# GNN Optimization

Yi Deng, Muhao Guo, Qian Wang
Advisor: Mohammad Saeed

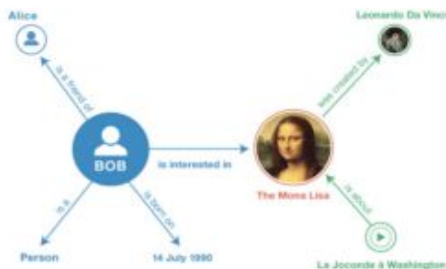USC University of Southern California

Contents

- Introduction
  - GNN application
  - Tasks
- GNN in homogeneous graph
  - Dataset (citation full)
  - Model (GCN)
  - Performance
- GNN in Heterogeneous graph
  - Dataset (netflix dataset)
  - Model (pinsage)
  - Performance
- Conclusion
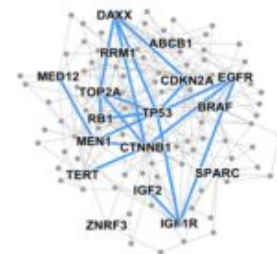 - Pros and cons
  - Expect to do

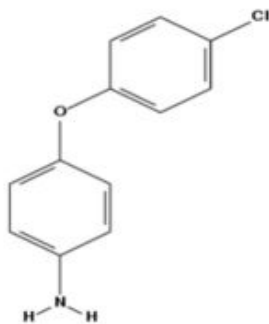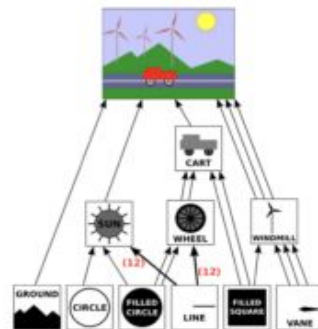# Introduction | GNN application
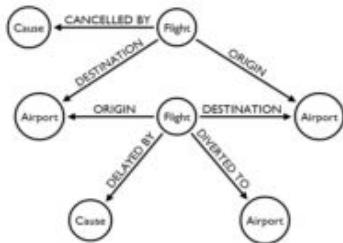
Event Graphs

Knowledge Graphs

Disease pathways

Molecules

Scene Graphs

Cell-cell similarity networks

Introduction | Tasks

- Homogeneous graph
    - Data: citation full dataset
    - Model: GCN

- Heterogeneous graph
    - Data: Movielen Dataset, Netflix dataset
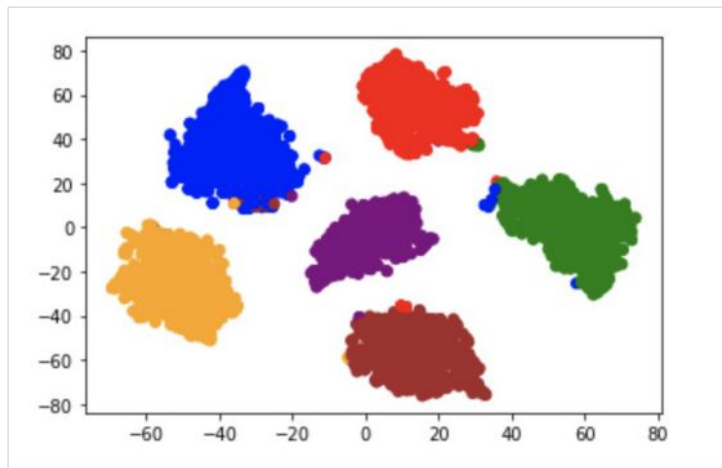   - Model: PinSage

GNN in homogeneous graph | Citation full dataset

dataset	#edges	#nodes	size of node features	#num_classes
cora	126842	19793	8710	70
cora_ml	16316	2995	2879	7
citeseer	10674	4230	602	6
dblp	105734	17716	1639	4
pubmed	88648	19717	500	3

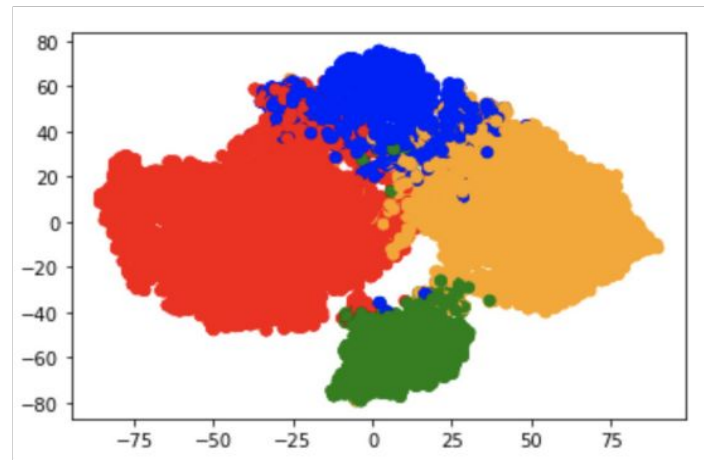# Gnn in homogeneous graph | Model
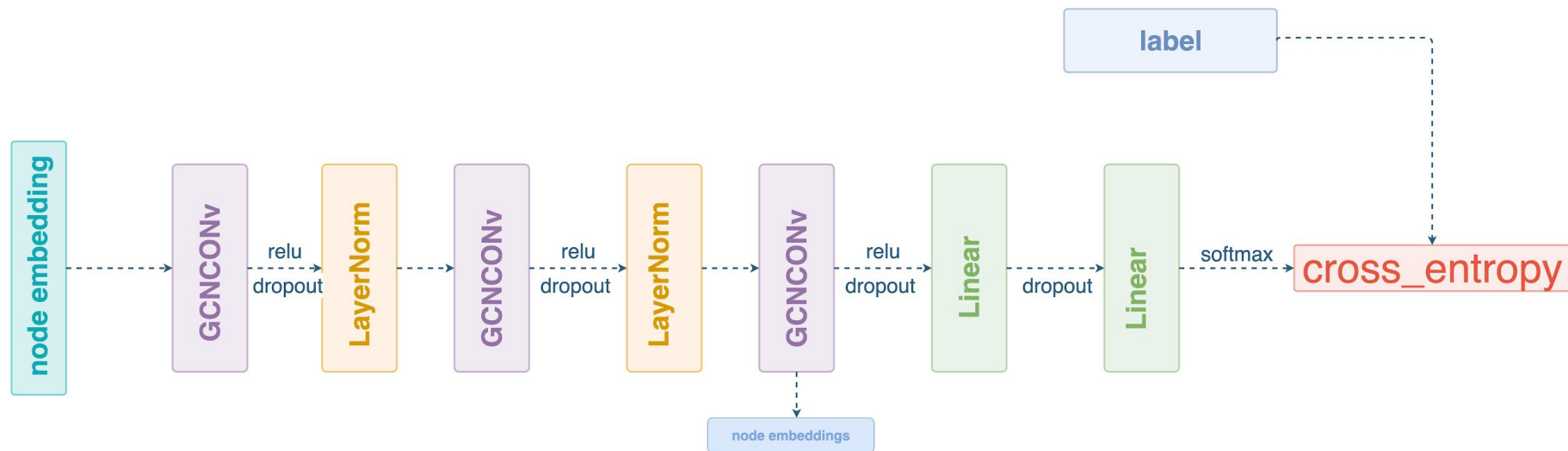
GNN in homogeneous graph | Performance

citeseer

dblq

dataset	cora	cora_ml	citeseer	dblp	pubmed
accuracy	0.8480	0.9790	0.9960	0.9434	0.9428

GNN in heterogeneous graph | movie recommendation
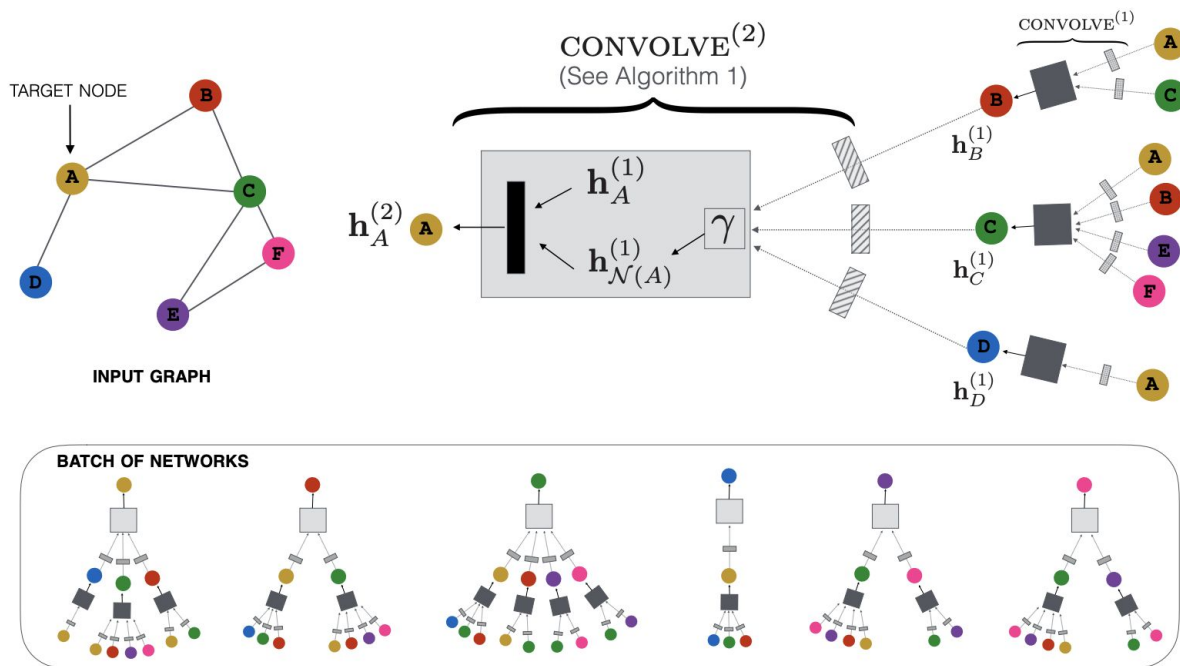
Pinsage

 For

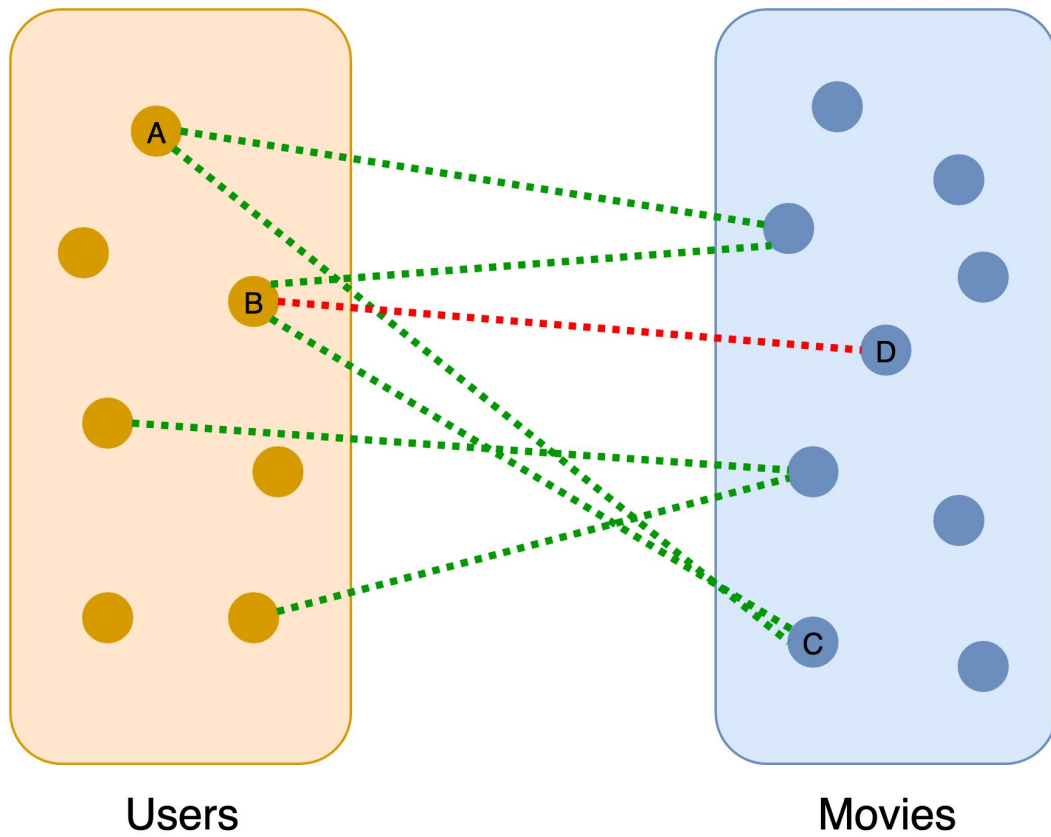movie recommendation
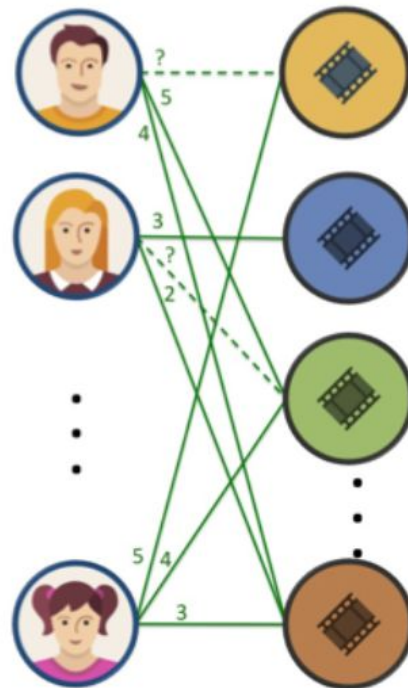
GNN in heterogeneous graph | Model

GNN in heterogeneous graph | random walk

Users   Movies

GNN in heterogeneous graph | movielens

#ratings:1,000,209,# movies:4000, #users: 6,040

#ratings: 100 million, # movies:17770, #users: 6,040

- Movie data example:
  - 1::Toy Story (1995)::Animation|Children's|Comedy
  - 2::Jumanji (1995)::Adventure|Children's|Fantasy
  - 3::Grumpier Old Men (1995)::Comedy|Romance

- Users data example:
  - 1::F::1::10::48067
  - 2::M::56::16::70072
  - 3::M::25::15::55117

- Rating data example:
  - 1::1193::5::978300760
  - 1::661::3::978302109
  - 1::914::3::978301968

- Movie data example:
  - 1, 2003, Dinosaur Planet
  - 2, 2004, Isle of Man TT 2004 Review
  - 3, 1997, Character
  - 4, 1994, Paula Abdul's Get Up & Dance
  - 5, 2004, The Rise and Fall of ECW

- Rating data example:
  - 10662, 2400203, 5, 2005-07-11
  - 11312, 1711268, 2, 2002-03-13
  - 12132, 1657481, 3, 2004-04-07
  - 11307, 1844276, 4, 2004-08-13
  - 11309, 1734805, 1, 2001-01-30

GNN in heterogeneous graph | pipeline

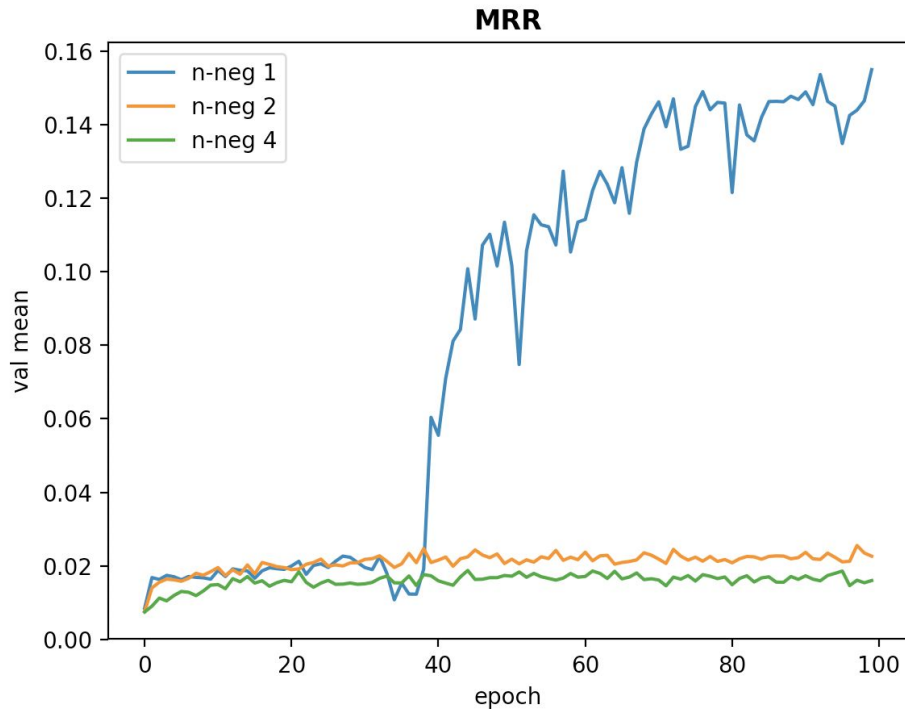# GNN in heterogeneous graph | Metric

Ground truth movie id :1, 3

movie_candidates	user_id	Cosine similarity	rank	Reciprocal rank
1	1	0.3	3	1/3
2	1	0.6	2	1/2
3	1	0.7	1	1
4	1	0.1	4	1/4

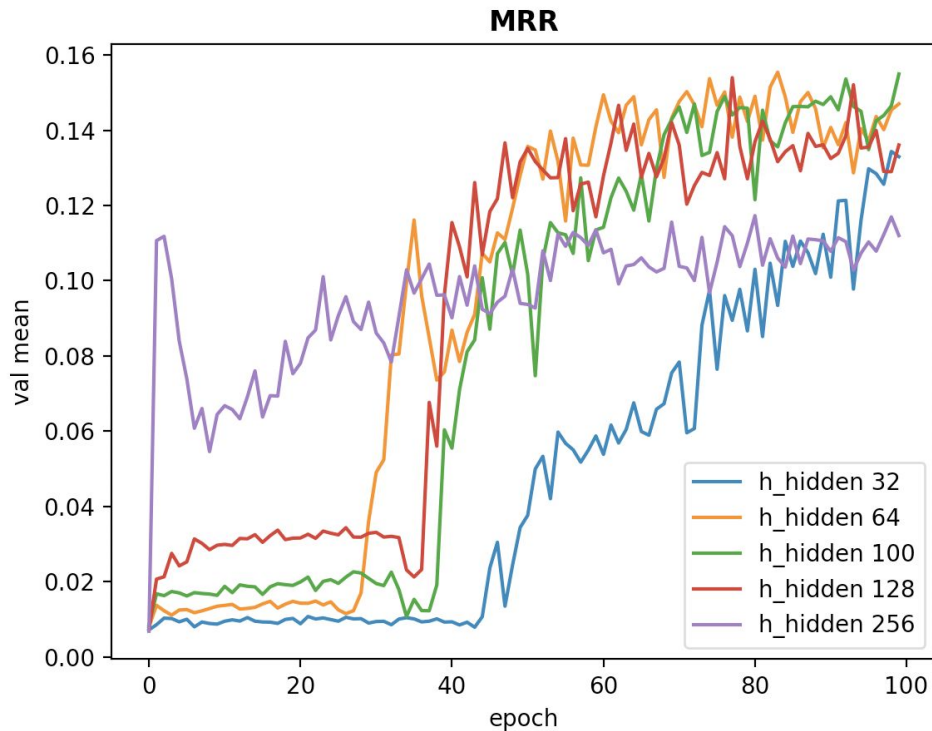# GNN in heterogeneous graph | Performance on MovieLens
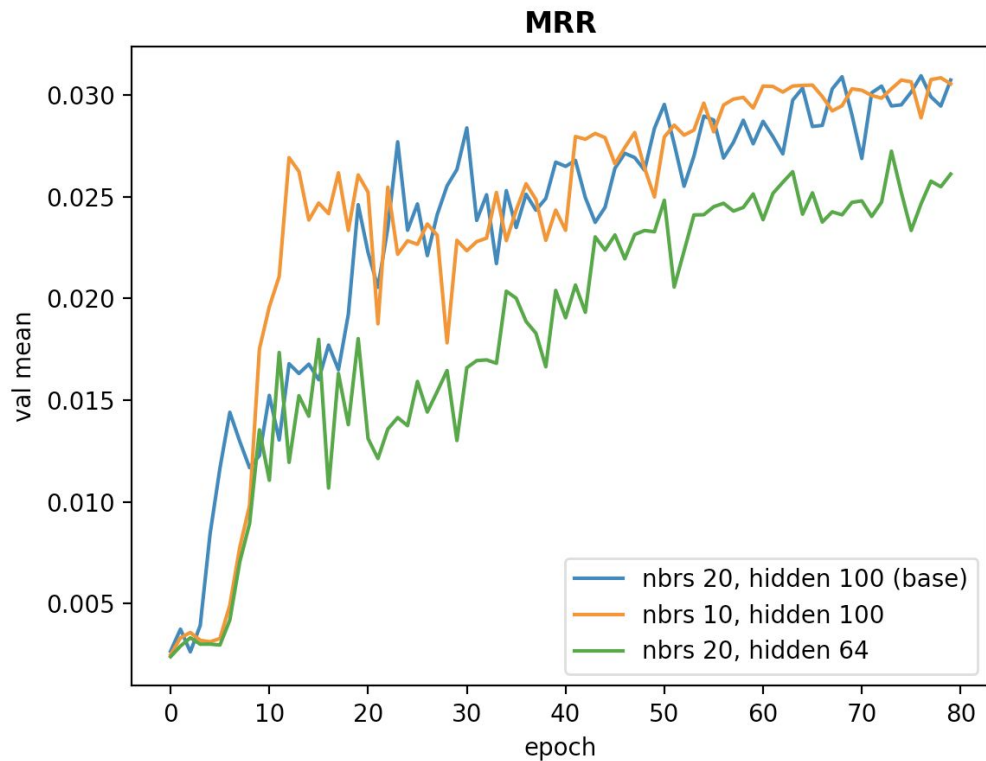
GNN in heterogeneous graph | Performance on Netflix dataset
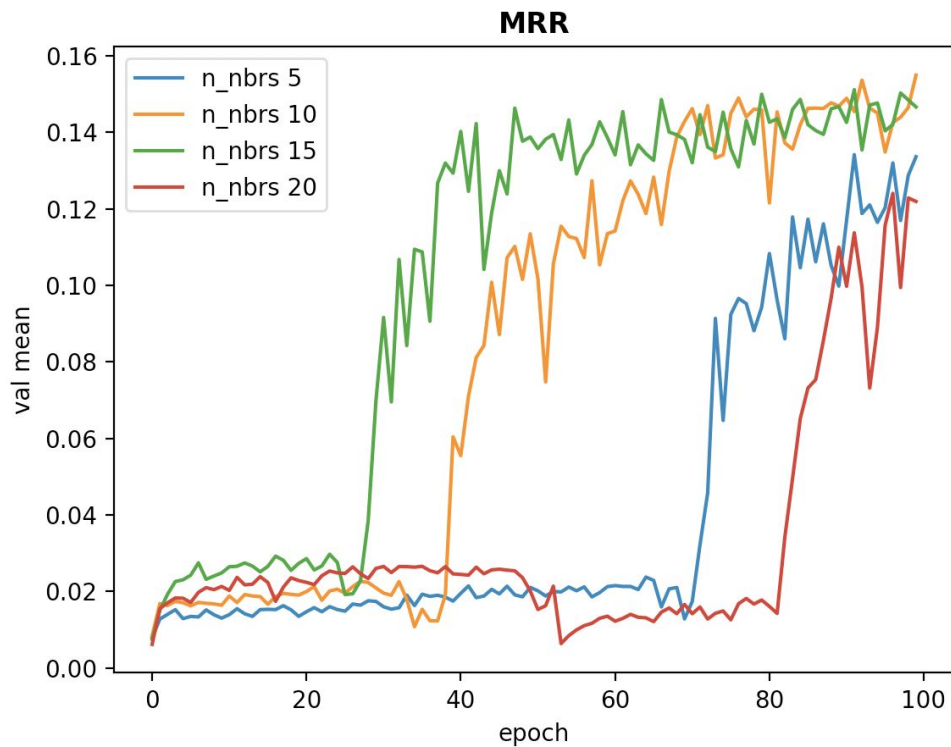
● #Negtive sampling

GNN in heterogeneous graph | Performance on Netflix dataset
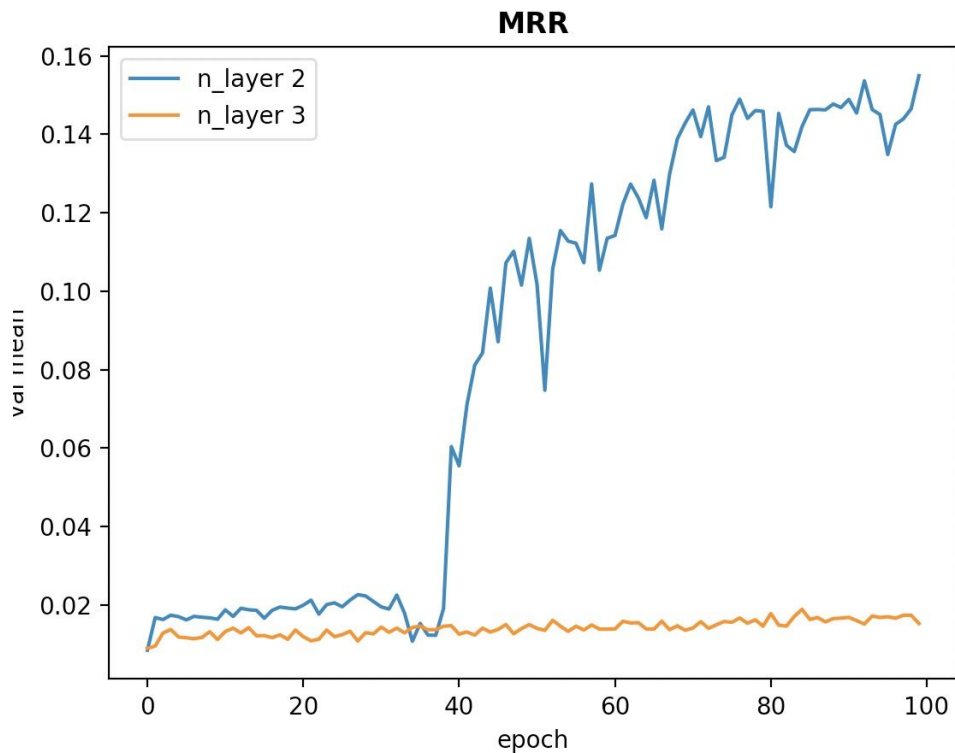
● #embedding_size(h_hidden)

GNN in heterogeneous graph | Performance on Netflix dataset

● # n_nbrs

GNN in heterogeneous graph | Performance on Netflix dataset

● # n_layer

Conclusion | Expect to do

- Train the model on the entire Netflix dataset

- Make a prediction on how the users rate the movies

Thank you !