Lecture 1: Introduction Machine Learning with Graphs

Graphs are a general language for describing and analyzing entities with relations/interactions. Examples of graphs include computer networks, disease pathways, and food webs. Also, many types of data are graphs (databases are graphs too!). The main question we ask in this class is how do we take advantage of the relational structure of data/graphs for better predictions. Currently, many of our deep learning tools are made for simple sequences and grids but data/networks are much more complex. Networks have arbitrary size and complex topological structure, no fixed node ordering or reference point, and dynamic and multimodal features.

Heterogeneous Graphs: Defined by G = (V, E, R, T)

- V - nodes with nodes types
- E - edges with relation types
- $T(v_i)$ - node type
- r - relation type

Nodes and edges have attributes/features.

When working with graphs, we must ask ourselves what our nodes and edges are. Choice of the proper network representation of a given domain/problem determines our ability to use networks successfully.

Directed vs Undirected Graphs: The difference lies in the links of the nodes. In an undirected graph, there is no direction of the edges, whereas in a directed graph there is a direction. When looking at the difference of these graphs, we must consider weights, properties, types, and attributes.

Bipartite Graph: a graph whose nodes can be divided into two disjoint sets U an V such that every link connects a node in U to one in V. That is, these are independent sets.

Applications of Graph ML:

- Node-level tasks

Task: to characterize the structure and position of a node in the network. Examples include protein folding, where for each node, its 3D coordinates are predicted.

- Edge-level prediction

Task: make a prediction for a pair nodes. This can be either trying to find missing links or finding new links as time progresses. Examples include graph-based recommender systems and drug side effects.

- Sub-graph-level prediction or Graph-level prediction

Task: Predict for an entire subgraph or graph. Examples include traffic prediction, drug discovery, physics simulation, and weather forecasting.