Isfahan University of Technology – Faculty of Computer Engineering

Project Proposal for Graph Data Mining Course

Instructor: Dr. Zeinab Maleki

Submission Date: 12/8/25

Students' Information:

Names: Kimia Nikooei - Mobina Massah
Student IDs: 40132933 – 40133803
Emails: k.nikooei@gmail.com - m.massah@ec.iut.ac.ir

Project Title

Graph-Based Recommendation Systems Using Graph Neural Network (GNN) Algorithms.

Abstract

Traditional recommendation methods like collaborative filtering and matrix factorization often struggle to capture deeper patterns in user–item relationships, especially those that span multiple hops in a network. In this project, we propose building a graph-based recommendation system powered by Graph Neural Networks (GNNs). By representing users and items as nodes and their interactions as edges, the model can naturally learn from the structure of the data. We plan to experiment with GCN, GraphSAGE, and recommender-focused models such as NGCF and LightGCN to learn rich graph embeddings. The system will be evaluated on real datasets like MovieLens—and Amazon Books if needed—using ranking metrics such as Recall@K and NDCG. We expect that incorporating graph structure will lead to more accurate and personalized recommendations compared to traditional approaches.

Problem and Motivation

Recommendation systems form a core component of many modern platforms such as Spotify, Amazon, and Netflix, and their quality directly impacts user experience and satisfaction. Traditional approaches—such as collaborative filtering and matrix factorization—primarily capture only first-order relations between users and items. As a result, they are limited in modeling deeper structural patterns such as multi-hop interactions, shared neighborhoods, or indirect similarities. These limitations cause valuable latent information in the data to remain underutilized.

Modeling interactions as a user–item bipartite graph provides a natural framework for analyzing such relationships. Within this structure, key graph mining concepts such as centrality, structural similarity, and especially link prediction become highly relevant, as the fundamental goal of a recommendation system is essentially to predict new edges or future user–item interactions. Graph Neural Networks (GNNs), through their message-passing mechanisms, enable the propagation of multi-hop information across the graph, allowing the model to learn rich and meaningful latent embeddings for users and items—capabilities that traditional methods lack.

Based on this motivation, our central research question is:

"To what extent can GNN-based models improve recommendation quality compared to classical collaborative filtering methods?"

Addressing this question is important both scientifically and practically. In real-world datasets such as MovieLens or Amazon Books, leveraging graph-based representations may lead to more accurate personalization, increased user engagement, and more effective recommendation systems overall.

Objectives

- Objective 1: Convert the original recommender dataset into a user–item bipartite graph and explore its structural characteristics, such as node degrees and connectivity patterns, to better understand the underlying interaction network.

- Objective 2: Build and experiment with GNN-based models—specifically GCN and GraphSAGE—to learn expressive embeddings for users and items using message passing over the graph.

- Objective 3: Compare the performance of GNN-based models with traditional recommendation methods such as matrix factorization.

- Objective 4: Evaluate the models using standard recommender metrics including Recall@K, Precision@K, and NDCG, and analyze the degree of improvement relative to baseline methods.

Related Work

Numerous studies have demonstrated the effectiveness of Graph Neural Networks in recommendation systems. The *Graph Convolutional Matrix Completion (GC-MC)* framework is one of the earliest models to apply GCNs for link prediction in user–item graphs. Another influential work, *Neural Graph Collaborative Filtering (NGCF)*, shows how multi-hop neighborhood information in interaction graphs can be leveraged to propagate embeddings and improve recommendation accuracy. Large-scale industry platforms such as Spotify and Pinterest have also adopted lightweight GraphSAGE-based architectures for scalable recommendation.
Building on these approaches, our project aims to implement simplified versions of these models on publicly available datasets and compare their performance with traditional recommendation methods.

Proposed Methodology

Dataset(s)

We use one of the following datasets (final choice based on size and preprocessing needs):

- MovieLens 100K (943 users, 1682 items, 100K ratings)
- Amazon Books subset (≈200K interactions)

Graph Type: Bipartite, undirected (or weighted if rating values included).

Preprocessing Steps:

- Removing users and items with very sparse interactions
- Converting explicit ratings into implicit feedback (e.g., rating ≥ 4 as a positive interaction)
- Splitting the data into training/validation/test sets using the **leave-one-out** strategy
- Normalizing edge weights when necessary

Techniques and Algorithms

Graph Construction:

- Nodes: users + items
- Edges: interactions (ratings, clicks, purchases)

Algorithms:

- GCN (Graph Convolutional Network)

Learns user and item embeddings through message passing and neighborhood aggregation across multiple layers, enabling the extraction of multi-hop structural patterns in the graph.

- GraphSAGE

Performs neighbor sampling and localized aggregation to improve scalability and support inductive representation learning on larger graphs.

- NGCF / LightGCN (if applicable)

NGCF models higher-order connectivity within the user–item interaction graph, while LightGCN removes nonlinear transformations and weight matrices to provide a simpler and more efficient architecture tailored for recommendation tasks.

- Link Prediction Layer
  dot-product or MLP to predict user–item edges.

Tools/Libraries:

- PyTorch Geometric (PyG) for implementing GNN models
- NetworkX for analyzing the graph structure
- Scikit-learn for baseline models such as matrix factorization

Evaluation Plan

The performance of the models will be measured using standard recommendation metrics such as Recall@K, Precision@K, and NDCG@K. The AUC metric may also be used optionally to evaluate link prediction quality. Baseline methods include matrix factorization and item-based collaborative filtering. The final analysis will compare overall performance, quantify the percentage of improvement, and examine how the graph structure influences the learned embeddings.

Challenges and Resources

GNN models can be computationally expensive on dense graphs, and tuning hyperparameters (layers, sampling, learning rate) may require experimentation. To mitigate this, we begin with smaller datasets (MovieLens 100K) and apply sampling-based GNNs such as GraphSAGE. Course materials, PyG documentation, and relevant papers will support implementation.

References

[1] Wang, X., He, X., Wang, X., Feng, F., & Chua, T.-S. “Neural Graph Collaborative Filtering.” 2019.

[2] He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., & Wang, M. “LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation.” arXiv, 2020.

[3] van den Berg, R., Kipf, T. N., & Welling, M. “Graph Convolutional Matrix Completion.” arXiv, 2017.

[4] Wittenbrink, B. “Spotify Track Neural Recommender System.” Stanford CS224W, 2023.

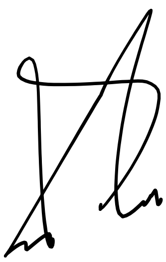
[5] Alexander, B., Chou, J.-P., & Bansal, A. “Implement Your Own Music Recommender with Graph Neural Networks (LightGCN).” Stanford CS224W, 2021.

[6] Haas, L., & Skręta, M. “Alexa, Queue That Banger! Recommending Spotify Playlist Tracks with Graph Machine Learning.” Stanford CS224W, 2022.

Student Signature(s)

Date:1404/8/18

Kimia Nikooei



Mobina Massah