

Haoyu Liu

Ph.D. (expected 2027-Jan) in Computer Science | Graph Learning Algorithms & Graph-LLMs

Ph.D. Candidate at NTU | haoyu.liu@ntu.edu.sg | [hyliu-777.github.io](https://github.com/hyliu-777)

PROFILE

Ph.D. candidate at Nanyang Technological University (NTU), advised by Prof. Siqiang Luo. Research focuses on scalable graph algorithms and learning, spanning Personalized PageRank computation, efficient and lightweight graph learning, and Graph-LLM integration. My work combines theoretical guarantees with scalable learning methods, resulting in (co-)first-author publications at PODS, PVLDB, KDD, and ICDE, including the Best Newcomer Research Paper Award at SIGMOD/PODS 2026. Previously interned at Samsung Research AI Lab, where I developed embodied AI agents for Instruction-following household tasks and wins 3rd place in the CVPR 2022 ALFRED Challenge.

EDUCATION

Nanyang Technological University, Singapore

Jan 2023 – Present (Expected Jan 2027)

Ph.D. Candidate in Computer Science, College of Computer and Data Science (CCDS); Advisor: Prof. Siqiang Luo

Renmin University of China, Beijing, China

Sep 2017 - Jun 2021

B.S. in Statistics and Big Data; GPA: 3.83/4.00; Xinshan Scholarship (4%); First Prize Scholarship; Outstanding Graduate.

INTERNSHIP

Samsung Research AI Lab | Embodied AI Intern

March 2021 - Aug 2022

- **Instruction-following VLA agents:** Designed a BART-based planner that combines step-by-step household instructions with scene priors to generate inspectable sub-step and object sequences, separating language understanding from action execution. Based on this, built dynamic 2D semantic maps and object-location memory from pretrained depth and segmentation model prediction, then used FMM replanning and atomic manipulation policies to execute each plan in unseen household scenes. Placed 3rd in the CVPR 2022 ALFRED Challenge.
- **Human-Guided Attention for Atari RL agents:** Constructed simulated human coarse-grained pre-guided (SHCP) maps from task-relevant Atari entities identified through game manuals and template matching, then fused them with learned top-down self-attention. Integrated fusion attention shows great interpretability with an improved average reward by 49%-199% over the self-attention baseline.

PROJECTS

Scalable Machine Learning Methods

[3 papers published in SIGMOD26, KDD26, ICDE25]

- **Fraud Video Detection:** Developed a unified learning framework (SSTGNN) for fraud video detection by graph modeling for integrating spatial, spectral, and temporal cues under a lightweight learning design. It utilizes only 42X fewer parameters with superior performance.
- **Heterophily GNN:** Designed an efficient heterophilous GNN model (SIGMA) to tackle the scalability issue in graph learning with heterophily.
- **Spectral GNN Benchmark:** Contributed to a spectral GNN benchmark of 35 popular GNN models, covering efficiency, memory, and effectiveness evaluation over comprehensive large-scale graph datasets. Design and implement the major experimental code platform.

Scalable Personalized PageRank (PPR)-based Graph Queries

[3 papers published in PODS26, PVLDB24&26]

- **Fundamental PPR Theory:** Established near-optimal upper and lower bounds for Single-Source Personalized PageRank (SSPPR) queries, closing key complexity gaps for a core primitive in ranking, recommendation, and graph analytics stuck for decades.
- **Scalable Bidirectional PPR for Bipartite Graph:** Developed efficient bidirectional hidden PPR algorithms on weighted bipartite graphs for similarity search tailored for bipartite search recommendation scenarios. Designed algorithm scales to billion-scale graphs with 10X speed-up.
- **Scalable PPR-based Graph Similarity:** Developed the first scalable PPR-based random walk kernel algorithm with clear relative error guarantee. Such kernel is effective in grasping graph similarities and supports downstream tasks such as GNNs and 3D-cloud transformers.

LLM + Graph Integration

[ongoing]

- **Temporal GraphRAG:** Co-Developed a training-free retrieval method that builds a compact, time-aligned rule graph and applies PPR-based propagation to retrieve temporally consistent evidence for LLM-QA. Reduced token usage up to 97.0% with superior recall accuracy.
- **Benchmark Graph tokenization for LLM-based GFMs:** Benchmarked graph tokenizers across local, global, and diffusion-based methods to study how graph representation choice controls structural visibility, infers the scaling law of LLM-based GFM and accuracy-token trade-offs.

SELECTED PUBLICATIONS & AWARDS (full list at [Google Scholar](https://scholar.google.com/citations?user=haoyu.liu); * equal contribution; † alphabetical order following theory convention)

Changan Liu*, Haoyu Liu*, Siqiang Luo and Laks VS Lakshmanan. Scalable Graph Kernel Query with Error Guarantees. to appear in **PVLDB 2026**.

Haoyu Liu† with Xinpeng Jiang†, Siqiang Luo†, and Xiaokui Xiao†. Near-Optimality for Single-Source Personalized PageRank. **PODS 2026**.

Best Newcomer Award of SIGMOD/PODS 2026.

Haoyu Liu*, Chaoyu Gong*, Mengke He, Jiatao Li, Kai Han, and Siqiang Luo. When Deepfake Detection Meets Graph Neural Network: A Unified and Lightweight Learning Framework. **KDD 2026**.

Ningyi Liao, Haoyu Liu, Zulun Zhu, Siqiang Luo and Laks VS Lakshmanan. A Comprehensive Benchmark on Spectral GNNs: The Impact on Efficiency, Memory, and Effectiveness. **SIGMOD 2026**. **PREMIA Best Student Paper Award 2025 in Singapore.**

Haoyu Liu, Ningyi Liao, and Siqiang Luo. SIGMA: An Efficient Heterophilous Graph Neural Network with Fast Global Aggregation. **ICDE 2025**.

Haoyu Liu and Siqiang Luo. BIRD: Efficient Approximation of Bidirectional Hidden Personalized PageRank. **PVLDB 2024**.

Haoyu Liu, Yang Liu, etc. Exploring Coarse-Grained Pre-Guided Attention to Assist Fine-Grained Attention Reinforcement Learning Agents. **IJCNN 2022**.

Haoyu Liu, etc. Nightlight as a proxy of economic indicators: Fine-grained GDP inference around Chinese mainland. **Remote Sensing 2021**.