

Yuntian He

PhD Student in Computer Science, The Ohio State University

heyuntian.cn@gmail.com

(614) 598-7496

[LinkedIn] [Twitter] [Google Scholar]

EDUCATION

The Ohio State University

Ph.D. in Computer Science and Engineering, Advisor: Dr. Srinivasan Parthasarathy
(Expected to graduate in December 2024)

Columbus, OH

2019–Present

University of Science and Technology of China

M.Eng. in Computer Science and Technology, Advisor: Dr. Kai Han

Hefei, China

2016–2019

Xi'an Jiaotong University

B.Eng. in Computer Science and Technology

Xi'an, China

2012–2016

SKILLS

- **Programming Languages:** Python, C, C++, Java
- **ML Frameworks:** Tensorflow, PyTorch, Keras, Jupyter, Google colab
- **HPC & Other Toolkits:** Slurm, OpenMP, CUDA, MPI, Horovod, Git, Django, HTML5

RESEARCH EXPERIENCE

• Applied Machine Learning

- Explored representation learning of graph and text data in darknet authorship attribution (EMNLP '21)
- Used transformer and graph embedding approaches to learn representations
- Gained up to **220% improvement** on the prediction scores in the classification task

• Efficient and Scalable ML

- Studied efficient and scalable graph embedding based on hierarchical frameworks
- Leveraged Horovod, CUDA and OpenMP on high-performance computing resources (HiPC '21)
- Tested the proposed framework on computing clusters and achieved **28× speedup** over the baselines
- Implemented a web tool that scales graph embedding scripts in different languages (VLDB '22)

• Ethics and Fairness in Machine Learning

- Proposed novel frameworks for fairness-aware graph representation learning (WebConf '23, EAAMO '22/'23)
- Leveraged ML libraries including TensorFlow and PyTorch to build GNN modules
- Enhanced the efficiency by **up to 100×** w.r.t SOTA while achieving competitive utility scores

WORK EXPERIENCE

Lawrence Berkeley National Laboratory

Student Assistant

Berkeley, CA

May 2023 - Aug 2023

- Developed an HDF5 data compression software for scientific computing projects at LBNL
- Explored HDF5 I/O, data conversion, and filter plugin development in C and Python
- Tested the developed software and achieved high compression ratios (**up to 245**)

Samsung Research America

Machine Learning Research Intern

Mountain View, CA

May 2022 - Aug 2022

- Researched fairness in heterogeneous graph embedding
- Benchmarked existing baselines and designed two new approaches based on parallel random walk and heterogeneous GNNs respectively
- Improved the fairness by **up to 45%** in node classification tasks

Nokia Bell Labs, Ireland
Machine Learning Intern

Dublin, Ireland
June 2020 - Aug 2020

- Developed two approaches for anomaly detection in call flows using skip-gram model and GNNs
- Achieved 93% model accuracy on call flow data collected by Nokia

SELECTED PUBLICATIONS

- [1] **Y. He**, S. Gurukar, and S. Parthasarathy, “Efficient fair graph representation learning using a multi-level framework”, in *Companion Proceedings of the Web Conference 2023*, 2023.
- [2] **Y. He**, S. Gurukar, and S. Parthasarathy, “Fairmile: A multi-level framework for fair and scalable graph representation learning”, in *The 3rd ACM conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO)*, 2023.
- [3] Y. Zhang, **Y. He**, S. Gurukar, and S. Parthasarathy, “Heteromile: A multi-level graph representation learning framework for heterogeneous graphs”, *Under review*, 2023.
- [4] S. Current, **Y. He**, S. Gurukar, and S. Parthasarathy, “FairEGM: Fair link prediction and recommendation via emulated graph modification”, in *The 2nd ACM conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO)*, 2022.
- [5] **Y. He**, Y. Zhang, S. Gurukar, and S. Parthasarathy, “WebMILE: Democratizing network representation learning at scale”, in *The 48th International Conference on Very Large Databases (VLDB '22, Demo Track)*, 2022.
- [6] **Y. He**, S. Gurukar, P. Kousha, H. Subramoni, D. K. Panda, and S. Parthasarathy, “DistMILE: A distributed multi-level framework for scalable graph embedding”, in *The 28th IEEE International Conference on High Performance Computing, Data, and Analytics (HiPC)*, 2021.
- [7] P. Maneriker, **Y. He**, and S. Parthasarathy, “SysML: Stylometry with structure and multitask learning: Implications for darknet forum migrant analysis”, in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021.
- [8] K. Han, **Y. He**, K. Huang, X. Xiao, S. Tang, J. Xu, and L. Huang, “Best bang for the buck: Cost-effective seed selection for online social networks”, *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol. 32, no. 12, pp. 2297–2309, 2020.
- [9] K. Han, F. Gui, X. Xiao, J. Tang, **Y. He**, Z. Cao, and H. Huang, “Efficient and effective algorithms for clustering uncertain graphs”, in *Proceedings of the 44th International Conference on Very Large Data Bases (VLDB)*, 2019.
- [10] K. Han, **Y. He**, X. Xiao, S. Tang, F. Gui, C. Xu, and J. Luo, “Budget-constrained organization of influential social events”, in *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, IEEE, 2018, pp. 917–928.