ZHIYU NI

Email: zhiyuni@berkeley.edu Tel: (+1)323-633-1447

EDUCATION

University of California, Berkeley (UC Berkeley) *From Spring 2025*
PhD student, *Computer Science*, Advisor: Pierluigi Nuzzo

University of Southern California (USC) *Los Angeles, CA*
PhD student, *Computer Engineering* *2022.8 - 2024.9*

University of Science and Technology of China (USTC) *Hefei, China*
B.S. *Physics*, Outstanding Graduates *2018.8 - 2022.6*

PUBLICATIONS

Analyzing Adversarial Vulnerabilities of Graph Lottery Tickets (ICASSP 2024 **Oral**)
Zhiyu Ni*, Subhajit Dutta Chowdhury*, Qingyuan Peng, Souvik Kundu, Pierluigi Nuzzo

Finding Adversarially Robust Graph Lottery Tickets (TMLR)
Zhiyu Ni*, Subhajit Dutta Chowdhury*, Qingyuan Peng, Souvik Kundu, Pierluigi Nuzzo

Let Me Grok for You: Accelerating Grokking via Embedding Transfer from a Weaker Model (Under review of ICLR 2025)
Zhiwei Xu*, **Zhiyu Ni***, Wei Hu, Yixin Wang

Differential Privately Embeddings Generation for GNNs (Under review of DAC 2025)
Zhiyu Ni*, Subhajit Dutta Chowdhury, Akshay Ahah, Pierluigi Nuzzo

RESEARCH EXPERIENCE

Adversarially Robust Graph Lottery Ticket (GitHub)

- Systematically analyzed the robustness of graph lottery tickets (GLT) against adversarial attacks.
- Integrated self-training and developed a new loss function to prune the graph edges and model weights, largely improving GLT's robustness against adversarial attacks.

LLMs for Anomaly Detection (GitHub)

- Exploring the capabilities of LLMs (ChatGpt, Llama, etc.) in terms of anomaly detection and encoding graph information into natural language.
- Designing in-context learning flow to enable LLMs to identify fraud and numerically evaluate performance in various datasets (YelpChi, AmazonChi).

CoT-guided Defense against Prompt Injection Attack of LLMs

- Analyzed how prompt injection attacks mislead LLMs to mix instruction prompt and user prompt.
- Generating positive and negative CoT-based prompt pairs to align LLMs to defend against PIA.

WORK EXPERIENCE

NLP Intern (Iflytek *AI Research Institute*)

- Independently carried out machine translation tasks utilizing seamless bidirectional translation.
- Surpassed Google Translation performance by attaining a superior BLEU score.

SKILLS

Python, C, PyTorch, Linux