



DISSERTATION DEFENSE

Yujun Yan



Towards Generalizable Neural Networks for Graph Applications

Monday, November 7, 2022

12:00-2:00 pm

Virtual – [Zoom](#)

ABSTRACT: Graph-based deep learning has been successful in various industrial settings and applications. However, deep models can hardly achieve generalization, because graphs from different domains may exhibit different properties, and may have significant noise. These challenges limit the usage of graph models in various areas.

In this thesis, I present a variety of theoretical and empirical analyses towards generalizable neural networks for graphs. I consider two types of generalizability for graph neural networks (GNNs): (1) data generalizability, where a graph model has the expressive power to effectively handle various graphs with different properties; (2) size generalizability, where a graph model can learn from small-sized graphs and generalize to larger graphs. In the first part of my thesis, I study the data generalizability problem at the node and subgraph level. At the node level, I analyze when the performance of GNNs degenerates for nodes with different properties (e.g., degree, label distribution of neighboring nodes), and propose effective theory-grounded designs that alleviate the degradation. At the subgraph level, I consider the situation when the data is limited and noisy, and propose the use of clustering to enable GNNs to overcome these issues and find meaningful patterns. In the second part of my thesis, I study the size generalizability problem at the graph level. Specifically, I consider graphs with different sizes and study how to transfer knowledge from small graphs to larger graphs. I first show that the GNNs that are based on spectral properties may suffer from the correlation of spectrum and graph sizes, which limits their ability to generalize over sizes. Then I propose a technique that learns to eliminate the size-related component, which improves the size generalizability of GNNs. Moreover, I study the transformer model, which is related to GNNs but does not depend on the graph spectrum. In this case, I find that the vanilla transformer model fails to generalize to larger sequences and graphs due to its attention mask gradually losing fidelity with larger inputs. Based on my findings, I introduce a learned conditional masking mechanism, which enables strong generalization far outside the model's training range.

CHAIR: Prof. Danai Koutra