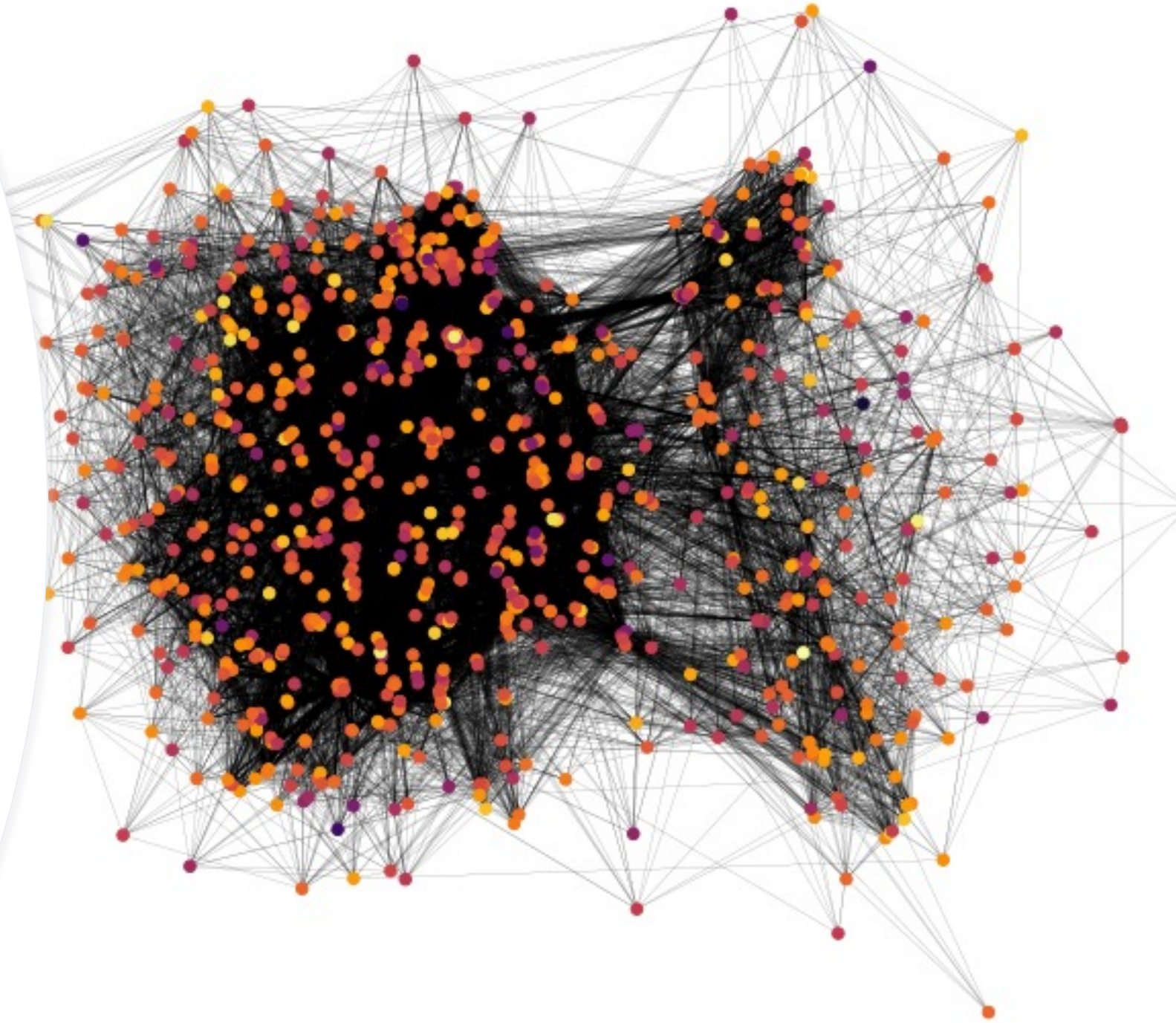
Movie Recommender Exploration

Arthur Babey, Stanislas Ducotterd,
Nessreddine Loudiy, Lamyae Omari



Dataset

- 943 users
- 1682 movies
- 100’000 ratings

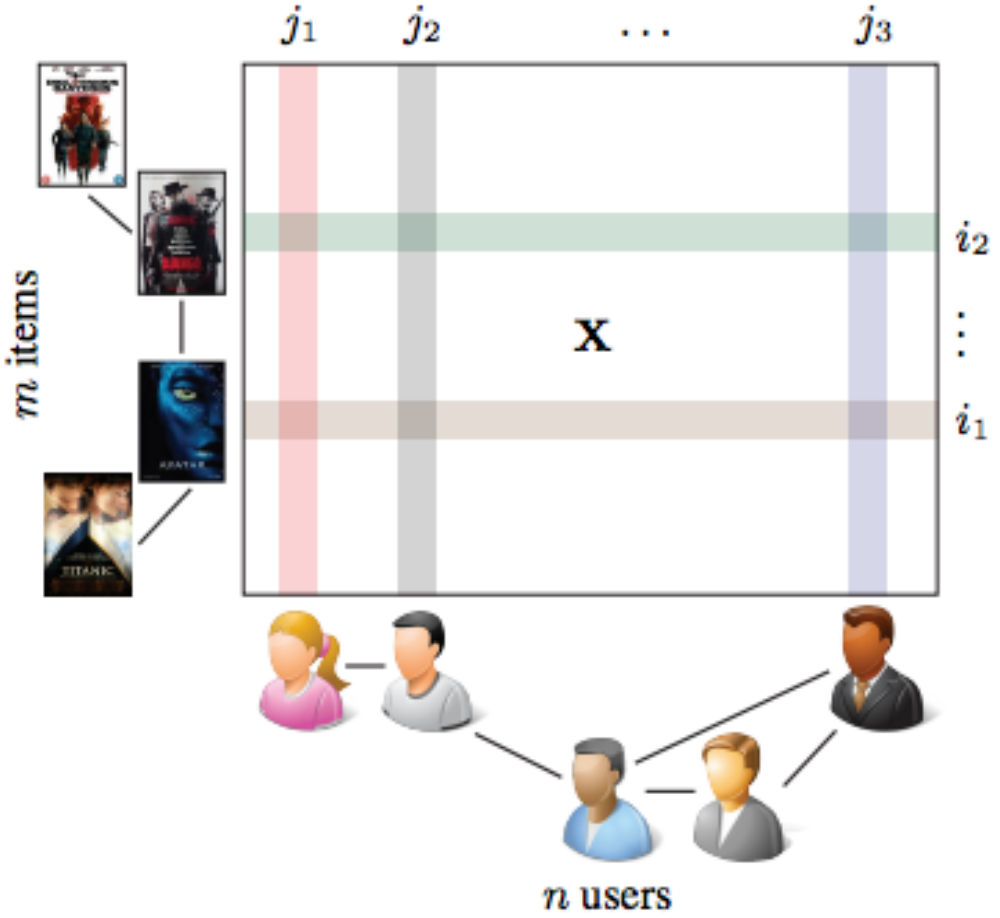


Figure taken from Federico Monti, Michael M. Bronstein, and Xavier Bresson. Geometric matrix completion with recurrent multi-graph neural networks. NIPS, 2017.

Features

Users features :

- Age
- Gender
- Occupation
- Zip Code

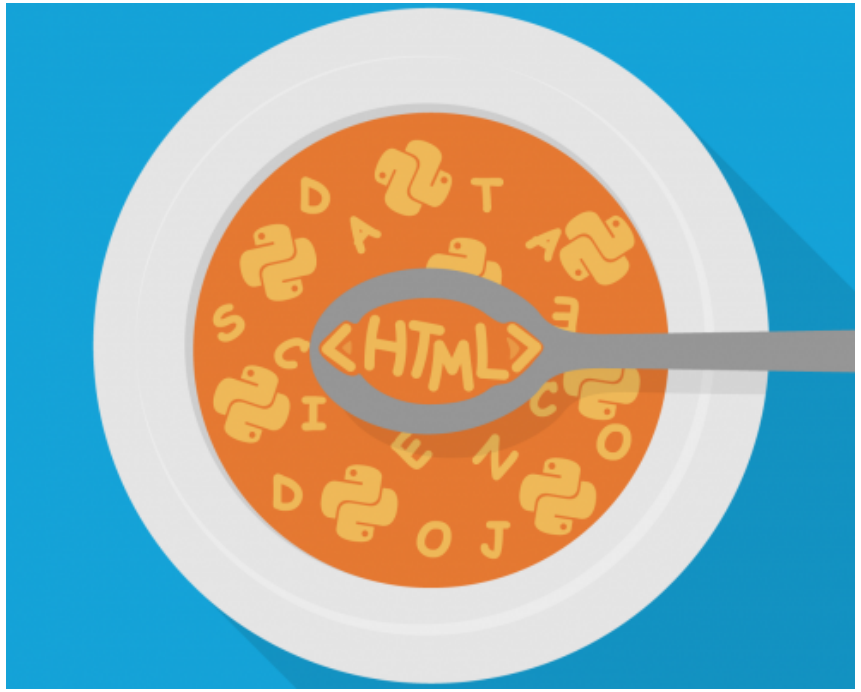
Movies features :

- Title
- Release date
- Genre

Scraping

- URL from the dataset was not up-to-date as the dataset is from 1998

- We automatically made queries on the IMDb website with the title of the movie in the database to get the IMDb id of each movie



Making the scraping robust

- Most of the titles contain additional information between parenthesis which can disrupt the result of the query on IMDb
- We tried an alternative query on IMDb without the parenthesis and the stuff between them
- We compared both of the result and took the best one
- When even the best result was too different from the dataset, we manually retrieved the IMDb id

Extended Features

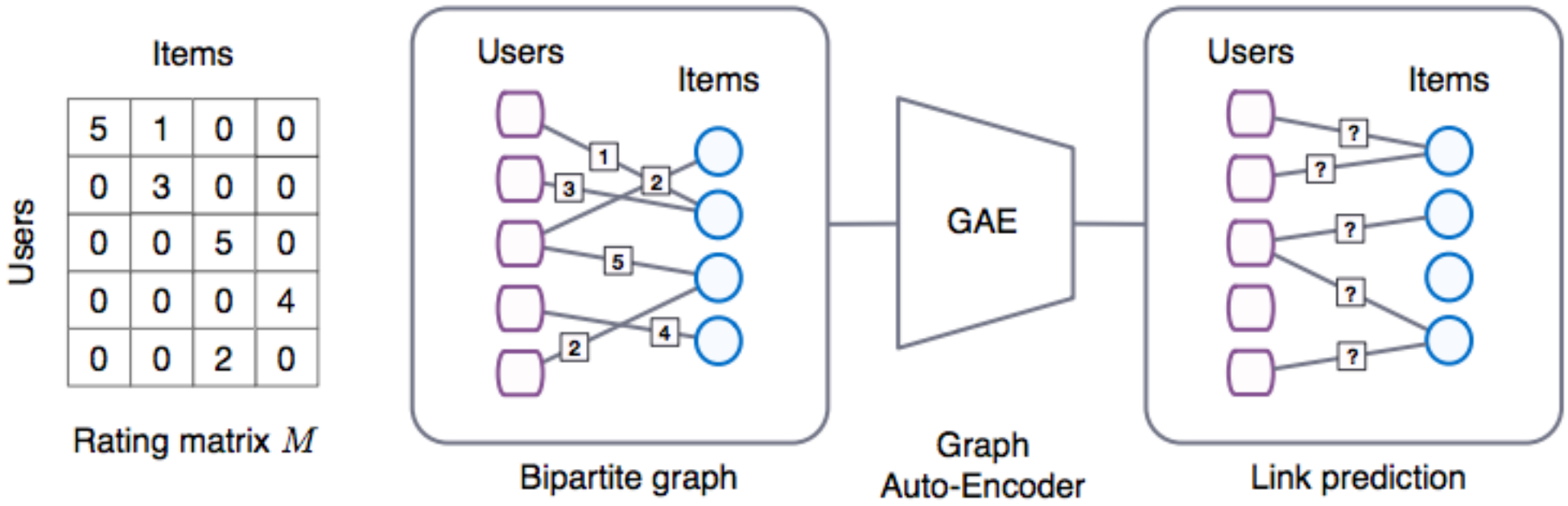
Users features :

- US state
- Median income

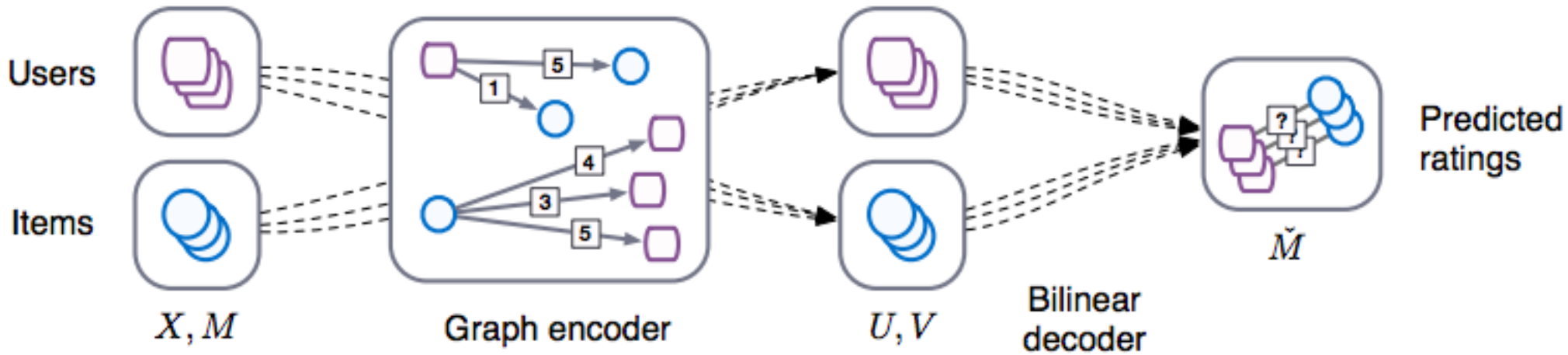
Movies features :

- Movie length in minutes
- Average rating from IMDb
- The age of the movie in 1998

The model


Items
Users
5 1 0 0
0 3 0 0
0 0 5 0
0 0 0 4
0 0 2 0
Rating matrix M
Users
Items
1
3
2
5
4
2
Bipartite graph
GAE
Graph
Auto-Encoder
Users
Items
?
?
?
?
?
Link prediction

The model

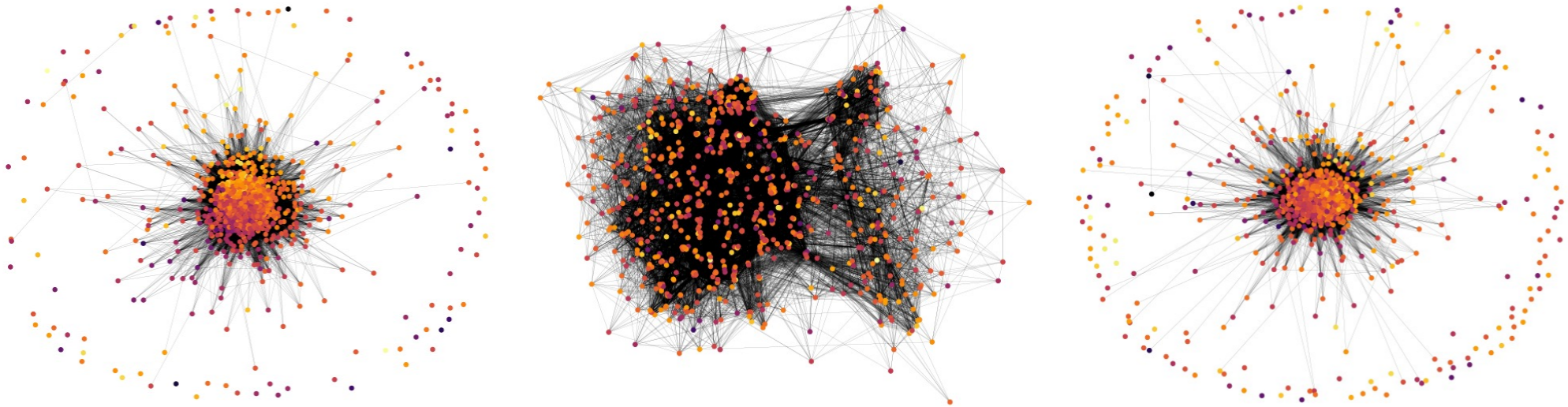

Users
Items
X, M
5
1
4
3
5
Graph encoder
U, V
Bilinear
decoder
?
$\check{M}$
Predicted
ratings

The results

Model	RMSE
GC-MC without feature	0.910
GC-MC+Feature	0.905
GC-MC+ Extended Feature	0.900

Exploring the embeddings

- We built three epsilon-similarity graphs
  - One from the features
  - One from the embedding space without the features
  - One from the embedding space with the features



Summary statistics for each graph

Graph	#Nodes	#Edges	Average degree	Degree standard deviation	#Connected components	Clustering coefficient	Diameter
Features	943	42400	89.93	62.38	1	0.66	3
Embedding without features	943	136929	290.41	214.20	93	0.71	infinite
Embedding with features	943	125491	266.15	217.88	120	0.70	infinite

Highest degrees, highest Katz centrality and highest betweeness centrality give different top nodes

Conclusion