A Network Tour of Millenial Movies

Team 49: Milena Filipovic, Kristijan Lopatichki, Jelena Malic, Davor Todorovski

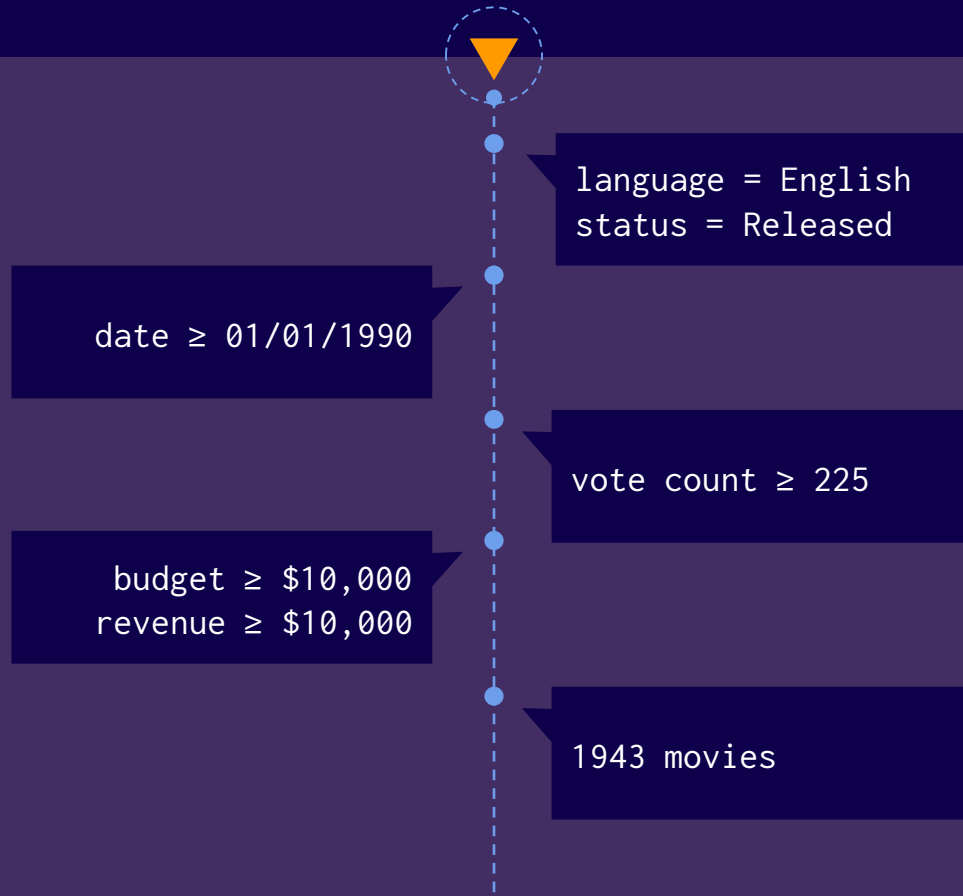# Introducion

"Movies can and do have tremendous influence in shaping young lives in the realm of entertainment towards the ideals and objectives of normal adulthood."

TMDb Dataset

- since 2008
- users from over 180 countries
- over 437,000 movie entries

Data Cleaning

language = English
status = Released

date ≥ 01/01/1990

vote count ≥ 225

budget ≥ $10,000
revenue ≥ $10,000

1943 movies
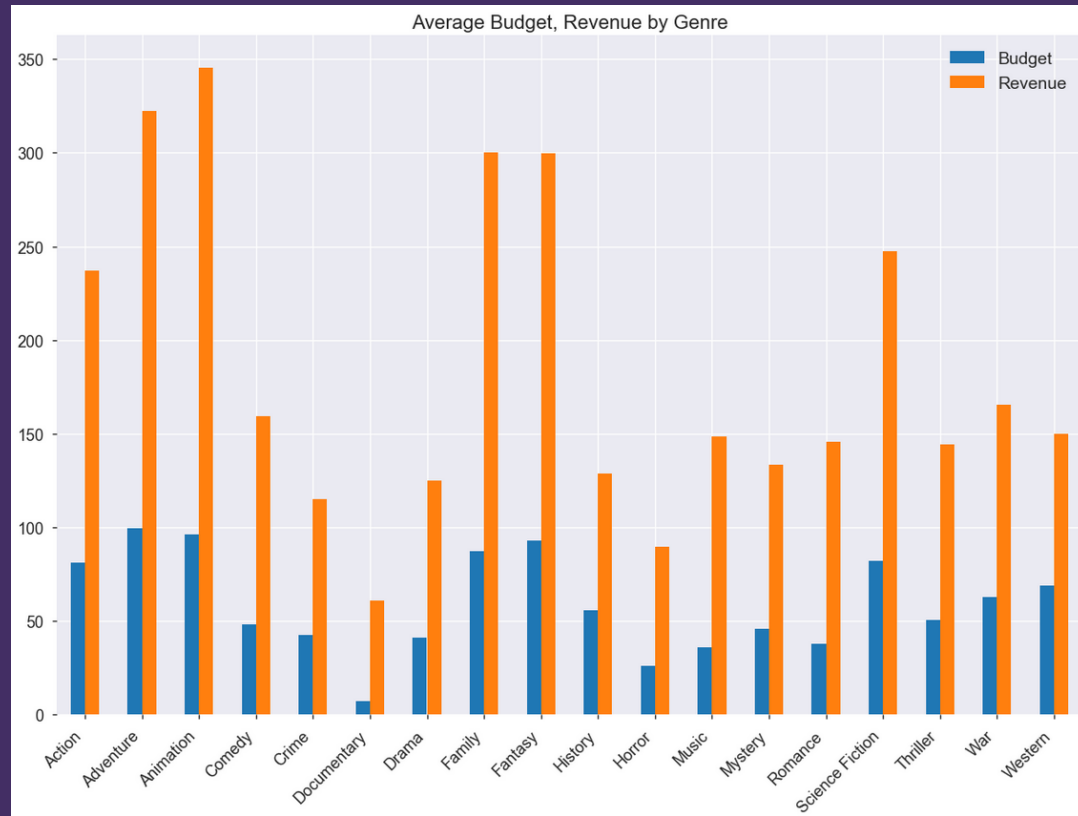
Comparison of different Budget and Revenue values by genre

Correlation between Number of votes and Vote averages

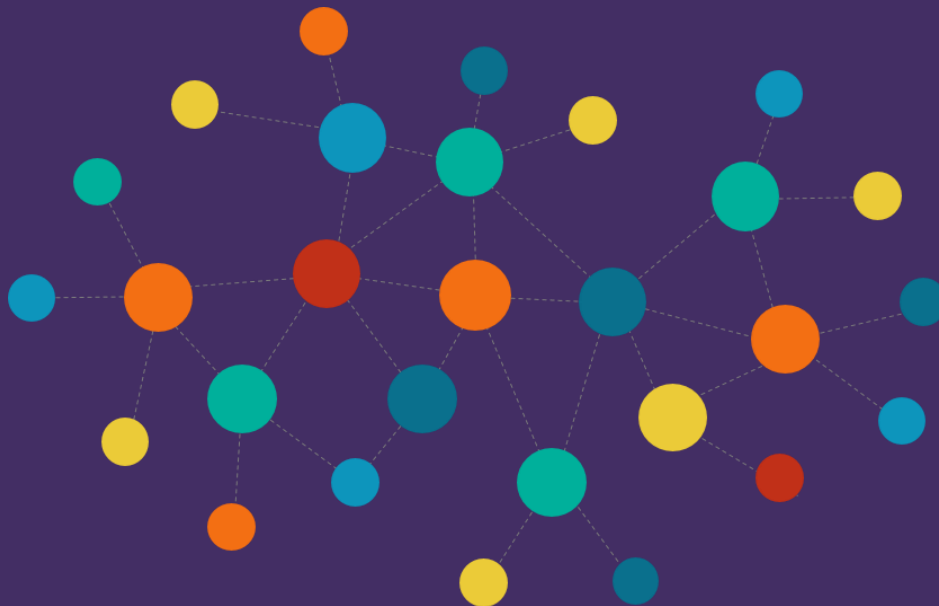# Cast and Crew Co-Occurrence Graph

Nodes
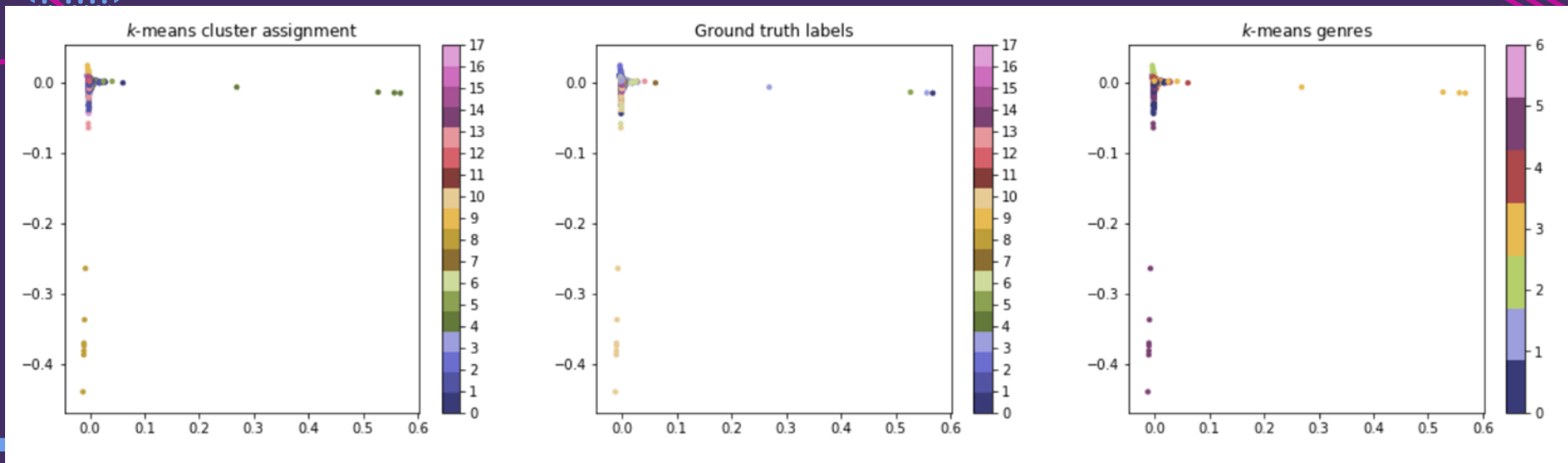
1942 movies

Weights

of shared C&C

Edges

209,095 created from C&C co-occurrence

k = 18

d = 18

Drama
Action, Comedy,
Animation, Horror,
Adventure, Science
Fiction

Percentage of
correctly
labeled nodes:
38.52 %

Cast and Crew: Spectral Clustering

Keyword Co-Occurrence Graph
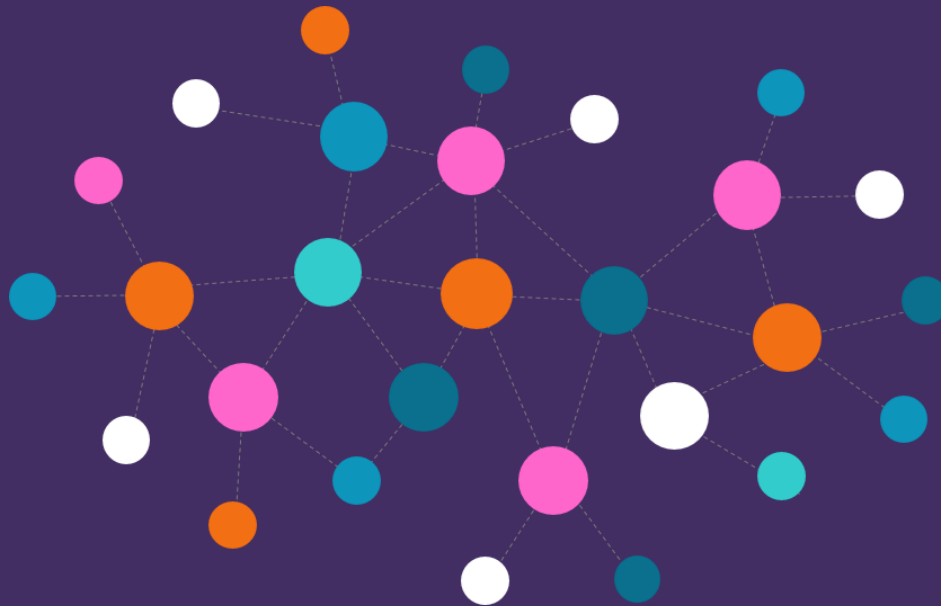
Nodes

1924 movies

Weights

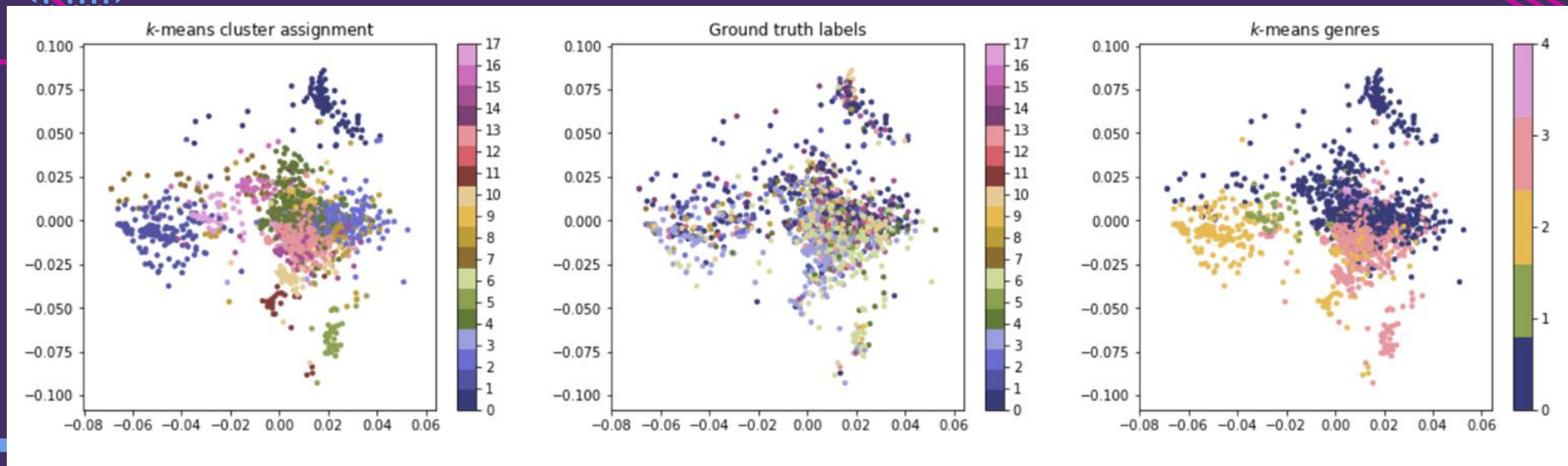# of shared keywords

Edges

created from keyword co-occurrence

k = 18
d = 18

Action,
Drama, Comedy,
Adventure,
Horror

Percentage of
correctly
labeled nodes:
32.95 %

Keywords: Spectral Clustering

Input graph signals
e.g. bags of words

Feature extraction
Convolutional layers

Classification
Fully connected layers

Output signals
e.g. labels

Graph signal filtering
1. Convolution
2. Non-linear activation

$0 = \lambda_1 < \lambda < \lambda_{M_{l-1}}$

Graph coarsening
3. Sub-sampling
4. Pooling

CNN: Pre-Processing

Divide into training (60%) and testing (40%) sets

lower case + whitespaces & stopwords removed

movie keywords → bag-of-words

movies with < 5 words removed

word embeddings (666 words)

normalized
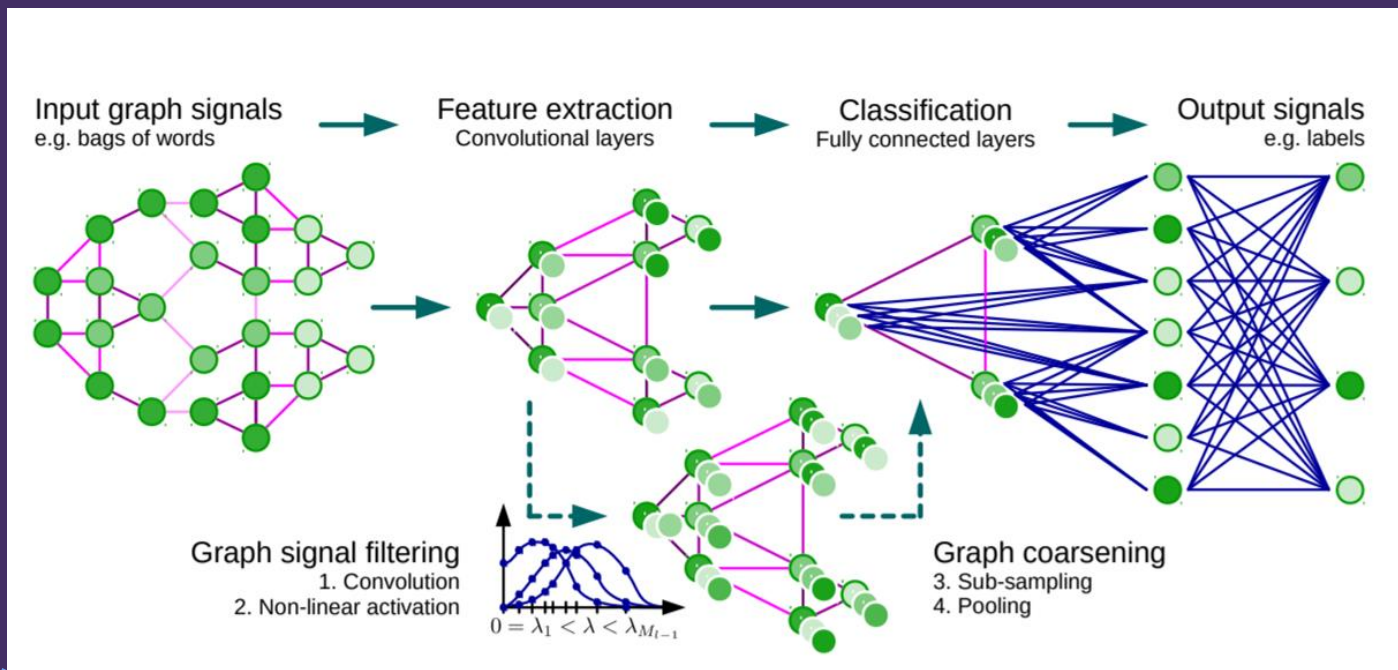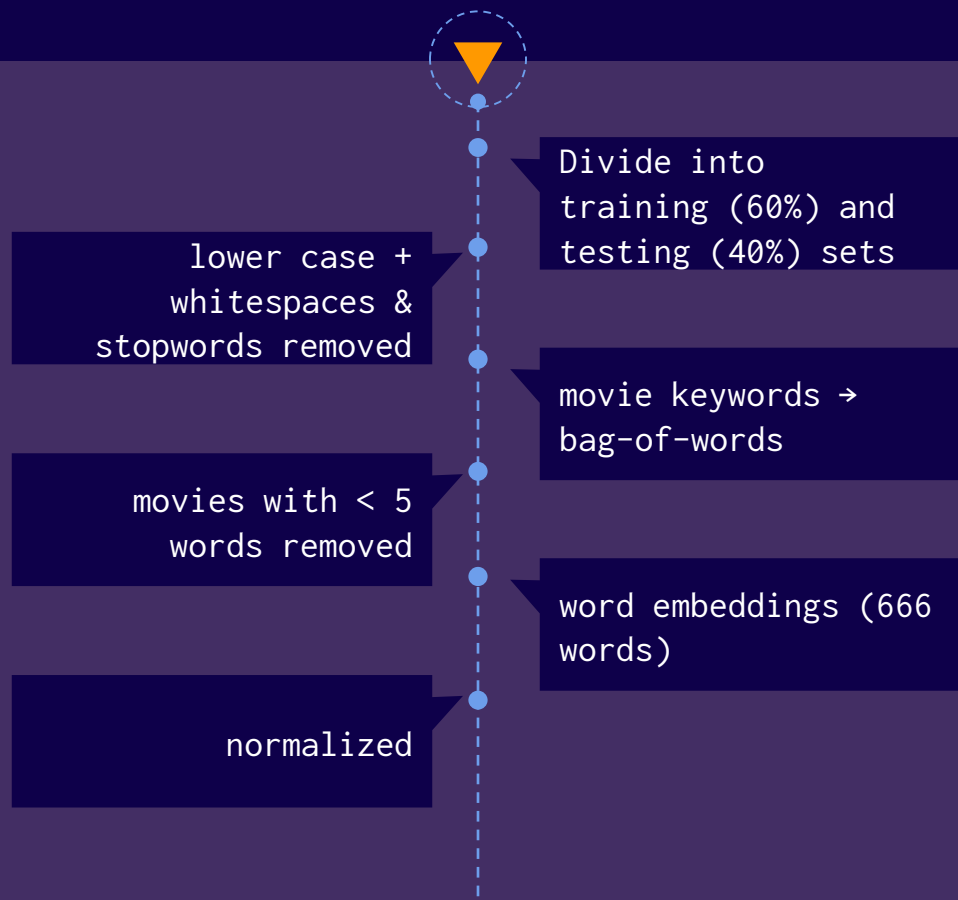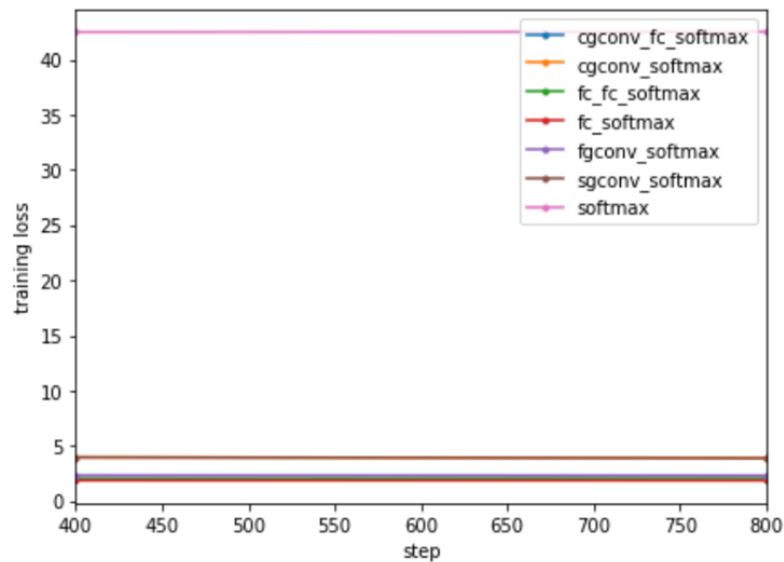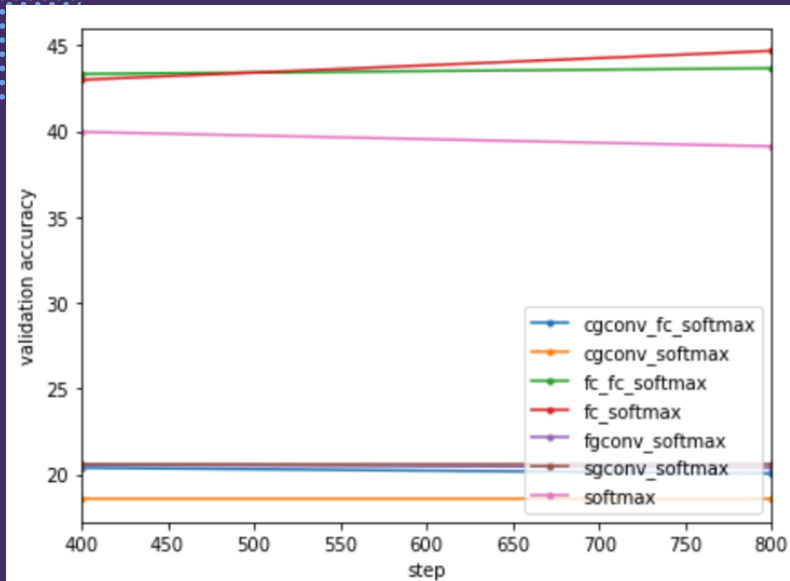
CNN: Results

CNN: Results

Model	train accuracy (%)	test accuracy (%)
softmax	98.36	39.12
fc_softmax	69.96	44.69
fc_fc_softmax	91.45	43.68
fgconv_softmax	20.18	20.40
sgconv_softmax	20.18	20.57
cgconv_softmax	18.86	18.55
cgconv_fc_softmax	18.31	20.07

Recommender Systems
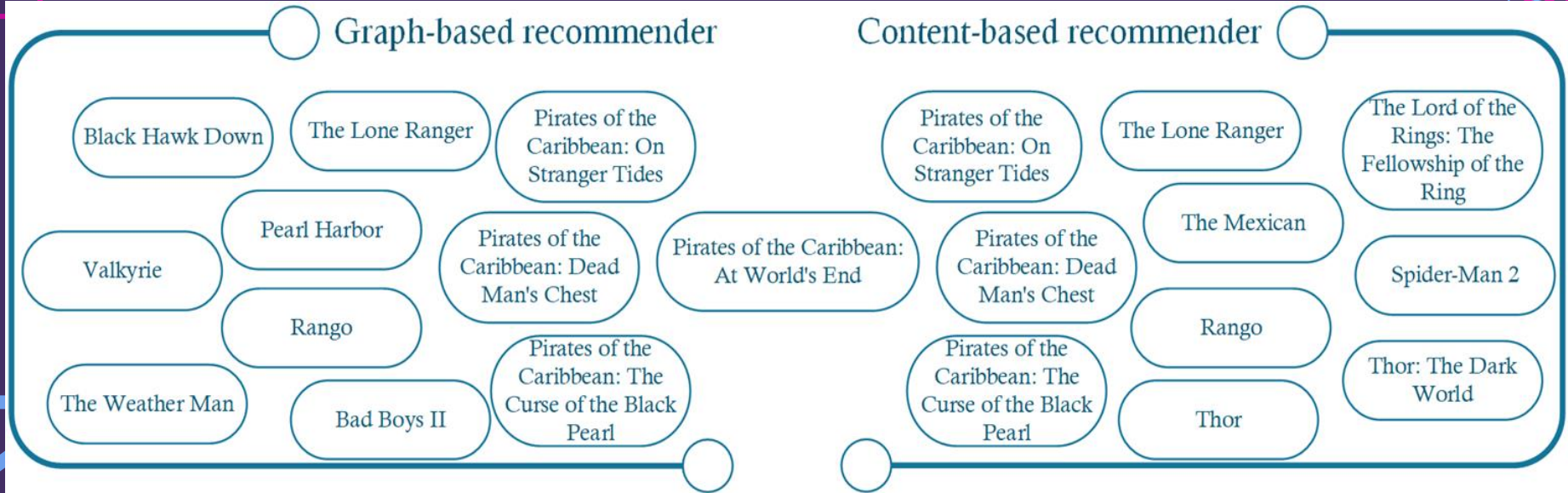
3 Recommender systems:

- Graph - based recommender system
- Content - based recommender system
- Collaborative recommender system

Graph vs. Content recommender system

Collaborative recommender

- Used additional dataset with user input
- User oriented - specific user preferences and tastes taken into account
- Measure of accuracy

Conclusion

Summary

- Spectral graph analysis
- CNN
- Recommender systems

Future work

- Alternative labels
- Genre grouping

Thank you for your attention!

Any questions?