

Where Multi-View Learning Meets Graph: Collaborative Graph Mixture of Experts

Appendix

A Experiment

A.1 Supplementary Experiment

To further validate the effectiveness and scalability of the proposed MvCGE, we also test it on two large-scale graphs, including a single-view graph OGBN-arXiv, and a multi-view graph Freebase, and record the Micro F1 scores. For the multi-view graph Freebase, we introduce RGCN [Battaglia *et al.*, 2018], HGT [Hu *et al.*, 2020b] for comparisons. Figures 1 (a) and (b) illustrate the performance of MvCGE as well as the comparison algorithms on OGBN-arXiv and Freebase, respectively. It is observed that MvCGE can scale well to larger graphs and still maintains the leading performance, as the time complexity is linear to the number of nodes and edges. It is worth noting that we can control the number of graph experts to achieve efficient inference.

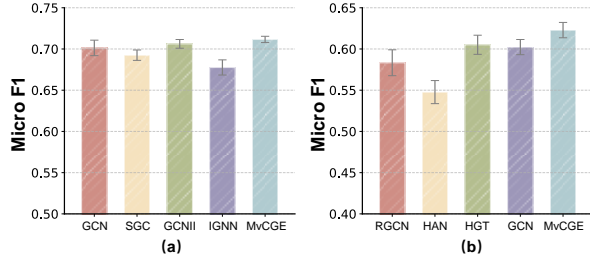


Figure 1: The classification performance of MvCGE on two large-scale graphs.

A.2 Dataset Description

Table 1: Summary of datasets used in experiments

	Datasets	# Nodes	# Edges	# Node types	# Views	# Classes
Multi-view	ACM-M	8,994	25,922	3	2	3
	DBLP	18,405	67,946	4	3	4
	IMDB	12,722	19,120	4	3	3
	YELP	3,913	72,132	4	3	3
	AMINER	55,783	153,676	3	2	6
	Freebase	180,098	1,645,725	8	7	7
Single-view	ACM	3,025	13,255	1	2	3
	Citeseer	3,327	4,732	1	2	6
	CoraFull	19,793	65,311	1	2	70
	Flickr	7,575	239,738	1	2	9
	UAI	3,067	28,311	1	2	19
	arXiv	169,343	1,166,243	1	2	40

The statistics of the datasets used are summarized in Table 1. Detailed descriptions of each dataset are provided as follows:

ACM-M [Wang *et al.*, 2019] is an academic graph dataset comprising three types of nodes: paper (P), author (A), and subject (S). It includes four types of edges (PA, AP, PS, SP), and the task is to classify papers based on their research domains. All nodes are leveraged to construct citation networks, paper content, and other data integration studies. We employ the meta-path set {PAP, PSP} for experiments.

DBLP [Wang *et al.*, 2019] is a graph extracted from the DBLP network, consisting of paper (P), author (A), and conference (C) nodes. All the nodes are classified into four categories, i.e., author, paper, term, and conference. The meta-path set {APA, APCPA, APTPA} is employed to conduct experiments.

IMDB [Wang *et al.*, 2019] is a movie dataset containing four types of nodes, i.e., movie (M), actor (A), director (D), and year (Y). Nodes are divided into three classes according to the movie genre. We perform experiments using the meta-path set {MAM, MDM, MYM}.

AMINER [Hu *et al.*, 2019] is a scholarly heterogeneous graph constructed from paper (P), author (A), and reference (R) nodes. There are two types of metapaths {PAP, PRP} that are leveraged, with the classification task focused on paper domains.

YELP [Lu *et al.*, 2019] is a subset derived from the merchant review website with four types of nodes, i.e., business (B), user (U), service (S), and level (L). We generate the meta-path set {BUB, BLB, BSB} for our experiments.

Freebase [Lv *et al.*, 2021] is a large-scale knowledge graph that includes eight node types: book (B), film (F), location (L), music (M), person (P), sport (S), organization (O), and business (U). We take the meta-path set {BB, BFB, BLMB, BPB, BPSB, BOFB, BUB} for the experiment.

ACM [Wang *et al.*, 2020] is a paper network. Different from the above citation networks, the edges denote the co-author relationships between any two papers. Another difference is that the node features are bag-of-words representations of papers' keywords.

Citeseer [Yang *et al.*, 2016] is also a well-known citation network with nodes, node features, and edges having the same meanings as Cora. Similarly, the papers are grouped into 6 classes.

CoraFull [Bojchevski and Günnemann, 2017] is a citation network containing a number of machine learning papers. Each node represents a paper, and the edges represent the citation relationships between papers. Node features are the bag-of-words representations of papers, and all papers are divided into 70 categories according to their domains.

Flickr [Huang *et al.*, 2017] is a social network, where nodes represent users and edges represent that a user has added another user as a contact. All users are grouped into 9 categories according to their interests.

UAI [Wang *et al.*, 2018] is a webpage network that has been used to test GCN for community detection. Nodes represent web pages and each edge represents a citation between two pages.

OGBN-arXiv [Hu *et al.*, 2020a] dataset is a citation network with 169,343 computer science arXiv papers, where each node is an arXiv paper and each edge indicates that a paper cites another paper. Each paper has a 128-D feature vector, which is obtained by averaging the embeddings of words in its title and abstract.

B Complexity Analysis

The graph experts and routers are two main components of MvCGE. Using a simple first-order graph filter-GCN layer to implement each graph expert, and assuming the input feature dimension $D \gg D_l$ for any l , the computational cost of each graph expert is $\mathcal{O}(ND^2 + |\mathcal{E}|D)$. Then it is $\mathcal{O}(NDE + |\mathcal{E}|D)$ for the routing mechanism. Since we only route each sample to K graph experts, we have $\mathcal{O}(KND^2 + K|\mathcal{E}|D)$ for each MvCGE layer when $D \gg E$. When constructing L layers and processing V views, the total computational complexity is $\mathcal{O}(LVKND^2 + LVK|\mathcal{E}|D)$. Therefore, when the number of experts E is not too large, properly increasing it and controlling K can expand the model's capacity while preserving efficiency.

References

Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*, 2018.

- Aleksandar Bojchevski and Stephan Günnemann. Deep gaussian embedding of graphs: Unsupervised inductive learning via ranking. *arXiv preprint arXiv:1707.03815*, 2017.
- Binbin Hu, Yuan Fang, and Chuan Shi. Adversarial learning on heterogeneous information networks. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 120–129, 2019.
- 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. In *Advances in neural information processing systems*, volume 33, pages 22118–22133, 2020.
- Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. Heterogeneous graph transformer. In *Proceedings of the Web Conference*, pages 2704–2710, 2020.
- Xiao Huang, Jundong Li, and Xia Hu. Label informed attributed network embedding. In *Proceedings of the ACM International Conference on Web Search and Data Mining*, pages 731–739, 2017.
- Yuanfu Lu, Chuan Shi, Linmei Hu, and Zhiyuan Liu. Relation structure-aware heterogeneous information network embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4456–4463, 2019.
- Qingsong Lv, Ming Ding, Qiang Liu, Yuxiang Chen, Wenzheng Feng, Siming He, Chang Zhou, Jianguo Jiang, Yuxiao Dong, and Jie Tang. Are we really making much progress? revisiting, benchmarking and refining heterogeneous graph neural networks. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1150–1160, 2021.
- Wenjun Wang, Xiao Liu, Pengfei Jiao, Xue Chen, and Di Jin. A unified weakly supervised framework for community detection and semantic matching. In Dinh Phung, Vincent S. Tseng, Geoffrey I. Webb, Bao Ho, Mohadeseh Ganji, and Lida Rashidi, editors, *Advances in Knowledge Discovery and Data Mining*, pages 218–230, 2018.
- Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. Heterogeneous graph attention network. In *Proceedings of the International World Wide Web Conference*, pages 2022–2032, 2019.
- Xiao Wang, Meiqi Zhu, Deyu Bo, Peng Cui, Chuan Shi, and Jian Pei. Am-gcn: Adaptive multi-channel graph convolutional networks. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1243–1253, 2020.
- Zhilin Yang, William Cohen, and Ruslan Salakhudinov. Revisiting semi-supervised learning with graph embeddings. In *Proceedings of the International conference on machine learning*, pages 40–48, 2016.