Statement of Purpose
Xuan Yang

My research interests mainly lie in learning and mining on highly-structured data as well as machine learning, which aims to discover useful information inside graph-structured data or infer the hidden relational information among temporal data and text data. Graph-structured data, which stores relationships amongst data, is ubiquitous in both science and industry. For instance, social network, a common type of graph data, can be leveraged to explore the potential interactions between two individuals and to understand the generation or evolution process of social relationships. Despite its ubiquity, applying machine learning methods to graph-structured data is challenging because of the complex topological structure and non-i.i.d. nature of graph structures, which I find highly motivating and attractive from an academic perspective. During the past years, the increasing research attention on highly-structured data also brings out new research problems in the field (e.g., explainability, fairness) and motivates novel real-world applications where graphs structures can be applied (e.g., knowledge graphs, causal inference, drug discovery). Thus, I began to actively think about how to exploit my research skills to develop novel machine learning techniques benefit the real world. I am devoted to conducting in-depth research on highly-structured data mining and learning to promote the frontiers of the field.

To date, my research experience during my undergraduate and master studies have laid a solid foundation for me to pursue a Ph.D. degree. My interests in machine learning can be traced back to my undergraduate study. Attending National University of Singapore (NUS) and Stanford University in summer 2018 and winter 2019, respectively, I participated as a research intern in machine learning projects. Throughout these projects, I have found out the ability of computer technology, especially machine learning, in making significant societal contributions. I was particularly impressed by the NUS project I was involved. In this project, we used deep learning models to enable automatic food intake recording from photos uploaded by the users, which could further help users to live a healthier lifestyle. Such experiences further motivated me to conduct research in machine learning related areas.

After obtaining my bachelor's degree, I joined the DCD Lab at Zhejiang University to pursue my Master's degree under the supervision of Dr. Yang Yang. In my first year, I joined Alibaba as a research intern working on improving its recommender system. During my internship, I found out that its recommender system tended to be unfair for items with different popularity, which led to the well-known "Matthew Effect". To alleviate the problem, it was critical to identify the promising items in the long-tail items (i.e., rising stars). However, challenges lay on sharp fluctuations in the sales trend of rising stars, which cannot be simply solved by existing sales prediction models. Having observed that there was rich user interest information in the user sharing behaviors, I proposed to improve the prediction of 'rising stars' with the help of social networks among users. This further raised the problem of how to quantify the interaction information contained in the graph structure and how to incorporate the quantitative measurement into the temporal data. To solve this, I proposed a novel framework named RiseNet that adopts a coupled mechanism to capture the dynamic interplay between time-series data and graphs, as well as a specially designed GNN-based module to quantify the user interest in dynamic graphs. This work is published at TKDE, which is a top-tier data mining journal, and currently being deployed on Taobao (the biggest e-commerce website in China) for item recommendation.

After getting my very first first-authored paper published and making real-world impacts with my research, I became more enthusiastic by machine learning. After outbreaks of COVID-19, in collaboration with Dr. Chenhao Tan from the University of Chicago, I conducted a project studying the how our society recovered from the lockdown influence. Different from my previous research experience, all I had at first was a bunch of noisy electricity data. And we struggled to understand how the recovery process due to the bad data quality (e.g., noisy and incomplete data) and lack of existing ways to verify our conclusion formally. Rather than being discouraged by these challenges, I became more active in reading related articles and analyzing data with different strategies to find information that might potentially benefit our research. After that, I proposed to decompose the recovery of entire society as the recovery of major industries. Furthermore, I developed a graph learning model that aided the modeling of recovery trends. In this research experience, though I faced many difficulties (e.g., data privacy and different writing style), these difficulties benefit me a lot in my research: they broadened my thinking about data analysis, and made me realize the importance of trustworthiness of the deep learning models.

My most recent research studies multi-modal learning with graphs. I found that graphs are often coupled with multi-modal data. Thus, it is critical to inject the relational information on graphs to other modalities to achieve better learning performance, e.g., injecting social interactions between individuals to user posts and uploaded photos for better recommendations. In collaboration with Dr. Xiang Ren from University of Southern California (USC), we study how to fuse graphs with multi-modal data. It is worth noting that I created the first social media multi-modal dataset with graph information to facilitate future research in this direction.

All my research experiences collectively shape my research skills and motivate me to pursue a Ph.D. degree. With my experiences in graph mining and learning , alongside research skills gained through these experiences, I hope to continue this line of research to make graph learning techniques useful and trustworthy to be reliably applied to real-world problems. I find that UCLA has an immense number of experts that would all be very instructive for my interested research areas. Specifically, I would be excited to work with Dr. Yizhou Sun, Dr. Baharan Mirzasoleiman, Dr. Wei Wang. Dr. Yizhou Sun's works on graph learning attract me a lot. Dr. Baharan Mirzasoleiman's study on improving robustness and data-efficient training of machine learning models is very inspiring to me. I am also highly interested in Dr. Wei Wang's research on intelligent computing for large-scale data. I very much hope to be considered for Ph.D. study in the next academic year.