Project Topics

Below is a list of possible project topics. Some of these are open-ended, meaning that you are required to come up with a new algorithm or model, and formulate it yourselves. Such projects may require more effort, but they will also be graded based on the effort, as well as the final result. Others are more straight-forward, you would need to obtain a complex dataset and apply algorithms on this dataset. There are also more theoretical projects, and more practical ones, so you can pick depending on your preference.

You will also have to present in class one paper related to your project. For some of the projects, one or more papers are suggested. You can present one or two of the suggested papers, or one or two relevant papers of your choice.

Papers also vary in difficulty and scope. For experimental papers, that just report results of experimental studies, we expect that you just present and explain the main findings. Since such papers require less effort, you will be asked to present two such papers.

You also need to create a GitHub page for the project (including the final report and dataset used).

Projects should be done **in teams of at most two** students.

Deliverables and Timeline:

Wednesday, 20/12/2023	A one-page project proposal outlining what you plan to do. This should include the topic (and papers) of your presentation. A 15' presentation of the project proposal
Wednesday, 10/1/2024 Wednesday, 17/1/2024	Paper presentations 1 Paper presentations 2
Wednesday, 7/2/2024	Submit GitHub page: ○ Source code of the project ○ Datasets used ○ Project report

Topic 1

Fairness of diffusion in real social networks

Project:

The main goal is to consider the problem of fairness of information diffusion in a real social network. You need to define and evaluate at least two notions of diffusion fairness and evaluate whether specific networks are fair based on your definition. Below are some suggested notions of diffusion fairness and how to evaluate it.

1. For randomly selected seed nodes compute the fraction of activated nodes for each group and test if there is demographic fairness. Repeat the measurements with the nodes with the highest Pagerank.
2. Consider as seed nodes, randomly selected nodes from the different groups and compare the influence they have as seeds (number of activated nodes). Repeat with the nodes with the highest Pagerank per group.
3. Run the Greedy algorithm for different values of k and measure the representation of the groups in the selected seed set.

Other notions of diffusion fairness can be found in the papers suggested below. Perform experiments on 4 different real datasets and report your results. Propose a mitigation strategy to guarantee that there is no unfairness for some of the cases above.

Papers:

Tsang et al., Group-Fairness in Influence Maximization.

Stoica et al., Seeding Network Influence in Biased Networks and the Benefits of Diversity

Topic 2

Link Recommendation for reducing polarization.

Project

In the paper by Matakos et al., [1] a metric is defined for measuring the polarization in a social network with opinions. Propose algorithms for the problem of suggesting links to reduce the polarization metric. The recommendations should take into account the probability of a recommendation to be accepted.

Papers:

[1] Antonis Matakos, Evimaria Terzi, Panayiotis Tsaparas: *Measuring and moderating opinion polarization in social networks.* Data Min. Knowl. Discov. 31(5): 1480-1505 (2017)

[2] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2017. *Reducing Controversy by Connecting Opposing Views*. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining (WSDM '17). ACM, New York, NY, USA, 81-90.

Topic 3

Using fair random walks in graph embeddings

Project

When nodes in a network belong to different groups (e.g., female/male), we would like all groups to be fairly represented in the embeddings. Previous work focused on the node2vec embedding and proposed a modified fair walk to achieve equal representation of groups in the produced embeddings [2]. In this project, you will replace random walk in node2vec by the residual fair random walk proposed by Tsioutsiouliklis et al [1]. You will evaluate the produced embeddings for the link recommendation problem.

Papers:

[1] Sotiris Tsioutsiouliklis, Evaggelia Pitoura, Panayiotis Tsaparas, Ilias Kleftakis, Nikos Mamoulis: *Fairness-Aware PageRank*. WWW 2021: 3815-3826

[2] Tahleen A. Rahman, Bartlomiej Surma, Michael Backes, Yang Zhang*: Fairwalk: Towards Fair Graph Embedding.* IJCAI 2019: 3289-3295

Topic 4

Fair community detection

Project

In this project, you will study community detection in terms of fairness. For the definition of fairness, we will use the definition from [2], where given groups of nodes (where groups are formed based on some protected attribute (e.g., gender, race)), we ask that each group is equally represented in each community. For example, when the protected attribute is gender, we may ask for an equal representation of women and men in each community. You will study the fairness of at least 3 of the community algorithms that were presented in class. For the evaluation, you need to use at least 4 networks and synthetic datasets following the synthetic data generation in [2]. In addition, derive and evaluate a post-processing method that tries to achieve fairness with minimum changes from the initial clustering.

Papers:

[1] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, Sergei Vassilvitskii: *Fair Clustering Through Fairlets.* NIPS 2017: 5029-5037

[2] Kleindessner et al., Guarantees for Spectral Clustering with Fairness Constraints

Topic 5

Fair node classification

Project

In this project, you will study node classification in terms of fairness. For the definition of fairness, you will use a statistical group-based definition that looks at the classification errors for each group (see, Verma and Rubin, below). You will study the fairness of at least 4 of the embedding methods for the classification problem using at least 5 datasets. In addition, derive and evaluate a post-processing method that tries to achieve fairness with minimum changes from the initial classification.

Paper

Verma, S., Rubin, J.: *Fairness definitions explained*. In: Proceedings of the International Workshop on Software Fairness, FairWare@ICSE 2018, pp. 1–7. ACM (2018)

Topic 6

Information Diffusion and Opinion Formation

In class we saw the information cascade model for item diffusion, and the DeGroot and Friedkin and Johnson models for opinion formation. Define a model that combines the two: It assumes that network users have opinions in the [-1,1] range, and items also have a polarity in the same range. It assumes that propagated items affect the opinions of users, and the opinions of users affect the probability of propagating the items. Perform simulations with this model, and consider the problem of selecting influencers for maximizing either diffusion or opinion.

Paper

Sijing Tu, Stefan Neumann: A Viral Marketing-Based Model for Opinion Dynamics in Online Social Networks. WWW 2022: 1570-1578

Topic 7-11

The following projects are inspired by the CS224W class at Stanford and the goal of the projects is to use Graph Neural Networks in practice. Through your projects, you will gain experience using PyG (PyTorch Geometric), the most popular graph deep learning framework built on top of PyTorch. PyG is suitable for quickly implementing GNN models, with abundant graph models already implemented and datasets ready to load. Furthermore, GraphGym, available as part of PyG, allows a training/evaluation pipeline to be built in a few lines of code.

Below are some application areas for GNNs. It is also acceptable to propose your own application, or datasets.

For each of the following applications, either present the selected paper, one of the suggested ones in the CS224W list, or a paper of your choice.

For the project, you are asked to design and experiment with a baseline GNN approach (not the one presented in the paper). For example, for project B, you can use a classifier as in Assignment 2.

A. Recommender systems

Graphs:

- Nodes: Users, items
- Edges: User-item interactions

Tasks:

- Predicting the edge ratings, for example, user-movie ratings. Metric: RMSE
- Predicting edge existence, for example user-movie selection. Metric: Recall@K

Public datasets: MovieLens, a dataset from Recsys repository

Paper: LightGCN

B. 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

Datasets: Facebook, Google+, Twitter

Paper: GraFRank (paper)

C. Knowledge graphs

Graphs:

- Nodes: Entities
- Edges: Knowledge triples

Tasks:

- Predicting missing triples. Metric: Mean Reciprocal Rank (MRR)

Datasets: FB15k-237, WN18RR

Papers: TransE, DistMult, ComplEx, RotatE

D. Paper citation graphs

Graphs:

- Nodes: Papers

- Edges: Paper citations

Tasks:

- Predicting subject areas of papers. Metric: Classification accuracy

Datasets: ogbn-arxiv

Paper: SIMTEG: A FRUSTRATINGLY SIMPLE APPROACH IMPROVES TEXTUAL GRAPH LEARNING

Taken from here: leaderboard

E. 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

Datasets: Bitcoin Fraud Dataset (only use labeled data!)

Papers:

Ma et al., Towards Graph-level Anomaly Detection via Deep Evolutionary Mapping, KDD 2023

Yu et al., Group-based Fraud Detection Network on e-Commerce Platforms, KDD 2023

Taken from here: https://github.com/safe-graph/graph-fraud-detection-papers

Topic 12

Prepare a survey article on how LLMs can be used for graphs. You need to study at least 6 papers. You may use the following paper as a starting point:

Bahare Fatemi, et. al., Talking like a graph: Encoding Graphs for Large Language Models

Topic 13

Suggest a project of your own, based on what you have seen in the class so far, questions you may have thought of, and things that are related to your research area. In this case you should create a project proposal (initially just a paragraph or an idea) and contact us to discuss it.