Kernelized Probabilistic Matrix Factorization

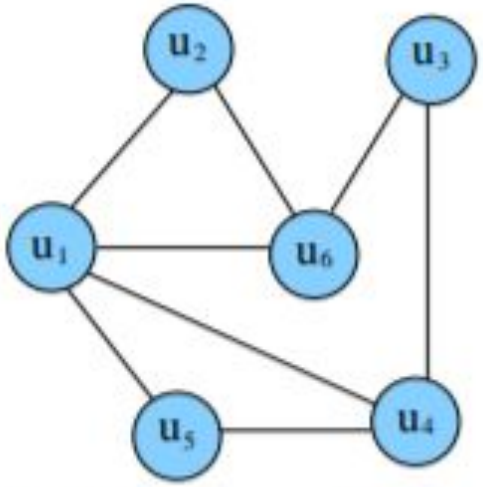
Alexeychuk Nikita
Belikova Kristina
Gazdieva Milena
Kuzin Mikhail
Maksimov Ivan

Problem statement

Skoltech
Skolkovo Institute of Science and Technology

Matrix completetion problem

1. Recommendation systems
2. Image restoration



(a) Social network graph

	i_1	i_2	i_3	i_4
u_1	3	5	?	5
u_2	?	1	?	4
u_3	3	?	4	1
u_4	?	?	5	5
u_5	5	?	?	2
u_6	?	4	2	?

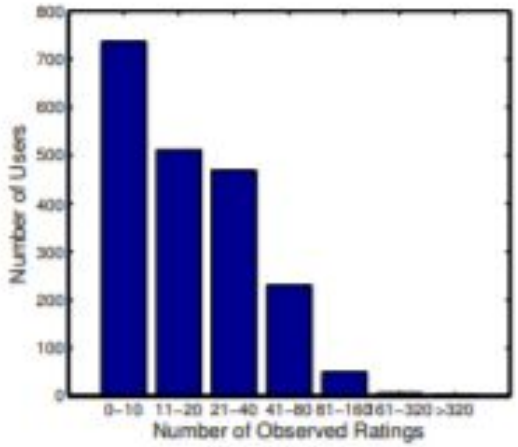
(b) Rating matrix

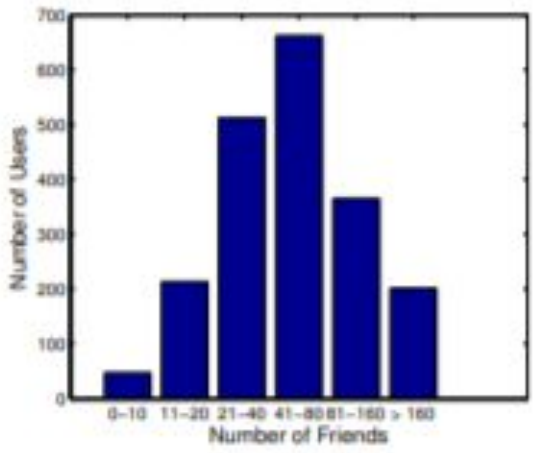
Methodology - Dataset

Skoltech
Skolkovo Institute of Science and Technology

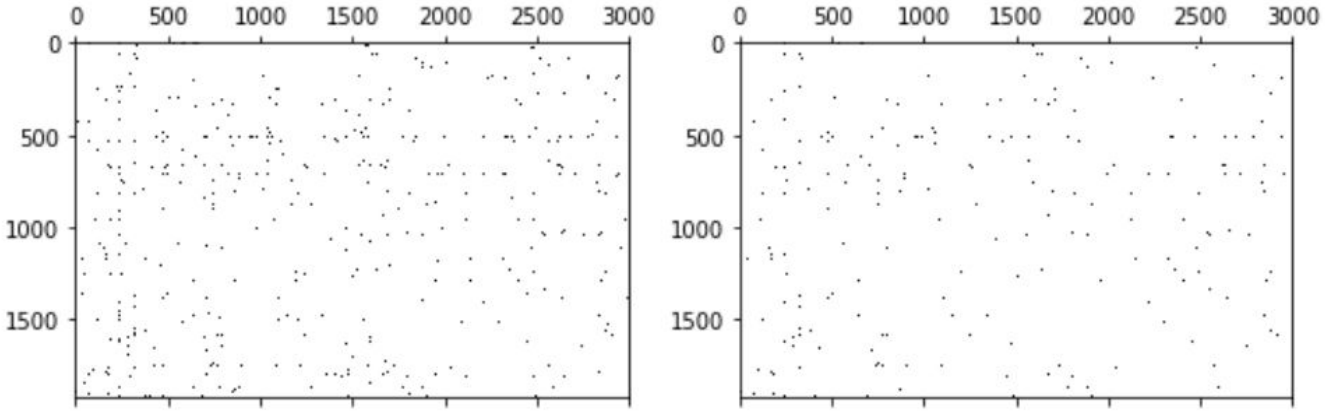
Dataset: Epinion

Number of Users	2 000
Number of Items	3 000
Number of Rating	60 485
Number of Relations	32 548
User-Item matrix density	1.01%

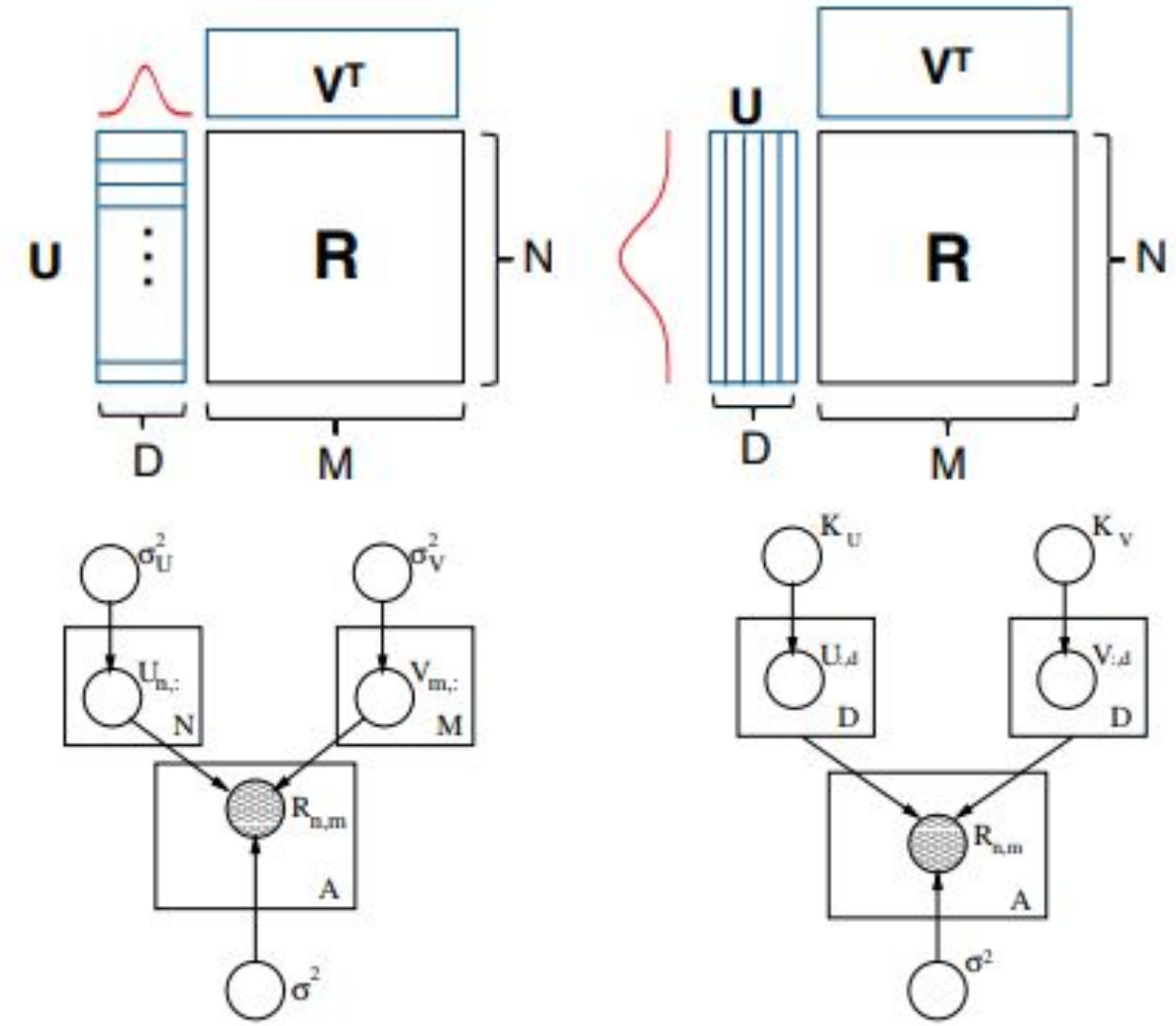




1. Get User-Item matrix (X_full)
2. Drop randomly 25% of its elements (X_trun)
3. Run decomposition on X_trun, trying to impute missing values => X_trun_imputed
4. RMSE(X_trun; X_trun_imputed)



Methodology - PMF vs KPMF



- PMF
  - U is sampled in a “row-wise” manner
  - independent latent v. for each row
- KPMF
  - U is sample in a “column-wise”
  - latent vectors spanning all rows
  - uses side information

KPMF: Maximize the log-posterior

Skoltech
Skolkovo Institute of Science and Technology

$$
\begin{aligned}
&\log p(U, V | R, \sigma^2, K_U, K_V) \\
&= -\frac{1}{2\sigma^2} \sum_{n=1}^{N} \sum_{m=1}^{M} \delta_{n,m} (R_{n,m} - U_{n,:} V_{m,:}^T)^2 \\
&\quad - \frac{1}{2} \sum_{d=1}^{D} U_{:,d}^T S_U U_{:,d} - \frac{1}{2} \sum_{d=1}^{D} V_{:,d}^T S_V V_{:,d} \\
&\quad - A \log \sigma^2 - \frac{D}{2} (\log |K_U| + \log |K_V|) + C,
\end{aligned}
$$

Methodology - Graph Kernels

Skoltech
Skolkovo Institute of Science and Technology

Define a similarity measure for users' taste

- Diffusion kernel

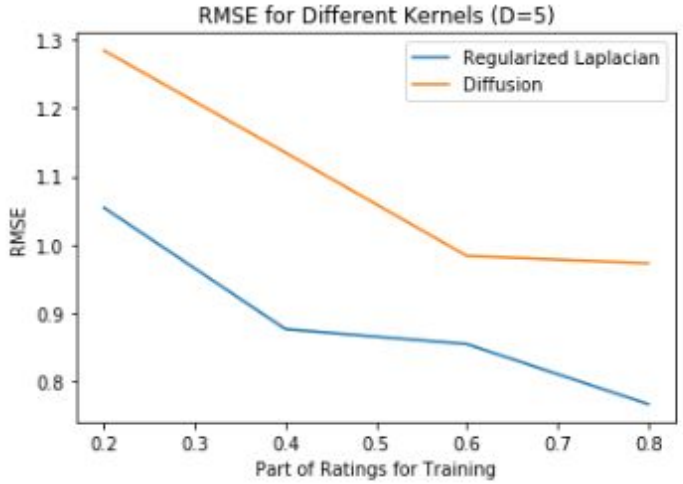
$$K_D = \lim_{n\to\infty}\left(1-\frac{\beta L}{n}\right)^n = e^{-\beta L},$$

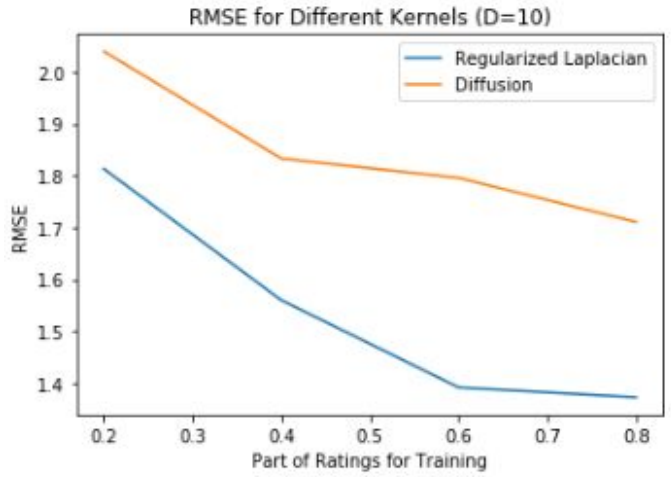
- Regularized laplacian

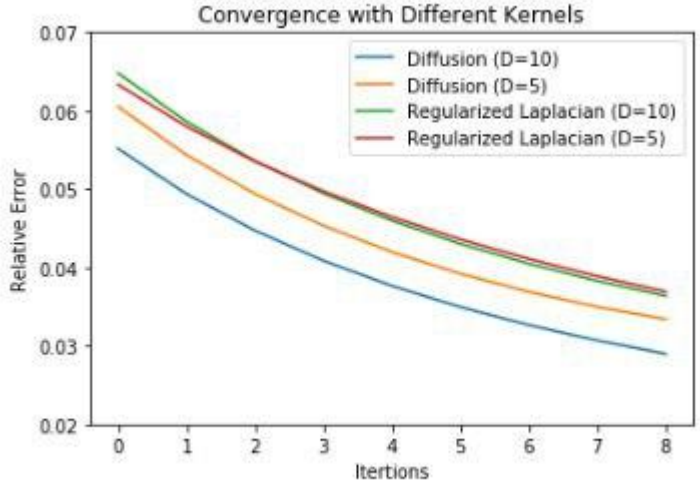
$$K_{RL} = (I+\gamma L)^{-1},$$

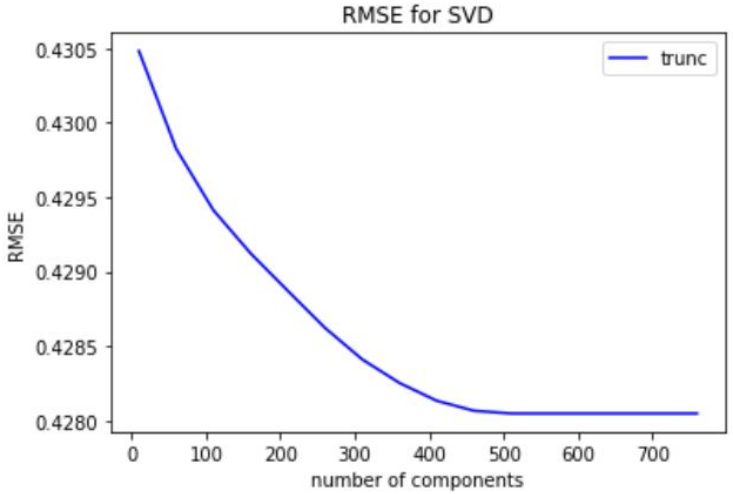
Results

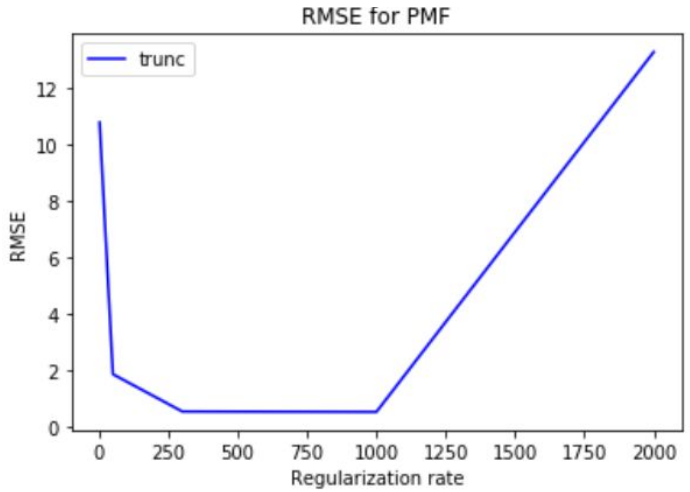
Skoltech
Skolkovo Institute of Science and Technology











Model with top parameters	RMSE
KPMF (D = 5; Kernel = 'Regularized Laplacian') **(need more iteration!)**	0.763
PMF (regularization_rate = 300)	0.431
SVD (n_comp = 550)	0.423