CAREER: ToGFM: Towards Efficient Computation for Large Graph Foundation Model on GPUs

---

Overview:

The PI's career goal is to advance and promote the research and education on high-performance computing for graph-based machine learning that has been widely deployed in critical applications, such as molecular property prediction, drug discovery, material design, and cyber threat detection. Though with the successful applications, existing graph-based machine learning approaches face a major challenge of low generability. That is, most graph learning approaches are tailored to train from scratch for a single task on a particular graph, which requires data collection and deployment for each individual graph and task. Inspired by the success of existing foundation models in natural language processing (NLP) and computer vision (CV) community, the graph foundation model (GFM) has been proposed recently with the goal of developing a graph model capable of generalizing across different graphs and tasks. As GFM usually has a complex structure with a large scale of parameters and often requires training on large dataset, the computation time becomes a major blocker. Therefore, in this project, we propose to design TOGFM, the first system towards efficient computation for large GFMs.

As a crucial step towards the PI's career goal, this project proposes research and education coherent objectives. In particular, the research objective is to pioneer the effort of uniting into one integrated system, named TOGFM leveraging the Graphics Processing Unit (GPU) computing infrastructure, including exascale computing to handle large graphs in terms of both size and count. The educational objective is to use this proposed research to attract and train a diverse group of participants, including K-12, undergraduate, female, and underrepresented minority (URM) populations in the Science, Technology, Engineering, and Mathematics (STEM) field.

Intellectual Merit:

This project will lead to an efficient computation system that advances the computation for large graph foundation models with the following innovations.

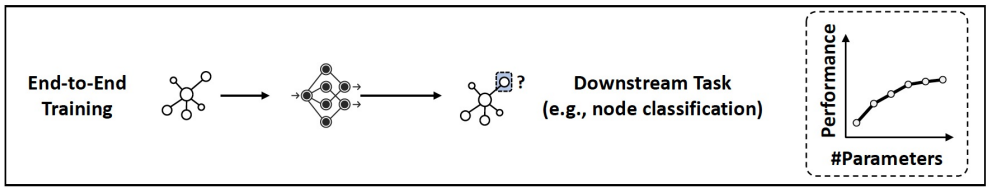
(i) Efficient Graph Neural Network-based Backbone Model Computation, which introduces novel computation techniques tailored for the pertaining on diverse graphs and tasks.

(ii) Scalable Graph Transformer-based Backbone Model Computation, which designs scalable computation techniques to support graph transformer-based backbone model computation.

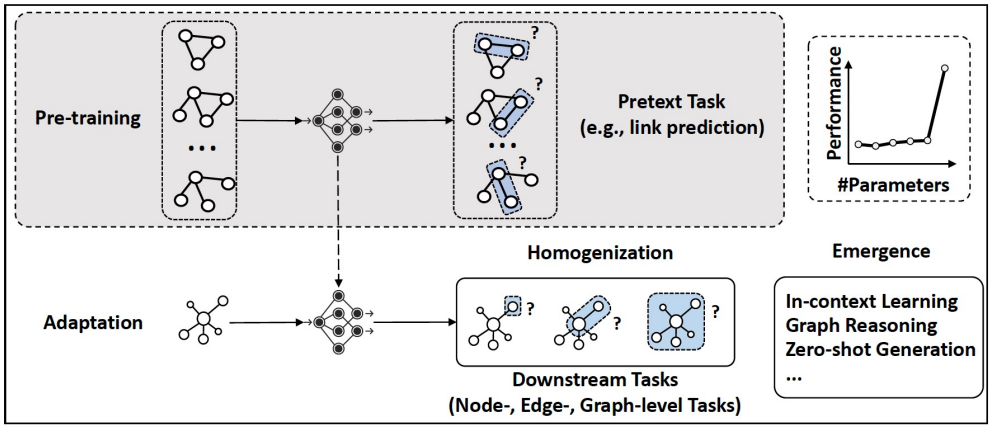
(iii) Fast Adaptation for New Graphs and Tasks, which offers fast designs for adapting the pretrained GFM to new graphs and tasks.

Broader Impact:

This project will result in an efficient computation system for GFM that serves as a foundational tool for fellow science and engineering practitioners from academia, national laboratories and industry. With a commitment to helping K-12, undergraduate, female, and URM populations in the STEM field through the proposed investment and rewarding education plan, this project provides a comprehensive road map to prepare the next-generation of software security professional workers and researchers. This project will contribute to the US national goal to increase participation in science and engineering, which is of paramount value to America's success in addressing global challenges, building a stronger and more diversified workforce and meeting the needs of the global innovation economy. This project will also lead to new courses and redesigned core courses for the PI's home department which is the process of improving the curriculum. To benefit the society at large, the PI will share project data, open source software, and publications with the broader research community.



(a) Deep Graph Learning.



(b) Graph Foundation Models.

Figure 1: Comparison between (a) conventional deep graph learning, and (b) graph foundation model.