Sepideh Maleki

Email: Sepid.maleki@gmail.com *Cell:(512) 781 6221*
Google scholar: https://scholar.google.com/citations?user=JSwHqtYAAAAJ&hl=en
Website: https://www.cs.utexas.edu/~smaleki/
Work Authorization: U.S. Permanent Resident (Green Card)

Interests

My research focuses on the convergence of foundation models, graph theory, and machine learning, with a special emphasis on advancing biological and drug discovery applications. I am particularly engaged in developing machine learning methods that efficiently leverage foundation models—such as large language models and single-cell models—to tackle complex challenges in molecular and genetic perturbation.

Selected Projects

Multi-modal foundation models for predicting molecular perturbation: *I have been working on fine-tuning single-cell foundation models (FMs), pre-trained on tens of millions of single cells, to address molecular perturbation prediction. I introduced a drug-conditional adapter that allows efficient fine-tuning by training less than 1% of the original foundation model, thus enabling molecular conditioning while preserving the rich biological representation learned during pre-training. The proposed strategy allows not only the prediction of cellular responses to novel drugs, but also the zero-shot generalization to unseen cell types.single cell foundation models such as scGPT to optimize it for the task of predicting molecular perturbation.*

Developing a contrastive approach to explain (hyper)graph neural networks: *I have been working on developing a novel framework, grXAI, for graph representation explainability. grXAI generalizes existing score-based graph explainers to identify the subgraph most responsible for constructing the latent representation of the input graph. This framework can be easily and efficiently implemented as a wrapper around existing methods, enabling the explanation of graph representations through connected subgraphs, which are more human-intelligible.*

Representation Learning on Hypergraphs: *I have been working on applying machine learning techniques to solve prediction and classification tasks on large hypergraphs. I developed and designed a hierarchical hypergraph embedding system that can generate an embedding of a hypergraph with millions of nodes and hyperedges in just a few minutes while current approaches take days to finish. These embeddings then can be used in classification and prediction tasks on hypergraphs.*

Classification and Prediction on hyperedges using transfer learning: *I introduced the novel task of hyperedge classification, i.e. the problem of predicting categorical labels on hyperedges. I developed a simple and effective approach to solving this task. Building on this framework, I demonstrated that by using hyperedge classification as a pretext task with labels obtained from a hierarchical community detection algorithm, we can perform hyperedge prediction. This self-supervised approach improves the state of the art in hyperedge prediction by a significant margin on standard benchmarks.*

Education

PhD student at the University of Texas at Austin, August 2022
Master of Science in Computer Science, Texas State University, May 2016
Bachelor of Science in Computer Science, Kharazmi University, Iran, June 2011

Work Experience

***Postdoctoral Associate** – Genentech (2022 - Present)*

***Graduate Research Assistant** – University of Texas at Austin (2016 - 2022)*
Advised by Professor Keshav Pingali

***Biomedical Data Science**– Stanford University (Summer 2020)*
Advised by Professor Dennis P Wall

***Software Engineering Intern** – Google, Inc. (Summer 2017)*
Developed a benchmark that is representative of the kind of code that Google Executes

***Researcher** – Texas State University (Summer 2016)*

***Graduate Research Assistant** – Texas State University (2015, Spring 2016)*
Advised by Professor Martin Burtscher
***Graduate Instructor Assistant** – Texas State University (Fall 2014)*
Digital Logic, Foundations of Computer Science
***Hardware/Software Tester** – Sanjesh Afzar Asia (2012 - 2013)*
Hardware/software testing of handheld devices capable of reading digital metering

Qualifications

Programming Languages
Python, C/C++, Java, CUDA, C#.NET, R, Matlab
Libraries & Frameworks
Pytorch, Tensorflow, LangGraph, WandB, Scikit-Learn, Numpy, Pandas
Machine Learning & AI
(Un)Supervised Learning, Foundation Models, LLMs, GNNs, Explainability

Honors, Awards, and Activities

***ICLR'24 workshop organizer** – 1st Machine Learning for Genomics Exploration Workshop (MLGenX) ICLR 2024*

***Graduate School fellowship** – The Graduate School at the U. of Texas at Austin (2022)*

***AMS Mathematics Research on Models and Methods for Sparse (Hyper)Networks** – funded participant (2022)*

***Spring school** – Lake Como School of Advanced Studies; Complex Networks (2022)*

***PSB session organizer** – Graph Representations and Algorithms in Biomedicine (2022)*

***PPOPP artifact evaluation committee** – (2021)*

***Best Workshop paper** – WOSET (2019)*

First Place at the Conference of Southern Graduate Schools (CSGS) competition (2017)

***Outstanding Master's Student Award** – College of Science and Engineering at Texas State U. (2016)*

***Computer Science Endowed Scholarship** – Department of Computer Science at Texas State University (2016)*
***Graduate Research Excellence Award** – Department of Computer Science at Texas State University (2015)*