LINGXIAO ZHAO

CONTACT

lingxiao@cmu.edu | 412-880-9109
github.com/LingxiaoShawn | Hamburgh Hall 2005
linkedin.com/in/lingxiaozhao | 4800 Forbes Ave
http://lingxiaozhao.com/ | Pittsburgh, PA
Google Scholar | United States, 15213

EDUCATION

Carnegie Mellon University — Pittsburgh, PA, USA
Ph.D. in Machine Learning (joint Public Policy) — Aug. 2018 to Now
Advisors: Prof. Leman Akoglu, Prof. Aarti Singh
Ph.D. Committee: Prof. Leman Akoglu, Prof. Aarti Singh, Prof. Andrej Risteski, Dr. Neil Shah

Carnegie Mellon University — Pittsburgh, PA, USA
M.S. in Electrical and Computer Engineering, GPA: 3.90/4.00 — Aug. 2016 to Dec. 2017

Xi'an Jiaotong University — Xi'an, Shaanxi, China
B.S. in Electrical Engineering, Rank: 7/370 — Aug. 2012 to July 2016

EXPERIENCE

Apple Inc., Research Intern. with *Navdeep Jaitly*. — May 2023 to Aug. 2023
- Working on improving *pre-training* convergence and generalization of large language model.

Nvidia, Reserach Intern. with *Haggai Maron*. — May 2022 to Aug. 2022
- Working on graph generative model.

Snap Inc., Reserach Intern. with *Neil Shah*. — May 2021 to Aug. 2021
- Working on improving expressivity of graph neural network.

IBM Research, Research Intern. with *Charu Aggarwal*. — May 2020 to Aug.2020
- Working on graph-level anomaly detection, with GNN based approach.

RESEARCH INTEREST

Graph is a powerful representation that captures interactions between identities. My work focuses on solving some fundamental problems of Graph Neural Networks, such as breaking depth-limitation, relaxing homophily requirement, learning better structure, and improving expressiveness. My recent work is focusing on discrete-state diffusion, graph generation, and graph pretraining, with a final goal of building **a generative foundation model on graph**. In future, I'm interested in 1) combining graph models and LLMs to **boost LLMs' reasoning and emergent abilities** in different stages, and towards a multimodal foundation model; 2) understanding LLM's reasoning limitation within single step, and **improving LLM planner** with graph; 3) leveraging graph and relational structure to improve RAG. I am also interested in **scientific discovery** and AI for Science with LLM and graph.

AWARDS

- Daniel and Rise Nagin **Dissertation Fellowship** — Aug. 2023
- NeurIPS 2022 Travel Grant — Oct. 2022
- SIGIR Student Travel Grant — Sep. 2018
- Outstanding Student Pacesetter (Only **10** students in **3800** undergraduates) — Oct. 2015

- National Scholarship (Rank **4/370** and **5/370**, respectively) Oct. 2015, Oct. 2014
- **Outstanding Winner** (highest) In MCM/ICM[1] (**0.19%** worldwide) April. 2014

PUBLICATIONS

** denotes equal contribution.*

Preprint

1. Improving Training Stability and Generalization of Transformer
   Lingxiao Zhao's intern project at Apple
2. Unifying Discrete- and Continuous-Time Diffusion on Categorical Data
   Lingxiao Zhao*, Xueying Ding*, Leman Akoglu.
3. Autoregressive Graph Generation with Local Diffusion
   Lingxiao Zhao, Xueying Ding, Leman Akoglu.
4. Descriptive Kernel Convolution Network with Improved Random Walk Kernel
   Jeremy Lee*, Lingxiao Zhao*, Leman Akoglu.
5. End-to-End Augmentation Hyperparameter Tuning for Self-Supervised Anomaly Detection
   Jaemin Yoo, **Lingxiao Zhao**, and Leman Akoglu

Refereed Conference Publications

1. Anomaly Detection of Attributed Multi-graphs with Metadata: A Unifieds GNN Approach
   2023 IEEE BigData
   Konstantinos Sotiropoulos*, Lingxiao Zhao*, Pierre Jinghong Liang, Leman Akoglu
2. DSV: An Alignment Validation Loss for Self-supervised Outlier Model Selection,
   2023 ECML PKDD
   Jaemin Yoo, Yue Zhao, **Lingxiao Zhao**, and Leman Akoglu
3. Sign and Basis Invariant Networks for Spectral Graph Representation Learning, **2023 ICLR**
   Derek Lim*, Joshua Robinson*, **Lingxiao Zhao**, Tess Smidit, Suvrit Sra, Haggai Maron, Stefanie Jegelka.
4. A Practical, Progressively-Expressive GNN, **2022 NeurIPS**
   Lingxiao Zhao, Louis Härtel, Neil Shah, Leman Akoglu
5. Hyperparameter Sensitivity in Deep Outlier Detection: Analysis and a Scalable Hyper-Ensemble Solution, **2022 NeurIPS**
   Xueying Ding, **Lingxiao Zhao**, Leman Akoglu
6. Graph-level Anomaly Detection with Unsupervised GNNs, **2022 ICDM Short**
   Lingxiao Zhao, Saurabh Sawlani, Arvind Srinivasan, Leman Akoglu
7. From Stars to Subgraphs: Uplifting Any GNN with Local Structure Awareness, **2022 ICLR**
   Lingxiao Zhao, Wei Jin, Leman Akoglu, Neil Shah
8. Graph Condensation for Graph Neural Networks, **2022 ICLR**
   Wei Jin, **Lingxiao Zhao**, Shichang Zhang, Yozen Liu, Jiliang Tang, Neil Shah
9. Fast Attributed Graph Embedding via Density of States , **2021 ICDM**
   Saurabh Sawlani, **Lingxiao Zhao**, Leman Akoglu

[1]The Mathematical Contest in Modeling/The Interdisciplinary Contest in Modeling: comap.com

10. Graph unrolling networks: Interpretable neural networks for graph signal denoising
2021 IEEE Transactions on Signal Processing
Siheng Chen, Yonina C. Eldar, **Lingxiao Zhao**

11. Beyond Homophily in Graph Neural Networks: Limitations and Effective Designs, **2020 NeurIPS**
Jiong Zhu, Yujun Yan, **Lingxiao Zhao**, Mark Heimann, Leman Akoglu, Danai Koutra

12. Connecting Graph Convolutional Net and Graph-Regularized PCA, **2020 ICML Workshop**
Lingxiao Zhao, Leman Akoglu

13. PairNorm: Tackling Oversmoothing in GNNs, **2020 ICLR**
Lingxiao Zhao, Leman Akoglu

14. A Quest for Structure: Jointly Learning Graph and Semi-Supervised Classification, **2018 CIKM**
Xuan Wu*, Lingxiao Zhao*, Leman Akoglu. (* equal contribution)

Refereed Journal Publications

1. Heterophily and Graph Neural Networks: Past, Present and Future
2023 IEEE Data Engineering Bulletin, June 2023
Jiong Zhu, Yujun Yan, Mark Heimann, **Lingxiao Zhao**, Leman Akoglu, Danai Koutra

2. Density of States for Fast Embedding Node-Attributed Graphs
2022 Knowledge and Information System Joural
Lingxiao Zhao, Saurabh Sawlani, Leman Akoglu

3. On Using Classification Datasets to Evaluate Graph Outlier Detection: Peculiar Observations
2021 Big Data Journal
Lingxiao Zhao, Leman Akoglu

ACADEMIC SERVICE

Program Committee Member, Conference Reviewer

- International Conference on Machine Learning (ICML) 2022
- Conference on Neural Information Processing Systems (NeurIPS) 2021-2023
- International Conference on Learning Representations (ICLR) 2022-2024
- Learning on Graphs Conference (LoG) 2022

TEACHING

- Teaching assistant for CMU 95-869 Big Data and Large Scale Computing (2020, 2021, 2022)
- Teaching assistant for CMU 95-828 Machine Learning for Problem Solving (2021, 2022)
- Teaching assistant for CMU 90-812 Introduction to Programming with Python (2019)
- Teaching assistant for CMU 10-725 Convex Optimization (2019)

SKILLS

Languages: Python, MATLAB, Java, C, C++, LaTeX, Bash
Frameworks: Pytorch, Tensorflow
Large-Scale Computing: MapReduce, Hadoop, Spark, Pig
Web: HTML, CSS, Bootstrap, Jekyll

SELECTED PROJECTS

Generative Pretraining on Graphs Jan. 2023 to Now

Proposed an autoregressive based diffusion model, it builds the foundation of graph pretraining.

- Unified continuous- & discrete-time discrete-state diffusion which supports diverse noises.
- Proposed an local diffusion model that generates graph block-by-block, autoregressively.
- Working on using the designed blockwise diffusion model to perform generative pretraining.

From Stars to Subgraphs: Uplifting Any GNN Mar. 2021 to Oct. 2021

Proposed the first general framework to boost expressiveness of any GNN

- Designed a general framework to uplift any GNN, by generalizing MPNN's aggregation field from stars to subgraphs and encoding subgraphs with the (base) GNN.
- Proved the expressiveness is strictly better than 1-WL, while being not less powerful than 3-WL.
- Designed effective and efficient realizations: different encodings and distance-to-centroid feature.
- Designed SubgraphDrop that greatly reduces memory cost and still keeps same performance.

Tackling oversmoothing problem in GNNs March 2019 to Sept. 2019

Proposed the first normalization layer in GNNs, for solving oversmoothing

- Studied oversmoothing thoroughly with SGC, which decouples the effect of oversmoothing on performance drop of increasing layers from overfitting and vanishing gradient problems.
- Proposed an efficient and effective normalization layer, PairNorm, that "pushes" oversmoothed node representations away from each other. PairNorm greatly improves robustness of depth.
- Studied a new setting: Semi-Supervised Learning with large percentage missing feature problem, where PairNorm greatly boost performance for all three types of GNN (SGC,GCN,GAT).

COURSEWORK

Carnegie Mellon University

- 10-716: Advanced Machine Learning
- 10-725: Convex Optimization
- 10-708: Probabilistic Graphical Model
- 10-705: Intermediate Statistics
- 10-605: Machine Learning in Large Datasets
- 11-785: Introduction to Deep Learning
- 15-213: Introduction to Computer System
- 15-214: Pinciples of Software Construction
- 15-780: Artificial Intelligence
- 15-659: Probability and Computing
- 15-650: Algorithm and Data Structure
- 36-731 & 36-732: Causal Inference
- 36-707: Regression Analysis
- 15-859: Spectral Graph Theory
- 10-703: Deep Reinforcement Learning