THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Collaborative Filtering with Awareness of Social Networks

Xianshi Yu

Collaborators: Bing-Yi Jing, Ningchen Ying & Ting Li

Collaborative Filtering with Awareness of Social Networks

Contents

Introduction and Review

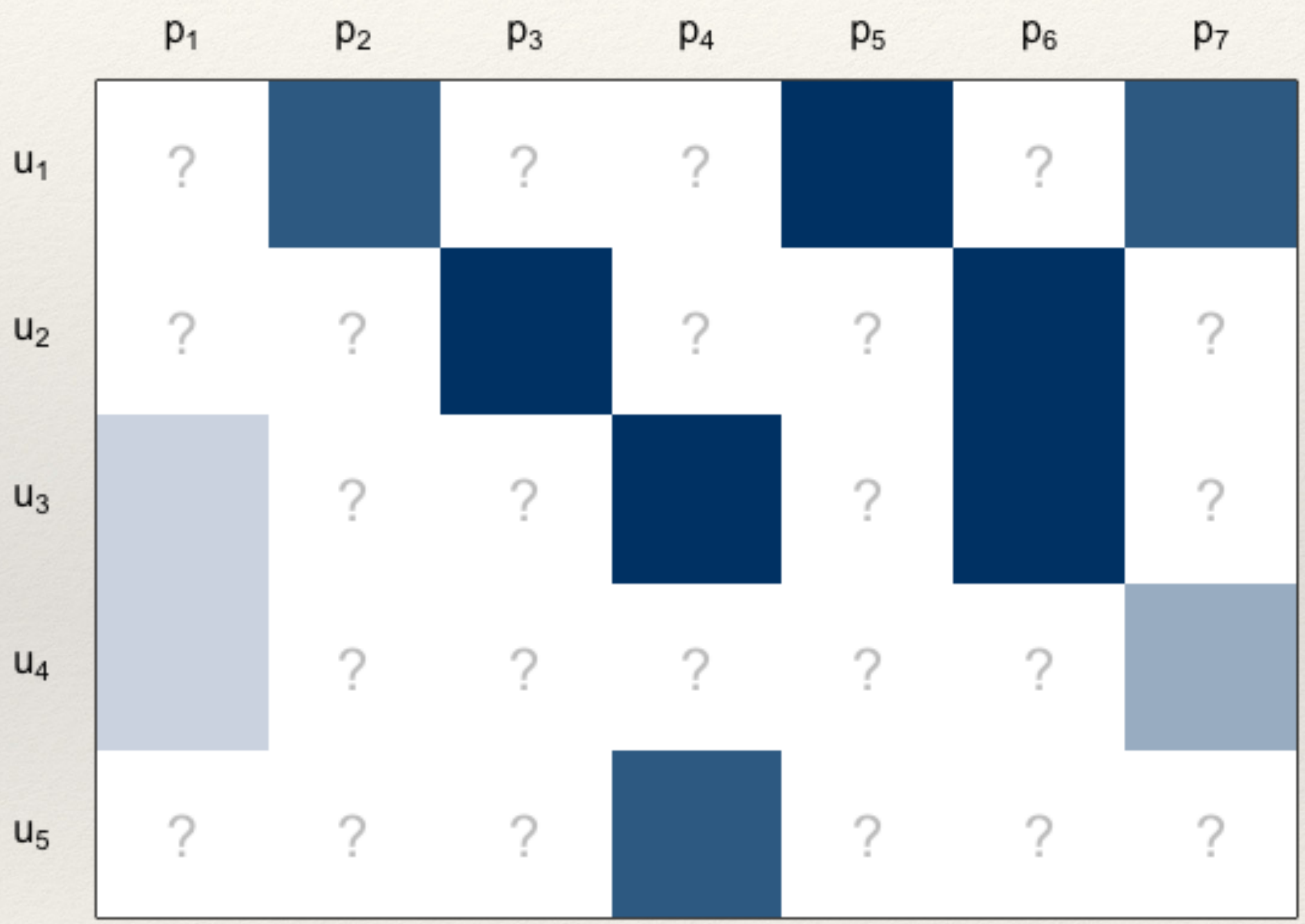
- collaborative filtering (CF)
- using social network information in CF

NetRec Method

- objective function
- algorithm
- numerical results
- theoretical results

Collaborative Filtering (CF)

Collaborative Filtering

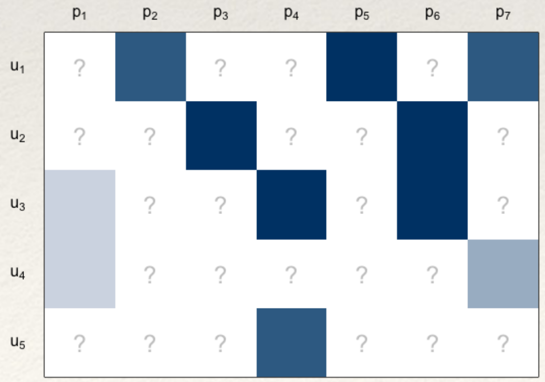


$\mathbf{R}$ is observed on $\Omega \in [n_1] \times [n_2]$

Collaborative Filtering

- Memory Based Methods
- Model Based Methods

Low-Rank Assumption



Collaborative Filtering

$$\text{rank}(\,\cdot\,) \xrightarrow{\text{convex relaxation}} \|\cdot\|_*$$

Candès and Recht 2009

$$\text{minimize } \|X\|_* \\ \text{subject to } \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(R)$$

Mazumder et al. 2010

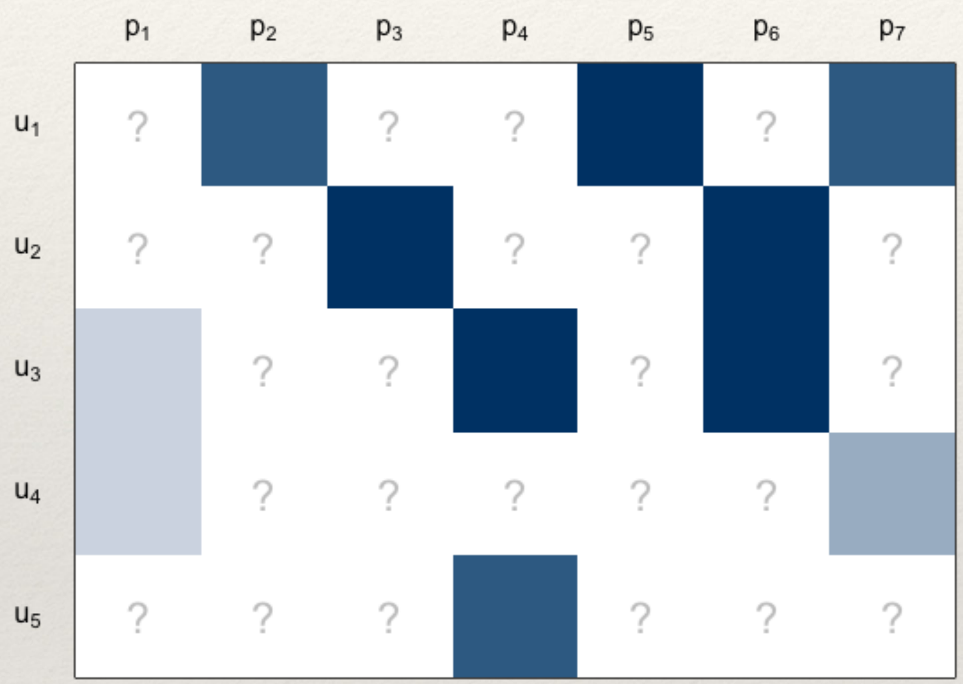
$$\underset{X}{\text{minimize}} \; \|\mathcal{P}_\Omega(X - R)\|_F^2 + \lambda \|X\|_*$$

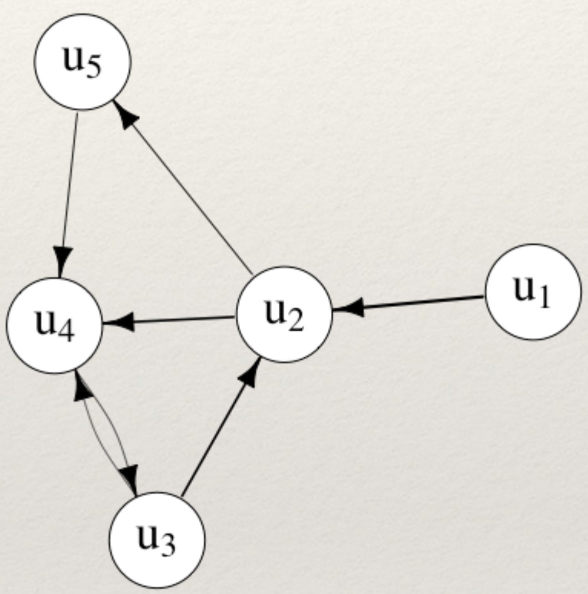
Cai et al. 2010
Mazumder et al. 2010

singular value soft-thresholding algorithm

Using Social Network Information in CF

Using Social Network Information





Using Social Network Information

Previous Methods:

$$R^o = U_{n_1 \times r} V_{n_2 \times r}^T$$

(U and V are respectively user/product feature matrices)

Jamali and Ester 2010
Ma et al. 2011
Yang et al. 2014

$$\underset{U,V}{\text{minimize}} \ \|\mathscr{P}_\Omega(UV^T - R)\|_F^2 + \mathrm{P}_{\text{net}}(U)$$

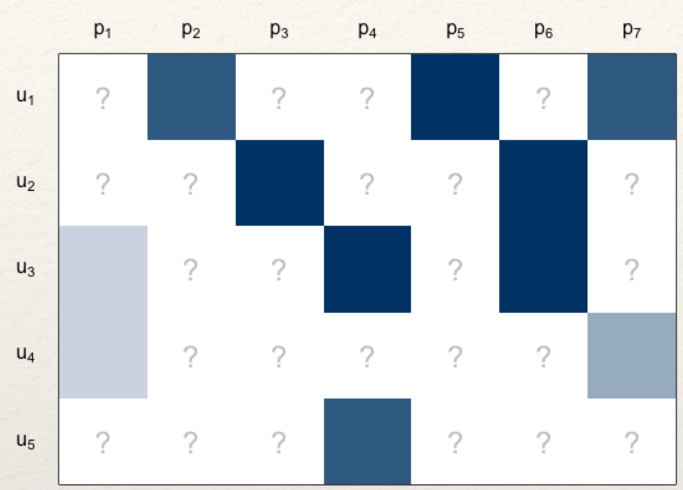
coordinate descent algorithm

- non-convex problem;
- no theoretical characterization about the effect of adding relational information

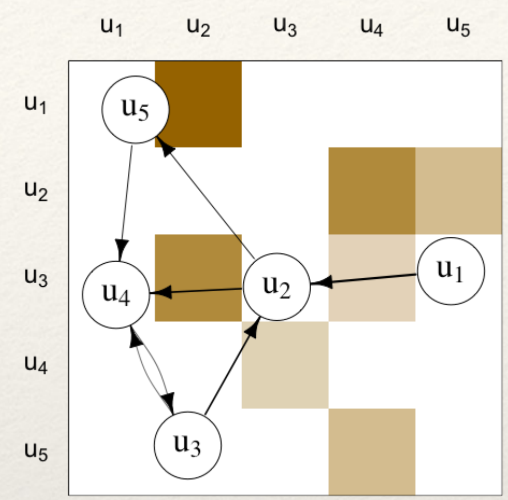
NetRec Method

- objective function
- algorithm
- numerical results
- theoretical results

Objective Function



R_Ω



A

$$\underset{X}{\text{minimize}} \; \|\mathscr{P}_\Omega(X - R)\|_F^2 + \lambda\|X\|_* + \gamma \mathrm{P}_1(X_R) \quad \text{or} \quad + \gamma \mathrm{P}_2(X_R)$$

NetRec1 $P_1(X) = \frac{1}{2}\sum_{i,j} A_{i,j}\|X_{i.} - X_{j.}\|_2^2$

NetRec2 $P_2(X) = \sum_{i=1}^{n_1} \|X_{i.} - \sum_{j=1}^{n_1} \frac{A_{i,j}X_{j.}}{\sum_{j=1}^{n} A_{i,j}}\|_2^2$

and $X_R = \mathscr{P}_{\Omega^c}(X) + \mathscr{P}_\Omega(R)$

Objective Function

why $P_i(X_R)$, not $P_i(X)$?

$$P_1(X) = \frac{1}{2}\sum_{i,j} A_{i,j}\|X_{i.} - X_{j.}\|_2^2$$

$$P_2(X) = \sum_{i=1}^{n_1}\|X_{i.} - \sum_{j=1}^{n_1}\frac{A_{i,j}X_{j.}}{\sum_{j=1}^{n} A_{i,j}}\|_2^2$$

- $P_i(X)$ introduces additional bias
- they encourage X with constant columns
- empirically $P_i(X)$ generates $\hat{X}$ close to zero

Algorithm

Algorithm

$$\underset{X}{\text{minimize}}\ \|\mathscr{P}_{\Omega}(X - R)\|_F^2 + \lambda\|X\|_* + \gamma\ \mathrm{tr}\ X_R^T G X_R$$

$$G = L_{(A+A^T)/2} \text{ or } L^T D^{-2} L, \text{ where } L = D - A, D = \mathrm{diag}(A\mathbf{1})$$

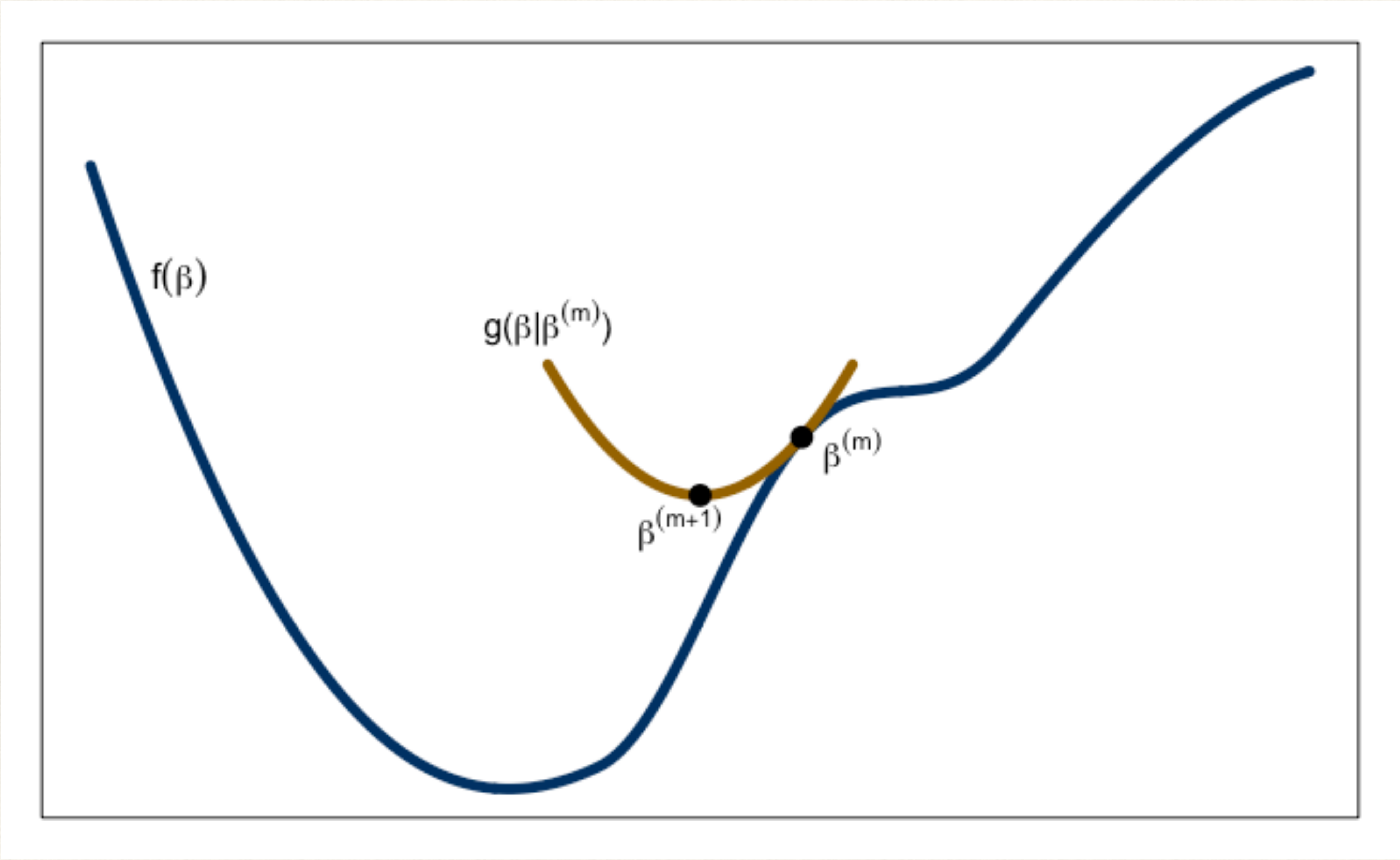
NetRec1 / NetRec2

singular value soft-thresholding algorithm (SVST) still available?

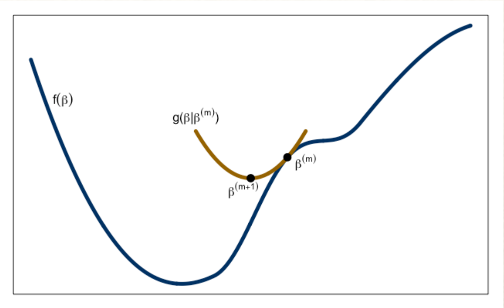
Majorization-Minimization (MM) Algorithm

Algorithm

Majorization-Minimization Algorithm



Algorithm



$$f(X) = \|\mathscr{P}_\Omega(X - R)\|_F^2 + \lambda \|X\|_* + \gamma \operatorname{tr} X_R^T G X_R$$

$$\text{if } \gamma \le \frac{1}{\|G\|} \qquad g(X \mid X^{(m)}) = \|X - h_1(X^{(m)})\|_F^2 + \lambda \|X\|_*$$

$$X^{(m+1)} = \operatorname{argmin} g(X \mid X^{(m)}) = S_{\frac{\lambda}{2}}(h_1(X^{(m)}))$$

$$S_\lambda(X) = U \Sigma_{\lambda +} V^T \; (X = U \Sigma V^T)$$ ← **SVST**

$$\text{if } \gamma > \frac{1}{\|G\|} \qquad X^{(m+1)} = S_{\frac{\lambda}{2\gamma\|G\|}}(h_2(X^{(m)}))$$

$$h_1(X^{(m)}) = \mathscr{P}_{\Omega^c}((I - \gamma G) X_R^{(m)}) + \mathscr{P}_\Omega(R)$$

$$h_2(X^{(m)}) = \mathscr{P}_{\Omega^c}\left(\left(I - \frac{G}{\|G\|}\right) X_R^{(m)}\right) + \mathscr{P}_\Omega\left(X^{(m)} + \frac{R - X^{(m)}}{\gamma \|G\|}\right)$$

Numerical Results

Numerical Results

NetRec1 & NetRec2

Soft-Impute — $\underset{X}{\text{minimize}}\ \|\mathcal{P}_{\Omega}(X - R)\|_F^2 + \lambda\|X\|_*$

RMSE $\sqrt{\underset{i,j \in \Omega^c}{\text{mean}}(\hat{X}_{i,j} - R^o_{i,j})^2}$

Numerical Results

data simulation scheme

$$n_1, n_2, r \longrightarrow R^o$$

- simulate $U_{n_1 \times r}$ and $V_{n_2 \times r}$ independently from uniform distributions of matrices satisfying $U^T U = I$ and $V^T V = I$
- $\text{diag}(\Sigma) \sim_{i.i.d.}$ Beta(1,5)
- $R^o = U \Sigma V^T$

$$R^o, p, \sigma \longrightarrow R_\Omega$$

- elements in $[n_1] \times [n_2]$ are selected to Ω with probability p independently
- $R_\Omega = R^o_\Omega + \varepsilon,$ with $\varepsilon \sim N(0, \sigma^2 I)$

$$R^o, d \longrightarrow A$$

NN(5)

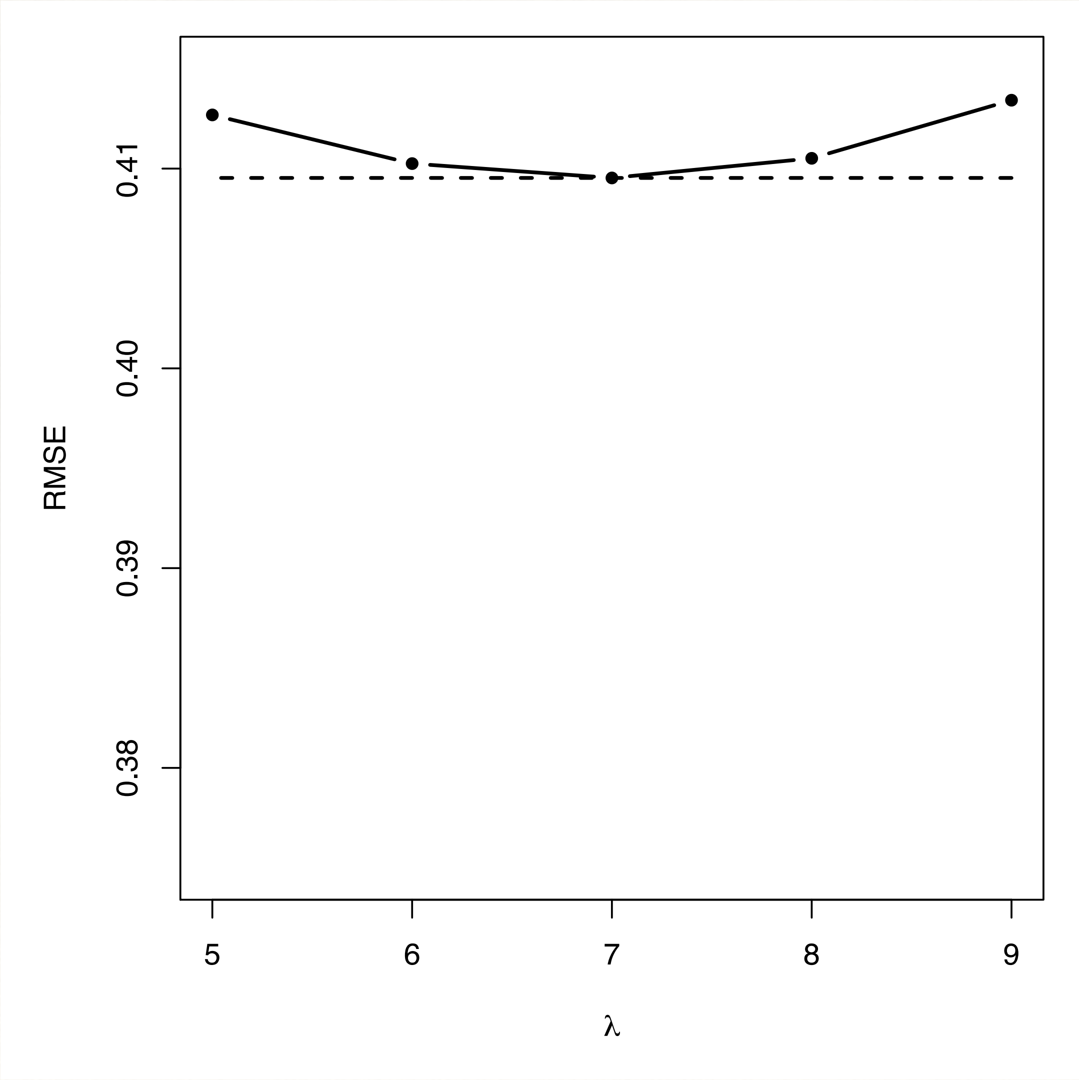
EXP(5)

$$P_{i,j} = \exp(-\alpha d_{i,j})$$

Numerical Results

$$(n_1, n_2, r, p, \sigma) = (2000{,}500{,}10{,}0.05{,}0.3)$$

Soft-Impute — $\underset{X}{\text{minimize}} \; \|\mathscr{P}_\Omega(X - R)\|_F^2 + \lambda \|X\|_*$

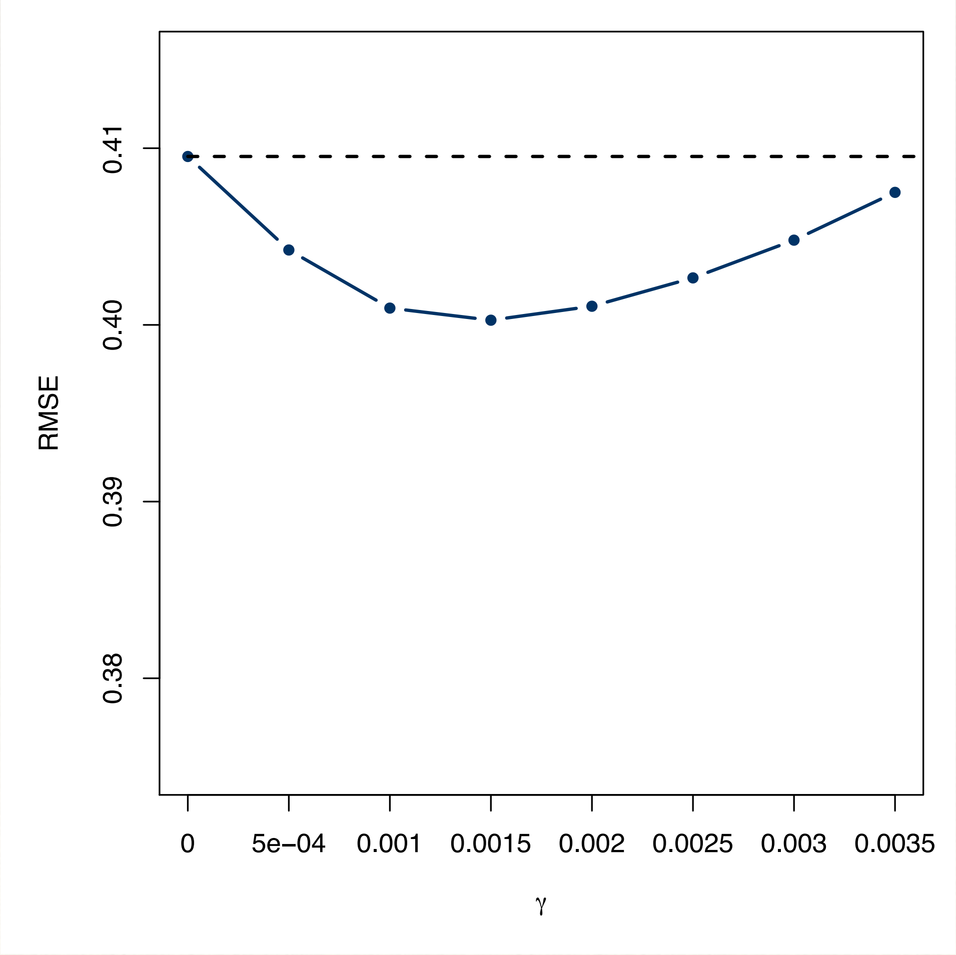


Numerical Results

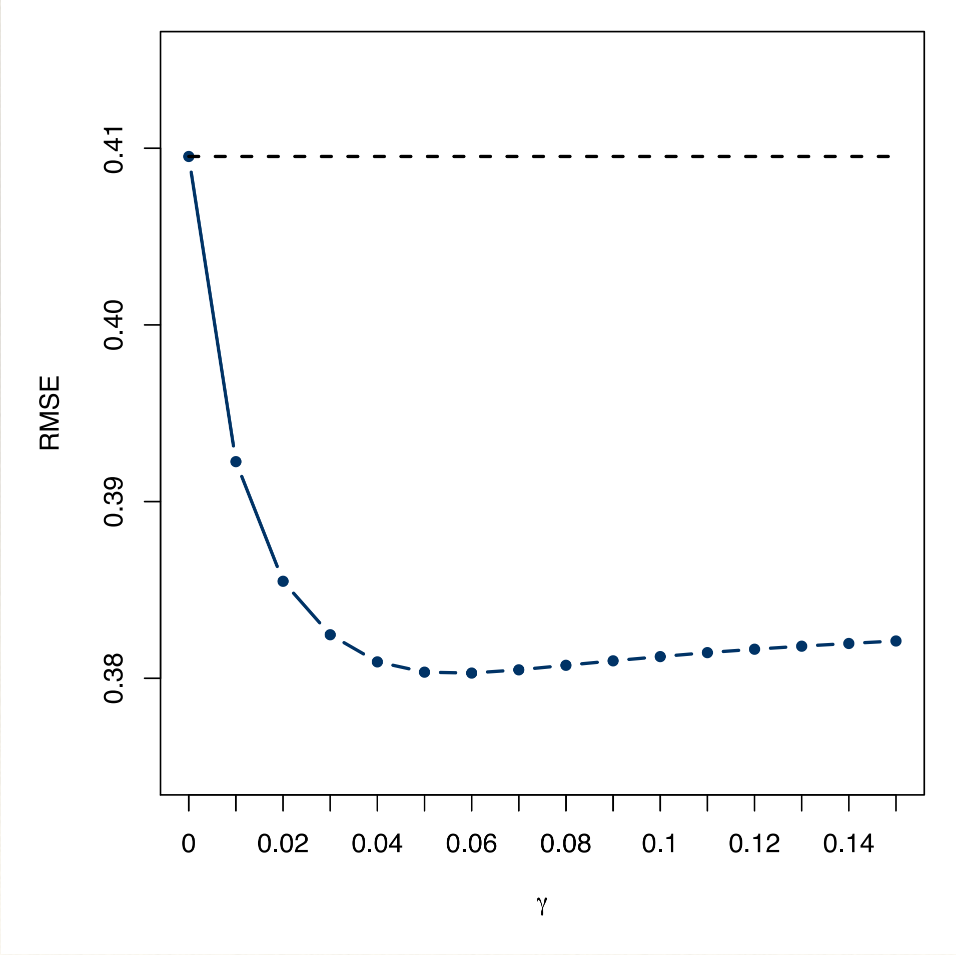
$(n_1, n_2, r, p, \sigma) = (2000{,}500{,}10{,}0.05{,}0.3)$; NN(8)

$$\underset{X}{\text{minimize}}\ \|\mathcal{P}_\Omega(X - R)\|_F^2 + \lambda\|X\|_* + \gamma \mathrm{P}_1(X_R) \text{ or } \gamma \mathrm{P}_2(X_R)$$

NetRec1



NetRec2

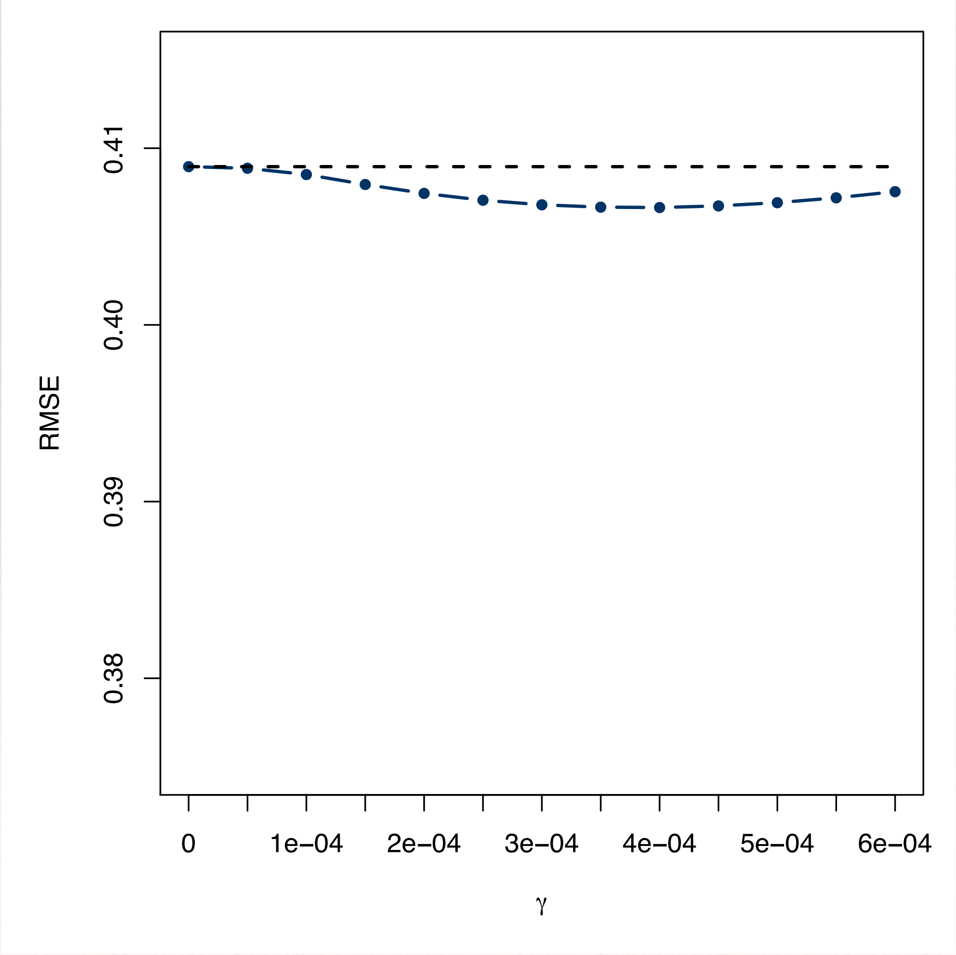


Numerical Results

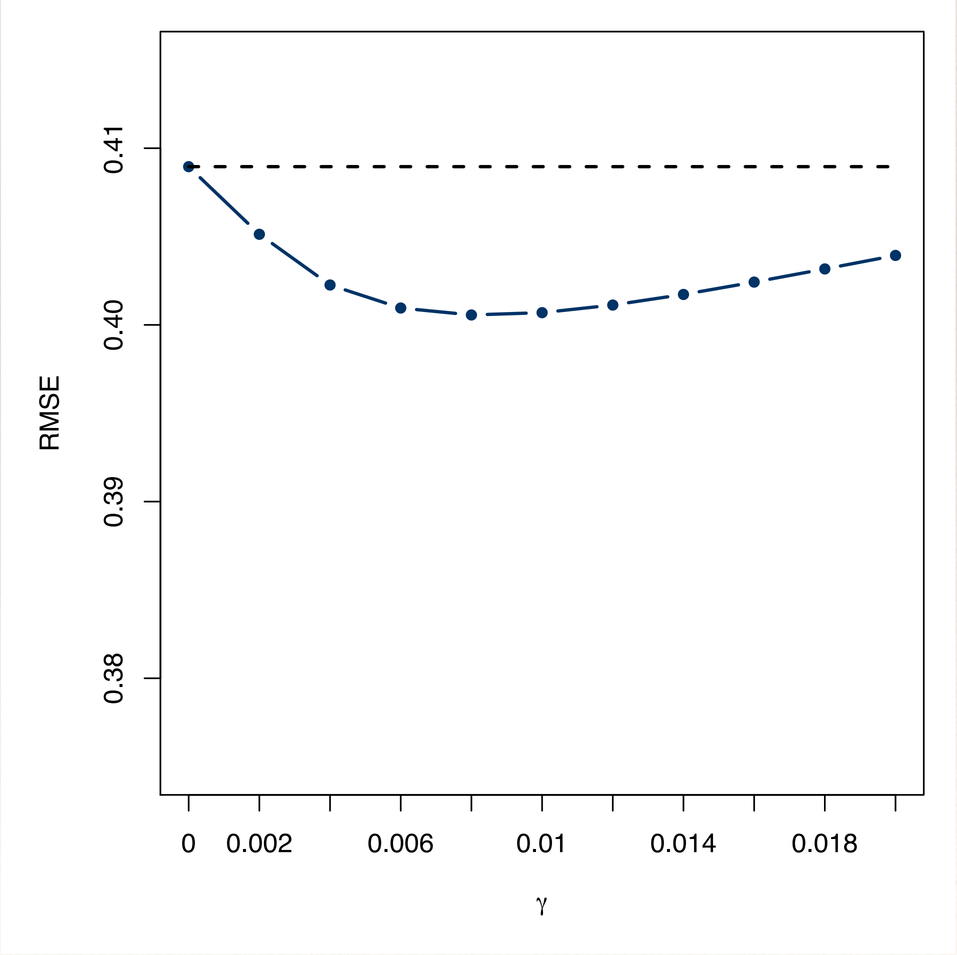
$(n_1, n_2, r, p, \sigma) = (2000{,}500{,}10{,}0.05{,}0.3)$; EXP(8,0.05)

$$\underset{X}{\text{minimize}}\ \|\mathscr{P}_\Omega(X - R)\|_F^2 + \lambda\|X\|_* + \gamma \mathrm{P}_1(X_R) \text{ or } \gamma \mathrm{P}_2(X_R)$$

NetRec1

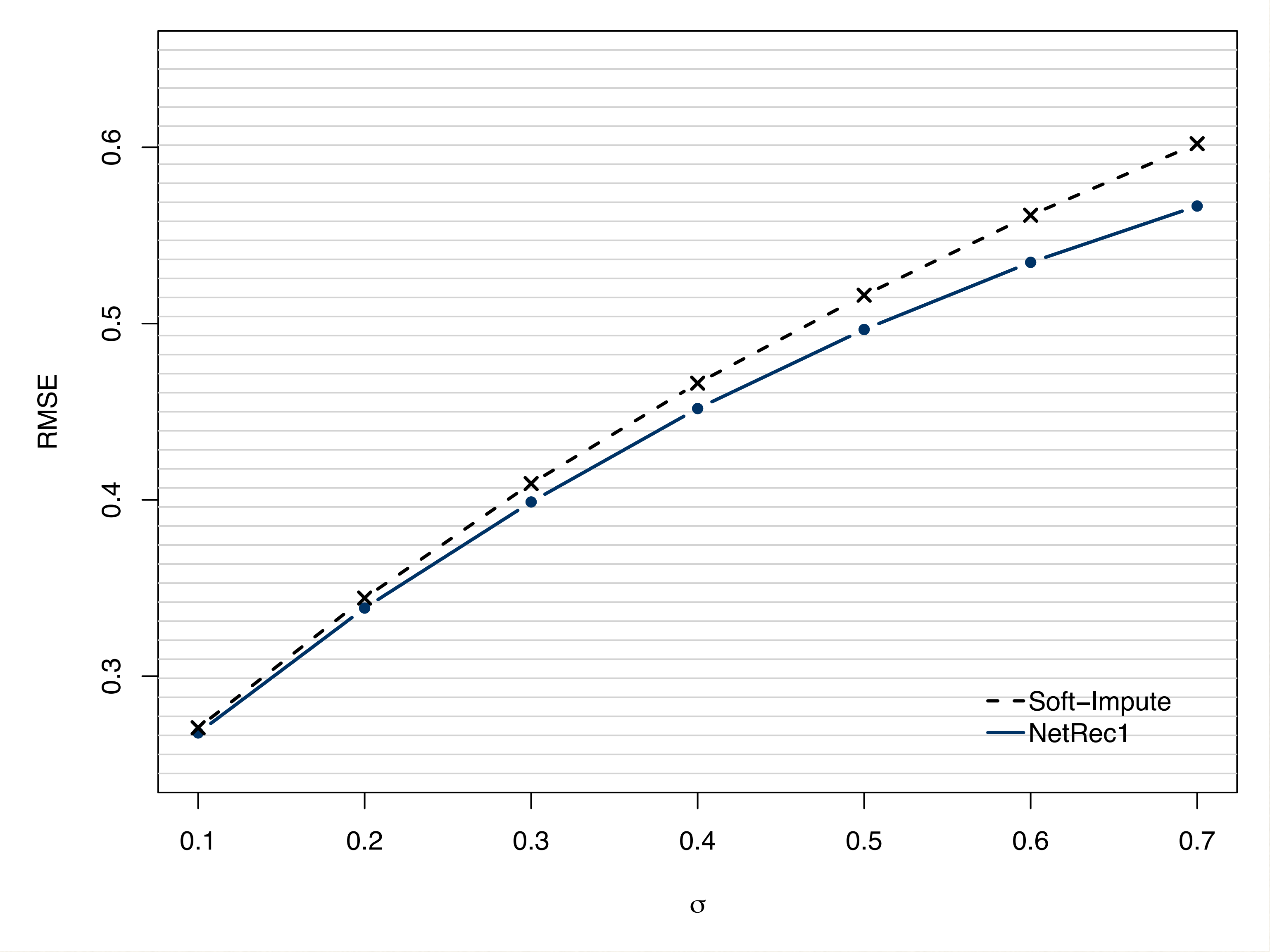


NetRec2



Numerical Results

when noise magnitude varies



$(n_1, n_2, r, p) = (2000{,}500{,}10{,}0.05)$; NN(8)

Numerical Results

www.yelp.com/dataset

data preprocessing

- select restaurants located in a city
- select reviews of these restaurants
- select customers who gave these reviews
- iteratively delete customers/restaurants with less than S reviews

city name	S	user count n_1	restaurant count n_2	review count	average friend count
Edinburgh	5	2029	1268	29317	7.13
Cleveland	6	2205	1196	34270	9.67

examine methods

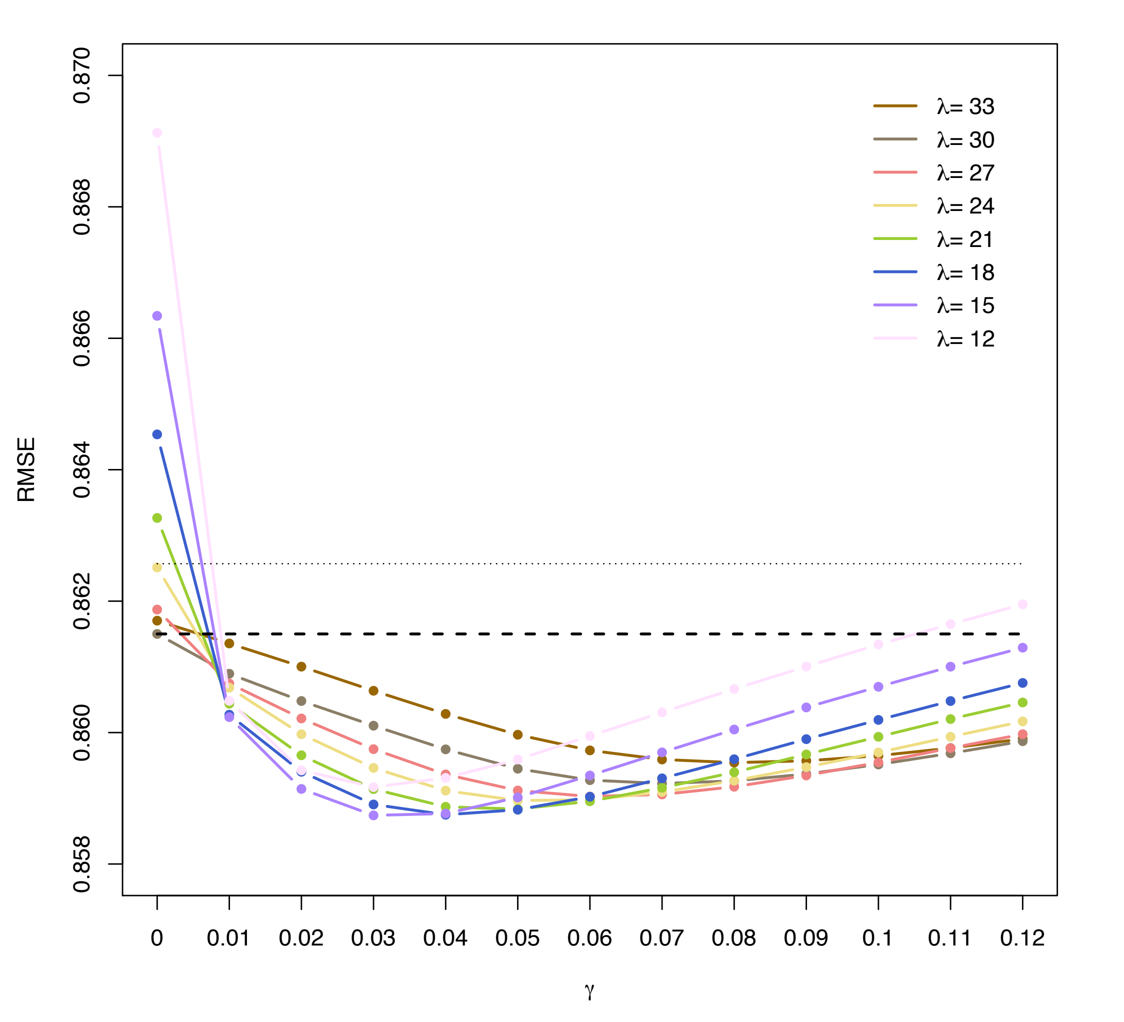
- training : testing = 8:2
- simultaneously center-scale rows and center columns
- apply NetRec/Soft-Impute
- reverse center-scale
- $\max\{\min\{\hat{R}_{i,j},5\},1\}$

NetRec Method

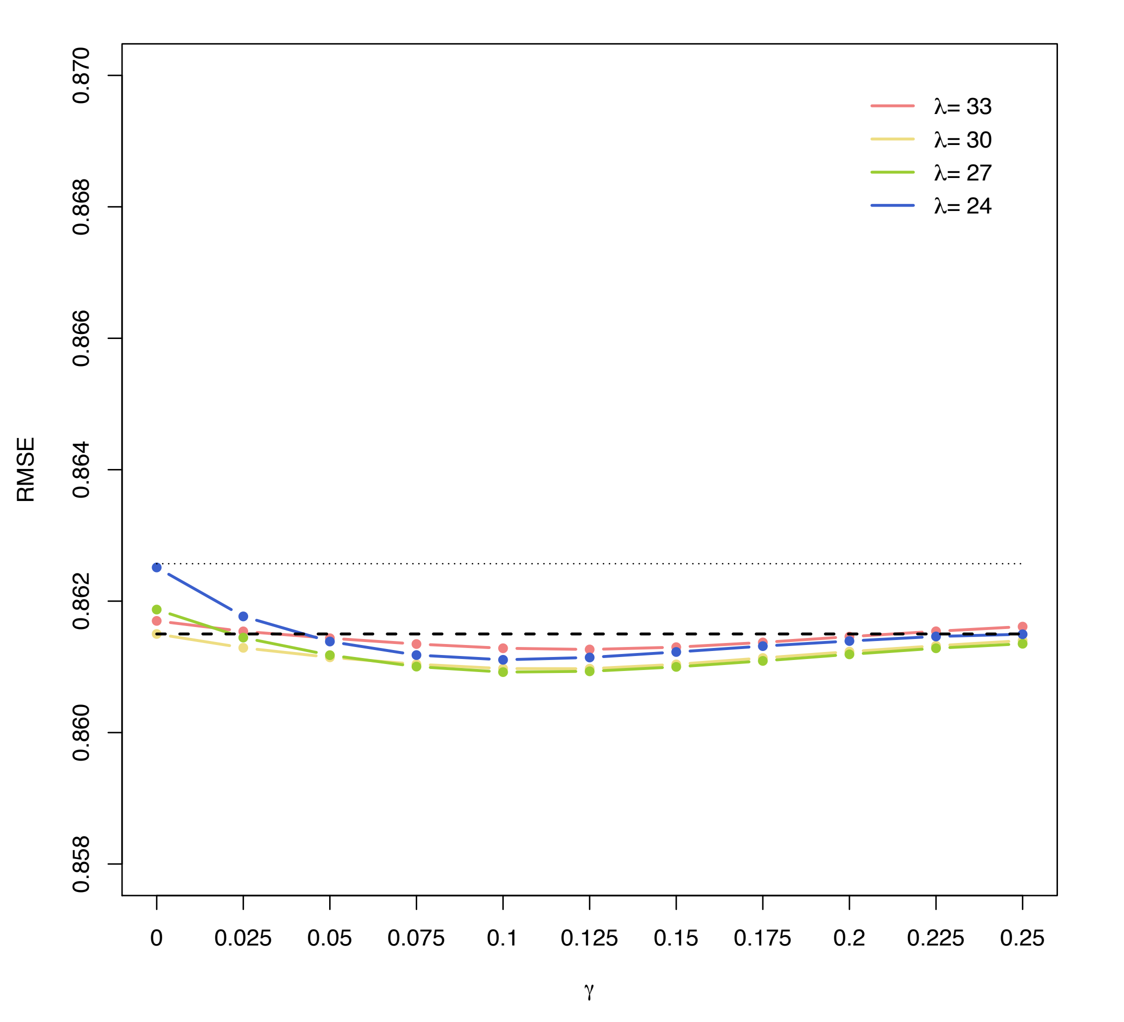
Numerical Results

Edinburgh

NetRec1



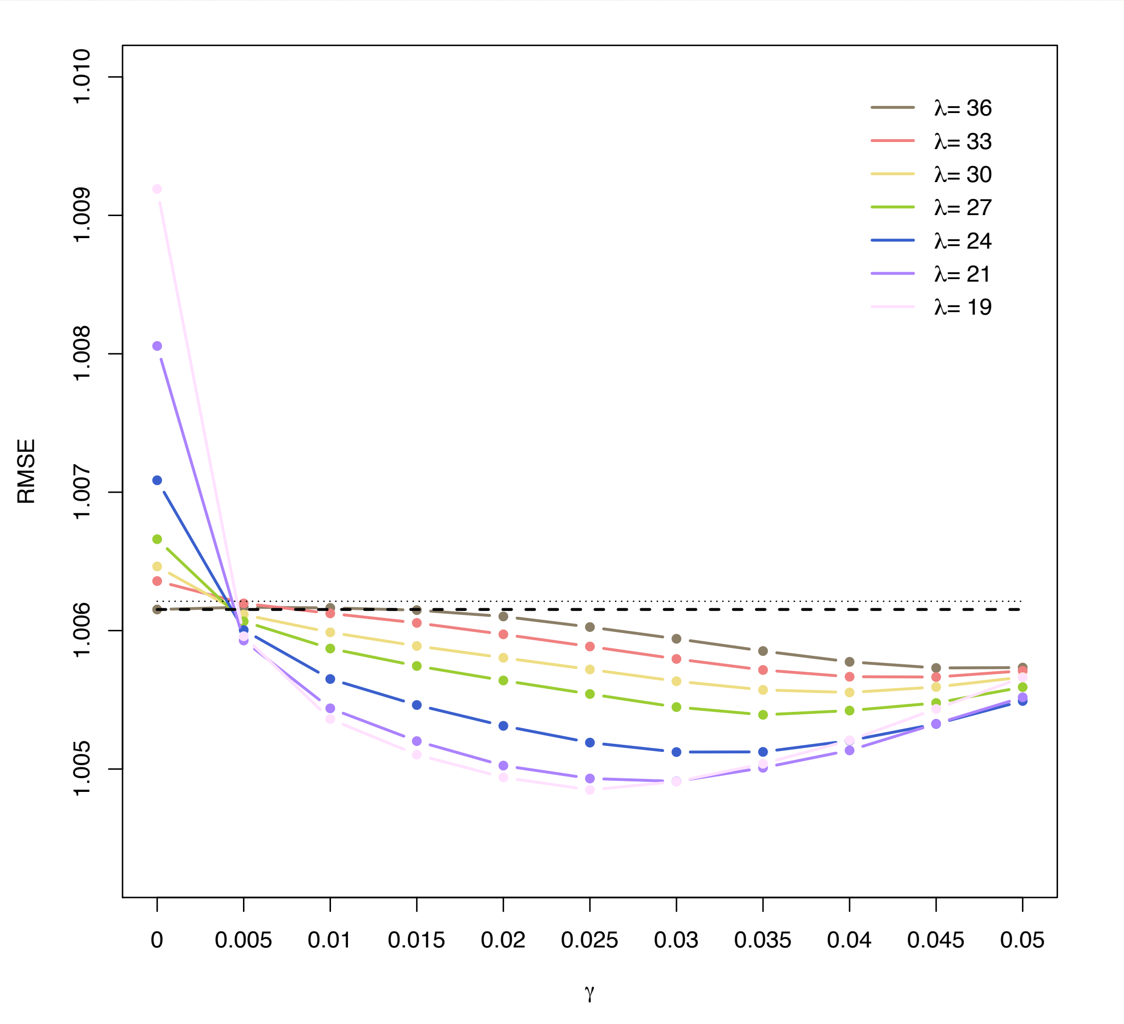
NetRec2



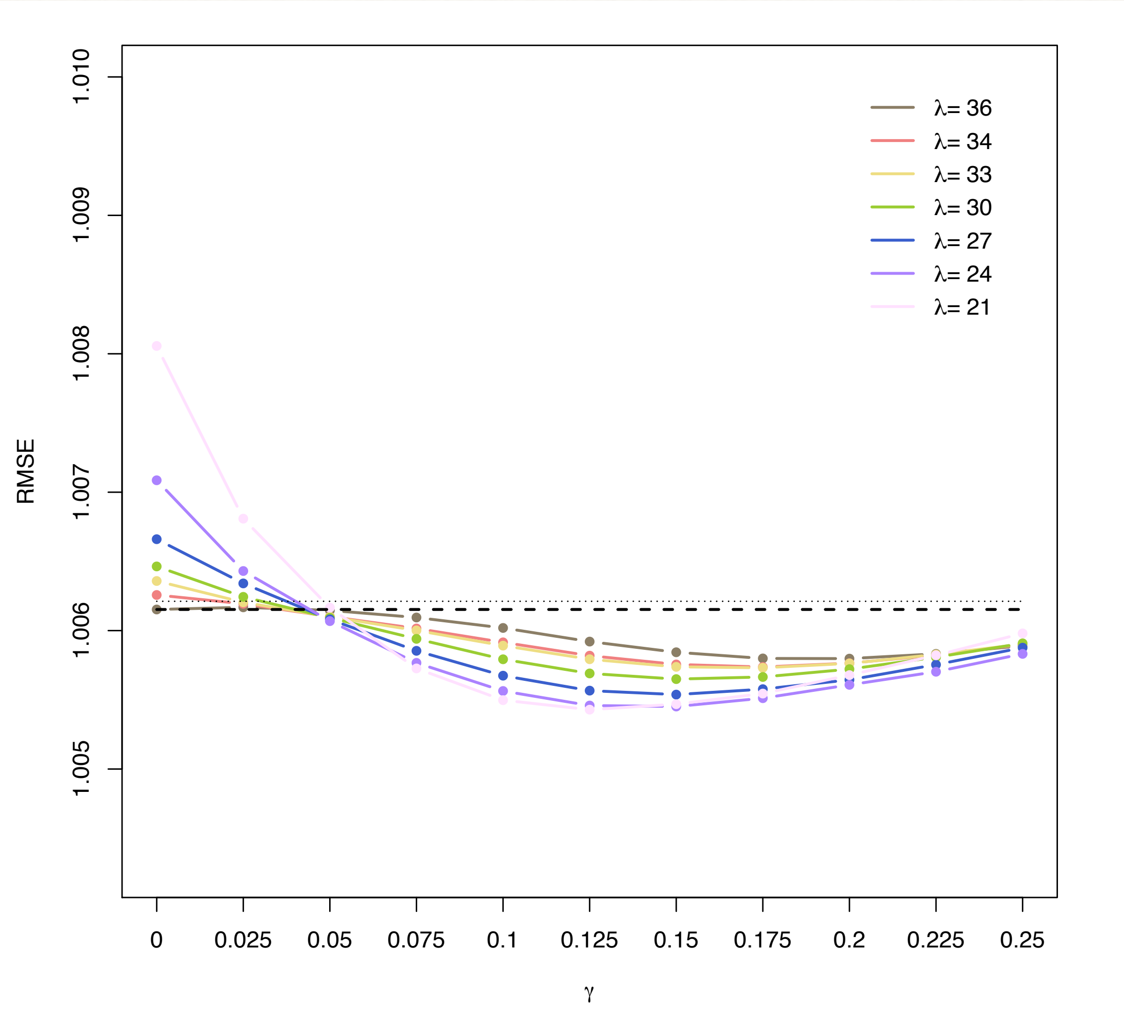
Numerical Results

Cleveland

NetRec1



NetRec2



Theoretical Results

Theoretical Results

Convergence

Theorem 1: the NetRec algorithm converges to the solution of the NetRec1/2 objective function.

Theoretical Results

Error Bound

$$R^o, p, \sigma \longrightarrow R_\Omega$$

- elements in $[n_1] \times [n_2]$ are selected to Ω with probability p independently
- $R_\Omega = R^o_\Omega + \varepsilon,$ with $\varepsilon \sim N(0, \sigma^2 I)$

Theoretical Results

Error Bound

$$\hat{X}_{net} = \underset{X}{\operatorname{argmin}} \|\mathscr{P}_\Omega(X - R)\|_F^2 + \lambda\|X\|_* + \gamma \mathrm{P}_1(X_R)$$

Theorem 2: If R^o obeys the strong incoherence property with parameter μ and $pn_1n_2 \geq C\mu^2 Nr \log^6 N$, let $\delta = \|\mathscr{P}_\Omega(\varepsilon)\|_F$, then with a proper choice of λ and γ, W.H.P.

$$\|\mathscr{P}_{\Omega^c}(\hat{X}_{net} - R^o)\|_F \leq e_{net}(R^o, A, \Omega, \delta)\,.$$

$$\hat{X}_v = \underset{X}{\operatorname{argmin}} \|\mathscr{P}_\Omega(X - R)\|_F^2 + \lambda\|X\|_*$$

Candès and Plan 2010 Under the same condition, with a proper choice of λ, W.H.P.

$$\|\mathscr{P}_{\Omega^c}(\hat{X}_v - R^o)\|_F \leq 4\sqrt{\frac{(2+p)\min(n_1, n_2)}{p}}\delta = e_v$$

Theoretical Results

Theorem 3: If $\frac{\eta}{\delta} < f(R^o, A, \Omega, p)$, then $e_{net} < e_v$ and $e_v - e_{net}$ is strictly increasing in δ.

$$\delta = \|\mathcal{P}_\Omega(\varepsilon)\|_F$$

$$\eta = \|LR^o\|_F, LR^o = \frac{1}{2}\frac{\partial P_1(X)}{\partial X}\bigg|_{R^o}$$

Conclusion

- optimization problem that leverages social network data in collaborative filtering
- convex objective function; SVST algorithm that finds the global optimum
- special effort to reduce the bias introduced by the network related term
- calibrate the improvement brought by relational information
- numerical experiments with the Yelp data

References

- Cai, J. F., Candès, E. J. and Shen, Z. (2010) A singular value thresholding algorithm for matrix completion. SIAM Journal on Optimization. 20(4) 1956-1982.
- Candès, E. J. and Recht, B. (2009) Exact matrix completion via convex optimiza- tion. Foundations of Computational mathematics. 9(6) 717.
- Candès, E. J. and Plan, Y. (2010) Matrix completion with noise. Proceedings of the IEEE. 98(6) 925-936.
- Hastie, T., Mazumder, R., Lee, J. D. and Zadeh, R. (2015) Matrix completion and low-rank SVD via fast alternating least squares. Journal of Machine Learning Research. 16 3367-3402.
- Jamali, M. and Ester, M. (2010, September) A matrix factorization technique with trust propagation for recommendation in social networks. In Proceedings of the fourth ACM conference on Recommender systems (pp. 135-142). ACM.
- Mazumder, R., Hastie, T. and Tibshirani, R. (2010) Spectral regularization algorithms for learning large incomplete matrices. Journal of machine learning research. 11(Aug) 2287-2322.
- Ma,H., Zhou,D., Liu,C.,Lyu, M.R.andKing,I.(2011,February) Recommender systems with social regularization. In roceedings of the fourth ACM international con- ference on Web search and data mining (pp. 287-296). ACM.
- Yang,X.,Guo,Y.,Liu,Y.andSteck,H.(2014)A survey of collaborative filtering based social recommender systems. Computer Communications. 41 1-10.
- Yu, X., Li, T, Ying, N. and Jing, B. (2018) Collaborative Filtering with Awareness of Social Networks, submitted.

Q&A