SUPPLEMENTARY MATERIAL: TEXT ATTRIBUTED GRAPH NODE CLASSIFICATION USING SHEAF NEURAL NETWORKS AND LARGE LANGUAGE MODELS

Anonymous authors
Paper under double-blind review

A DATASETS

This section provides a detailed introduction to the datasets used in the main content:

Cora Sen et al. (2008) dataset consists of 2,708 scientific publications categorized into seven classes: case-based, genetic algorithms, neural networks, probabilistic methods, reinforcement learning, rule learning, and theory. Each paper in the citation network cites or is cited by at least one other paper, resulting in a total of 5,429 edges. We use the collected dataset[1] with raw texts provided by TAPE He et al. (2023).

CiteSeer Giles et al. (1998) dataset consists of 3,186 scientific publications categorized into six classes: Agents, Machine Learning, Information Retrieval, Database, Human Computer Interaction, and Artificial Intelligence. Our task is to predict the category of a paper based on its title and abstract.

WikiCS Mernyei & Cangea (2020) dataset is a Wikipedia-based dataset designed for benchmarking Graph Neural Networks. It is constructed from Wikipedia categories, specifically featuring 10 classes corresponding to branches of computer science, exhibiting very high connectivity. The node features are derived from the text of the corresponding articles. We obtain the raw texts of each node from https://github.com/pmernyei/wiki-cs-dataset.

OGBN-ArXiv Hu et al. (2020) dataset is a directed graph representing the citation network among all computer science arXiv papers indexed by MAG Wang et al. (2020). Each node corresponds to an arXiv paper, and directed edges indicate citations from one paper to another. The objective is to predict the 40 subject areas of arXiv CS papers, such as cs.AI, cs.LG, and cs.OS. These subject areas are manually determined and labeled by the paper's authors and arXiv moderators.

Arxiv-2023, proposed in TAPE He et al. (2023), is a directed graph illustrating the citation network among all computer science arXiv papers published in 2023 or later. Like OGBN-ArXiv, each node represents an arXiv paper, and directed edges denote citations from one paper to another. The objective is to predict the 40 subject areas of arXiv CS papers, including cs.AI, cs.LG, and cs.OS. These subject areas are manually determined and labeled by the paper's authors and arXiv moderators.

OGBN-Products Hu et al. (2020) is characterized by a substantial scale, comprising 2 million nodes and 61 million edges. We utilize a node sampling strategy, following TAPE He et al. (2023), to obtain a subgraph containing 54k nodes and 74k edges, resulting in the OGBN-Products(subset) dataset. Each node in this dataset represents products sold on Amazon, and edges between two products indicate that the products are purchased together. The task involves predicting the category of a product in a multi-class classification setup, using the 47 top-level categories as target labels.

B BASELINES

In this section, we provide a detailed introduction to the baselines used:

- Traditional GNNs: We adopt three simple but widely used GNN models in this work, *i.e.*, GCN Kipf & Welling (2016), SAGE Hamilton et al. (2017), GAT Veličković et al. (2018).
- GraphFormers Yang et al. (2021) is a graph transformer nested with GNN in each layer, originally designed for link prediction tasks.

[1] https://github.com/XiaoxinHe/TAPE

Table 1: The interpretation comparison with LLM predictor methods on two cases. Case 1: a paper classification task on Cora dataset. The blue parts are the labels. The red parts are the wrong predictions, while the green parts are the correct predictions and the corresponding structurally relevant interpretations.

Text	0-hop	1-hop	Sheaf
(label: theory) Title: Constructive induction using a non-greedy strategy for feature selection.; Abstract:We present a method for feature construction and selection that finds a minimal set of conjunctive features that are appropriate to perform the classification task···	Based on the content of the paper, the most suitable category for this paper would be theoretical"case-based methods". The paper introduces a method for learning adaptation strategies in case-based reasoning systems, focusing on generating memory search plans through introspective reasoning···	The two papers are related by citation. The first paper presents a method for feature construction and selection that finds minimal conjunctive features for classification tasks, focusing on minimal multi-level Boolean expressions. The second paper explores constructive induction in unsupervised learning, proposing a theoretical model to distinguish and motivate the process of CI. Given the focus on feature selection and Boolean expressions, the first paper falls under the rule learning category. ···	Based on the content of the paper, the most appropriate category for this paper would be "theory". The paper presents a method for constructive induction using a non-greedy strategy for feature selection. It focuses on developing a theoretical framework and algorithmic approach to identify a minimal set ..... The paper also cites several references related to theoretical machine learning···

- NodeFormer Wu et al. (2022) is an efficient graph transformer for large graphs which develops a kernelized Gumbel-Softmax Jang et al. (2016) operator.
- Fintuned LMs: We adopt three widely used pre-trained language models: BERT Kenton & Toutanova (2019), SentenceBERT Reimers & Gurevych (2019), and DeBERTa He et al. (2020). The parameters of these models are fully fine-tuned in our experiments.
- GIANT Chien et al. (2021) is a cascading structure method with two stages: pretraining LMs and training GNNs. In the first stage, it enhances node representations by integrating structural information into LM pre-training. Then fine-tuned LM-generated node features serve as initial features for GNN training.
- GLEM Zhao et al. (2022) is an effective framework that fuses large language models and GNNs in the training phase through a variational EM framework. We used the source code[2] provided in the original paper.
- TAPE He et al. (2023) utilizes LLMs, like ChatGPT OpenAI (2023), to generate pseudo labels and explanations for textual nodes. Then it will finetune PLMs with the generated content and original texts. The enhanced features, derived from the fine-tuned PLMs, are used as initial node features for training GNNs.
- SimTeG Duan et al. (2023) is also a cascading structure method tailored for textual graphs. It employs a two-stage training paradigm, initially fintuning language models and subsequently training GNNs.
- Fine-Tuning methods: LoRA Hu et al. (2021), IA3 Liu et al. (2022), Prompt Tuning Lester et al. (2021), and LST Sung et al. (2022). These methods involve fine-tuning large language models to showcase experimental results on textual graphs.

Methods	Cora	CiteSeer	WikiCS
MLP	83.92±1.39	91.00±0.95	92.31±0.07
GCN	90.22±0.89	92.93±1.36	93.63±0.24
SAGE	88.25±0.88	91.68±1.08	95.93±0.20
GAT	89.70±1.72	91.95±0.90	93.25±0.13
SheaFormer	**94.55±0.51**	**96.27±0.72**	**98.43±0.28**

Table 2: Link prediction performance, as evaluated by AUC metric.

	Cora	CiteSeer	WikiCS	OGBN-ArXiv	ArXiv-2023	OGBN-Products	Ele-Photo
Learning rate	1e-3	1e-2	1e-2	1e-3	1e-2	1e-3	1e-3
Batch size	32	32	32	32	32	32	32
Optimizer	AdamW	AdamW	AdamW	AdamW	AdamW	AdamW	AdamW

Table 3: Hyper-parameters for fine tune of LLM baselines.

C LINK PREDICTION

Our method is not limited to node classification tasks, it can also be applied to edge-level or graph-level tasks. In this section, we conduct experiments on link prediction tasks. We split existing edges into train:val:test=0.85:0.05:0.1 for all datasets.

	Cora	CiteSeer	WikiCS	OGBN-ArXiv	ArXiv-2023	OGBN-Products	Ele-Photo
# Hidden size	64	64	64	64	64	64	64
# Layers	2	1	1	1	1	1	2
Norm	ID	ID	ID	ID	LN	LN	ID
Activation	ELU	ELU	ReLU	ELU	ELU	ELU	ELU
Dropout	0.5	0.5	0.5	0.2	0.2	0.2	0.5
# Epochs	200	200	200	200	200	200	200
Learning rate	5e-5	5e-5	1e-3	1e-3	1e-4	1e-2	1e-3
Optimizer	AdamW	AdamW	AdamW	AdamW	AdamW	AdamW	AdamW
Weight decay	5e-4	5e-4	1e-4	5e-4	5e-4	5e-4	5e-4
Early stop	True	True	True	True	True	True	True
Patience	500	500	500	50	50	20	20
Sampler	subGraph	subGraph	subGraph	subGraph	subGraph	subGraph	subGraph

Table 4: Hyper-parameters for GNN baselines. 'ID' means no norm layer(Identity), 'LN' denotes Layer Normalization. For sampler, 'RWR' means random walk sampler with restart, and 'subGraph' is one-hop subgraph sampler.

	Cora	CiteSeer	WikiCS	OGBN-ArXiv	ArXiv-2023	OGBN-Products
chatGLM3 +SheaFormer	92.93±0.45	79.04±0.68	82.20±0.77	78.25±0.41	80.70±0.37	81.36±0.62

Table 5: Experimental results of different LLM.

[2] https://github.com/AndyJZhao/GLEM

References

Eli Chien, Wei-Cheng Chang, Cho-Jui Hsieh, Hsiang-Fu Yu, Jiong Zhang, Olgica Milenkovic, and Inderjit S Dhillon. Node feature extraction by self-supervised multi-scale neighborhood prediction. In *arXiv preprint arXiv:2111.00064*, 2021.

Keyu Duan, Qian Liu, Tat-Seng Chua, Shuicheng Yan, Wei Tsang Ooi, Qizhe Xie, and Junxian He. Simteg: A frustratingly simple approach improves textual graph learning. In *arXiv preprint arXiv:2308.02565*, 2023.

C Lee Giles, Kurt D Bollacker, and Steve Lawrence. Citeseer: An automatic citation indexing system. In *Proceedings of the third ACM conference on Digital libraries*, 1998.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Proc. of NeurIPS*, 2017.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. Deberta: decoding-enhanced bert with disentangled attention. In *arXiv preprint arXiv:2006.03654*, 2020.

Xiaoxin He, Xavier Bresson, Thomas Laurent, and Bryan Hooi. Explanations as features: Llm-based features for text-attributed graphs. In *arXiv preprint arXiv:2305.19523*, 2023.

Edward J Hu, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. In *Proc. of ICLR*, 2021.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. *Proc. of NeurIPS*, 2020.

Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. In *Proc. of ICLR*, 2016.

Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proc. of AACL*, 2019.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

Brian Lester, Rami Al-Rfou, and Noah Constant. The power of scale for parameter-efficient prompt tuning. In *Proc. of EMNLP*, 2021.

Haokun Liu, Derek Tam, Mohammed Muqeeth, Jay Mohta, Tenghao Huang, Mohit Bansal, and Colin A Raffel. Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. *Proc. of NeurIPS*, 2022.

Péter Mernyei and Cătălina Cangea. Wiki-cs: A wikipedia-based benchmark for graph neural networks. *arXiv preprint arXiv:2007.02901*, 2020.

OpenAI. Gpt-4 technical report, 2023.

Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*, 2019.

Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI magazine*, 2008.

Yi-Lin Sung, Jaemin Cho, and Mohit Bansal. Lst: Ladder side-tuning for parameter and memory efficient transfer learning. *Proc. of NeurIPS*, 2022.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *International Conference on Learning Representations (ICLR)*, 2018.

Kuansan Wang, Zhihong Shen, Chiyuan Huang, Chieh-Han Wu, Yuxiao Dong, and Anshul Kanakia. Microsoft academic graph: When experts are not enough. *Quantitative Science Studies*, 2020.

Qitian Wu, Wentao Zhao, Zenan Li, David P Wipf, and Junchi Yan. Nodeformer: A scalable graph structure learning transformer for node classification. *Proc. of NeurIPS*, 2022.

Junhan Yang, Zheng Liu, Shitao Xiao, Chaozhuo Li, Defu Lian, Sanjay Agrawal, Amit Singh, Guangzhong Sun, and Xing Xie. Graphformers: Gnn-nested transformers for representation learning on textual graph. In *NeurIPS'21*, volume 34, pp. 28798–28810, 2021.

Jianan Zhao, Meng Qu, Chaozhuo Li, Hao Yan, Qian Liu, Rui Li, Xing Xie, and Jian Tang. Learning on large-scale text-attributed graphs via variational inference. In *Proc. of ICLR*, 2022.