Boosting ResGAT for Node Classification of Citation Network with Multi-hop and Motif-Based Attention

Yixiao Chen
Department of Statistics and Data Science
Yale University
New Haven, CT 06511
yixiao.chen@yale.edu

1 Problem Definition

In our project, we aim to solve the node classification problem of citation network using deep learning with graph-based methods. Accurate classification of academic papers is fundamental for efficient information retrieval and bibliometric analysis (Debackere, 2022; Ye, 2022). Recent studies have highlighted the advantages of leveraging inter-paper relationships, and many researchers have applied a combination of graph-based and attention-based learning methods for improved paper classification performance (Veličković et al., 2017; Zhang et al., 2022; Huang et al., 2024).

The graph-based paper classification problem can be formally defined as follows.

- **Citation Graph** - The citation graph is defined as $G = (V, E)$, where $V = \{v_1, v_2, \cdots, v_n\}$ represents a list of papers and $E \subset V \times V$ is the edge set. The nodes in the graph represent papers, and the edges represent relationships between the papers.
- **Node Features** - Given a paper $v_i \in V$, its features are defined as X_i, which may include information such as the paper's title, abstract, keywords, or other metadata.
- **Node Labels** - The label y_i for each paper, which indicates the classification of paper v_i. The label is derived based on specific categories or topics assigned to the paper.
- **Edges** - The edge set E represents relationships between papers, such as citation links or co-authorship, and is typically static for this problem.

The existing ResGAT model proposed by Huang et al. (2024) integrates a multi-head attention mechanism and a residual network structure to improve academic paper classification by effectively handling multi-level citation relationships. However, ResGAT's ability to capture rich, contextual information from multiple levels of the graph is limited. This project aims to develop a more accurate and generalized graph neural network model for paper classification by incorporating a multi-hop attention mechanism (Wang et al., 2020) and explicitly separates and processes relationships across different levels of aggregation. To further enhance the model's performance in learning complex graph structures, we incorporate a motif-based hybrid information matrix (Sheng et al., 2024) that take into account both first-order and higher-order structural information to compute the multi-hop attention computation.

Preprint. Under review.

2 Datasets

In this project, we will implement our novel approach with the two citation benchmark datasets used by Bojchevski and Günnemann (2017): CiteSeer and CORA ML (will try some larger dataset like CORAFULL if time permitted). The summary statistics of both dataset are shown in table 1, and the degree distribution and class distribution can be visualized through the following figures 1 and 2.

Dataset Statistics	CiteSeer Dataset	CORA ML Dataset
Graph Size		
Total number of nodes	4,230	2,995
Total number of edges	5,337	8,158
Number of isolated nodes	0	0
Node Classification		
Number of labeled nodes	4,230	2,995
Number of non-labeled nodes	0	0
Number of node categories	6	7

Table 1: Statistics of the CiteSeer Dataset and the CORA dataset.

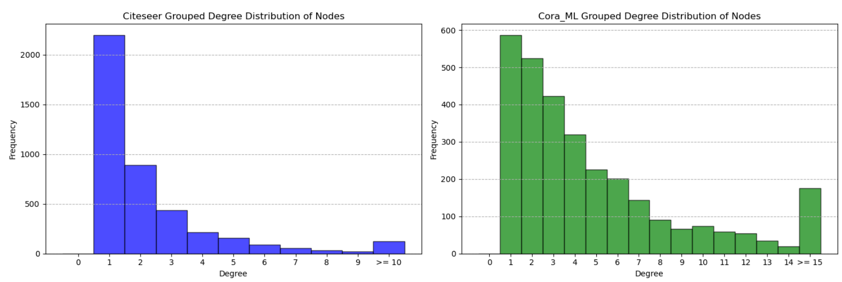

Figure 1: Grouped Degree Distribution Comparison for Citeseer and Cora_ML Datasets.

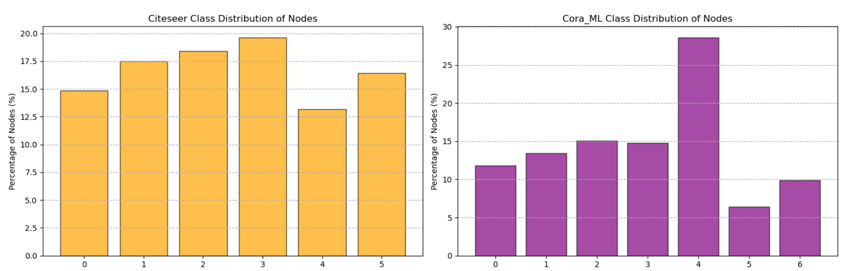

Figure 2: Class Distribution Comparison for Citeseer and Cora_ML Datasets.

3 Related Works

- **Graph Neural Networks:** Graph Neural Networks (GNNs) have recently achieved significant advancements in graph representation learning. Significant efforts on designing the architecture of GNN have been made to propagate messages differently. Graph Convolutional Networks (GCN) (Kipf and Welling, 2016) aggregate feature information from a node's neighbors using convolution operations for semi-supervised learning on graphs. GraphSAGE (Hamilton et al., 2017) extends GCN by sampling a fixed-size neighborhood and applying learnable aggregation functions to generate node embeddings, enabling inductive learning on large-scale graphs. Unlike these GNNs, which assume equal influence from neighboring nodes, the Graph Attention Network (GAT) (Veličković et al., 2017) utilizes an attention mechanism to learn the relative importance between connected nodes.

Recent works (Li and Ouyang, 2024; Zhang et al., 2024) attempt to incroporate LLMs into the process of graph representation learning to further foster the performance and the generalized ability of GNNs.

- **GAT in Paper Classification & ResGAT:** Graph Attention Networks (GAT) are particular useful for paper classification with applying attention mechanisms to weigh the importance of neighboring nodes during feature aggregation. This mechanisms allow the model to focus more on influential citations while dynamically adjusting these weights. The effectiveness of GAT in paper classification is evidenced by numerous studies (Kim and Oh, 2019; Verma et al., 2023; Huang et al., 2024). Specifically, Huang et al. (2024) proposed ResGAT incorporates a multi-head attention mechanism along with a residual network structure to enhance academic paper classification by efficiently managing multi-level citation relationships.
- **Multi-hop Attention:** Wang et al. (2020) propose Multi-hop Attention Graph Neural Network (MAGNA), which incorporates multi-hop context information into each layer of GNN attention computation. MAGNA uses a diffusion-based approach to capture large-scale structural information effectively, achieving state-of-the-art performance on node classification and knowledge graph completion tasks. This technique are widely adopted in recent works (Wang et al., 2023; Deng and Huang, 2024; Gong et al., 2024) to improve the performance of GAT for various tasks.
- **Motif-based Attention:** After Sankar et al. (2017) proposed to use motif to create novel spatial GCN to capture high-order structural information, Peng et al. (2018) exploit motifs to capture local stationary and spatial structures of graphs to improve the performance of GNNs in graph classification tasks. Furthermore, Sheng et al. (2024) introduce demonstrate that incorporating motifs also improves the expressive power and robustness of graph neural networks in various node classification tasks.

4 Proposed Approaches

4.1 Multi-hop Attention Mechanism

Objective: We want to enhance the ability of the model to capture complex network relationships by incorporating multi-hop context in the attention computation. Multi-hop attention mechanism allows nodes to receive and aggregate information from a broader set of neighbors.

Methodology: The **Multi-hop Attention Mechanism** extends the traditional attention computation of the Graph Attention Network (GAT) by integrating a multi-hop context Wang et al. (2020). In Multi-hop Attention Graph Neural Network (MAGNA), attention is diffused across multiple hops, increasing the receptive field of each node layer without requiring deeper network architectures that might face over-smoothing problems. Specifically, the multi-hop attention diffusion layer allows the computation of attention scores between nodes that are not directly connected but are within a multi-hop neighborhood. This ensures that the model effectively aggregates both direct and indirect relationships in the graph, improving long-range interaction capabilities.

Mathematically, the diffusion-based approach is used to compute an extended attention matrix, which aggregates the influence of neighboring nodes over multiple hops. This results in more informative attention scores that incorporate both immediate and extended neighborhoods, thereby expanding the receptive field for each node.

4.2 Motif-Based Hybrid Information Integration

Objective: To improve the model's capacity to learn from higher-order structural features by explicitly incorporating network motifs into the attention mechanism.

Methodology: Drawing from the **MGAT** and **MGATv2** models presented by Sheng et al. (2024), we introduce **motif-based hybrid information matrices** into our proposed model. The purpose of these matrices is to incorporate both first-order neighbor relationships and higher-order structures (such as triadic motifs) that frequently occur in real-world graphs. The motif-based hybrid information matrix H is defined as:

$$H = \beta \cdot A + (1 - \beta) \cdot A_M \tag{1}$$

where:

- A is the adjacency matrix representing first-order relationships,
- A_M is the motif-based adjacency matrix that captures the occurrence of higher-order structures,
- β is a hyperparameter that balances the importance between low-order and high-order graph information.

This hybrid information matrix is used in computing the attention scores to ensure that the model can effectively utilize both the immediate attributes of neighboring nodes as well as the richer structural context provided by motifs.

4.3 Research Questions and Hypotheses

- **RQ1: How does incorporating multi-hop attention improve classification accuracy compared to traditional GAT?**
  - **Hypothesis:** The multi-hop attention mechanism will enhance classification by providing more comprehensive contextual information, capturing a richer structure across multiple levels of the network.
- **RQ2: How does the integration of motif-based hybrid information matrices influence the model's performance?**
  - **Hypothesis:** Integrating motif-based information will improve node classification performance by leveraging higher-order structural features, allowing the model to capture complex graph structures more effectively.

5 Evaluation Metrics and Expected Results

Evaluation Metrics:

- **Accuracy:** To assess overall classification performance.
- **Macro F1-Score:** To address class imbalance.
- **Node Classification AUC:** To evaluate the model's performance across different aggregation levels.

Expected Results: The proposed enhancements are expected to yield improved accuracy for paper classification but increased the computational load.

6 Timeline

Detailed Project Timeline:

- **Week 7 (10.7 - 10.13):** Implementation of the data preprocessing module, and get familiarize with required package and methods.
- **Week 8 & Week 9 (10.14 - 10.27):** Implementation of GCN, GRAPHSAGE, GAT, and RESGAT for these two dataset as baseline.
- **Week 10 (10.28 - 11.03):** Refinement of the methodology (incorporate with new acquired techniques, revise the proposing approach, start implementation).
- **Week 11 & Week 12 (11.04 - 11.17):** Implementation of the methodology in cluster; performing optimization and tuning and gathering results. Finishing the most part of experiments of the project.
- **Week 13 & Week 14 & Week 15 (11.17 - 12.9):** Implementation of the methodology in a larger dataset if time permitted and finishing the report writing.

References

Bojchevski, A. and Günnemann, S. (2017). Deep Gaussian Embedding of Graphs: Unsupervised Inductive Learning via Ranking. *arXiv*.

Debackere, W. G. &. K. (2022). Various aspects of interdisciplinarity in research and how to quantify and measure those. *Scientometrics*, 127(9):5551–5569.

Deng, P. and Huang, Y. (2024). Edge-featured multi-hop attention graph neural network for intrusion detection system. *Computers & Security*, 148:104132.

Gong, C., Wei, Z., Wang, X., Wang, R., Li, Y., and Zhu, P. (2024). Subgraph-based attention network for multi-hop question answering. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9.

Hamilton, W. L., Ying, R., and Leskovec, J. (2017). Inductive Representation Learning on Large Graphs. *arXiv*.

Huang, X., Wu, Z., Wang, G., Li, Z., Luo, Y., and 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.

Kim, D. and Oh, A. (2019). Supervised Graph Attention Network for Semi-Supervised Node Classification. [Online; accessed 4. Oct. 2024].

Kipf, T. N. and Welling, M. (2016). Semi-Supervised Classification with Graph Convolutional Networks. *arXiv*.

Li, X. and Ouyang, J. (2024). Related Work and Citation Text Generation: A Survey. *arXiv*.

Peng, H., Li, J., Gong, Q., Wang, S., Ning, Y., and Yu, P. S. (2018). Graph Convolutional Neural Networks via Motif-based Attention. *arXiv*.

Sankar, A., Zhang, X., and Chang, K. C.-C. (2017). Motif-based Convolutional Neural Network on Graphs. *arXiv*.

Sheng, J., Zhang, Y., Wang, B., and Chang, Y. (2024). MGATs: Motif-Based Graph Attention Networks. *Mathematics*, 12(2):293.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. (2017). Graph Attention Networks. *arXiv*.

Verma, A. K., Saxena, R., Jadeja, M., Bhateja, V., and Lin, J. C.-W. (2023). Bet-GAT: An Efficient Centrality-Based Graph Attention Model for Semi-Supervised Node Classification. *Appl. Sci.*, 13(2):847.

Wang, G., Ying, R., Huang, J., and Leskovec, J. (2020). Multi-hop Attention Graph Neural Network. *arXiv*.

Wang, L., Li, Z.-W., You, Z.-H., Huang, D.-S., and Wong, L. (2023). MAGCDA: A Multi-Hop Attention Graph Neural Networks Method for CircRNA-Disease Association Prediction. *IEEE J. Biomed. Health Inf.*, 28(3):1752–1761.

Ye, H. Z. &. M. J. &. Y. Y. &. H. J. &. T. (2022). Recommendation method for academic journal submission based on doc2vec and XGBoost. *Scientometrics*, 127(5):2381–2394.

Zhang, J., Chen, J., Maatouk, A., Bui, N., Xie, Q., Tassiulas, L., Shao, J., Xu, H., and Ying, R. (2024). LitFM: A Retrieval Augmented Structure-aware Foundation Model For Citation Graphs. *arXiv*.

Zhang, L., Sun, B., Shu, F., and Huang, Y. (2022). Comparing paper level classifications across different methods and systems: an investigation of Nature publications. *Scientometrics*, 127(12):7633–7651.