Towards Trustworthy Graph Machine Learning

Cheng-Te Li

Institute of Data Science, National Cheng Kung University, Taiwan

Abstract

Graph Machine Learning (GraphML) is an emerging field that has shown great promise in solving many real-world problems, such as social network analysis, recommendation systems, and precision medicine. However, as with any machine learning application, GraphML also faces several challenges related to robustness and privacy. Robustness issues arise due to the vulnerability of GML models to adversarial attacks, which can lead to biased predictions and incorrect decisions. In addition, GraphML models are often trained on datasets that may contain noisy or incomplete information, which can further impact their robustness. Privacy concerns are also a significant challenge in GraphML, as the data used to train these models often contains sensitive information that must be protected. GraphML models may inadvertently leak information about individuals or organizations, which can lead to significant privacy breaches. In this talk, we will discuss our solutions to developing trustworthy GML methods that are robust and privacy-preserving. First, we will introduce adversarial defenses against Graph Neural Network-based privacy attacks, and present a graph perturbation-based approach, NetFense, to achieve the goal. NetFense can simultaneously keep graph data unnoticeability (i.e., having limited changes on the graph structure), maintain the prediction confidence of targeted label classification (i.e., preserving data utility), and reduce the prediction confidence of private label classification (i.e., protecting the privacy of nodes). Second, we aim at building a holistically robust GNN against four different types of graph noise, including adversarial attacks, edge sparsity, noisy labels, and high heterophily, in the presence of label scarcity. We will present our novel GNN framework, Holistic Robust Graph Neural Networks (HRGNN) that can fulfill the goal. Last, we will also highlight some key research directions of trustworthy GraphML.

Keywords: Graph Machine Learning; Graph Neural Networks; Trustworthy AI; Node Classification; Model Robustness; Adversarial Attacks; Data Noise; Label Scarcity