Project Deliverable

Team Information

Team Name: DSC_recomender

Team Members:

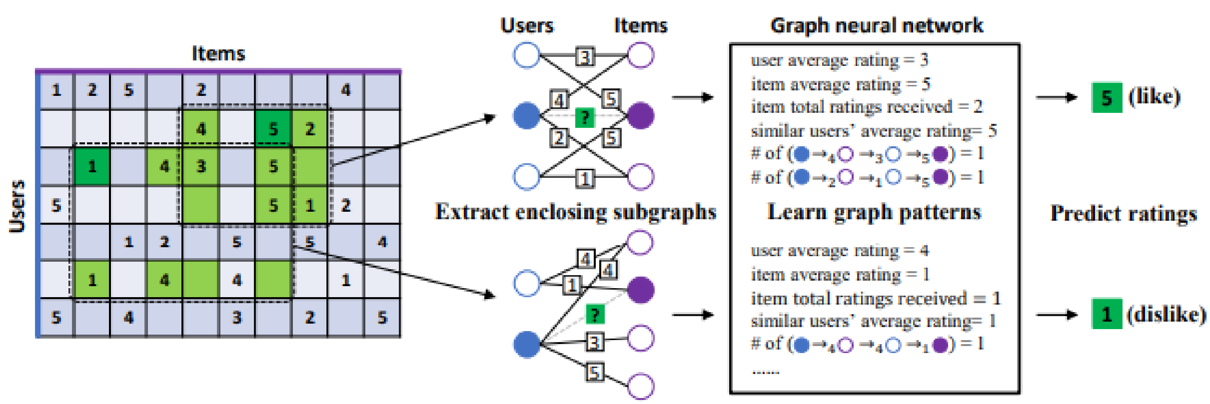
- Hamza Shafee Aldaghstany - h.aldaghstany@innopolis.university
- Yazan Alnakri - y.alnakri@innopolis.university
- Osama Orabi - o.orabi@innopolis.university

Github

Summary: Inductive graph-based matrix completion

The project "Graph Neural Network-Based Movie Recommender System," explores how Graph Neural Networks (GNNs) can enhance movie recommendation systems by leveraging graph-based representations of user-item interactions. Traditional recommendation systems often rely on collaborative filtering or content-based filtering, which can struggle with issues like cold starts or sparsity in data. GNNs address these limitations by modeling users and items as nodes in a graph and their interactions (e.g., ratings or views) as edges.

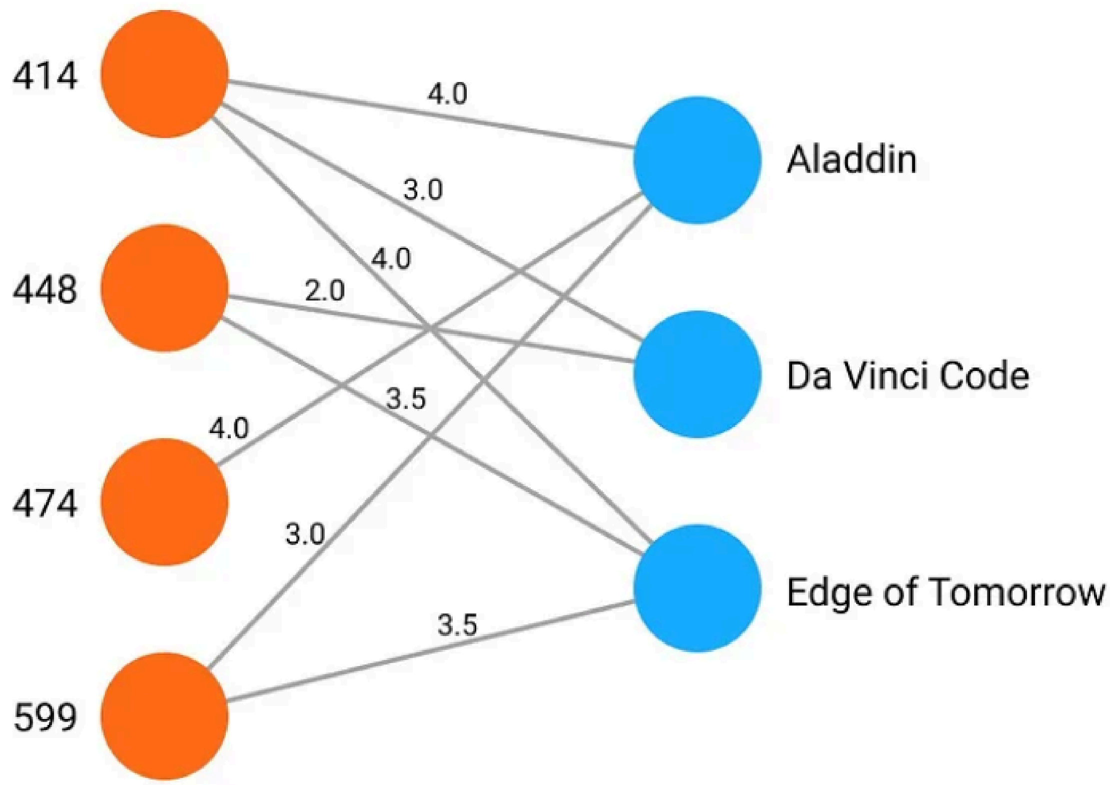
The approach IGMC:



IGMC is an inductive matrix completion model leveraging graph neural networks (GNNs) without relying on side information such as user demographics or item attributes. Unlike traditional matrix factorization approaches, which are transductive and depend on low-dimensional latent embeddings tied to specific rows (users) and columns (items), IGMC achieves inductiveness purely through local subgraph structures.

Key Features:

No Side Information Required: IGMC operates solely on the bipartite graph derived from the rating matrix, avoiding the need for potentially unavailable or hard-to-extract content information.



Subgraph-Based Learning: The model uses GNNs to map local subgraphs around (user, item) pairs to their corresponding ratings, enabling robust generalization.

Inductive Generalization: IGMC can handle unseen users/items during training, provided interactions are available, and can transfer knowledge to new datasets or tasks.

Cross-Dataset Transferability: The model trained on MovieLens data successfully predicts ratings in the Douban dataset, showcasing its practical versatility.

This project explores the potential of IGMC as a standalone inductive model, emphasizing its ability to generalize and adapt without embedding dependency or reliance on global matrix/task-specific information. For more details, refer to the original paper.

Key Highlights:

1. **Graph Representation**:

    - The project uses a bipartite graph to represent users and movies, where edges represent user-item interactions, such as ratings.
    - Additional features, such as user demographics or movie metadata, can be incorporated to enrich the graph.
2. **GNN Model**:

    - GNN architecture “IGMC” is used to aggregate information from neighbors (e.g., users influencing movie embeddings and vice versa).
    - The aggregated node embeddings are optimized to predict ratings or recommend items.
3. **Performance Metrics**:

    - Metrics like Mean Absolute Error (MAE) or Root Mean Square Error (RMSE) are used to evaluate the system's predictive performance compared to baseline methods.
    - We were able to get 0.95 RMSE in two epochs of training
4. **Challenges**:

    - Addressing data sparsity and scalability to large datasets.
    - Fine-tuning hyperparameters to balance learning for users and items.
5. **Results**:

    - GNN-based systems generally outperform traditional methods by capturing complex user-movie relationships and leveraging higher-order graph structures.

This project demonstrates the versatility of GNNs in recommendation tasks and their ability to generalize across domains with graph-structured data.