



DISSERTATION DEFENSE



Jiong Zhu

Advancing Graph Neural Networks
for Complex Data: A Perspective
Beyond Homophily

Tuesday, June 4, 2024

10:00am – 12:00pm

4901 Beyster

Hybrid – [Zoom](#) Passcode: 766344

ABSTRACT: Graph Neural Networks (GNNs) have demonstrated significant potential in extending the empirical success of deep learning from Euclidean spaces to non-Euclidean, graph-structured data. These models operate on versatile relational networks to extract meaningful representations, enabling a wide range of downstream applications such as friend recommendations, fraud detection, and bioinformatics. A key principle underlying many real-world networks---and often implicitly leveraged by GNN models---is homophily, whereby linked nodes often belong to the same class or have similar features (“birds of a feather flock together”). However, real-world settings also exist where “opposites attract,” resulting in networks characterized by heterophily, where linked nodes are likely from different classes or possess dissimilar features. How do GNNs perform in settings where the homophily is weak? Is the reduced accuracy of GNNs under heterophily related to their limitations in other performance aspects such as robustness, fairness, and scalability?

This dissertation aims to push forward the state-of-the-art in GNNs by addressing these questions. In Part I, I focus on understanding and improving GNN accuracy beyond homophily.

I first examine the limitations of existing GNN models under heterophily for semi-supervised node classification tasks, introducing key design strategies and new methods that significantly enhance learning from the graph structure on such datasets. Furthermore, I extend this analysis to link prediction tasks by formalizing the definitions of non-homophilic link prediction based on feature similarity, analyzing how different link prediction encoders and decoders adapt to varying levels of feature similarity, and introducing designs for improved performance. In Part II, I explore the implications of heterophily on other critical GNN research objectives beyond accuracy, including adversarial robustness, algorithmic fairness, and distributed scalability. My findings reveal that addressing heterophily not only enhances GNN accuracy but also improves their robustness and fairness, making them more suitable for deployment in complex real-world applications. Additionally, I show that heterophily is key to streamlining distributed training of GNNs on massive graphs, reducing communication overhead and improving efficiency and scalability.

Overall, this dissertation advances the field of GNNs by moving beyond homophily, offering new insights and methodologies for handling heterophilous graph datasets, and demonstrating the broader benefits of these advancements in terms of robustness, fairness, and scalability.

CHAIR: Prof. Danai Koutra