



## Yi Fang

Research Assistant  
SFSC of AI and DL  
New York University(Shanghai)

+86-17771845817

✉ yf2722@nyu.edu

🐙 GitHub Profile

🌐 LinkedIn Profile

## EDUCATION

### Wuhan University

Bachelor of Computer Science

Sept 2017 — Jun 2021

Cumulative GPA: 3.61/4.00

### Syracuse University

Master of Computer Science

Sept 2022 — Present

## RESEARCH INTERESTS

Graph Foundation Models, Multimodal Graphs Learning, Text-Structure Alignment

## RESEARCH EXPERIENCES

### Research Assistant at UIUC

June 2024 — Present

Supervisor: Prof. Jiawei Han

### Research Assistant at NYU Shanghai

August 2023 — June 2024

Supervisor: Prof. Qiaoyu Tan

Shanghai, CN

## PUBLICATIONS

### GAugLLM: Improving Graph Contrastive Learning for TAGs with LLMs

*Yi Fang, Dongzhe Fan, Daochen Zha, Qiaoyu Tan*

Accepted by ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2024)

### UniGLM: Training One Unified Language Model for Text-Attributed Graphs

*Yi Fang, Dongzhe Fan, Sirui Ding, Ninghao Liu, Qiaoyu Tan*

Reviewing by The 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024)

### GraphFM: A Comprehensive Benchmark for Graph Foundation Model

*Yuhao Xu, Xinqi Liu, Keyu Duan, Yi Fang, Yu-Neng Chuang, Daochen Zha, Qiaoyu Tan*

Reviewing by the Thirty-eighth Annual Conference on Neural Information Processing Systems (NeurIPS 2024)

## RESEARCH PROJECTS

### GAugLLM: Improving Graph Contrastive Learning for TAGs with LLMs

- **Utilized LLM for data augmentation:** Applied LLM to enhance graph contrastive learning from both feature and structure augmentation perspectives by using various prompt and structure information to enhance textual attributes and graph structure robustness.
- **Mixture-of-Prompt-Experts:** Integrates textual attributes revised by different prompts experts with node and dataset context information by adaptively aligning text attributes with structure in a learnable manner.
- **Collaborative edge modifier:** Utilized collaborative knowledge in the hidden space, input structure, and text attributes to identify valuable and malicious connections for edge augmentation.

### UniGLM: Training One Unified Language Model for Text-Attributed Graphs

- **Generalist Embedding Model:** Explored the development of a generalist embedding model for TAGs and introduce UniGLM a novel language model pre-training framework tailored for a set of TAGs. To the best of our knowledge, UniGLM is the first graph embedding foundation model for TAGs.
- **Adaptive Positive Sample Selection:** Proposed an adaptive positive sample selection method for sampling positive samples of each node for contrastive learning. Unlike standard sampling strategies, our personalized scheme identifies positive samples based on nodes' local, global, and graph-related contexts, thereby unifying graph structures across various TAGs.
- **Dynamic Embedding Table:** Devised a simple yet effective dynamic embedding table scheme to encode sampled positive samples off-the-fly during mini-batch training. By maintaining an external memory bank to update and retrieval embeddings of positives examples, we accelerate the training process using historical embeddings as supervision.