Graph Neural Networks for Link Prediction

Alina Lazar
Youngstown State University
Youngstown, OH, USA
alazar@ysu.edu

Abstract

Graph Neural Networks (GNNs) belong to a class of deep learning methods that are specialized for extracting critical information and making accurate predictions on graph representations. Researchers have been striving to adapt neural networks to process graph data for over a decade. GNNs have found practical applications in various fields, including physics simulations, object detection, and recommendation systems. Predicting missing links in graphs is a crucial problem in various scientific fields because real-world graphs are frequently incompletely observed. This task, also known as link prediction, aims to predict the existence or absence of links in a graph. This tutorial is designed for researchers who have no prior experience with GNNs and will provide an overview of the link prediction task. In addition, we will discuss further reading, applications, and the most commonly used software packages and frameworks.

Tutorial Description

Graph-structured data exist everywhere in the real world. Almost any problem can be modeled using graph representations. The range of practical applications is vast, from social networks to molecular structures and particle tracking. In this context, designing and evaluating advanced learning methods on graph-structured data is becoming important. GNNs [1], that extend the well-known deep neural network models to graph representations, offer researchers a new way to learn graph representations at the node, edge, and graph levels. For this tutorial, we will focus on the link prediction task.

The link prediction task [3] can be described as follows: given two nodes' internal representations in a GNN, the model can be utilized to predict how likely two nodes will be connected in the future. The link prediction framework works in three steps:

1. Subgraph extraction (edge embedding)
2. Node feature matrix construction
3. GNN learning

Depending on how the edge embedding is performed the GNN link prediction approaches can be divided into two categories: node2link and subgraph2link [2].

This tutorial will cover relevant GNN-related topics, including the basics of learning graph-structured data, graph embeddings, attention networks, aggregation functions, and examples of applications (node classification, predicting missing links, detecting communities, and graph matching). GNNs have achieved impressive performance on relatively small graph datasets for these applications. Unfortunately, most real-world problems rely on large graphs that do not fit into the available GPU memory of current hardware systems. We will also discuss ways to design, evaluate and scale GNN training and inference methods.

Outline

Part 1:

- Graphs and graph-structured data (15 min)
- Node vs edge level tasks (10 min)
- Graph embeddings (15 min)
- Graph convolutional neural networks (GCNNs) (10 min)

Part 2:

- Graph Attention Networks (10 min)
- Aggregation Functions in GNNs (10 min)
- Edge-level GNN Applications (10 min)
- Sampling for Link Prediction (5 min)

References

[1] Scarselli, F.; Gori, M.; Tsoi, A. C.; Hagenbuchner, M.; and Monfardini, G. 2008. The graph neural network model. *IEEE transactions on neural networks* 20(1):61–80.

[2] Tan, Q.; Zhang, X.; Liu, N.; Zha, D.; Li, L.; Chen, R.; Choi, S.-H.; and Hu, X. 2023. Bring your own view: Graph neural networks for link prediction with personalized subgraph selection. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, 625–633.

[3] Zhang, M., and Chen, Y. 2018. Link prediction based on graph neural networks. *Advances in neural information processing systems* 31.