



Dating recommendations using Spark

Felipe Chamma, Felipe Formenti, Ryan Speed



UNIVERSITY OF
SAN FRANCISCO

Agenda

- 1 – The problem
- 2 – Data set description
- 3 – Predicting Ratings
- 4 – The love_scoreTM feature
- 5 – JSON data output





1 - The problem

- Czech dating website
- Improve matchmaking
- Improve scalability
- Test Spark for real



2

Data set description



168,791 users

User_id, gender



17,359,346 ratings

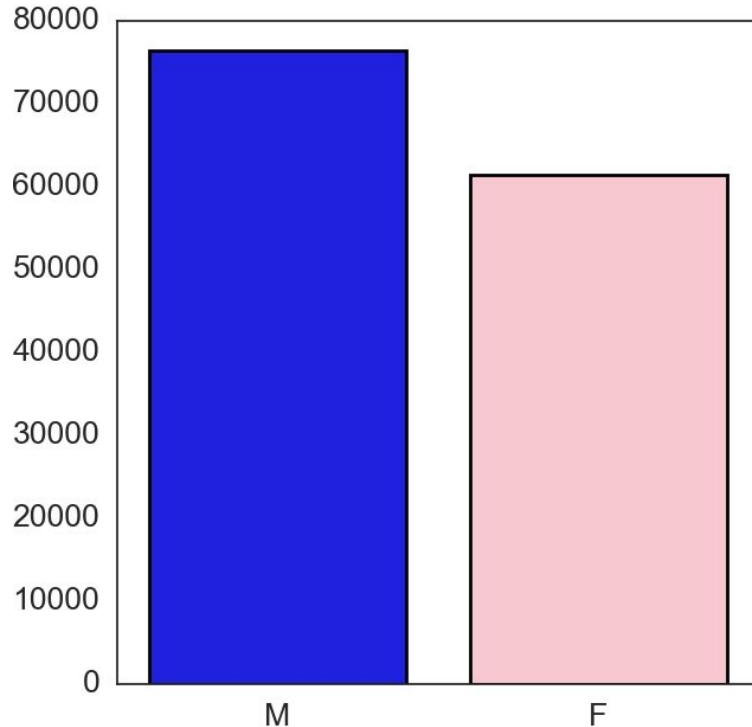
User_id_1, user_id_2, rating (0-10)



0.06%

Matrix sparsity

Count by gender



- Males are the majority, after cleaning data for unknowns (which was allowed back in 2006)

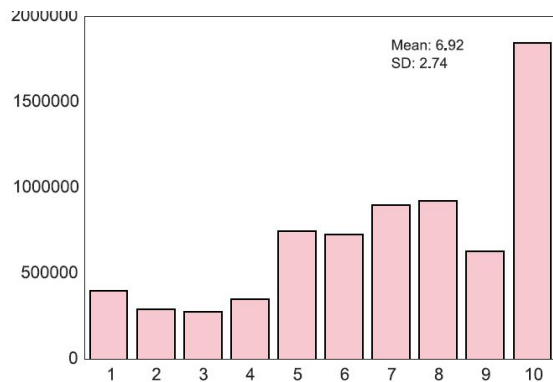


Ratings by gender

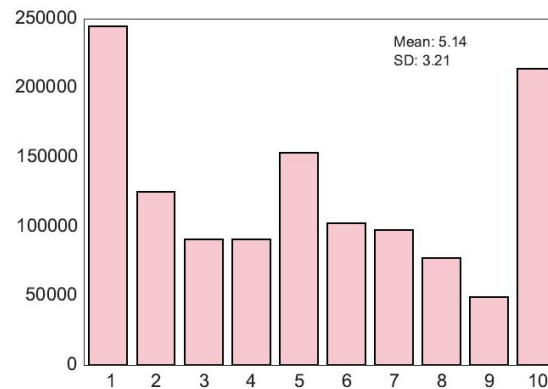
	Total Ratings	Mean	SD
F-M	7.10 M	6.92	2.74
M-F	3.23 M	5.48	3.12
F-F	1.24 M	5.14	3.21
M-M	683 k	4.46	3.32



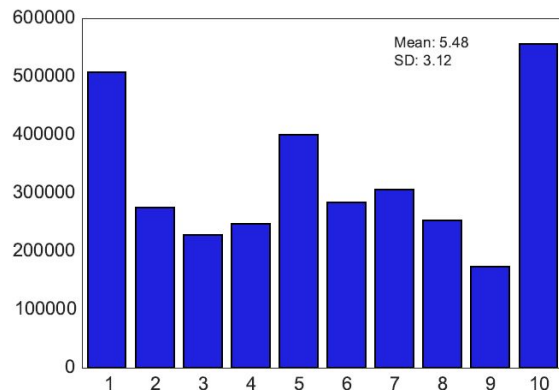
Ratings by gender



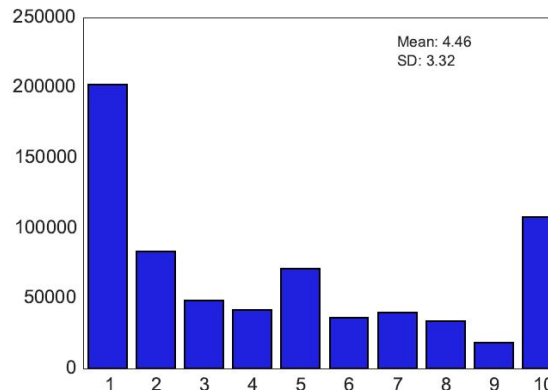
F - M



F - F



M - F



M - M

3

Predicting Ratings

Splitting users into gender/preference matrices

Male – Female

Items

Users

	1	2	3	...	N
1				...	
2				...	
3				...	
...				...	
N				...	

Female – Male

Items

Users

	1	2	3	...	N
1				...	
2				...	
3				...	
...				...	
N				...	

Collaborative filtering methods

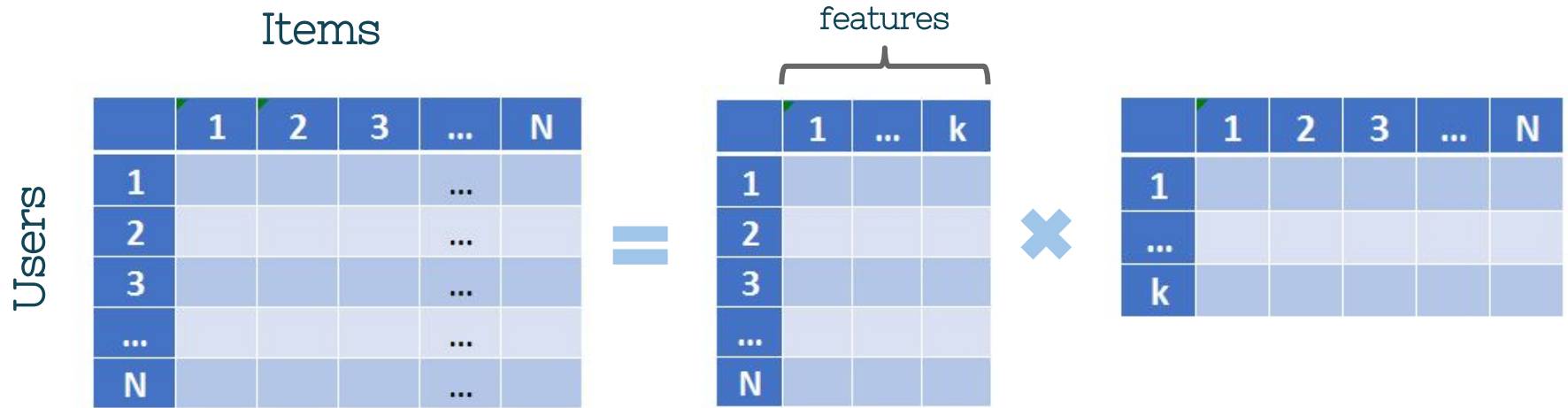
UV decomposition

- Good approach if appropriate k is chosen
- Much faster

Similarity matrix

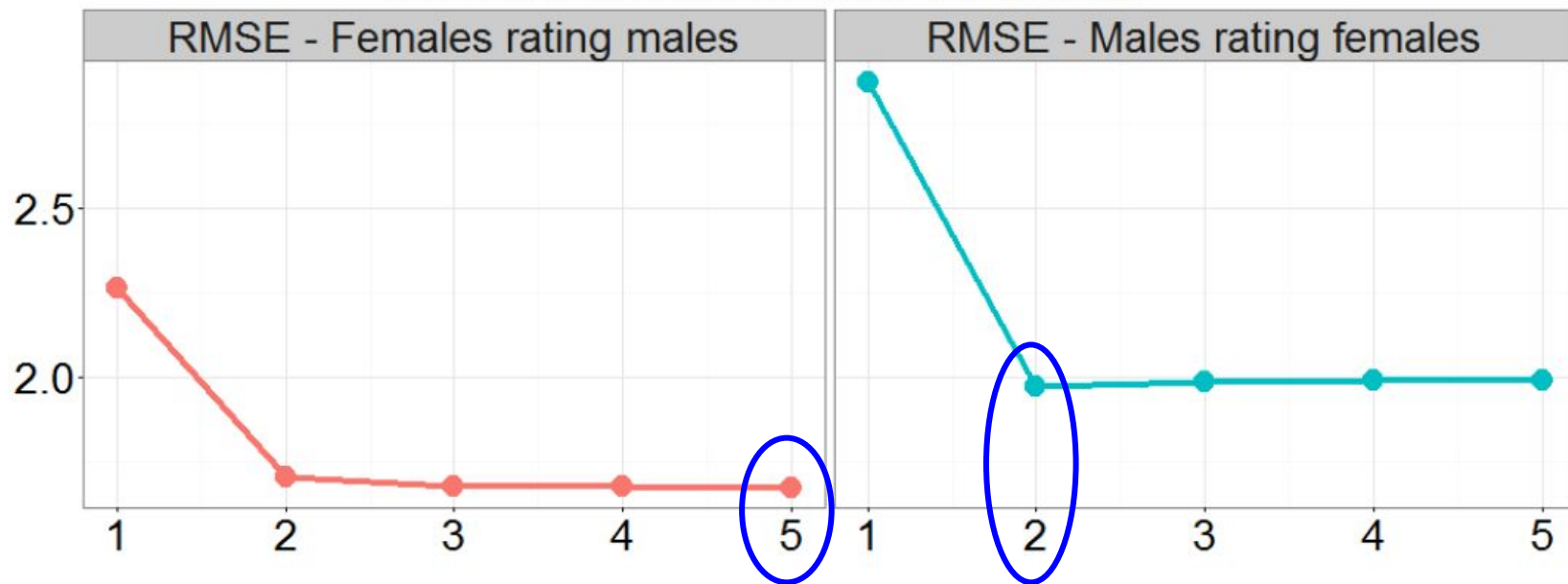
- Would take forever

UV Matrix Decomposition



Matrix decomposition

Root mean squared error vs. k value



4

The love_scoreTM feature



The metric

$$\textit{love_score}(U_i, U_j) = \lambda_1(10 - |\textit{Att}(U_i) - \textit{Att}(U_j)|) + \lambda_2\textit{Pref}(U_i, U_j) + \lambda_3\textit{Pref}(U_j, U_i)$$

$$\textit{Pref}(U_i, U_j) = \hat{f}(U_i, U_j) - \overline{U_{iG}}$$

$\hat{f}(U_i, U_j)$: predicted rating of user j by user i



5 - JSON data output

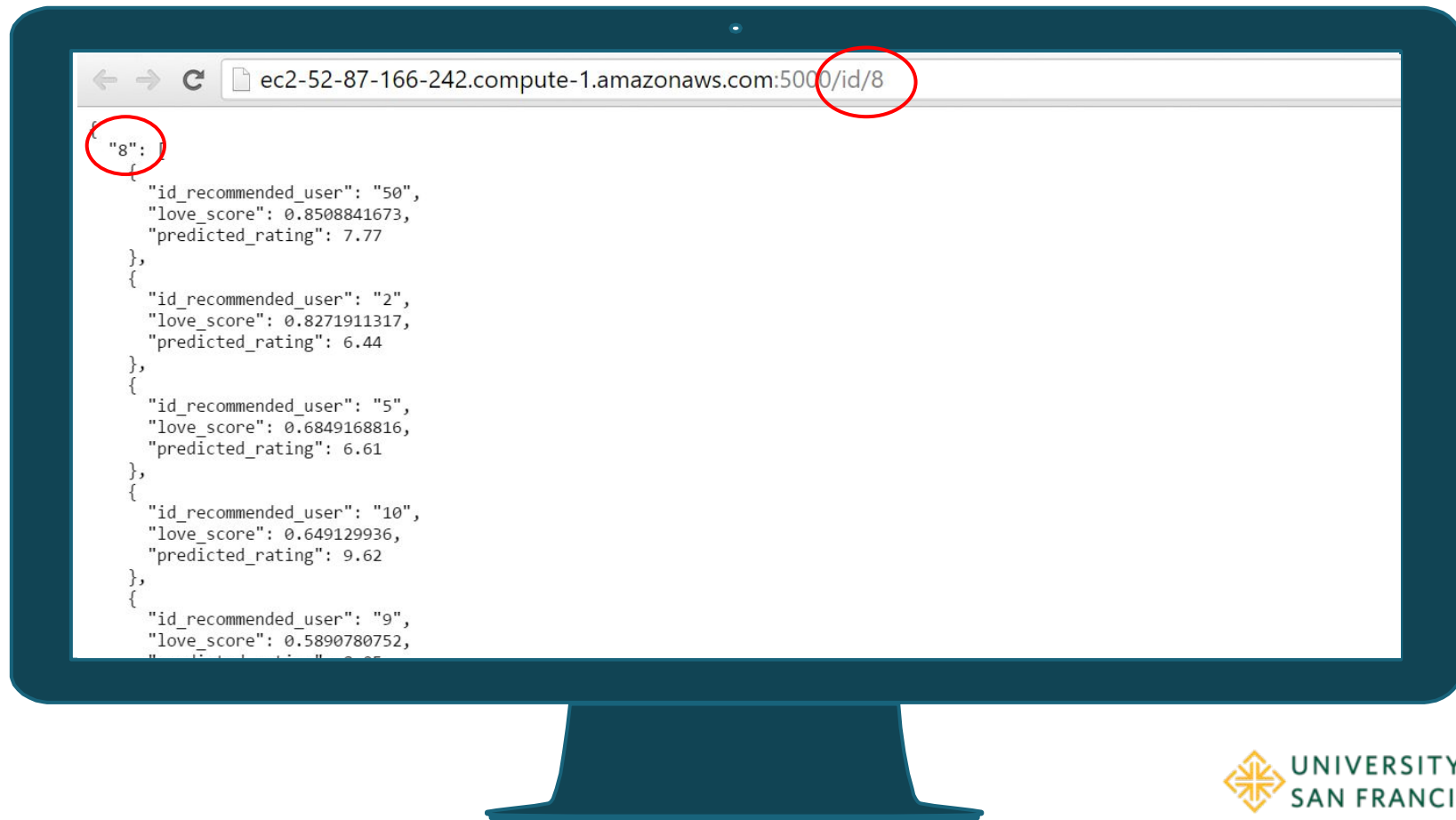
- Using Flask app running on EC2 server
- Data stored in Amazon RDS (postgresql)
- Request of recommendations by user_id:
<http://ec2-52-87-166-242.compute-1.amazonaws.com:5000/id/8>



How it looks



How it looks



Thanks

You can find us at LinkedIn:



Felipe Chamma - [/fchamma](#)



Felipe Formenti - [/felipeformenti](#)



Ryan Speed - [/speedryan](#)



UNIVERSITY OF
SAN FRANCISCO