# TEAM A

ANDREAS SCHÖRGENHUMER, MARKUS HIESMAIR, ONUR KIRKIZOGLU, RAPHAEL SULZER, TARIQ AL MOUSA

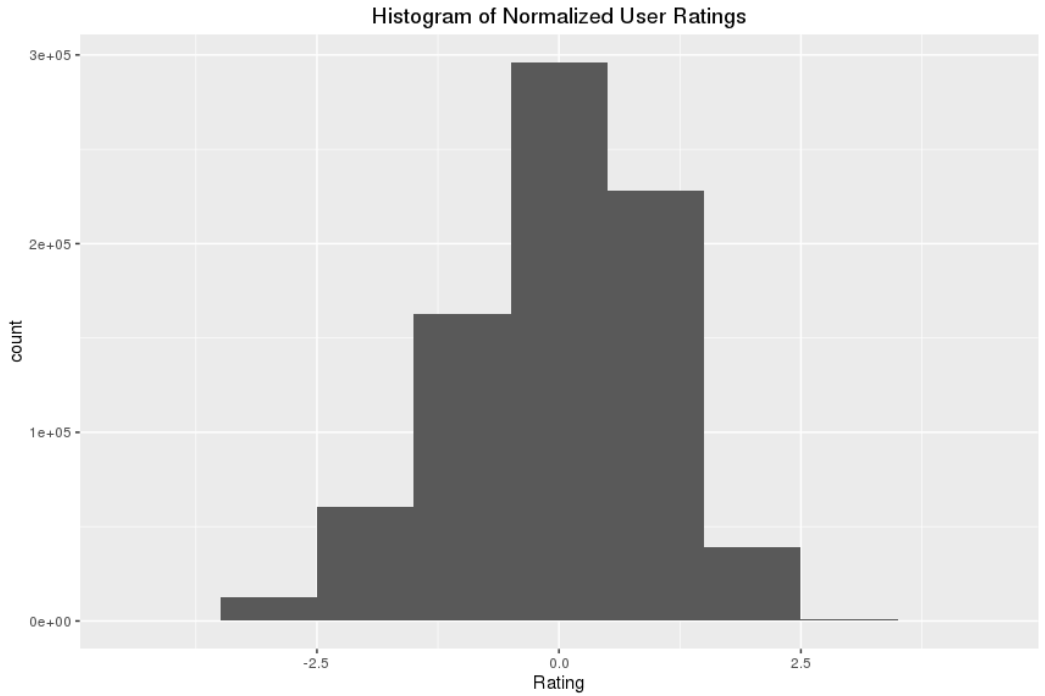
USER-GENERATED DATA, 19.04.2016

# Contents

- Data analysis
- Proposed methods
- Experimental setup
- Results
- Conclusions

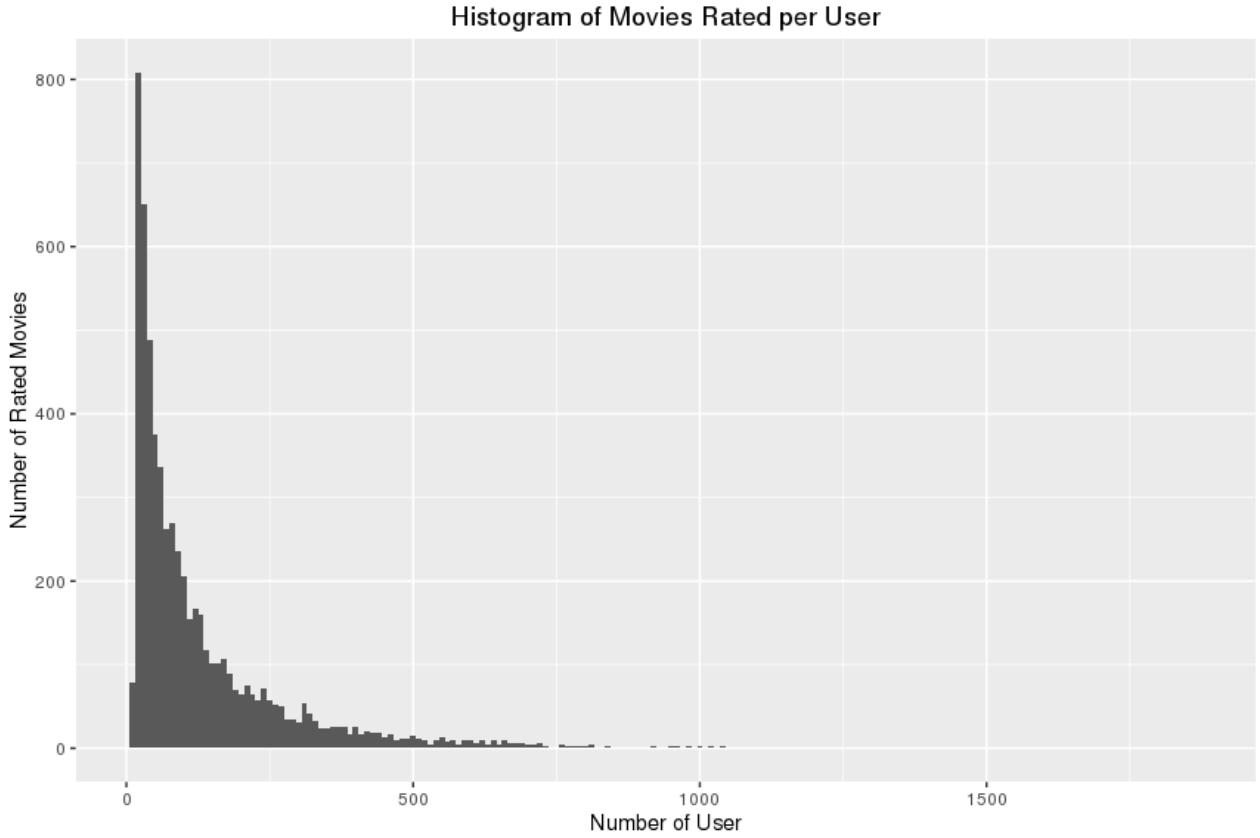
# Data analysis

- Ratings are integer-values from 1 – 5
- Normal distribution with highest density at the rating of 4



# Data analysis

- More than 50% of the movies have a rating of 4 or 5
- Algorithms should work good for users with a small amount of ratings, as well as with eventually unknown users or movies



# Proposed methods

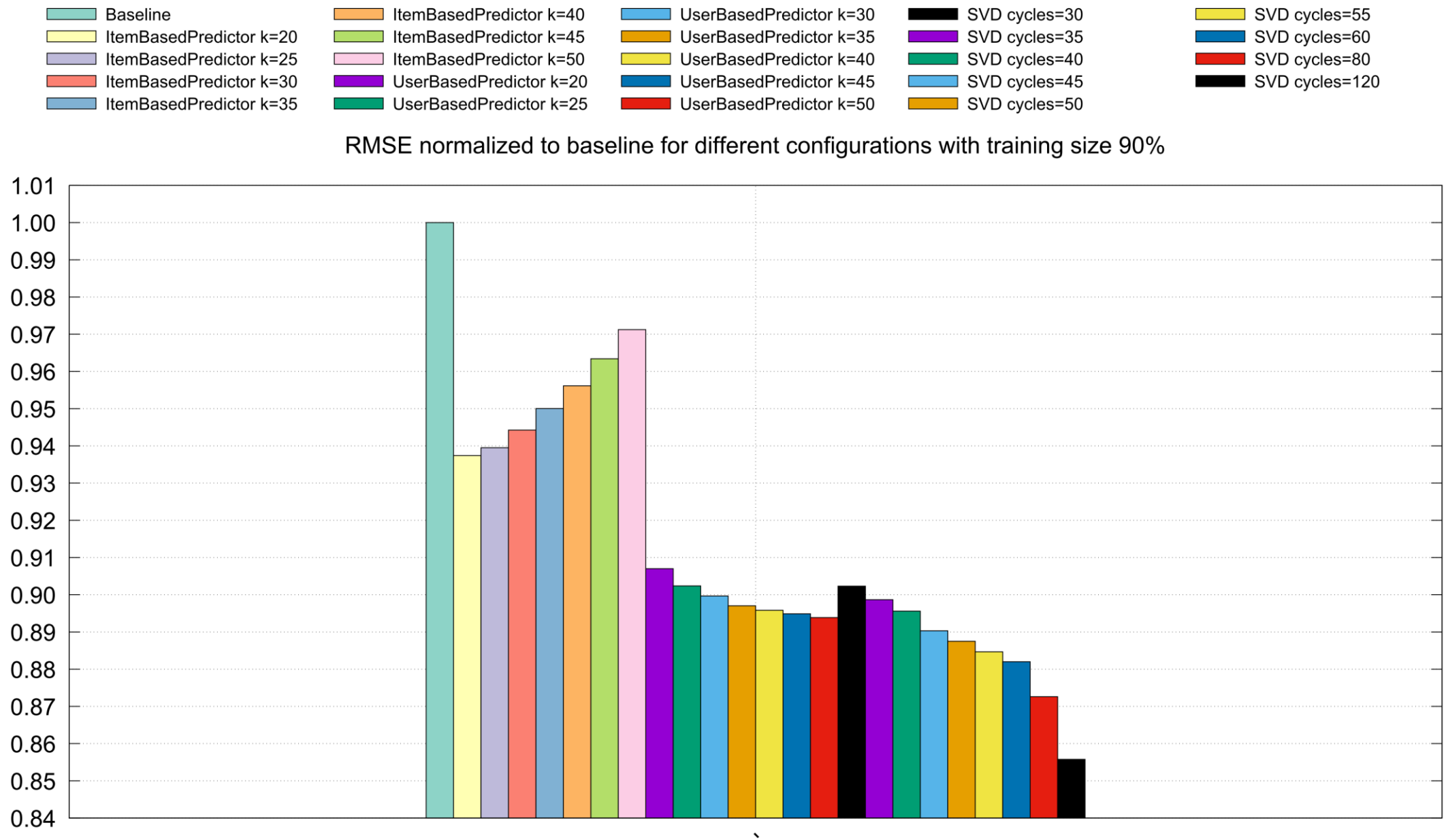
- Memory-based methods
  - User-based Collaborative Filtering Recommendation (UBCF)
  - Item-based Collaborative Filtering Recommendation (IBCF)

- Model-based methods
  - Singular Value Decomposition (SVD)

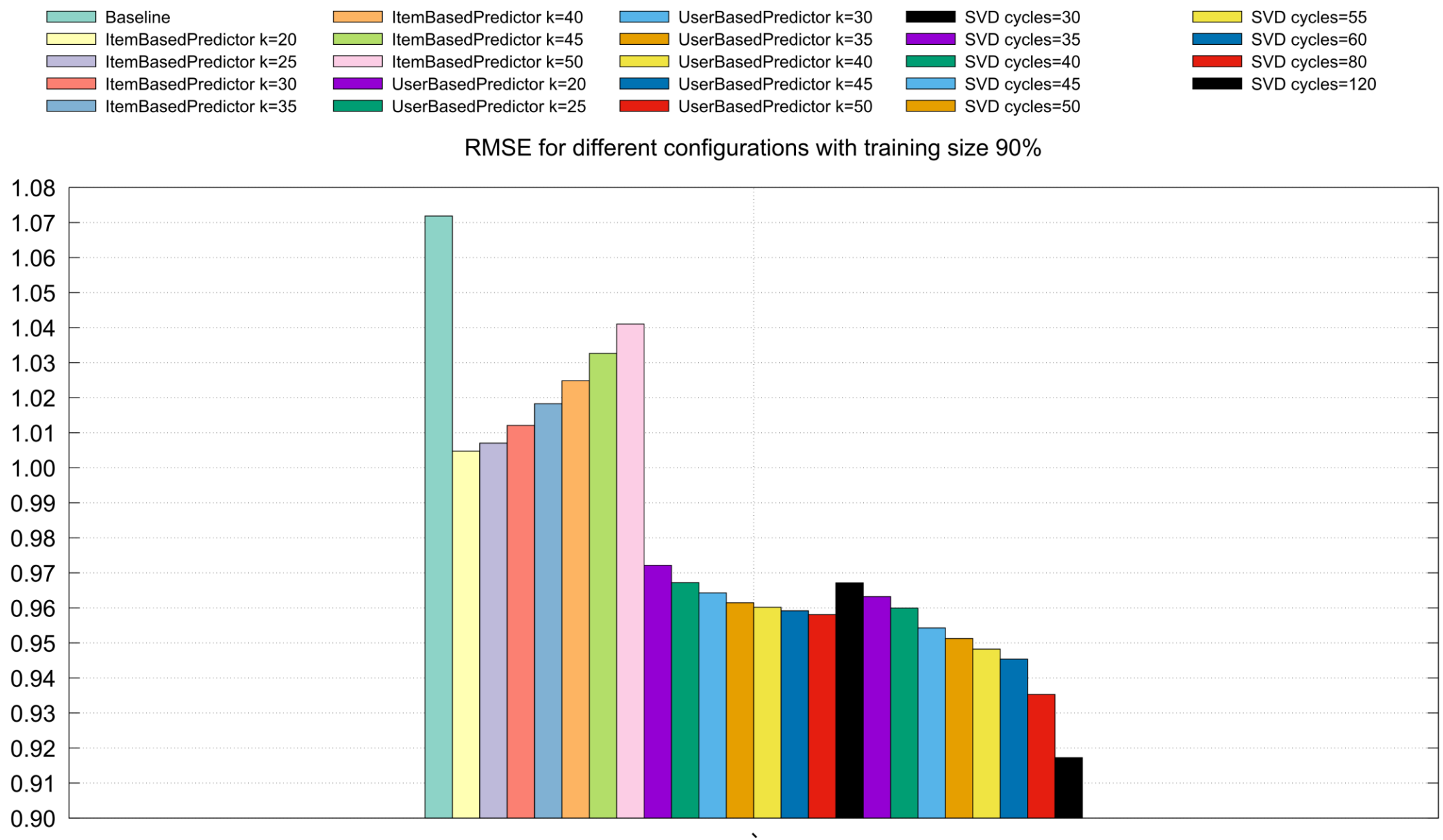
# Experimental setup

- Predictions are evaluated by Root-Mean-Squared-Error
- The configuration of every method were tested on a training-set size of 90% (10% test data, no overlap)
- UBCF & IBCF with configurations:
  k = 20 up to k = 50 with steps of 5
- SVD with configurations: cycles = 30, 35, 40, 50, 55, 60, 80, 120
- 3 iterations for better assessment
- Simple averaging-method as baseline

# Results


Baseline
ItemBasedPredictor k=20
ItemBasedPredictor k=25
ItemBasedPredictor k=30
ItemBasedPredictor k=35
ItemBasedPredictor k=40
ItemBasedPredictor k=45
ItemBasedPredictor k=50
UserBasedPredictor k=20
UserBasedPredictor k=25
UserBasedPredictor k=30
UserBasedPredictor k=35
UserBasedPredictor k=40
UserBasedPredictor k=45
UserBasedPredictor k=50
SVD cycles=30
SVD cycles=35
SVD cycles=40
SVD cycles=45
SVD cycles=50
SVD cycles=55
SVD cycles=60
SVD cycles=80
SVD cycles=120
RMSE normalized to baseline for different configurations with training size 90%
1.01
1.00
0.99
0.98
0.97
0.96
0.95
0.94
0.93
0.92
0.91
0.90
0.89
0.88
0.87
0.86
0.85
0.84

# Results


Baseline
ItemBasedPredictor k=20
ItemBasedPredictor k=25
ItemBasedPredictor k=30
ItemBasedPredictor k=35
ItemBasedPredictor k=40
ItemBasedPredictor k=45
ItemBasedPredictor k=50
UserBasedPredictor k=20
UserBasedPredictor k=25
UserBasedPredictor k=30
UserBasedPredictor k=35
UserBasedPredictor k=40
UserBasedPredictor k=45
UserBasedPredictor k=50
SVD cycles=30
SVD cycles=35
SVD cycles=40
SVD cycles=45
SVD cycles=50
SVD cycles=55
SVD cycles=60
SVD cycles=80
SVD cycles=120
RMSE for different configurations with training size 90%
1.08
1.07
1.06
1.05
1.04
1.03
1.02
1.01
1.00
0.99
0.98
0.97
0.96
0.95
0.94
0.93
0.92
0.91
0.90

# Conclusions

- Worst perfomance in all configurations with IBCF
- UBCF ≈ up to 45 cycles in SVD
- Starting with 50 cycles SVD achieved best results
- SVD with 120 cycles was the best configuration