL	0.510±0.010	0.670±0.016	0.550±0.011	0.720±0.017	0.590±0.012	0.720±0.017
ROLAND	0.500±0.010	0.680±0.016	0.530±0.011	0.750±0.018	0.564±0.012	0.793±0.019
SILD	0.470±0.009	0.590±0.013	0.500±0.010	0.600±0.013	0.530±0.011	0.690±0.016
DPPDCC(hard)	0.370±0.007	0.500±0.010	0.420±0.008	0.560±0.012	0.446±0.009	0.563±0.012
DPPDCC(soft)	0.348±0.007	0.470±0.010	0.395±0.008	0.525±0.011	0.432±0.009	0.542±0.012
DP-DHGNN	0.330±0.006	0.450±0.009	0.380±0.007	0.480±0.010	0.420±0.008	0.520±0.011
Improvement (%)†	5.2*	4.3*	3.8*	8.6*	2.8*	4.1*

3.3. Ablation experiments

Table 2 presents ablation results averaged over the three datasets; trends are consistent across CS, CHM, and PSY ($p<0.001$). The lifecycle-aware mechanism is the most critical component: replacing it with mean pooling increases average MALE by +12.5%, confirming the necessity of stage-adaptive feature aggregation. Intrinsic value disentanglement contributes the second largest gain (+8.5%), as its removal forces the model to rely on external signals and fails to identify "underrated papers" with strong content but low reputation. Removing CompGAT (+7.7%), social bias disentanglement (+7.4%), and propagation disentanglement (+6.9%) each degrade performance substantially. The orthogonal constraint has the smallest individual effect (+0.5%), attributable to the natural positive correlation between propagation and intrinsic value effects ($r=0.673$). Removing the entire disentanglement module causes the most severe degradation (+34.5%), validating the dual value of the framework for both accuracy and interpretability.

DOI: 10.54254/2755-2721/2026.32917

Table 2. Ablation study results (MALE ↓, averaged over three datasets and five runs)

Model Variant	Mean MALE	Drop
Full Model (DP-DHGNN)	0.377	–
– w/o Co-citation/Coupling Weight	0.401	+6.4%
– w/o CompGAT	0.406	+7.7%
– w/o Lifecycle Awareness (Mean Pooling)	0.424	+12.5%
– w/o Propagation Decoupling	0.403	+6.9%
– w/o Social Bias Decoupling	0.405	+7.4%
– w/o Intrinsic Value Decoupling	0.409	+8.5%
– w/o Orthogonal Constraint	0.379	+0.5%
– w/o Entire Decoupling Module (Single MLP)	0.507	+34.5%

3.4. Further analysis

Parameter sensitivity. DP-DHGNN is robust near the optimal configuration ($L=4$, $\lambda=0.375$, $\beta=0.5$), with performance fluctuation <2% for adjacent parameter values across all three datasets.

Disentanglement analysis. Over 1997–2017, the intrinsic value proportion declined from ~45% to 30% while social bias increased from 18% to 34% ($p<0.001$), reflecting the growing Matthew effect in academic publishing. Instant papers show significantly higher intrinsic value contributions (43% ±3.5% vs. 30%±2.9% for seen papers, $p<0.001$), confirming that cold-start papers rely more on content signals before citation relationships form.

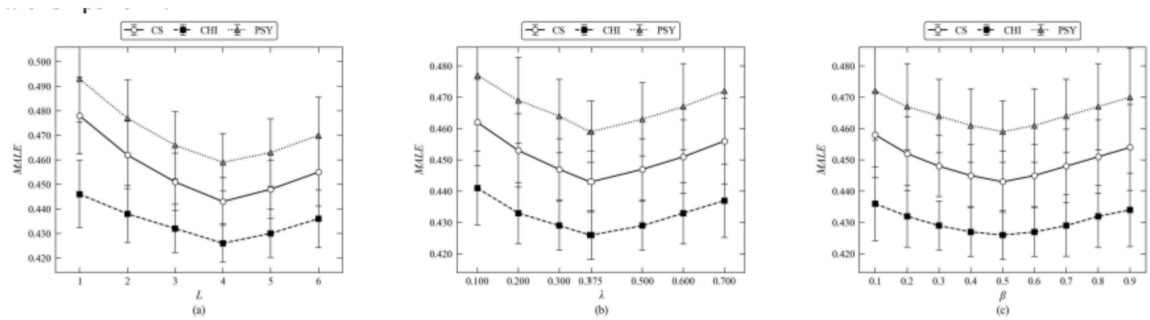

Figure 5. Sensitivity analysis of key hyperparameters. (a) Sensitivity Analysis of Network Depth L. (b)Effect Proportion Distribution across Life Cycle Stages(c) Effect Proportion Distribution by Citation Group

4. Conclusion

In this paper, a two-path dynamic heterogeneous graph neural network model DP-DHGNN is proposed to solve the problem of influence prediction in the early stage of academic papers. The model modeled citation relations and non-citation relations separately through dual-path coding mechanism, introduced life-cycle-aware snapshot readout mechanism to adapt to the growth pattern in different stages, and decoupled the citation increment into three explainable components: propagation effect, conformity effect and intrinsic contribution. Experiments show that DP-DHGNN is significantly better than the existing methods in three different discipline datasets, especially in the cold start scenario of new papers.

Acknowledgement

Foundation item: National Key Research and Development Program of China (2021YFF0600605)

References

[1] Amin Mazloumian et al. "Predicting Scholars' Scientific Impact." PLoS ONE, 7 (2012). https: //doi.org/10.1371/journal.pone.0049246.
[2] Yuxiao Dong et al. "Can Scientific Impact Be Predicted?." IEEE Transactions on Big Data, 2 (2016): 18-30. https: //doi.org/10.1109/tbdata.2016.2521657.
[3] Hirako, Jun, et al. "Realistic Citation Count Prediction Task for Newly Published Papers." Findings of the Association for Computational Linguistics: EACL 2023, edited by Andreas Vlachos and Isabelle Augenstein, Association for Computational Linguistics, 2023, pp. 1131–1141, Dubrovnik, Croatia. ACL Anthology, https: //aclanthology.org/2023.findings-eacl.84/.
[4] Wang, Dashun, Chaoming Song, and Albert-László Barabási. "Quantifying Long-Term Scientific Impact." Science, vol. 342, no. 6154, 2013, pp. 127–132. https: //doi.org/10.1126/science.1237825 WSB model
[5] Shen, Hua-Wei, Dashun Wang, Chaoming Song, and Albert-László Barabási. "Modeling and Predicting Popularity Dynamics via Reinforced Poisson Processes." arXiv preprint, 2014, https: //arxiv.org/abs/1401.0778.Poisson process
[6] Jiasheng Duan et al. "Long short-term enhanced memory for sequential recommendation." World Wide Web, 26 (2022): 561-583. https: //doi.org/10.1007/s11280-022-01056-9. LSTM
[7] Man Yao et al. "Attention Spiking Neural Networks." IEEE Transactions on Pattern Analysis and Machine Intelligence, 45 (2022): 9393-9410. https: //doi.org/10.1109/tpami.2023.3241201.Transformer
[8] Nallapati, Ramesh, Daniel McFarland, and Christopher Manning. "TopicFlow Model: Unsupervised Learning of Topic-Specific Influences of Hyperlinked Documents." Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, edited by Geoffrey Gordon et al., vol. 15, Proceedings of Machine Learning Research, PMLR, 2011, pp. 543–551. https: //proceedings.mlr.press/v15/nallapati11a.html.
[9] Walker, Dylan, Huafeng Xie, Koon-Kiu Yan, and Sergei Maslov. "Ranking Scientific Publications Using a Model of Network Traffic." Journal of Statistical Mechanics: Theory and Experiment, 2007, p. P06010. https: //doi.org/10.1088/1742-5468/2007/06/P06010.citeran.
[10] Sayyadi, Hassan, and Lise Getoor. "FutureRank: Ranking Scientific Articles by Predicting Their Future PageRank." Proceedings of the 2009 SIAM International Conference on Data Mining, SIAM, 2009, pp. 533–544. https: //doi.org/10.1137/1.9781611972795.46.
[11] Hu, Ziniu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. "Heterogeneous Graph Transformer." Proceedings of The Web Conference 2020, Association for Computing Machinery, 2020, pp. 2704–2710. https: //doi.org/10.1145/3366423.3380027.HGT
[12] Jiang, Song, Bernard Koch, and Yizhou Sun. "HINTS: Citation Time Series Prediction for New Publications via Dynamic Heterogeneous Information Network Embedding." Proceedings of the Web Conference 2021, Association for Computing Machinery, 2021, pp. 3158–3167. https: //doi.org/10.1145/3442381.3450107.
[13] He, Guoxiu, et al. "H2CGL: Modeling Dynamics of Citation Network for Impact Prediction." arXiv preprint, 2023, https: //arxiv.org/abs/2305.01572.
[14] You, Jiaxuan, Tianyu Du, and Jure Leskovec. "ROLAND: Graph Learning Framework for Dynamic Graphs." arXiv preprint, 2022, https: //arxiv.org/abs/2208.07239.
[15] Ma, Jianxin, et al. "Disentangled Graph Convolutional Networks." Proceedings of the 36th International Conference on Machine Learning, edited by Kamalika Chaudhuri and Ruslan Salakhutdinov, vol. 97, Proceedings of Machine Learning Research, PMLR, 2019, pp. 4212–4221. https: //proceedings.mlr.press/v97/ma19a.html.DisenGCN
[16] Zhang, Zeyang, et al. "Spectral Invariant Learning for Dynamic Graphs under Distribution Shifts." arXiv preprint, 2024, https: //arxiv.org/abs/2403.05026.SILD
[17] Zhikai Xue et al. "Predicting Scientific Impact Through Diffusion, Conformity, and Contribution Disentanglement." Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (2023). https: //doi.org/10.1145/3627673.3679546.
[18] Nils Reimers et al. "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks." ArXiv, abs/1908.10084 (2019). https: //doi.org/10.18653/v1/d19-1410.

DOI: 10.54254/2755-2721/2026.32917

[19] Hao Geng et al. "Modeling Dynamic Heterogeneous Graph and Node Importance for Future Citation Prediction." Proceedings of the 31st ACM International Conference on Information & Knowledge Management (2022). https: //doi.org/10.1145/3511808.3557398.
[20] Ming-Liang Lu et al. "From Newborn to Impact: Bias-Aware Citation Prediction." ArXiv, abs/2510.19246 (2025). https: //doi.org/10.48550/arxiv.2510.19246.
[21] Saberi-Movahed, Farshad, et al. "Deep Metric Learning with Soft Orthogonal Proxies." arXiv preprint, 2023, https: //arxiv.org/abs/2306.13055.Soft Orthogonality
[22] Dapogny, Arnaud, et al. "Adversarial Deep Multi-Task Learning Using Semantically Orthogonal Spaces and Application to Facial Attributes Prediction." 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG), IEEE, 2023, pp. 1–8. https: //doi.org/10.1109/FG57933.2023.10042750.Soft Orthogonality