Artificial Intelligence Initiative (Aii)

https://ai.ucf.edu/



Invited Speaker

Mr. Haoteng Yin, Purdue University

"Towards Scalable and Privacy-Preserving Graph Learning via System-aware Algorithm Design"

Thursday, July 10, 2025 • 1:00PM • TCII 222 • Zoom

Abstract:
Graph learning is a critical driver of advances in scientific discovery, business modeling, and AI-assisted decision-making. However, two fundamental roadblocks hinder its broader adoption: the scalability of powerful, subgraph-based learning methods, and the critical need to protect sensitive information within relational data. The complex dependencies inherent in graph structures render many traditional optimization techniques intractable.

This talk introduces a unified, system-aware algorithmic framework to dismantle these scalability and privacy barriers. To address scalability, I will present a novel family of subgraph -based frameworks, which decouple subgraph-level feature extraction from model training and inference by utilizing reduced forms such as random walks, sets, and hashes. This approach has enabled large-scale learning on graphs with billions of edges. The co-design principle is then extended to privacy-preserving graph learning. Here, the structural dependencies in graphs often violate the key assumption of standard privacy mechanisms like DP-SGD and lead to prohibitive memory overheads. I will introduce the first relational learning framework that offers rigorous privacy guarantees. We have successfully deployed this framework to fine-tune large language models on sensitive graph data and achieve a strong privacy utility trade off.

By holistically co-designing learning algorithms and their underlying system implementations, this research demonstrates the efficient and trustworthy application of graph-based AI to real-world problems. This work opens new avenues for leveraging complex structured data, a crucial step towards building the next generation of foundation models .

Bio:
Haoteng Yin is a Ph.D. candidate in Computer Science at Purdue University, advised by Prof. Pan Li. His research interests lie at the intersection of machine learning and systems, focusing on developing efficient, scalable, and trustworthy learning algorithms for graphs and foundational models on complex structured data. His work has been recognized with a Best Paper Award at SMP '16, a Best Poster Award at the 3rd RCAC Cybersecurity Symposium, a Top Reviewer Award at NeurIPS '22, and the Purdue Teaching Academy Graduate Teaching Award in 2024.