NJIT | INSTITUTE FOR DATA SCIENCE

Department of Data Science

Data Science Seminar Series

Graph Embedding with Uncertainty Quantification for Diverse Applications

Mengjia Xu, Ph.D.
Post-Doctoral Associate
Massachusetts Institute of Technology
Brown University

Date:	Wednesday, February 2nd, 2022
Time:	2:30 PM – 3:30 PM EST
Location:	Zoom Virtual Room
Web Link:	Zoom Meeting Room Link

Graphs are a universal language for describing and modeling complex, large-scale systems. Network data are ubiquitous across diverse application fields such as social networks, bank-asset networks, brain networks, molecular networks in drug discovery, protein-protein interaction networks in genetics, etc. Graph analytics has become an exciting and impactful research area in recent years. However, traditional methods are largely based on extracting handcrafted graph topological features directly from the adjacency matrices. When we apply these methods to large-scale network analysis in industrial systems, they may suffer from high computational cost and excessive memory requirements due to the challenging and inevitable high-dimensionality and emerging heterogeneous characteristics of the original networks. Recently, neural network-based graph embedding techniques have shown a remarkable capacity to convert high-dimensional sparse graphs into low-dimensional, dense, and continuous vector spaces, where graph structure properties are maximally preserved. The nonlinear and highly informative graph embeddings (or features) generated in the latent space can be readily used to address different downstream graph analytic tasks (e.g., node classification, link prediction, community detection, and recommender systems). The main aim of graph embedding methods is to encode nodes into a latent vector space based on neural networks, i.e., pack every node's properties into a vector with a smaller dimension. However, all recent methods in dynamic graph embedding focus on learning each node as a deterministic “vector”, thus, they cannot capture the important graph node hierarchical structure property and the crucial evolving node uncertainties in the latent space. I will present two graph embedding approaches with effective *uncertainty quantification*. The first one, Graph2Gauss (G2G), is a general *density function embedding*, while the second one is a *hyperbolic embedding*, more appropriate for hierarchical data such as knowledge graphs. The uncertainty is captured by the learned multivariate Gaussians assigned to each graph node. I will present applications of G2G for the static functional connectome in Alzheimer’s disease using fMRI and MEG data. I will also present more recent developments on extending these methods for dynamic graphs, and demonstrate scalability and state-of-the-art accuracy in eight different benchmarks from graphs involving 100 nodes to large-scale graphs of about 100,000 nodes.

Dr. Mengjia Xu is currently a postdoc associate working at McGovern Institute for Brain Research, Center for Brains, Minds and Machines (CBMM) at MIT with Prof. Tomaso Poggio, and with Prof. George Karniadakis from the Division of Applied Mathematics at Brown University. Before joining MIT and Brown, she completed her PhD degree at the Department of Computer Science, Northeastern University (China) and two-year visiting PhD at Brown University. Her current research focuses on the optimization of deep neural networks, spatio-temporal graph representation learning, uncertainty quantification, and biomedical imaging data analysis for diverse applications.