Deep Learning on Graphs: A Data-Centric Exploration

Wei Jin

Emory University
wei.jin@emory.edu

As powerful deep learning tools for graph data, graph neural networks (GNNs) have demonstrated remarkable performance in various applications. Despite the significant accomplishments of GNNs, recent studies have highlighted that their efficiency and effectiveness face significant challenges, which are fundamentally linked to data. To address the data-related issues, I developed a set of techniques to optimize the graph datasets for training GNNs. My research contributions can be summarized as follows.

Securing GNNs from a Data-Centric View. Despite the prosperity of GNNs, they have also exposed critical vulnerabilities as summarized in my survey paper (Jin et al. 2021). Concretely, an attacker can inject a small perturbation to the input graph, which is referred to as *adversarial attack*, and mislead the GNN model into giving wrong predictions. The lack of robustness can lead to severe consequences for safety-critical applications such as financial systems and risk management. I proposed ProGNN (Jin et al. 2020) for learning to optimize the graph to counteract adversarial attacks, which iteratively optimizes the noisy graph by restoring the violated graph properties. Different from ProGNN which targets defending training-time attacks, I proposed GTrans (Jin et al. 2023) to counteract test-time attacks by transforming the graph structure as well as node features to remove adversarial patterns. To deepen our understanding and immensely foster the research field of robust graph representation learning, I developed DeepRobust (Li et al. 2021), a comprehensive Python toolkit for generating adversarial attacks and building robust models. My other research works (Dai et al. 2022; Liu et al. 2021) in this direction have also tremendously promoted the reliability of GNNs and facilitated safety-critical applications.

Scaling up GNNs from a Data-Centric View. As large-scale graphs are prevalent in real-world scenarios (often on the scale of millions of nodes and edges), it poses significant challenges in storing datasets and training GNNs on them. To tackle these issues, I proposed GCond (Jin et al. 2022b), the first approach for synthesizing a small but informative graph with which we can train machine learning models sufficiently and efficiently. Remarkably, we are able to reduce the graph size by 99.9% while approximating the original test accuracy by 99.8%. However, its condensation process in GCond is computationally expensive as it requires multiple steps of gradient matching. Further, I proposed DosCond (Jin et al. 2022a) to address this limitation and facilitate graph condensation in large datasets by performing gradient matching for one single step.

GNNs for Interdisciplinary Research. I have integrated data-centric approaches with model-centric approaches to benefit interdisciplinary research, where I constructed graph data from the raw data in other fields and carefully designed graph ML models for them, including single-cell analysis (Wen et al. 2022), traffic forecasting (Wang et al. 2020), and e-commerce platform (Fan et al. 2022).

References

Dai, E.; Jin, W.; Liu, H.; and Wang, S. 2022. Towards robust graph neural networks for noisy graphs with sparse labels. In *WSDM*.

Fan, W.; Liu, X.; Jin, W.; Zhao, X.; Tang, J.; and Li, Q. 2022. Graph Trend Filtering Networks for Recommendation. In *SIGIR*.

Jin, W.; Li, Y.; Xu, H.; Wang, Y.; Ji, S.; Aggarwal, C.; and Tang, J. 2021. Adversarial attacks and defenses on graphs. *ACM SIGKDD Explorations Newsletter*, (2).

Jin, W.; Ma, Y.; Liu, X.; Tang, X.; Wang, S.; and Tang, J. 2020. Graph Structure Learning for Robust Graph Neural Networks. In *KDD*.

Jin, W.; Tang, X.; Jiang, H.; Li, Z.; Zhang, D.; Tang, J.; and Yin, B. 2022a. Condensing graphs via one-step gradient matching. In *KDD*.

Jin, W.; Zhao, L.; Zhang, S.; Liu, Y.; Tang, J.; and Shah, N. 2022b. Graph Condensation for Graph Neural Networks. In *International Conference on Learning Representations*.

Jin, W.; Zhao, T.; Ding, J.; Liu, Y.; Tang, J.; and Shah, N. 2023. Empowering Graph Representation Learning with Test-Time Graph Transformation. In *ICLR*.

Li, Y.; Jin, W.; Xu, H.; and Tang, J. 2021. Deeprobust: a platform for adversarial attacks and defenses. In *AAAI*.

Liu, X.; Jin, W.; Ma, Y.; Li, Y.; Liu, H.; Wang, Y.; Yan, M.; and Tang, J. 2021. Elastic Graph Neural Networks. In *ICML*.

Wang, X.; Ma, Y.; Wang, Y.; Jin, W.; Wang, X.; Tang, J.; Jia, C.; and Yu, J. 2020. Traffic Flow Prediction via Spatial Temporal Graph Neural Network. In *WWW*.

Wen, H.; Ding, J.; Jin, W.; Wang, Y.; Xie, Y.; and Tang, J. 2022. Graph neural networks for multimodal single-cell data integration. In *KDD*.