Statement of Purpose

Yuxin Yang

Machine learning on graphs provides powerful methods with many successful applications. However, a lack of theoretical support sometimes hinders our confidence in them and limits the usage of applications. From my experience in graph representation learning, I have become convinced that combining machine learning with well-founded theories is necessary for real-world applications. I am determined to apply to the Ph.D. program in Information Systems and Management at Heinz College, where I wish to further research graph mining with scalability and expressiveness. I hope these works will lead to broader applications in social networks, knowledge graphs and biomedicine.

My first research experience was participating in a project on light-field microscopy under the supervision of Prof. Jingtao Fan from the Department of Automation, Tsinghua University. I maintained the controlling software in this project and developed it to meet higher demands. I utilized CMake, C++, and CUDA programming to make many improvements, including buffered image saving, high-speed camera control, drop-frame handling and fast image rendering. Our work was accepted by *Nature Protocols* (2022). During this process, I became a proficient coder with an awareness of algorithmic efficiency. Moreover, to understand the principles of light-field microscopy, I became familiar with signal processing and Fourier Optics.

From February to June in 2022, I applied my knowledge of signal processing to non-Euclidean domains such as graphs. I joined Prof. Muhan Zhang's group to lead a project on adaptive graph neural networks. We found that local disparity in graphs has long been overlooked. I was driven by the interconnection between graph neural networks (GNNs) and graph signal processing to develop GNNs as adaptive graph filters. To draw information from different regions most flexibly, I proposed a GNN that could adapt to the most suitable aggregation structure for every node. GNN often concentrates locally on a graph, but graph signal processing has a global domain. I innovatively employed a global embedding-generating method to encode the graph structure, which improved performance significantly. Our overall pipeline improved node classification accuracy on graphs by up to 5%. To better demonstrate what our network truly captured, we used a metric to analyze the learning results on a dataset that showed the most significant differences among different regions. We also yielded the most improvement on this particular dataset, leading us to confidently conclude that capturing local patterns of graph datasets helps improve learning tasks on graphs. Our work was pioneering in applying dynamic structured neural networks to graph representation learning. It was a pleasure to introduce it when the work was accepted for oral presentation by ICML-22-DyNN-Workshop in Baltimore. In this project, I became familiar with the whole procedure of research. I also gained the confidence to lead a project by pushing it forward with new ideas. I learned from this experience that effective learning methods should be based on theories, leading me to explore future problems.

Interested in the theoretical support of graph representation learning, I set about understanding invariant and equivariant GNNs in physics tasks and expressive analysis, the Weisfeiler-Lehman graph isomorphism test in analyzing GNN expressiveness, information bottleneck and probabilistic theory in deciding the optimization target. Through this study, I noticed the deficiency in graph theory in proving GNN expressiveness on homophilous and heterophilous graphs. I then had the opportunity to work with Prof. Hanghang Tong and Ph.D. student Zhe Xu from UIUC to build theories in modeling these graph properties and designing powerful graph structure learners. Following previous works, I used a link predictor that captures important common neighbor information. I modeled graph properties into a probabilistic model and mathematically proved that common neighbor information helps our structure learner provide more accurate link prediction results. To constrain time and space complexity, I further proposed a sampling strategy based on empirical observations. Applying the structure learner to learn a better graph before performing learning tasks increased the accuracy of downstream node classification tasks by ~2% on several benchmarks. This project led me to think further about the restricted field of existing analysis, which may not cover the learning problems in the presence of various graph properties, thus motivating me to explore in this direction.

Throughout the first two years of my academic career, I was determined to research deeply in graph representation learning. I believe it is a promising topic because it is a shared challenge in many fields and has many open questions. For example, few learning methods are adaptable to non-homophilous graphs. Besides homophily, graph properties like motifs and density are still under-studied. I am interested in developing theories to analyze graph properties and design graph learning methods. With the establishment of theories and metrics, graph learning methods will be applied with confidence, and we will take a firm step toward transforming data into knowledge.

I am applying to the Ph.D. program in Information Systems and Management at Heinz College because of its high-quality courses and impactful research groups. I wish to take systematic courses about mathematical methods in machine learning. On the other hand, I wish to cooperate with the top-class professors here. Professor Leman Akoglu did interesting and impactful work on graph anomaly detection. I am interested in modeling graph anomalies by discovering new traits with more expressive GNNs. I am also interested in Professor Yiming Yang's research on knowledge graphs. Her recent work JAKET enables inductive inference in an elegant way.

As I am determined to work as a research scientist in the future, I look forward to the possibility of continuing my study and research in the vibrant and stimulating community of Heinz College.