COMP 4222

Movie Recommendation System Using GNN

YUEN, Tsz Cheuk (20957326)
DENG, King Ho (20960866)

COMP 4222

Table of Content

Datasets

MovieLens 100K

- User ID
- Moive ID
- Rating
  - In a scale of 1 to 5
- Timestamp
  - When the user rated the movie

	userId	movieId	rating	timestamp
0	196	242	3	881250949
1	186	302	3	891717742
2	22	377	1	878887116
3	244	51	2	880606923
4	166	346	1	886397596
...	...	...	...	...
99995	880	476	3	880175444
99996	716	204	5	879795543
99997	276	1090	1	874795795
99998	13	225	2	882399156
99999	12	203	3	879959583

100000 rows × 4 columns

Reference: https://grouplens.org/datasets/movielens/

COMP 4222

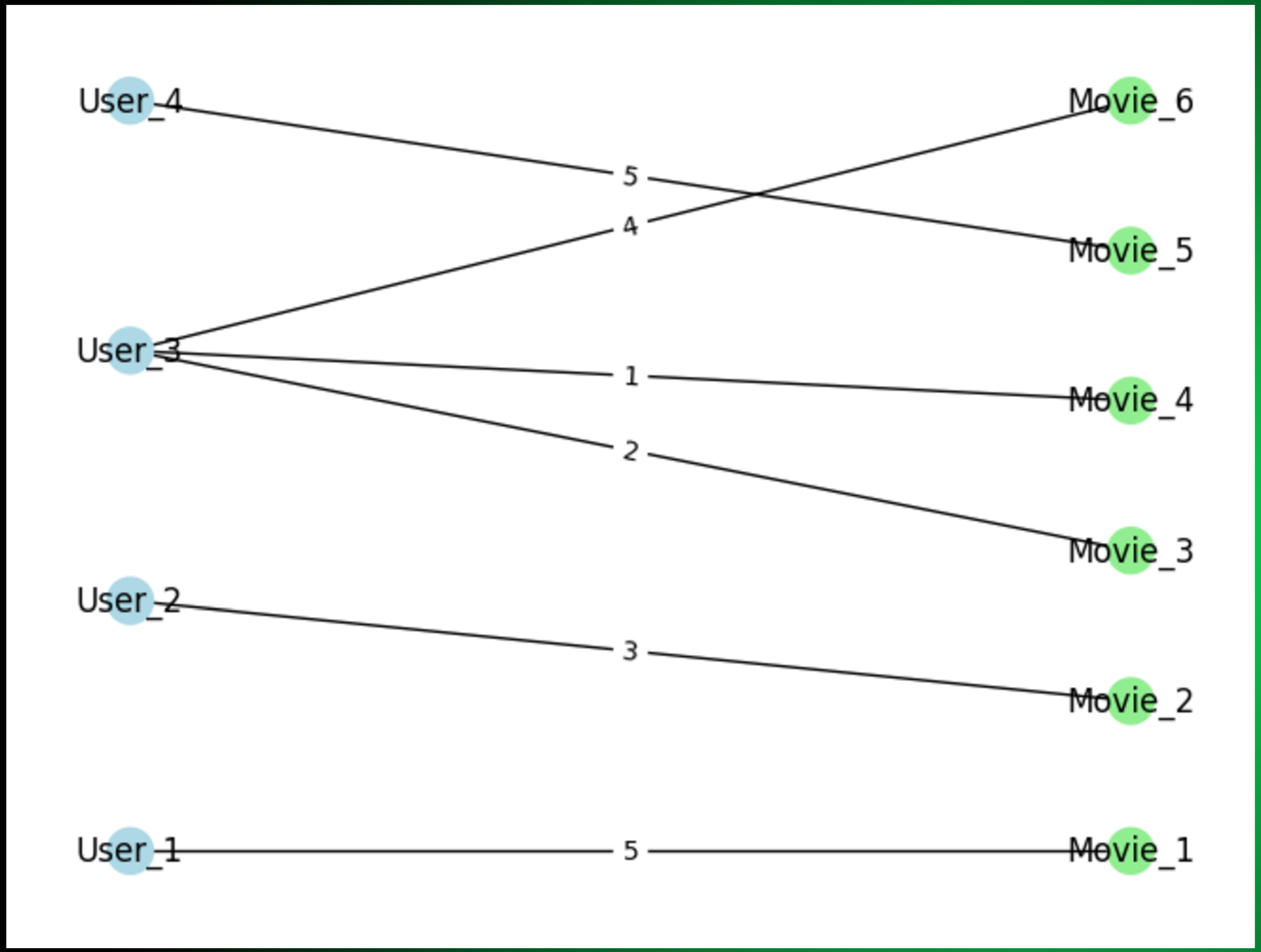
Graph Formtion

Node

- *User*
- *Movie*

Edge

- *Rating*



Reference: https://grouplens.org/datasets/movielens/

Data-preprocessing

Rating Distribution

rating	rating
1	6110
2	11370
3	27145
4	34174
5	21201

dtype: int64

'rating'>=3

Rating Distribution

rating	rating
3	27145
4	34174
5	21201

dtype: int64

Reference: https://grouplens.org/datasets/movielens/

80-20 Split

	userId	movieId	rating	timestamp
0	196	242	3	881250949
1	186	302	3	891717742
5	298	474	4	884182806
7	253	465	5	891628467
8	305	451	3	886324817
...	...	...	...	...
99992	721	262	3	877137285
99994	378	78	3	880056976
99995	880	476	3	880175444
99996	716	204	5	879795543
99999	12	203	3	879959583

82520 rows × 4 columns

	userId	movieId	rating	timestamp
0	770	250	5	875971902
1	169	331	5	891268491
2	327	143	4	888251408
3	85	1101	4	879454046
4	548	264	4	891043547
...	...	...	...	...
66011	807	177	4	892705191
66012	145	12	5	882182917
66013	602	748	3	888638160
66014	622	1078	3	882671160
66015	60	47	4	883326399

66016 rows × 4 columns

COMP 4222

Re-labeling

	userId	movieId	rating	timestamp	userId_index	movieId_index
0	770	250	5	875971902	769	249
1	169	331	5	891268491	168	329
2	327	143	4	888251408	326	142
3	85	1101	4	879454046	84	1086
4	548	264	4	891043547	547	263
5	279	456	3	875296924	278	449
6	280	86	4	891700475	279	85
7	653	87	4	878854332	652	86

Approach

WHY USE GNN?

GNNs effectively capture complex relationships and patterns in user-item interactions.

GRAPHSAGE

GCN

LIGHTGCN

VS

Traditional?

Hyper-parameters

Hyperparameter	GCNLinkPredictor	GraphSAGELinkPredicto	LightGCNLinkPredictor
Embedding Dimension	16	16	16
Hidden Channels	8	8	Not applicable
Number of Layers	2 (GCNConv)	2 (SAGEConv)	3
Activation Function	ReLU	ReLU	None
Learning Rate	0.01	0.01	0.01
Number of Epochs	200	200	200
Loss Function	Mean Squared Error (MSE)	Mean Squared Error (MSE)	Mean Squared Error (MSE)
Negative Sampling	Matches positive edges	Matches positive edges	Matches positive edges
Weight Initialization	PyTorch default	PyTorch default	Xavier uniform
Optimizer	Adam	Adam	Adam

GCN

Key Characteristics:

- Layer-Wise Propagation: Aggregates neighbor information through layers.
- Feature Transformation: Uses weights and activations (e.g., ReLU).
- Homogeneous Graphs: Excels with consistent node features.

Advantages:

- Theoretical Foundation: Rooted in spectral graph theory.
- Captures Local Patterns: Aggregates meaningful neighborhood data.
- Utilizes Node Features: Supports demographics or item attributes.

LIGHTGCN

Key Characteristics:

- No Nonlinear Activations: Faster and computationally efficient.
- Neighborhood Aggregation: Directly aggregates embeddings from neighbors.
- Less Complex Model: Fewer parameters, reducing risk of overfitting.

Advantages:

- Efficiency: Handles large datasets quickly, ideal for movie recommendations.
- Good for Collaborative Filtering: Directly leverages user-movie interactions.
- Simpler and Scalable: Suitable for large user-item graphs.

GraphSAGE

Key Characteristics:

- Sampling-Based Aggregation: Scalable, samples fixed-size neighbor sets.
- Feature Transformation: Uses weights and activations (e.g., ReLU).
- Supports Node Features: Effectively leverages attributes like demographics or genres.

Advantages:

- Utilizes Node Features: Enhances recommendations with rich attributes.
- Expressive: Captures complex relationships via nonlinear transformations.
- Adaptive Aggregation: Focuses on relevant interactions for better predictions.

Results

GCN:

MSE: 1.246371067606549

MAE: 0.889264211000151

GraphSAGE:

MSE: 1.1780590104434243

MAE: 0.8357644481780635

LightGCN:

MSE: 1.8424100625363011

MAE: 1.0925375528126515