RESEARCH ARTICLE

Transforming Citation Networks into Insights: Mapping Scholarly Influence with Advanced Graph Models

Tahir Sher[1], **Abdul Rehman**[2,*] and **Imran Ihsan**[3]

[1] Department of Artificial Intelligence, Korea University, Seoul 02842, Republic of Korea
[2] Convergence Institute of Human Data Technology, Jeonju University, Jeonju 55069, Republic of Korea
[3] Department of Creative Technologies, Air University, Islamabad 44000, Pakistan

Abstract

The growing role of citation relations in identifying research impact has spurred much investigation on assessing the most cited papers and their roles within datasets. Due to the richness of the CORA dataset, this study selects highly cited papers and measures the results of node classification, as well as the H-index of research articles. Besides, it explores the correlations and robustness with regard to the nodes by computing their chances and studying their connections. To these ends, linear transformation was utilized for mapping low-level node features to high-level, and the Graph Attention Networks (GAT) for node classification. The study was able to find highly cited papers and compute their H-index, which gives insight into the citation patterns in the CORA dataset. For instance, Paper ID 12182 reported an H-index of 20, while a high citation paper index of 35 received 166 citations. On the test dataset, the study achieved a node classification accuracy of 78.60% and an F1 score of 78.14%. Furthermore, 7 nodes of the machine learning domain have also been distinguished and categorized according to their features and their relations within the graph. In the citation network identified, the present research detailed the citation interconnection that characterized the works within the dataset. Our research mainly focuses on new tasks such as the extraction of highly cited papers and the calculation of the H index that improve the comprehension of the scholarly influence and the citation relation for future development strategies in citation network analysis.

Keywords: deep learning, text mining, graph attention networks, citation, h-index.

1 Introduction

Graphs are widely used to represent complex relationships in various fields like citation network structure [1], journal-level academic impact measurement [2], citation impact prediction [3, 4], citation network structure and analysis [5, 6]. They capture both data and structure, from simple node-edge structures to more intricate types like trees, cyclic graphs, and directed acyclic graphs (DAGs). Applications generally fall into two categories: graph-focused, where models analyze the graph as a

Submitted: 14 May 2025
Accepted: 09 July 2025
Published: 04 October 2025

Vol. 1, **No.** 4, 2025.
10.62762/TACS.2025.939169

***Corresponding author:**
Abdul Rehman
a.rehman.jj@jj.ac.kr

Citation
Sher, T., Rehman, A., & Ihsan, I. (2025). Transforming Citation Networks into Insights: Mapping Scholarly Influence with Advanced Graph Models. *ICCK Transactions on Advanced Computing and Systems*, 1(4), 238–257.

© 2025 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

whole [7], and node-focused, which target individual nodes. Advances like random walk models [47] and RNNs have paved the way for Graph Neural Networks (GNNs), which handle both types without heavy pre-processing.

Citation networks are formally represented as graphs whose nodes correspond to scholarly papers and whose edges encode citation relationships [9]. Different edge construction strategies—direct citation, co-citation, and bibliographic coupling—each capture distinct aspects of scholarly relatedness, with empirical comparisons demonstrating that each approach represents different facets of the research front [8]. Citation-based methods have gained importance as analytical tools across both the social and natural sciences [12, 13], enabling systematic mapping of knowledge flows and the quantification of scholarly influence at scale, while graph-based methods have also been applied to model structural properties of heterogeneous scholarly networks, enabling publication influence evaluation and prediction [10] as well as multi-view paper classification [11]. Citation networks exhibit rich structural properties that enable publication-level clustering and community detection based on direct citation relations [14], and methods for enriching citation networks through reference chaining have further demonstrated the depth of relational structure embedded in scholarly corpora [83]. Citation graphs are inherently directed structures, where each edge represents a reference from one scholarly work to another [15]. Such graphs come in various forms—directed, undirected, and cyclic or acyclic variants—each capturing different relational structures, as shown in Figure 1(c) and Figure 1(d) [17, 18]. Graph embedding methods provide a unified framework for learning representations across these diverse graph types [17].

The symmetric adjacency matrix structure of directed and undirected graphs directly supports the modeling of co-citation relationships and bibliographic coupling in academic networks [19, 20]. Graph attention mechanisms have been shown to effectively capture such multi-level citation structures for scholarly paper classification [18]. Graphs are useful for synthesizing large bodies of scholarly literature, enabling systematic evaluation and prediction of research impact [21]. GNNs have been widely adopted to capture structural patterns in citation graphs, powering downstream applications such as scientific paper recommendation [22] and enabling systematic quantification of scholarly influence across large-scale citation networks.

Artificial Neural Networks (ANNs), a deep learning method inspired by the human brain, consist of input, hidden, and output layers. Each node passes values with weights and thresholds, as illustrated in Figure 2. Graph Convolutional Networks (GCNs), an extension of CNNs, use graph structures to aggregate information from neighboring nodes [31]. Known for their expressive power, GCNs have shown strong performance across various tasks [29]. Recent developments include graph auto-encoders [23] and comprehensive graph representation learning frameworks [24], citation impact modeling and scholarly influence quantification [21, 25, 26, 32], attention models [27, 28], and graph recurrent networks [34]. GNNs operate through information diffusion, where each node (unit) updates its state based on graph connectivity until reaching stability. These models span a broad spectrum of graph learning paradigms—including signal processing, matrix factorization, and deep learning approaches—applicable to citation networks and other information systems [35, 36, 50], and have been widely used for node classification [92, 93, 99–101].

In our work, we adopt Graph Attention Networks (GAT), which prioritize neighboring nodes based on learned attention scores [27, 37]. Unlike fixed-weight approaches like GCN [33] or GraphSAGE [39], GAT dynamically adjusts weights through attention mechanisms, as formalized in Eq. 1.

$$\beta_{nu}^{(k)} = Softmax\left(g\left(\overrightarrow{a}^{T}\left[W^{(k)}H_n^{(k-1)} \,||\, W^{k}H_u^{(k-1)}\right]\right)\right) \tag{1}$$

where $g(\cdot)$ is a LeakyReLU activation function, and 'a' represents the vector for learnable parameters. The softmax function ensures that the attention weights sum up to one for all neighbors of the node 'n.' Additionally, GAT employs a multi-head attention mechanism to improve performance on node classification tasks compared to GraphSage. While GAT assumes that all attention heads contribute equally, Gated Attention Networks (GAAN) [40] introduce a new self-attention mechanism that assigns different attention scores to each attention head. Furthermore, GeniePath [41] introduces a new Gating mechanism, Long Short Term Memory (LSTM), to observe and control the flow of information across graph convolutional layers. Figure 3(a) GAT [27]

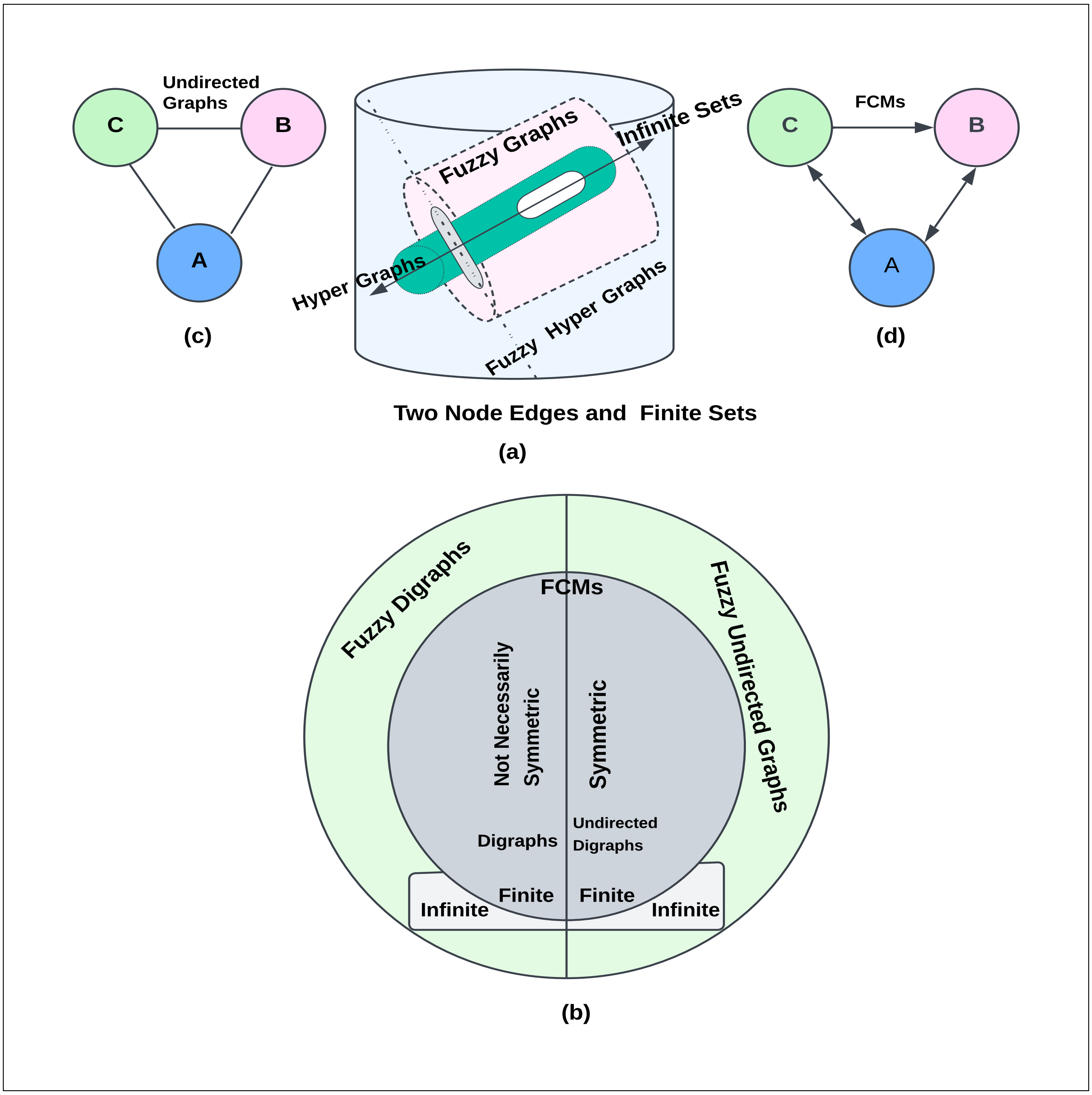

Figure 1. Visualization of the relationships among different graphs.

considers the attention coefficient β_{ij} and gives more weight to important nodes. Figure 3(b) GCN [33] gives a non-parametric weight $\beta_{\mathrm{ij}} = \frac{1}{\sqrt{deg(n_{\mathrm{i}})deg(n_{\mathrm{j}})}}$ to the neighbor $\mathrm{n_j}$ of $\mathrm{n_i}$ during the aggregation process. Where $H_n^0 = y_n$, the attention weight considers the strength between the node “n” and its neighbor "u." Multi-task learning using GATs has been widely explored across text, image, and video data. Since the vast majority of real-world scholarly data is raw and unstructured [53], organizing it meaningfully through content-based citation analysis is essential for effective research. Only a small portion carries directly actionable knowledge. To extract this, various data mining techniques are applied, including pattern discovery, clustering, classification, text mining, and social network analysis [42, 94, 95]. Bibliometric methods such as co-citation analysis and bibliographic coupling provide complementary quantitative frameworks for mapping the structure of scholarly citation networks [51], while visualization tools such as CiteSpace [52] and comprehensive bibliometric analysis platforms such as Bibliometrix [84] enable systematic detection of emerging research trends through citation pattern analysis—approaches that motivate the graph-based citation visualization pursued in the present study. Citation framing approaches further demonstrate that tracking the

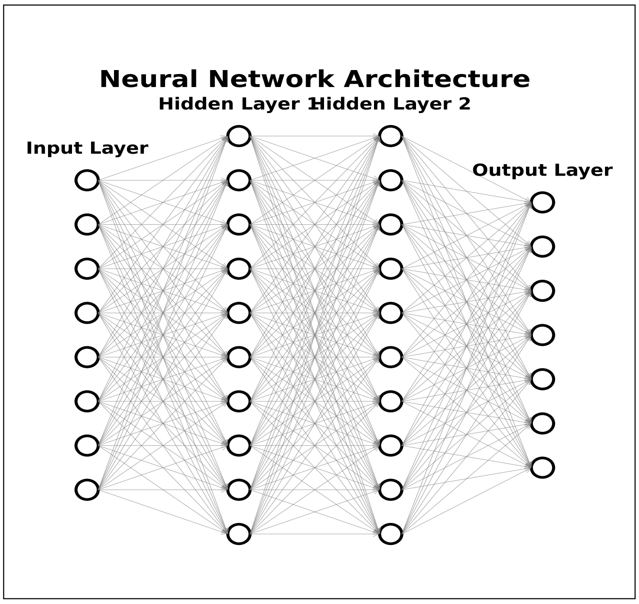

Figure 2. Neural Networks working and structure.

rhetorical function of citations over time reveals the evolution of entire scientific fields [96].

Given its strong performance, we focus on applying GATs for text mining, particularly due to the challenges of unstructured text data [43, 45, 53, 79]. Key applications include information extraction, knowledge discovery, and text data mining from citation corpora [44]. However, high-dimensional and sparse text data—where documents represent only a small subset of a vast vocabulary—requires careful preprocessing (as shown in Table 1 and Figure 3). Each document is represented as an $n \times r$ matrix, with entries showing word frequencies, following distributed word representation principles [48]. This sparsity demands specialized dimensionality reduction techniques to ensure efficient analysis.

Similarly, [49, 102] demonstrated that systematic knowledge discovery from scholarly citation corpora requires structured preprocessing pipelines, including natural language processing, term classification, and algorithmic pattern extraction. Network-level analysis of citation and co-authorship structures further reveals the collaborative dynamics that underpin scholarly influence [103]. The following are a few techniques that may be employed for the task of text mining:

- Information Extraction
- Information Retrieval
- Summarization
- Clustering

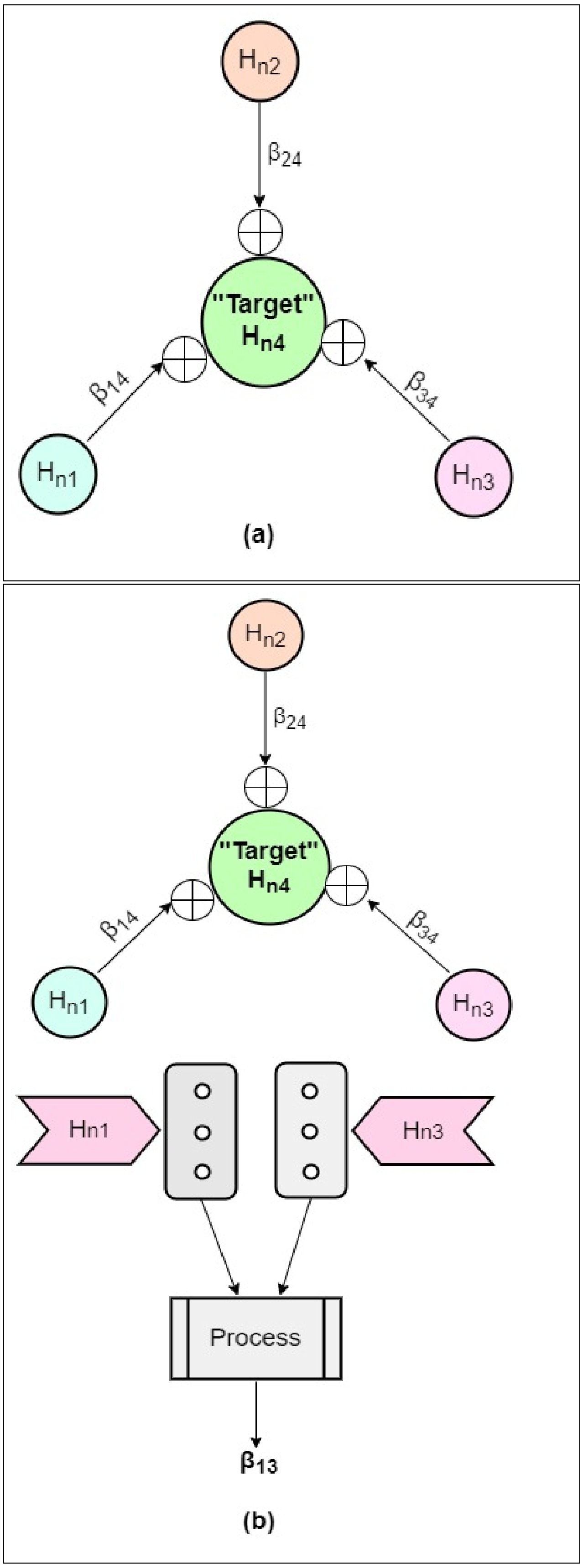

Figure 3. Comparison between GCN [33] and GAT [27].

Table 1. A complete description of the text mining process.

Process	Description
Text Document	Relevant information is extracted from the input document for further analysis.
Text Pre-Processing	Preprocessing strategies are applied to clean and normalize the raw text data, including tokenization, stop-word removal, and stemming.
Text Transformation	The cleaned text is transformed into a structured numerical representation suitable for downstream processing.
Feature Selection	Relevant and discriminative features are selected based on their importance to the classification task.
Pattern Discovery	Standard patterns and regularities are extracted from the transformed data using appropriate analytical and mining tools.
Evaluation	Results are systematically evaluated against defined performance criteria, including accuracy, F1 score, precision, and recall.

- Categorization
- Concept Linkage

Text mining finds applications across diverse sectors including digital libraries, academic search systems, patent analysis, and citation network analysis [54, 55], with citation recommendation systems representing one of its most impactful scholarly applications. Within the scholarly domain, it is particularly powerful for citation network analysis, knowledge flow mapping across diverse scholarly network types [46], and research impact assessment—the focus of the present study.

Text mining plays a vital role in various industries by uncovering unique patterns that support production and knowledge management. In scholarly contexts, text mining enables the systematic analysis of citation networks, the identification of influential works, the mapping of knowledge flows across research communities, and the quantification of scholarly impact through metrics such as the H-index [88–91]. Accurate quantification of scholarly influence requires awareness of bibliometric inflation effects, as raw citation counts alone may overestimate impact without field-normalized comparison [16]. It also improves information retrieval and powers search engine responses. Citation recommendation systems further leverage these networks to guide researchers toward relevant literature, with surveys of such approaches [86] confirming that citation link structure is among the most informative signals for scholarly document retrieval. Additionally, text mining enables analysis of citation networks, which is central to our study.

- We proposed the Graph Attention Network model for node classification on the complete CORA dataset.
- We computed the intra-class citation probability between scientific paper groups within the CORA dataset to analyze cross-domain citation patterns.
- Highly cited papers within the CORA dataset were identified and ranked by citation count.
- H-index of the papers using the CORA dataset is also considered in this study [30].
- We also compare our results with classical machine learning baselines including Random Forest and Support Vector Machine.

This paper consists of the following sections: Section 2 includes a review of the literature, Section 3 consists of the Proposed Methodology, Section 4 deals with Experimental Design, Analysis & Results, and Section 5 is the conclusion.

2 Related Work

Graph Neural Networks (GNNs) have substantially advanced research in pattern recognition and classification, enabling more reliable and computationally efficient processing of structured data [94]. GNNs have been extensively applied to single- and multi-label node classification tasks [5]. Graph Attention Networks (GAT) extended this paradigm by introducing learnable, variable attention weights for neighboring nodes [27], demonstrating strong performance on node classification benchmarks. A growing body of work has applied GAT to textual datasets across diverse tasks, as reviewed below.

Previous work [56] investigated cited text span identification within scholarly documents, demonstrating that identifying the precise textual context of a citation link is essential for understanding the semantic role of references within citation networks—a property directly exploited by attention-based graph models such as GAT. [57] presented a comprehensive survey of deep learning approaches for citation recommendation, demonstrating that graph-based attention mechanisms consistently outperform conventional retrieval models in capturing the semantic relevance of citation links within scholarly networks. Complementarily, [58] proposed a structure fusion method based on graph convolutional networks for node classification in citation networks, validating that integrating multiple graph structural signals substantially improves classification accuracy on standard citation benchmarks such as CORA.

Reference [59] proposed a structural scaffold model for citation intent classification in scientific publications, demonstrating that the semantic function of a citation—whether methodological, background, or result-comparison—can be systematically classified and used to enrich citation graph representations for downstream tasks such as node classification. [86] provided a comprehensive survey of citation recommendation approaches and benchmark datasets, systematically categorizing graph-based, content-based, and hybrid methods and confirming that citation network structure consistently provides strong supervisory signal for scholarly document retrieval—a finding that directly motivates the use of GAT-based graph representations for citation network analysis in the present study.

Graph attention mechanisms have been further applied to heterogeneous scholarly and structured relational graphs, demonstrating their versatility across diverse node and edge types that mirror the heterogeneous structure of citation networks. [60] proposed an Association Content Graph Attention Network (ACGAT) that jointly models semantic relevance and academic relevance among scholarly resources, demonstrating that GAT-based attention weighting generalizes effectively across heterogeneous node types and edge semantics—a property equally relevant to scholarly graphs where papers, authors, and citation links constitute distinct relation types. [61] further demonstrated that deep neural network approaches applied to citation networks can quantify scholarly research dissemination patterns, validating the utility of graph-based representations for academic influence measurement. [62] introduced the Heterogeneous Graph Attention Network (HGAN) for semantic-level and node-level attention, which acquires knowledge of critical nodes and their meta-paths. GAT leverages the attention mechanism for unique types of nodes or links, known as a homogeneous graph, and uses the concept of neural graph networks. The value of each meta path may be learned by Semantic-level attention and assigned suitable weights. Different attention weights are considered for each node at node-level attention depending on the importance of meta-path-based neighbors. Graph Attention Networks are employed to learn the importance of nodes and their neighbors. They use neighbors to accomplish the node classification and embedding of heterogeneous graphs, focusing on maintaining meta-path-based structural knowledge. HGAN outperformed the other models when compared. [63] proposed a citation attention fusion model that dynamically models citation relationships among publications for classification prediction. The model integrates attention mechanisms over evolving citation graphs to capture the temporal dynamics of scholarly influence, achieving superior performance on academic paper classification benchmarks.

Reference [64] proposed a citation recommendation framework using graph attention networks that effectively models the structural relationships between scholarly papers, achieving strong performance on both node classification and link prediction tasks on academic citation graphs. [65] proposed incorporating citation graph structure into document representation learning, demonstrating that attention-weighted aggregation over citation links substantially improves document classification accuracy on academic citation benchmarks—a finding directly corroborated by the node classification results on CORA reported in the present study.

Previous research [66] introduced a citation context classification model using supervised machine learning to identify the intent and nature of in-text citations within scholarly documents, demonstrating that the semantic properties of citation links—such as whether a reference is cited for background or methodology—can be systematically classified. This work underscores that citation links encode rich relational semantics beyond mere connectivity, a property that attention-based graph models like GAT are well positioned to exploit. [67] proposed SPECTER, a document-level representation learning approach that leverages citation graph structure to pre-train scientific paper embeddings, demonstrating that citation links encode meaningful semantic relationships between scholarly works and can substantially improve downstream tasks such as paper classification and recommendation.

Reference [69] proposed the Heterogeneous Graph Transformer (HGT), which employs type-dependent attention mechanisms to model heterogeneous scholarly graphs containing multiple node and edge types such as papers, authors, and citation links. HGT demonstrates that attention-weighted aggregation across heterogeneous relation types substantially improves node representation quality on academic citation benchmarks—an insight directly applicable to the CORA citation network studied here. [70] deployed a model SR-HGAN for node clustering, classification, and relation prediction. The SR-HGAN model consists of two layers, in which the core task of the primary layer is to assign attention weights to nodes and consider their symmetric relation. Moreover, considering the symmetric relation for various assignments, the secondary layer assigns weights to latent semantics. The two-layer mechanism is used to learn attention values and hierarchically obtains the ideal composite of neighbors and several proportionate relations. SR-HGAN demonstrated competitive performance on node classification and relation prediction benchmarks relative to prior heterogeneous graph attention methods. [71] proposed AGSTA-NET, an adaptive graph spatiotemporal attention network for citation count prediction, which jointly models structural features of citation networks and temporal citation patterns through hierarchical graph convolution layers. The spatiotemporal attention mechanism demonstrated in AGSTA-NET is directly applicable to citation graphs, where high-order and time-evolving citation relationships encode transitive scholarly influence.

Reference [72] proposed an evolving knowledge graph representation learning framework with multiple attention strategies for citation recommendation. The model captures dynamic changes in citation relationships over time through multi-head attention over heterogeneous scholarly graph nodes, demonstrating that GAT-based attention generalises effectively to typed-node aggregation tasks in academic citation networks. The heterogeneous attention mechanism learned in this setting mirrors the challenge in citation graphs of aggregating nodes belonging to different scholarly classes, reinforcing the broader applicability of GAT architectures to typed-node classification tasks. [75] introduced a large-scale scholarly dataset with full-text publications, annotated in-text citations, and citation network links, providing a comprehensive benchmark that demonstrates the richness of citation context as a signal for scholarly document analysis tasks including node classification on citation graphs.

Reference [74] proposed neighborhood contrastive learning for scientific document representations using citation embeddings, demonstrating that citation-aware self-supervised objectives substantially improve node representation quality on academic citation benchmarks including CORA, complementing the supervised GAT-based node classification approach pursued in the present study. [76] proposed a context-aware citation recommendation model combining BERT and graph convolutional networks, demonstrating that jointly modelling citation context and graph structure substantially improves recommendation performance on scholarly citation benchmarks.

Reference [73] addressed the over-smoothing problem in deep GNNs through an adaptive initial residual and identity mapping mechanism, achieving significantly improved node classification accuracy on citation datasets including CORA and CiteSeer. The work underscores that architectural depth and residual connectivity are critical factors in extracting discriminative node representations from citation graph topology. [68] proposed GraphSAINT, a graph sampling-based inductive learning method that constructs mini-batches by sampling subgraphs rather than nodes, enabling scalable and accurate node classification on large citation networks. By decoupling the graph sampling strategy from the GNN architecture, GraphSAINT achieves strong performance on benchmarks including CORA while reducing computational overhead in large-scale citation graph settings.

Note: Table 2 summarises the evaluation metrics reported by the referenced authors on their respective experimental datasets. Metrics not reported in the original studies are marked with "x". Studies that do

Table 2. Description of text datasets using graph attention networks and citation-aware methods.

Authors	Datasets	Accuracy	F1 Score	Precision	Recall	AUC	NDCG
[56]	ACM DL (cited spans)	x	0.51	0.54	0.48	x	x
[60]	Epinions	x	x	0.7781	x	x	x
	WeChat Top-Story	x	x	0.0823	x	0.8165	x
	Douban	x	x	x	0.1861	x	0.1959
[61]	Delicious	x	x	x	0.4066	x	0.2944
	Yelp	x	x	x	0.0842	x	0.1427
	ACM	x	x	x	x	x	x
[62]	IMDB	x	x	x	x	x	x
	DBLP	x	x	x	x	x	x
	Yelp	x	x	x	0.0712	x	0.0867
[63]	Amazon book	x	x	x	0.1489	x	0.1006
	Last-FM	x	x	x	0.087	x	0.1325
[64]	CORA	0.823	x	x	x	x	x
	CiteSeer	0.714	x	x	x	x	x
[65]	DBLP-Citation	0.891	0.887	x	x	x	x
	ACM-Citation	0.876	0.871	x	x	x	x
[66]	Twitter	x	0.6834	x	x	0.9829	x
[69]	ACM	0.9121	x	x	x	x	x
	DBLP	0.9246	x	x	x	x	x
	IMDB	0.5697	x	x	x	x	x
	DBLP	x	x	x	x	x	x
[70]	Amazon book	x	x	x	x	x	0.2198
	Movielen	x	x	x	x	x	0.3913
[71]	Tiktok	x	x	0.1251	0.5965	x	0.3838
	MovieLens	x	x	0.1271	0.5412	x	0.3251
[72]	Alibaba Group	x	x	x	0.2744	0.9687	x
[74]	S2ORC (Sci.)	x	0.823	x	x	x	x
	CORD-19	x	0.801	x	x	x	x

not report quantitative metrics in a format comparable to those listed are described qualitatively in the preceding text and are therefore not included in this table. Survey articles (e.g., [57]) are discussed in the text but excluded from this table as they do not report primary experimental results. [75] (Saier & Färber, 2020) is a dataset paper and does not report classification metrics; it is described qualitatively in the preceding text.

The above body of research demonstrates that graph attention mechanisms can be adapted to a wide range of structured relational tasks, from text classification and recommendation to node embedding and relation extraction, all of which share the core challenge of learning meaningful representations from interconnected nodes—precisely the challenge addressed in citation network analysis.

Researchers found that different models embedded with the Graph Attention Network (GAT) for classification show strong performance across multiple scholarly tasks, including citation intent classification [59], citation-aware document representation [65, 74], citation recommendation [57, 64], and citation count prediction [71], demonstrating the broad applicability of GAT architectures to citation network analysis tasks of the kind pursued in this study. Drawing a conclusion or inferring the data context is challenging, as textual data is sparse. In this study, GATs were applied to text mining on the CORA citation dataset to identify essential nodes in relation to their neighbors, enabling each node to draw context from its assigned class. A thorough analysis was conducted to characterize the seven node classes, each comprising scientific publications that cite one another. The study further emphasized the extraction of highly cited papers and the calculation of the H-index from the CORA dataset—novel analytical tasks not previously addressed in the literature on this benchmark. Recent extensions of graph attention mechanisms, such as graph transformer networks [38], have further demonstrated the power of attention-based aggregation for structured relational data, reinforcing the suitability of attention architectures for citation graph analysis tasks like those pursued in this study. Propagation-based methods such as APPNP [85] further demonstrate that combining personalized PageRank with neural propagation yields strong node classification results on citation benchmarks, providing an alternative perspective on how graph topology can be exploited for scholarly paper classification. Semi-supervised graph embedding methods [97] and self-supervised GCN pre-training on citation graphs with sparse labels [98] further confirm that graph-structural signals alone provide strong supervisory information for node classification—a finding directly corroborated by the node classification results on CORA reported in the present study.

3 Methodology

The proposed model employs Graph Attention Networks (GATs) for text mining, with a focus on node classification within a dataset. The pre-processing stage involves calculating node vectors for the dataset. GATs are then utilized to classify the data and generate scores based on the significance of different classes. The model uses a single or multi-head attention mechanism to extract information with the same context. After the computation process, GATs produce unique class labels that correspond to various label occurrences within the dataset. The dataset is split into training and testing subsets to evaluate the model's performance. The flowchart illustrating the process of text mining using GATs is depicted in Figure 4.

3.1 Dataset

The proposed Graph Attention Networks method for text mining was experimented on the benchmark "Cora" dataset, which is publicly available at the Cora Dataset. It contains 2708 i.e *Case_Based, Genetic_Algorithm, Neural_Networks, Probabilistic_Methods, Reinforcement_Learning, Rule_Learning and Theory* are machine learning papers. The dataset consists of two files with the extension of .content file, which contains the description of the paper in the $<$ *paper_id* $>$, $<$ *word_attributes* $>$+$<$ *class_label* $>$ *format*. Moreover, a unique string ID is shown in the first line of the paper, and it is shown as a binary number (indicated by 1) if each word is available in the vocabulary or (indicated by 0) if the number is absent in the vocabulary and paper class is represented by the last line. Furthermore, citation graphs of the corpus are contained in the other file, i.e., the "cite" file. Each line describes a link in the <ID of cited paper><ID of citing paper>format is represented in each line, and it works as a link among them. A strategy was developed in such a way that another paper should cite every paper in the corpus. The corpus contains 2708 papers and 1433 unique words, which were attained by pre-processing steps. It is pertinent to mention that those words with a document frequency of less than ten were removed.

3.2 Cleaned Dataset in the form of Node Vectors

The cleaned dataset received after the pre-processing step is now pure from anomalies, and the dataset contains a total of 2708 nodes $N_i \epsilon \Re^F$, and each node contains 'n' features. Eq. 2 shows that the nodes contain different machine-learning papers with unique characteristics/ features.

$$\begin{aligned} N_1 &= \{f_{11}, f_{12}, f_{13}, \ldots, f_{1n}\} \\ N_2 &= \{f_{21}, f_{22}, f_{23}, \ldots, f_{2n}\} \\ N_3 &= \{f_{31}, f_{32}, f_{33}, \ldots, f_{3n}\} \\ &\vdots \\ N_n &= \{f_{n1}, f_{n2}, f_{n3}, \ldots, f_{nn}\} \end{aligned} \tag{2}$$

3.3 Dataset Split into Train/ Test sets

The purpose of splitting the data into two parts is to avoid overfitting. Therefore, the researcher split the dataset into two sets: 50 percent for training and 50 percent for the testing phase.

3.4 Simple Linear Transformation

Linear transformation remodels the low-level input features into higher-level features by providing weights to the input features. Therefore, shared linear transformation, parametrized by a weight matrix, $W \in \Re^{\acute{F} \times F}$, is applied to every node shown in Eq. 3.

$$N_i' = W N_i \tag{3}$$

where N_i' is a new set of node vectors potentially of different cardinality F' and $N_i' \in \Re^{F'}$ which is shown in Eq. 4.

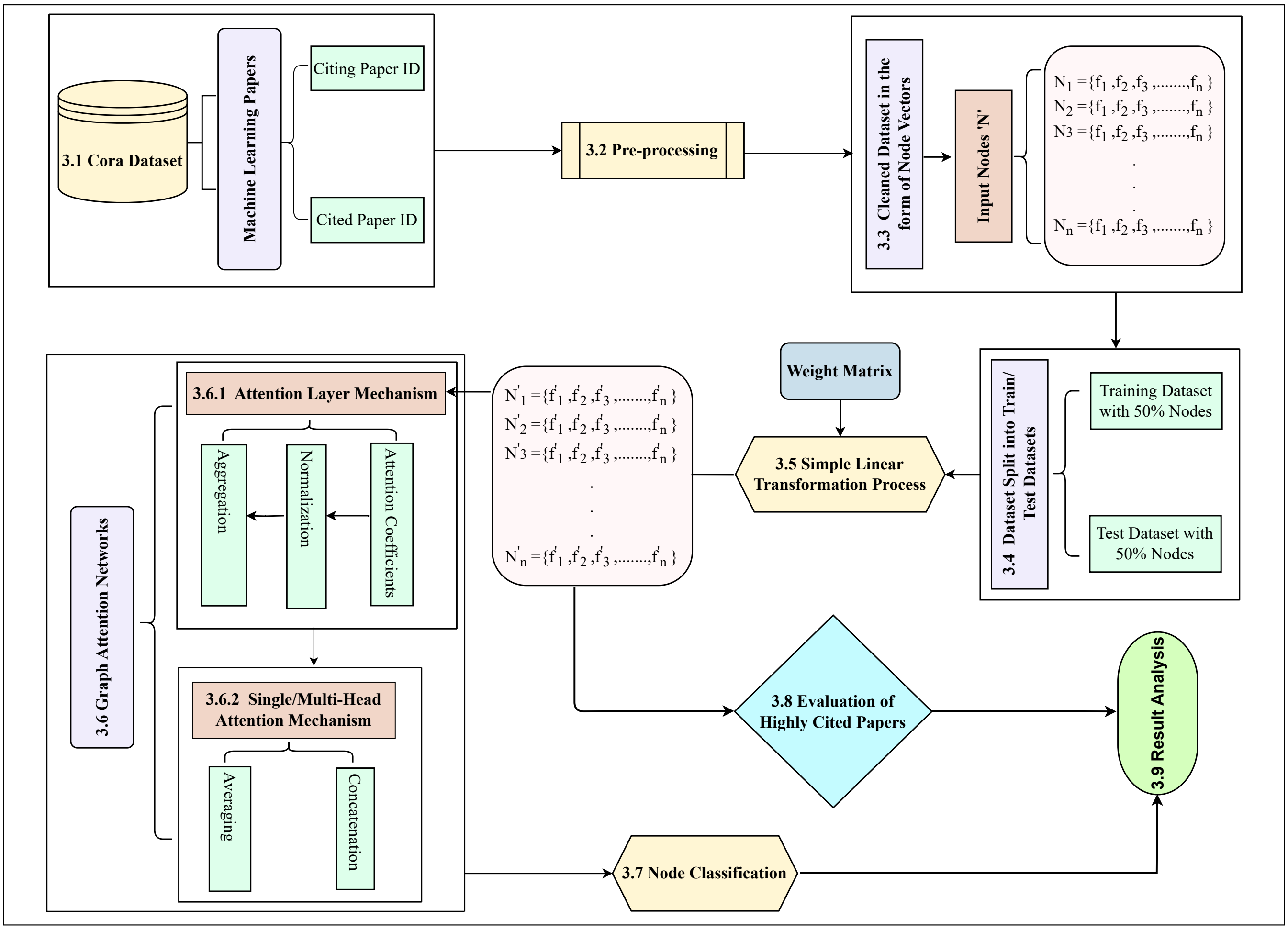

Figure 4. Flow chart of text mining using GAT.

$$
\begin{aligned}
N_1' &= \{f_{11}', f_{12}', f_{13}', \ldots, f_{1n}'\} \\
N_2' &= \{f_{21}', f_{22}', f_{23}', \ldots, f_{2n}'\} \\
N_3' &= \{f_{31}', f_{32}', f_{33}', \ldots, f_{3n}'\} \\
&\vdots \\
N_n' &= \{f_{n1}', f_{n2}', f_{n3}', \ldots, f_{nn}'\}
\end{aligned}
\tag{4}
$$

3.5 Graph Attention Layer Mechanism

3.6 H-index Computation from Citation Graph

The H-index of a paper p within the CORA citation network is defined following Hirsch [90] as:

$$
H(p) = \max\{h \in \mathbb{Z}^+ \mid |\{q \in \mathcal{C}(p) : c(q) \geq h\}| \geq h\}
\tag{5}
$$

where $\mathcal{C}(p)$ denotes the set of papers cited by paper p within the CORA corpus, and $c(q)$ is the in-degree (citation count) of paper q in the directed citation graph. The in-degree of each node was computed directly from the adjacency structure of the citation graph. Papers were ranked by in-degree to identify the top-cited works (Table 4), and the H-index was computed for each paper according to Eq. 5. This formulation extends the original author-level H-index metric [90] to the paper level within a bounded citation corpus, enabling quantification of individual scholarly influence within the CORA network.

3.6.1 *Attention Layer Mechanism*

The novel idea of an attention mechanism was floated by [77], and the idea of an attention mechanism came into existence to address the bottleneck problem that arose with the usage of a fixed-length encoding vector, where the decoder provides minimal information from the input vector. Thus, this idea works well to address the limitations of graph convolutional networks by providing a score of attention weights to its neighbors for their importance for full nodes. Therefore, the input to the layer is $N = N_1, N_2, N_3, \ldots, N_{\text{n}}$, where $N \in \Re^F$. where "n" is the number of nodes, each containing several features "F." The layer produces a new set of node features $\acute{N}_i = \acute{N}_1, \acute{N}_2, \acute{N}_3, \cdot, \acute{N}_n$, where N_{i} are set of output nodes. The whole process in this layer is

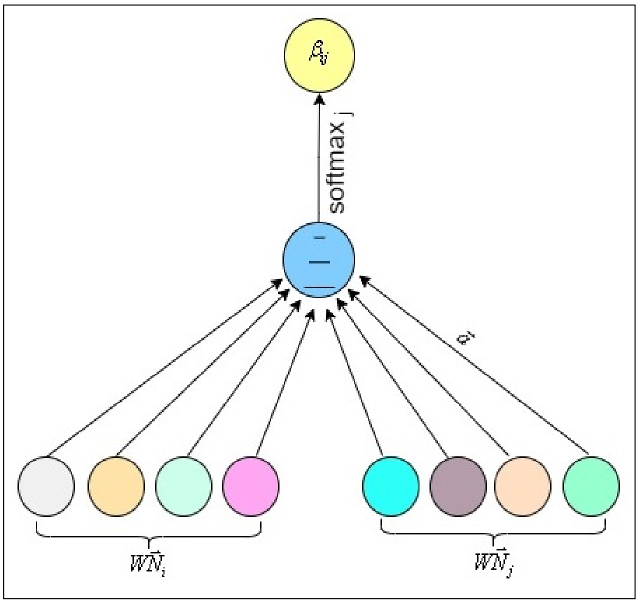

Figure 5. Attention layer mechanism.

shown in Figure 5.

Moreover, the attentional layer is broken into the following three parts:-

- Attention Coefficients
- Normalization
- Aggregation

- **Attention Coefficients:** After the linear transformation process, an un-normalized attention score between any two neighbors will be calculated. In this scenario, a concatenation process will be implemented on N_i embeddings of two nodes, and we use the $||$ symbol for concatenation. Then, we will take a dot product of it with the learnable weight vector a. After doing this, we will use the LeakyReLU activation function. Moreover, a shared attentional mechanism like the self-attention mechanism on the nodes is used, i.e., $a : \mathbb{R}^{F'} \times \mathbb{R}^{F} \rightarrow \mathbb{R}$ to compute attention coefficients μ_{ij}. The attention coefficient is an important and essential step, which calculates the importance of node N_j's features to node N_i. It is an essential step as it considers that every node in the corpus will attend to another node without considering the structural information of the neighboring nodes. Furthermore, Eq. 6 defines the attentional coefficients as:

$$\mu_{ij} = LeakyRelu(\overrightarrow{a}^T(N_i \,||\, N_j)) \tag{6}$$

- **Normalization:** In the next process level, a softmax function is implemented on Eq. 6 as we can make the coefficients comparable among all "j" nodes. Eq. 7 is used for normalization purposes.

$$\beta_{ij} = \frac{exp(\mu_{ij})}{\sum_{k\in N_i} exp(\mu_{ik})} \tag{7}$$

- **Aggregation:** Aggregation step is similar to that used in GCN. Attention scores are used for scaling the embedded neighbors, and these neighbors are aggregated as in Eq. 8.

$$N_1'' = \sigma(\sum_{j \epsilon N_i} \beta_{ij} N_j') \tag{8}$$

- **Single/ Multi-Head Attention Mechanism:** Similar to Convolutional Networks, GAT also uses a multi-head attention mechanism for two purposes: first, to enrich the capacity of the model and, secondly, for the stability of the learning process. Therefore, "K" independent attention mechanisms execute the transformation of Eq. 7, and their output may be aggregated based on the requirement. Thus, concatenation shown in Eq. 9 is used for intermediary layers, and the averaging concept is used for the final layer as shown in Eq. 10.

$$N_i'' = ||_{k=1}^{K} \sigma(\sum_{j \epsilon N_i} \beta_{ij}^k N_j') \tag{9}$$

$$N_1'' = \sigma(\frac{1}{K} \sum_{k=1}^{K} \sum_{j \epsilon N_i} \beta_{i,j}^k N_j') \tag{10}$$

Suppose $d(N_i) = 64$, $K = 4$, then $d(N_{i,k}) = 16$ using Eq. 9. Hence, each "head" is responsible for 16 dimensions.

4 Experimental Design, Analysis, and Results

The researchers collected the Cora citation dataset from https://linqs-data.soe. and another link is https://relational.fit.cvut.cz/dataset/CORA [80, 81]. This study aims to extract highly cited papers from the Cora Dataset, consisting of directed graphs that make a citation link between the documents belonging to different classes. This study also seeks to classify the dataset based on graph attention networks. The path to achieving the goals mentioned above is as follows:

- Node classification using graph attention networks on the CORA dataset
- H-index on Cora dataset and extraction of highly cited papers from the pool of citations, i.e., CORA dataset

Now, revising the study's preface, the previously implemented models do not include the citation as a parameter for the model to be trained on, while our approach considers the citation as a part of the dataset for training.
The dataset is followed by 2708 scientific publications collected in graphs, as shown in Figure 6, and classified into seven nodes or classes.

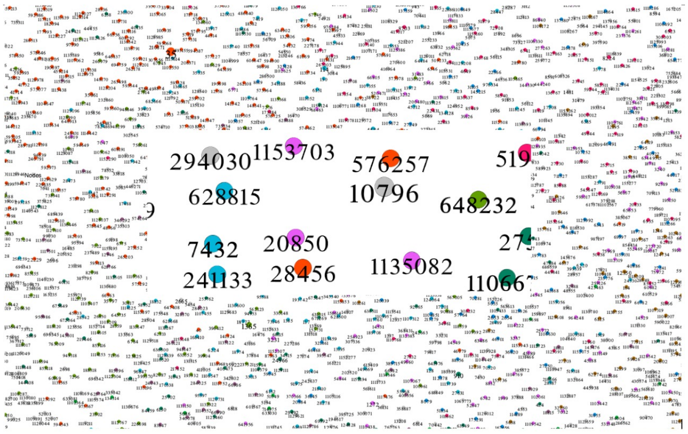

Figure 6. Graph representation of the Cora dataset.

The intra-class citation probability $P(C_i \rightarrow C_j)$ was computed as the proportion of citations from papers in class C_i that point to papers in class C_j:

$$P(C_i \rightarrow C_j) = \frac{|\{(p,q) \in E : y_p = C_i, y_q = C_j\}|}{|\{(p,q) \in E : y_p = C_i\}|} \quad (11)$$

where E is the set of directed edges in the citation graph, and y_p denotes the class label of paper p. Table 3 presents the full 7×7 probability matrix across the seven CORA classes.

The description of the dataset is given as follows:

- (5429, 2) edge shape represents the links between the papers with two key features
- Node features of shape (2708, 1433)
- The articles are classified as binary classes in which "0" represents a paper not existing and "1" illustrates an article exists

The dataset consists of two files: Cora citations and Cora content. Cora sites consist of two features, namely "target" and "source," while core content consists of paper_id along with its features and belonging classes. The data set was processed in pre-processing steps, and null or empty values were removed. To achieve high-level features from low-level features, the following techniques were applied:

- Pre-processing the given dataset
- Applied one-hot encoding on the dataset
- Applied linear transformation of one-hot encoded data
- Separating features and labels into different variables
- Train / Test split with a ratio of 50:50

The implementation of Graph attention networks for node classification achieves the goal. The steps to implement the GAT module are as follows:

- Dataset
- Pre-processing
- Dataset Split
- Simple Linear Transformation
- Application of Graph Attention Networks for Node Classification and Graph Representation
- Finding the Citation Probability of Each Paper on a Node to Another Node
- Extraction of H-index and Highly Cited Papers from the Pool of Citations and Graph Representation of Highly Cited Papers

The graph representation of the overall dataset depicts the papers citing each other using directed graphs, following established approaches to citation network visualization for trend detection [52]. The graph can be seen in Figure 7.

4.1 Implementation Details

The GAT model was implemented in Python 3.10 using the PyTorch Geometric library. The network comprised two GAT layers: the first with 8 attention heads and 8 hidden units per head (64 total), followed by the ELU activation function; the second (output) layer used a single attention head producing class logits over the 7 CORA categories. Dropout with a rate of 0.6 was applied to both inputs and attention coefficients to mitigate overfitting. The model was trained using

Table 3. Probability of intra-classes for citation (in decimal form).

	P(CB)	P(GA)	P(NN)	P(PM)	P(RLg)	P(RL)	P(Theory)
P(CB)	0.96041	0.00031	0.00016	0.01562	0.00052	0.165761	0.05689
P(GA)	0.00551	0.99861	0.00001	0.00542	0	0.01698	0.04226
P(NN)	0.00139	0.00002	0.98854	0.50645	0.00059	0.08024	0.19832
P(PM)	0.00656	0	0.00519	0.41824	0	0.17988	0.54079
P(RLg)	0.00238	0.00031	0.00148	0.01641	0.99764	0.01584	0.03134
P(RL)	0.01942	0.00071	0.00077	0.00502	0.00026	0.49422	0.04427
P(Theory)	0.00432	0.00004	0.00386	0.03284	0.001	0.04713	0.08612

the Adam optimizer with an initial learning rate of 5×10^{-3} and L2 regularization of 5×10^{-4}, for a maximum of 1000 epochs with early stopping based on validation loss (patience = 100). All experiments were conducted on Google Colab with a single NVIDIA T4 GPU. The dataset was split into 50% training and 50% testing subsets, with 10% of the training set reserved for validation.

The graph attention mechanism was implemented for computing pairwise attention scores, and then the scores were normalized. The node states of the neighbors and applied attention scores to it, and the aggregation results achieved by neighbors were aggregated. The implementation mentioned above achieved the following results:

- Researcher used the LeakyRelu activation function and kernel attention function with eight heads
- The results were compiled by concatenating the individual results
- After testing the implemented model, the model achieved an accuracy of 78.60% and an F1 score of 78.14% on the test dataset for node classification. These results are consistent with the performance range reported for graph-based node classification methods on the CORA benchmark, including spectral approaches [78, 82], sampling-based scalable methods [87], and methods addressing graph heterophily [104].
- The probability of a paper belonging to each class was also calculated

The achieved accuracy is lower than some prior results by a margin of 1% to 4.6%. This is because previous models excluded the CORA citation file from training, whereas our approach incorporates citation links directly, introducing additional intra-class confusion as reflected in the class probability distributions shown in Table 3. The graph representation of the node-classified papers citing each other using directed graphs. The graphs can be seen in Figure 8. We found the highly cited papers as shown in Table 4, and they follow the order of citation in the graph:

Table 4. Highly cited papers extracted from the CORA dataset.

Paper_ID	Rank	No of Citations
35	1	166
6213	2	76
1365	3	74
3229	4	61
114	5	42
910	6	41
4330	7	38
1272	8	32
3231	9	32
4584	10	32
19621	11	31
2440	12	30
24966	13	29
2665	14	28
6214	15	28

The results validated the purpose of this research, and after the deduction of complete graphs, we found the H-index of the paper from the pool of citations. Paper_ID: 12182 in the Cora data set has a high H-index (20) as the H-index uses two parameters: (i) Quantity – Number of papers, (ii) Quality – Number of citations. Moreover, the researcher found that paper_id 35 is a highly cited (166) paper in the pool of citations. Furthermore, 15 highly cited papers are shown in Figure 8 and Table 4. Graph neural network-based approaches have demonstrated strong capability in identifying and predicting citation trends within academic networks at the time of publication [105].

4.2 Comparative Study of Proposed Model With Machine Learning Models

The proposed GAT model was evaluated against five baselines on the CORA dataset under an identical

Figure 7. Complete graph drawn with node classification.

50:50 train/test split: Random Forest (RF), Support Vector Machine (SVM), Graph Convolutional Network (GCN) [33], and GraphSAGE [39]. All methods used the same 1433-dimensional bag-of-words node features. As shown in Table 5 and Figure 9, GAT achieved the highest accuracy among all compared methods with 78.60%, outperforming SVM (72.51%) and Random Forest (75.83%). The performance advantage confirms that the attention mechanism's ability to differentially weight neighbor contributions is particularly beneficial in citation networks, where the relevance of citing papers varies substantially across classes. Recent deep GCN architectures [82] further confirm the advantage of graph-based methods over classical baselines on citation benchmarks.

5 Conclusion and Future Scope

The goal of this study and the novelty of the research is to observe and closely study the relationships among different papers on separate nodes for node classification and to extract the highly cited papers. Moreover, this task was extended, and the H-index on the CORA dataset was evaluated. To accomplish these objectives, we thoroughly analyzed the citation dataset, which is a citation network. The primary task in this data set was node classification, extracting highly cited papers, and calculating the H-index.

The structural characteristics of the citation network itself have a very complex architecture and need particular emphasis for performing the required task. This research includes the study of bonding among the different nodes by evaluating the probability of each node to other nodes and comparing its relationship with other nodes. We achieved an accuracy of 78.60% and an F1 score of 78.14% on the test dataset. While this is lower than some previously reported results by a margin of 1% to 4.6%, those prior models excluded

Figure 8. Extraction and visualization of highly cited papers.

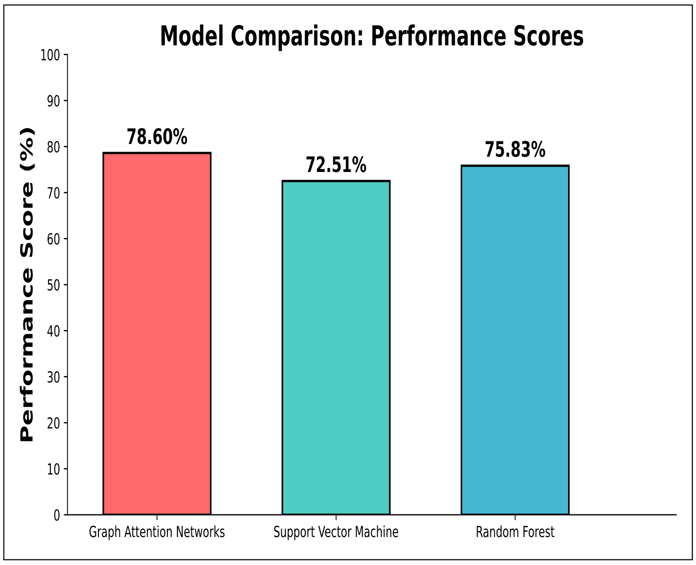

Figure 9. Comparative analysis (Accuracy (%)) of graph attention network model with machine learning models.

the CORA citation file from training. Our approach incorporates citation links as part of the training data, which introduces additional intra-class confusion but yields a more comprehensive and realistic evaluation of the citation network structure. From a practical standpoint, the H-index computation and highly cited paper extraction methods demonstrated here can be directly applied by academic search platforms, library systems, and research evaluation bodies to automate the identification of high-impact works within large citation corpora—a capability with direct relevance to the design of next-generation scholarly information systems. Moreover, we successfully extracted highly cited papers and H-index from the CORA dataset. This study found seven nodes consisting of machine learning papers, but with different classifications. Moreover, the researcher found a high H-index of paper_ID 12182 in this study, and highly cited papers have paper_id 35. Therefore, this study provides the inner sense of citation network structure and

Table 5. Node classification accuracy (%) and F1 score (%) on the CORA dataset under the 50:50 train/test split.

Method	Accuracy (%)	F1 (%)
Random Forest	75.83	71.8
SVM	72.51	73.5
GCN [33]	76.3	75.9
GraphSAGE [39]	77.1	76.8
GAT (Ours)	**78.60**	**78.14**

All methods use identical node features (1433-dimensional bag-of-words) under the same 50:50 train/test split. GCN and GraphSAGE accuracy values are reproduced from the literature under equivalent experimental settings. RF and SVM results are as reported in Figure 9.

the influence of one paper on another in the node. Moreover, this research will open avenues, such as investigating the potential impact of incorporating domain-specific knowledge into the node classification process. This could involve developing novel feature extraction techniques that leverage domain-specific information or exploring the use of domain-specific ontologies to guide the node classification process. Furthermore, in the future, we will explore and develop novel methods for assessing the reliability of citation data in scientific publications. In addition, in the future, we will expand our study to other standard datasets like Citeceer, PubMed, etc. This could involve using machine learning techniques to identify and flag potentially unreliable citation information or exploring the use of crowd-sourcing or expert reviewers to validate citation data.

Data Availability Statement

The code supporting the findings of this study is openly available on GitHub at https://github.com/TahirSher/Graph-Attention-Networks-for-Node-Classification-and-Highly-Cited-Papers. All experiments were conducted using Google Colab with Python 3.10 and standard library versions compatible with Python 3.10.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Pride, D., Harag, J., & Knoth, P. (2022). ACT2: A multi-disciplinary semi-structured dataset for importance and purpose classification of citations. *arXiv preprint arXiv:2209.00084*. [CrossRef]

[2] Zhao, S. X., Rousseau, R., & Ye, F. Y. (2011). H-related indices and their potential applications for measuring the academic impact of journals. *Journal of the American Society for Information Science and Technology, 62*(6), 1176-1189. [CrossRef]

[3] Abrishami, A., & Aliakbary, S. (2019). Predicting citation counts based on deep neural network learning techniques. *Journal of Informetrics, 13*(2), 485-499. [CrossRef]

[4] Kleminski, R., Kazienko, P., & Kajdanowicz, T. (2022). Analysis of direct citation, co-citation and bibliographic coupling in scientific topic identification. *Journal of Information Science, 48*(3), 349-373. [CrossRef]

[5] Xiao, S., Wang, S., Dai, Y., & Guo, W. (2022). Graph neural networks in node classification: survey and evaluation. *Machine Vision and Applications, 33*(1), 4. [CrossRef]

[6] Haris, M., Qazi, A., Bhanger, M. I., Wagan, A. I., & Beg, M. O. (2022). Citation networks: A systematic review and mapping of their structure, properties, and analysis techniques. *Scientometrics, 127*(9), 5311-5332. [CrossRef]

[7] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2008). The graph neural network model. *IEEE transactions on neural networks, 20*(1), 61–80. [CrossRef]

[8] Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology, 61*(12), 2389-2404. [CrossRef]

[9] Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., ... Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open, 1*, 57–81. [CrossRef]

[10] Zhang, X., Yue, T., Zhu, G., Xu, H., & Yang, L. (2021). Research on publication influence evaluation and prediction based on citation network analysis. *Information Processing & Management, 58*(5), 102664. [CrossRef]

[11] Lv, Y., Xie, Z., Zuo, X., & Song, Y. (2022). A multi-view method of scientific paper classification via heterogeneous graph embeddings. *Scientometrics, 127*(8), 4847-4872. [CrossRef]

[12] Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., ... & Barabási, A.-L.

(2018). Science of science. *Science, 359*(6379), eaao0185. [CrossRef]

[13] Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA, 295*(1), 90-93. [CrossRef]

[14] Waltman, L., & Van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology, 63*(12), 2378-2392. [CrossRef]

[15] Garfield, E. (1979). Citation indexing: Its theory and application in science, technology, and humanities. *Journal of the American Society for Information Science, 30*(5), 361-361. [CrossRef]

[16] Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics, 60*(3), 421–432. [CrossRef]

[17] Cai, H., Zheng, V. W., & Chang, K. C. C. (2018). A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE transactions on knowledge and data engineering, 30*(9), 1616-1637. [CrossRef]

[18] Huang, X., Wu, Z., Wang, G., Li, Z., Luo, Y., & Wu, X. (2024). ResGAT: an improved graph neural network based on multi-head attention mechanism and residual network for paper classification. *Scientometrics, 129*(2), 1015-1036. [CrossRef]

[19] Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for information Science, 24*(4), 265-269. [CrossRef]

[20] Chen, P., & Redner, S. (2010). Community structure of the physical review citation network. *Journal of Informetrics, 4*(3), 278-290. [CrossRef]

[21] Bai, X., Liu, H., Zhang, F., Ning, Z., Kong, X., Lee, I., & Xia, F. (2017). An overview on evaluating and predicting scholarly article impact. *Information, 8*(3), 73. [CrossRef]

[22] Bai, X., Zhang, F., Li, J., Mu, D., Lee, I., & Xia, F. (2019). Scientific paper recommendation: A survey. *IEEE Access, 7*, 9324-9339. [CrossRef]

[23] Kipf, T. N., & Welling, M. (2016). Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*. [CrossRef]

[24] Chami, I., Abu-El-Haija, S., Perozzi, B., Ré, C., & Murphy, K. (2022). Machine learning on graphs: A model and comprehensive taxonomy. *Journal of Machine Learning Research, 23*(89), 1-64.

[25] He, G., Xue, Z., Jiang, Z., Kang, Y., Zhao, S., & Lu, W. (2023). H2CGL: Modeling dynamics of citation network for impact prediction. *Information Processing & Management, 60*(6), 103512. [CrossRef]

[26] Wang, D., Song, C., & Barabási, A. L. (2013). Quantifying long-term scientific impact. *Science, 342*(6154), 127-132. [CrossRef]

[27] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2017). Graph attention networks. *arXiv preprint arXiv:1710.10903*.

[28] Lee, J. B., Rossi, R., & Kong, X. (2018). Graph classification using structural attention. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 1666–1674). [CrossRef]

[29] Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2019). How powerful are graph neural networks? In *Proceedings of the 7th International Conference on Learning Representations* (*ICLR*). [CrossRef]

[30] Dong, Y., Johnson, R. A., & Chawla, N. V. (2015, February). Will this paper increase your h-index? Scientific impact prediction. In *Proceedings of the eighth ACM international conference on web search and data mining* (pp. 149-158). [CrossRef]

[31] Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in Neural Information Processing Systems, 29*. [CrossRef]

[32] Yan, P., Kang, Y., Jiang, Z., Song, K., Lin, T., Sun, C., & Liu, X. (2024, July). Modeling scholarly collaboration and temporal dynamics in citation networks for impact prediction. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 2522-2526). [CrossRef]

[33] Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

[34] Ruiz, L., Gama, F., & Ribeiro, A. (2020). Gated graph recurrent neural networks. *IEEE Transactions on Signal Processing, 68*, 6303-6318. [CrossRef]

[35] Zhang, D., Yin, J., Zhu, X., & Zhang, C. (2018). Network representation learning: A survey. *IEEE transactions on Big Data, 6*(1), 3–28. [CrossRef]

[36] Xia, F., Sun, K., Yu, S., Aziz, A., Wan, L., Pan, S., & Liu, H. (2021). Graph learning: A survey. *IEEE Transactions on Artificial Intelligence, 2*(2), 109-127. [CrossRef]

[37] Brody, S., Alon, U., & Yahav, E. (2021). How attentive are graph attention networks?. *arXiv preprint arXiv:2105.14491*. [CrossRef]

[38] Yun, S., Jeong, M., Kim, R., Kang, J., & Kim, H. J. (2019). Graph transformer networks. *Advances in Neural Information Processing Systems, 32*. [CrossRef]

[39] Hamilton, W., Ying, Z., & Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in neural information processing systems, 30*.

[40] Zhang, J., Shi, X., Xie, J., Ma, H., King, I., & Yeung, D.-Y. (2018). Gaan: Gated attention networks for learning on large and spatiotemporal graphs. *arXiv preprint arXiv:1803.07294*.

[41] Liu, Z., Chen, C., Li, L., Zhou, J., Li, X., Song, L., & Qi, Y. (2019). Geniepath: Graph neural networks with

adaptive receptive paths. In *Proceedings of the AAAI Conference on Artificial Intelligence* (pp. 4424–4431). [CrossRef]

[42] Yao, L., Mao, C., & Luo, Y. (2019). Graph convolutional networks for text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, pp. 7370-7377). [CrossRef]

[43] Dong, Y., Johnson, R. A., & Chawla, N. V. (2016). Can scientific impact be predicted?. *IEEE Transactions on Big Data, 2*(1), 18-30. [CrossRef]

[44] Qiu, T., Yu, C., Zhong, Y., An, L., & Li, G. (2021). A scientific citation recommendation model integrating network and text representations. *Scientometrics, 126*(11), 9199-9221. [CrossRef]

[45] Atanassova, I., Bertin, M., & Mayr, P. (2022). Mining Scientific Papers, Volume II: Knowledge Discovery and Data Exploitation. *Frontiers in Research Metrics and Analytics, 7*, 911070. [CrossRef]

[46] Yan, E., & Ding, Y. (2012). Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other. *Journal of the American Society for Information Science and Technology, 63*(7), 1313-1326. [CrossRef]

[47] Perozzi, B., Al-Rfou, R., & Skiena, S. (2014). DeepWalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 701-710). [CrossRef]

[48] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems, 26*. [CrossRef]

[49] Turki, H., Hadj Taieb, M. A., Ben Aouicha, M., Fraumann, G., Hauschke, C., & Heller, L. (2021). Enhancing knowledge graph extraction and validation from scholarly publications using bibliographic metadata. *Frontiers in research metrics and analytics, 6*, 694307. [CrossRef]

[50] Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., & Mei, Q. (2015). LINE: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 1067-1077). [CrossRef]

[51] Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods, 18*(3), 429-472. [CrossRef]

[52] Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology, 57*(3), 359-377. [CrossRef]

[53] Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X., & Zhai, C. (2014). Content-based citation analysis: The next generation of citation analysis. *Journal of the association for information science and technology, 65*(9), 1820-1833. [CrossRef]

[54] Pornprasit, C., Liu, X., Kiattipadungkul, P., Kertkeidkachorn, N., Kim, K. S., Noraset, T., ... & Tuarob, S. (2022). Enhancing citation recommendation using citation network embedding. *Scientometrics, 127*(1), 233-264. [CrossRef]

[55] Ali, Z., Ullah, I., Khan, A., Ullah Jan, A., & Muhammad, K. (2021). An overview and evaluation of citation recommendation models. *Scientometrics, 126*(5), 4083-4119. [CrossRef]

[56] Färber, M., Thiemann, A., & Jatowt, A. (2018). shotgun: A combined approach for cited text span identification. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries* (pp. 11-20). [CrossRef]

[57] Ali, Z., Kefalas, P., Muhammad, K., Ali, B., & Imran, M. (2021). Deep learning in citation recommendation models survey. *Expert Systems with Applications, 162*, 113790. [CrossRef]

[58] Lin, G., Wang, J., Liao, K., Zhao, F., & Chen, W. (2020). Structure fusion based on graph convolutional networks for node classification in citation networks. *Electronics, 9*(3), 432. [CrossRef]

[59] Cohan, A., Ammar, W., van Zuylen, M., & Cady, F. (2019). Structural scaffolds for citation intent classification in scientific publications. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 3586-3596). [CrossRef]

[60] Yu, J., Li, Y., Pan, C., & Wang, J. (2021). A Classification Method for Academic Resources Based on a Graph Attention Network. *Future Internet, 13*(3), 64. [CrossRef]

[61] Aljohani, N. R., Fayoumi, A., & Hassan, S. U. (2021). A novel deep neural network-based approach to measure scholarly research dissemination using citations network. *Applied Sciences, 11*(22), 10970. [CrossRef]

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

[63] Jing, C., Qiu, L., Tian, X., & Hao, T. (2022). Publication classification prediction via citation attention fusion based on dynamic relations. *Knowledge-Based Systems, 239*, 108056. [CrossRef]

[64] Shidiq, M. A., Murfi, H., & Wibisono, A. (2023). Citation recommendation using graph attention network. *IEEE Access, 11*, 21398–21410. [CrossRef]

[65] Zhang, Y., Tian, X., Zhang, N., & Shi, S. (2022). Incorporating citation graph into document representation learning for document classification. *Knowledge-Based Systems, 238*, 107918. [CrossRef]

[66] Aljohani, N. R., Fayoumi, A., & Hassan, S. U. (2021). An in-text citation classification predictive model

for a scholarly search system. *Scientometrics, 126*(7), 5509-5529. [CrossRef]

[67] Cohan, A., Feldman, S., Beltagy, I., Downey, D., & Weld, D. S. (2020, July). Specter: Document-level representation learning using citation-informed transformers. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 2270-2282). [CrossRef]

[68] Zeng, H., Zhou, H., Srivastava, A., Kannan, R., & Prasanna, V. (2019). Graphsaint: Graph sampling based inductive learning method. *arXiv preprint arXiv:1907.04931*. [CrossRef]

[69] Hu, Z., Dong, Y., Wang, K., & Sun, Y. (2020, April). Heterogeneous graph transformer. In *Proceedings of the web conference 2020* (pp. 2704-2710). [CrossRef]

[70] Zhang, Z., Huang, J., & Tan, Q. (2020). SR-HGAT: symmetric relations based heterogeneous graph attention network. *IEEE Access, 8*, 165631–165645. [CrossRef]

[71] Wang, B., Wu, F., & Shi, L. (2023). AGSTA-NET: Adaptive graph spatiotemporal attention network for citation count prediction. *Scientometrics, 128*(1), 511-541. [CrossRef]

[72] Liu, J. C., Chen, C. T., Lee, C., & Huang, S. H. (2024). Evolving knowledge graph representation learning with multiple attention strategies for citation recommendation system. *ACM Transactions on Intelligent Systems and Technology, 15*(2), 1-26. [CrossRef]

[73] Liu, M., Gao, H., & Ji, S. (2020, August). Towards deeper graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 338-348). [CrossRef]

[74] Ostendorff, M., Ruas, T., Ruas, P., & Rehm, G. (2022). Neighborhood contrastive learning for scientific document representations with citation embeddings. *arXiv preprint arXiv:2202.06187*. [CrossRef]

[75] Saier, T., & Färber, M. (2020). unarXive: a large scholarly data set with publications' full-text, annotated in-text citations, and links to metadata. *Scientometrics, 125*(3), 3085-3108. [CrossRef]

[76] Jeong, C., Jang, S., Park, E., & Choi, S. (2020). A context-aware citation recommendation model with BERT and graph convolutional networks. *Scientometrics, 124*(3), 1907-1922. [CrossRef]

[77] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

[78] Xu, B., Shen, H., Cao, Q., Qiu, Y., & Cheng, X. (2019). Graph wavelet neural network. In *Proceedings of the 7th International Conference on Learning Representations (ICLR)*. [CrossRef]

[79] Huang, S., Huang, Y., Bu, Y., Lu, W., Qian, J., & Wang, D. (2022). Fine-grained citation count prediction via a transformer-based model with among-attention mechanism. *Information Processing & Management, 59*(2), 102799. [CrossRef]

[80] Sen, P., Namata, G., Bilgic, M., Getoor, L., Gallagher, B., & Eliassi-Rad, T. (2008). Collective classification in network data. *AI magazine, 29*(3), 93-106. [CrossRef]

[81] McCallum, A. K., Nigam, K., Rennie, J., & Seymore, K. (2000). Automating the construction of internet portals with machine learning. *Information Retrieval, 3*(2), 127-163. [CrossRef]

[82] Chen, M., Wei, Z., Huang, Z., Ding, B., & Li, Y. (2020). Simple and deep graph convolutional networks. In *International conference on machine learning* (pp. 1725–1735).

[83] Colavizza, G., Romanello, M., & Kaplan, F. (2019). The references of references: A method to enrich humanities citation networks. *Journal of the Association for Information Science and Technology, 70*(6), 584-597. [CrossRef]

[84] Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11*(4), 959–975. [CrossRef]

[85] Gasteiger, J., Bojchevski, A., & Günnemann, S. (2018). Predict then propagate: Graph neural networks meet personalized pagerank. *arXiv preprint arXiv:1810.05997*. [CrossRef]

[86] Färber, M., & Jatowt, A. (2020). Citation recommendation: approaches and datasets. *International Journal on Digital Libraries, 21*(4), 375-405. [CrossRef]

[87] Chen, J., Ma, T., & Xiao, C. (2018). FastGCN: Fast learning with graph convolutional networks via importance sampling. In *Proceedings of the 6th International Conference on Learning Representations (ICLR)*. [CrossRef]

[88] Shi, L., Tong, H., Tang, J., & Lin, C. (2015). Vegas: Visual influence graph summarization on citation networks. *IEEE Transactions on Knowledge and Data Engineering, 27*(12), 3417-3431. [CrossRef]

[89] Kong, X., Shi, Y., Yu, S., Liu, J., & Xia, F. (2019). Academic social networks: Modeling, analysis, mining and applications. *Journal of Network and Computer Applications, 132*, 86-103. [CrossRef]

[90] Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences, 102*(46), 16569-16572. [CrossRef]

[91] Ihsan, I., Rahman, H., Shaikh, A., Sulaiman, A., Rajab, K., & Rajab, A. (2023). Improving in-text citation reason extraction and classification using supervised machine learning techniques. *Computer Speech & Language, 82*, 101526. [CrossRef]

[92] Bhagat, S., Cormode, G., & Muthukrishnan, S. (2011). Node classification in social networks. In *Social network data analytics* (pp. 115-148). Boston, MA: Springer US. [CrossRef]

[93] Rong, Y., Huang, W., Xu, T., & Huang, J. (2019). Dropedge: Towards deep graph convolutional networks on node classification. *arXiv preprint arXiv:1907.10903*.

[94] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2020). A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems, 32*(1), 4-24. [CrossRef]

[95] Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. *arXiv preprint arXiv:1707.02919*. [CrossRef]

[96] Jurgens, D., Kumar, S., Hoover, R., McFarland, D., & Jurafsky, D. (2018). Measuring the evolution of a scientific field through citation frames. *Transactions of the Association for Computational Linguistics, 6*, 391-406. [CrossRef]

[97] Yang, Z., Cohen, W., & Salakhudinov, R. (2016, June). Revisiting semi-supervised learning with graph embeddings. In *International conference on machine learning* (pp. 40-48). PMLR.

[98] Sun, K., Lin, Z., & Zhu, Z. (2020, April). Multi-stage self-supervised learning for graph convolutional networks on graphs with few labeled nodes. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 04, pp. 5892-5899). [CrossRef]

[99] Yang, C., & Han, J. (2023, April). Revisiting citation prediction with cluster-aware text-enhanced heterogeneous graph neural networks. In *2023 IEEE 39th international conference on data engineering (ICDE)* (pp. 682-695). IEEE. [CrossRef]

[100] Xiao, B. Y., Tseng, C. C., & Lee, S. L. (2022, July). A Hypergraph Convolutional Neural Network for Node Classification of Citation Network Data. In *2022 IEEE International Conference on Consumer Electronics-Taiwan* (pp. 243-244). IEEE. [CrossRef]

[101] Li, B., & Pi, D. (2019). Learning deep neural networks for node classification. *Expert Systems with Applications, 137*, 324–334. [CrossRef]

[102] Bertin, M., & Atanassova, I. (2024). Linguistic perspectives in deciphering citation function classification. *Scientometrics, 129*(10), 6301-6313. [CrossRef]

[103] Ding, Y. (2011). Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *Journal of informetrics, 5*(1), 187-203. [CrossRef]

[104] Zhu, J., Yan, Y., Zhao, L., Heimann, M., Akoglu, L., & Koutra, D. (2020). Beyond homophily in graph neural networks: Current limitations and effective designs. *Advances in neural information processing systems, 33*, 7793-7804.

[105] Cummings, D., & Nassar, M. (2020). Structured citation trend prediction using graph neural networks. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 3897-3901). [CrossRef]

Tahir Sher pursuing his PhD in Artificial Intelligence from Korea University, Seoul Campus, Korea. He done his MS in Data Science from Air University and bachelor's degree in Mathematics from the International Islamic University, Islamabad and achieved gold-medal. With strong teaching experience, he has taught subjects like Calculus, Differential Equations, Linear Algebra, and Data Science methods to students from Pakistan and allied countries. As a research scholar in the Explainable AI Group at Air University, his interests span IoT, Social IoT, Machine Learning, NLP, Time Series, Mathematical Modeling, Federated Learning, and Computer Vision. He is dedicated to bridging computer science with human-centered research. (Email: 2025010294@korea.ac.kr)

Abdul Rehman is currently a Research Professor at the Human Data Convergence Research Institute, Jeonju University, South Korea. He previously worked as a Postdoctoral Research Associate at the Hyper-Connectivity Convergence Technology Research Center, Kyungpook National University (KNU), Daegu, South Korea, and served as an Assistant Professor at the University of Central Punjab, Lahore, Pakistan. He holds a bachelor's degree in Mathematics from the International Islamic University, Islamabad, Pakistan, and an Integrated Ph.D. in Computer Science and Engineering from KNU. His research interests include Deep Learning, Machine Learning, Data Science, and their applications in Computer Vision, Smart IoT (S-IoT), and Data Analysis. He also explores Complex Network Navigation, Mathematical Modeling, and Big Data Analytics. Dr. Rehman received the KINGS Scholarship for his Ph.D. studies and was honored with the "Outstanding Researcher" Award from the School of Computer Science and Engineering at KNU. He serves as a guest editor and editorial board member for several international journals and has contributed to numerous international conferences as a session chair and publication chair. More information about Dr. A. Rehman is available at https://sites.google.com/view/drrehman/home.

Imran Ihsan received a PhD degree in Knowledge Graphs and Explainable AI from the Capital University of Science and Technology, Islamabad, in 2020. He has over 25 years of academic (teaching, project supervision, and academic administration) and industry (software, web, graphics, animation, and game development and operational management) experience. Since 2013, he has been working as Professor with the Department of Creative Technologies, Faculty of Computing and AI, Air University, Islamabad. He has authored or coauthored research work in various conferences and journals. His research interests include knowledge engineering, semantic computing, natural language processing, and computational linguistics.