Lab 6. Graph ML Research Project Proposal

Identify a graph machine learning application or area of research that you are interested in for your project. Select an idea that you believe is interesting and potentially impactful.

For you project:

- Identify the problem to be solved. Examples, “apply graph machine learning based recommendation algorithms to patient-record data to identify similar patients”, “ implement a GCN from scratch to experiment with different aggregation and update methods”, “create a knowledge graph from Twitter data.”
- Explain why this problem interests you.
- Identify what methods have been used to solve this or related problems in the past.
- Identify at least one dataset you can use. If your project is more experimental, you can use synthetic data.
- What do you expect to learn from completing this project?
- It’s Ok to leverage existing work, but you need to make a contribution, i.e., create something new, apply an existing method on a distinctly different dataset, or perform exploratory research.

Expecting about 1 page max in bulleted format as described above.

Pytorch built-in datasets:
https://pytorch-geometric.readthedocs.io/en/latest/modules/datasets.html

The following pages include some project ideas and references.

A summary overview of GNNs Applications

Graph convolutional network/ graph attention network

A classic application of GNNs in NLP is Text Classification.

GNNs utilize the inter-relations of documents or words to infer document labels. GCN and GAT models are applied to solve this task. They convert text to graph-of-words, and then use graph convolution operations to convolve the word graph. The graph-of-words representation of texts has the advantage of capturing non-consecutive and long-distance semantics

Neural machine translation

Graph convolutional network/ gated graph neural network

The neural machine translation (NMT) is considered a sequence-to-sequence task. One of GNN's common applications is to incorporate semantic information into the NMT task.

Relation extraction

Graph LSTM/ graph convolutional network

Relation Extraction is the task of extracting semantic relations from the text, which usually occur between two or more entities. Traditional systems treat this task as a pipeline of two separated tasks, i.e., named entity recognition (NER) and relation extraction, but new studies show that end-to-end modeling of entity and relation is important for high performance since relations interact closely with entity information

Image classification

Graph convolutional network/ gated graph neural network

Image classification is a basic computer vision task. Most of the models provide attractive results when given a huge training set of labeled classes. The focus now is towards getting these models to perform well on zero-shot and few-shot learning tasks. For that, GNN appears quite appealing. Knowledge graphs can provide the necessary information to guide the ZSL (Zero-shot learning) task.

Object detection, Interaction detection, Region classification, Semantic segmentation

Graph attention network, Graph neural network, Graph CNN, LSTM, gated graph neural network, graph CNN, graph neural network

There are other applications of computer vision tasks like object detection, interaction detection, and region classification. In object detection, GNNs are used to calculate Region of interest features; in interaction detection, GNN is message-passing tools between humans and objects; in region classification, GNNs perform reasoning on graphs that connect regions and classes.

Physics

Graph neural network/ graph networks

Modeling real-world physical systems is one of the most basic aspects of understanding human intelligence. By representing objects as nodes and relations as edges, we can perform GNN-based reasoning about objects, relations, and physics in an effective way. Interaction networks can be trained to reason about the interactions of objects in a complex physical system. It can make predictions and inferences about various system properties in domains such as collision dynamics

Molecular fingerprints
Graph convolutional network
Molecular fingerprints are feature vectors that represent molecules. ML models predict the properties of a new molecule by learning from example molecules that use fixed-length fingerprints as inputs. GNNs can replace the traditional means that give a fixed encoding of the molecule to allow the generation of differentiable fingerprints adapted to the task for which they are required

Protein interface prediction
Graph convolutional network
This is a challenging problem with important applications in drug discovery. At a molecular level, the edges can be the bonds between atoms in a molecule or interactions between amino-acid residues in a protein. On a large scale, graphs can represent interactions between more complex structures such as proteins, mRNA, or metabolites

Combinatorial optimization
Graph convolutional network/ graph neural network/ graph attention network
Combinatorial optimization (CO) is a topic that consists of finding an optimal object from a finite set of objects. It is the base of many important applications in finance, logistics, energy, science, and hardware design. Most CO problems are formulated with graphs. In a recent work by DeepMind and Google, graph nets are used for two key subtasks involved in the MILP solver: joint variable assignment and bounding the objective value. Their neural network approach is faster than existing solvers on big datasets

Graph generation
Graph convolutional network/ graph neural network/ LSTM /RNN/ relational-GCN
Generative models for real-world graphs have drawn significant attention for their important applications including modeling social interactions, discovering new chemical structures, and constructing knowledge graphs. The GNN based model learns node embeddings for each graph independently and matches them using attention mechanisms. This method offers good performance compared to standard relaxation-based techniques

Projects with datasets:

Fraud Detection in Transaction Graphs

Graphs:
Nodes: Financial users (customers, banks)
Edges: Transaction (money and amount sent)
Tasks:
Edge classification - predict which edges are fraudulent. Metric: Hits@50
Example model(s): See https://github.com/safe-graph/graph-fraud-detection-papers
Datasets: Bitcoin Fraud Dataset (only use labeled data!)

Knowledge graph

Graphs:
Nodes: Entities
Edges: Knowledge triples
Tasks:
Predicting missing triples. Metric: Mean Reciprocal Rank (MRR)
Model(s): TransE, DistMult, ComplEx, RotatE
Datasets: FB15k-237, WN18RR

Molecule classification

Graphs:
Nodes: Atoms
Edges: Bonds
Tasks:
Predicting properties of molecules. Metric: ROC-AUC
Model(s): See the leaderboard
Datasets: ogbl-molhiv

Drug-Drug Interaction Network

Graphs:
Nodes: FDA-approved or experimental drug
Edges: interactions between drugs (joint effect of taking the two drugs together)
Tasks:
Predict drug-drug interactions - Metric: Hits@K
Example model(s): See the leaderboard
Datasets: ogbl-ddi

Recommender systems

Graphs:
Nodes: Users, items
Edges: User-item interactions
Tasks:
Predicting the edge ratings. Metric: RMSE
Predicting edge existence. Metric: Recall@K
Model(s): LightGCN
Public datasets: Movielens, Recsys repository

Protein-Protein Interaction Networks

Graphs:
Nodes: Gene nodes
Edges: Interaction between gene nodes
Tasks:
Node classification - protein function prediction. Metric: Classification accuracy
Example model(s): See methods on OGB node classification leaderboard
Datasets: https://snap.stanford.edu/biodata/datasets/10013/10013-PPT-Ohmnet.html

Friend recommendation

Graphs: social network
Nodes: users
Edges: potentially heterogeneous -- friend, follow, reply to, message, like, etc.
Tasks:
Recommending/ranking new friends for user -- metrics: Hits@K, NDCG@K, MRR
Example model(s): GraFRank (paper, GitHub)
Datasets: Facebook, Google+, Twitter

Ideas and references:

The following paper provides a review of available graph datasets and applications.
http://graphkernels.cs.tu-dortmund.de/

Geometric Deep Learning. Organized by a scientist from Cambridge. More challenging material.
https://geometricdeeplearning.com/

Stanford Graduate Projects from 2019. Stanford has a graduate level class in graph learning. Here are some past projects.
http://snap.stanford.edu/class/cs224w-2019/projects.html

Interesting graph neural network applications:
https://revolutionized.com/graph-neural-network/