Learning to Recommend

Hao Ma

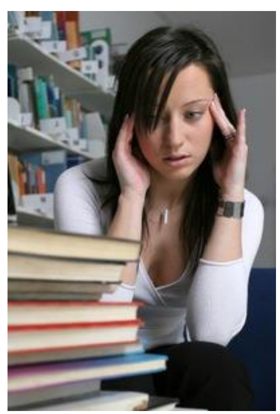
Supervisors: Prof. Irwin King and Prof. Michael R. Lyu

Dept. of Computer Science & Engineering
The Chinese University of Hong Kong

26-Nov-09

How much information is on the web?

Information Overload



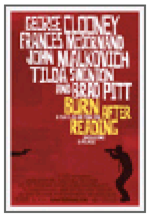
We Need Recommender Systems

YAHOO! MOVIES

My Movies: gabe_ma Edit Profile

Recommendations For You Receive Recommendations by Email

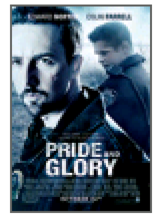
Movies in Theaters: 94089



Burn After Reading (R)
Showtimes & Tickets | Add to My Lists

Yahoo! Users:	B-	4794 ratings
The Critics:	B	14 reviews

Don't Recommend Again · Seen It? Rate It!



Pride and Glory (R)
Showtimes & Tickets | Add to My Lists

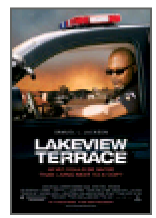
Yahoo! Users:	A-	59 ratings
The Critics:	C+	6 reviews

Don't Recommend Again · Seen It? Rate It!

Fight Club (R)
Showtimes & Tickets | Add to My Lists

Yahoo! Users:	B+	52392 ratings
The Critics:	B	12 reviews

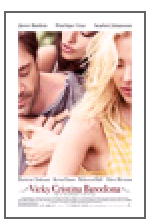
Don't Recommend Again · Seen It? Rate It!



Lakeview Terrace (PG-13)
Showtimes & Tickets | Add to My Lists

Yahoo! Users:	B	3229 ratings
The Critics:	C	12 reviews

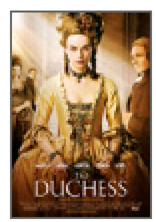
Don't Recommend Again · Seen It? Rate It!



Vicky Cristina Barcelona (PG-13)
Showtimes & Tickets | Add to My Lists

Yahoo! Users:	B	1923 ratings
The Critics:	B+	13 reviews

Don't Recommend Again · Seen It? Rate It!



The Duchess (PG-13)
Showtimes & Tickets | Add to My Lists

Yahoo! Users:	B+	953 ratings
The Critics:	B-	10 reviews

Don't Recommend Again · Seen It? Rate It!

See All Recommendations

iTunes 8

iTunes
DEA vs. Heroin Kingpin
DEA
0:00
-0:30
View
Search
LIBRARY
Music
Movies
TV Shows
Podcasts
Applications
Radio
Ringtones
STORE
iTunes Store
Purchased
Purchased on A...
PLAYLISTS
iTunes DJ
Genius
ADC on iTunes
Pink Floyd
TV Shows
60's Music
Music Videos
My Top Rated
Recently Played
Top 25 Most Pl...
00-lil_kim-the_...
8Mile
50 Cent - Gues...
Acquisition
Aerosmith-MTV...
Animals
Atmosphere
Atmosphere Ov...
Beastie Boys - ...
TV Shows
Genres
New
Anthony Bourdain: No...
Season 1
Anthony Bourdain: No...
Season 3
Black Gold
Season 1
Cops
Season 20
CSI: NY
Season 5
Play Selected
DEA
Easy Money
Season 1
The First 48
Season 7
Kitchen Nightmares
Season 1
Law & Order
Season 17
Law & Order: SVU
Season 10
Genius Sidebar
DEA
More Episodes From This Season
DEA, Season 1
Reality TV
21 Ratings
DEA vs. Heroin Kingpin
DEA, Season 1 $1.99 BUY
Episode 1
Operation Pill Grinder
DEA, Season 1 $1.99 BUY
Episode 3
Show 1 Remaining Episode
More From DEA
The Six Million Doll...
DEA, Season 2
Season 2, Ep. 1
Reality TV
$1.99 BUY
$2.99 BUY HD
Genius Recommendations
Preview
Carrier
Season 1, Ep. 1
Nonfiction
$1.99 BUY
American Gangster
Gangland, Season 1
Season 1, Ep. 1
Nonfiction
$1.99 BUY
Scorpio
Flashpoint, Seaso...
Season 1, Ep. 1
$1.99 BUY
11 TV shows, 8.9 hours, 8.74 GB
Computer

iLike™

Songs from friends and similar people

Play All | Buy all

Victims by The Oppressed
New! Traditional Byrd69

Skinhead Girl by The Oppressed
New! Traditional Byrd69

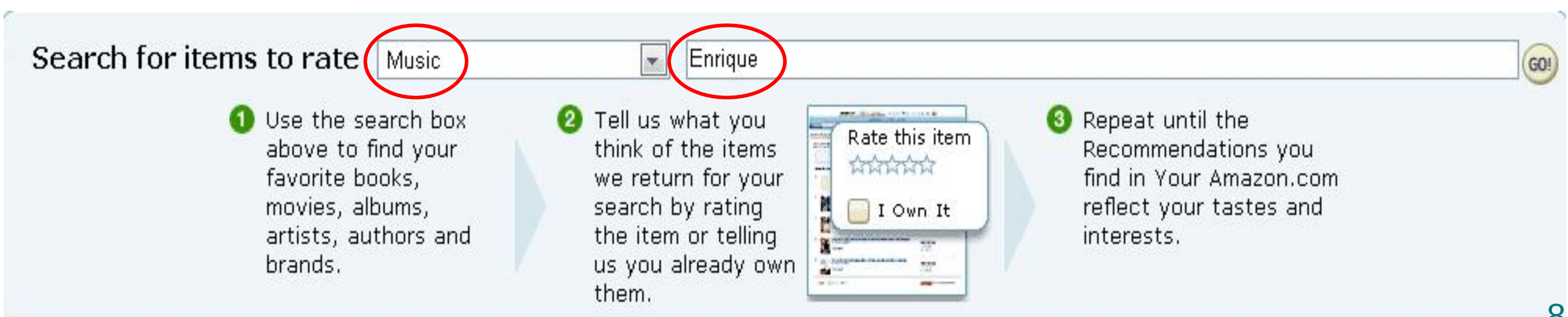
King Of The Jungle by Last Resort
New! Traditional Byrd69

Violence In Our Minds by Last Resort
New! Traditional Byrd69

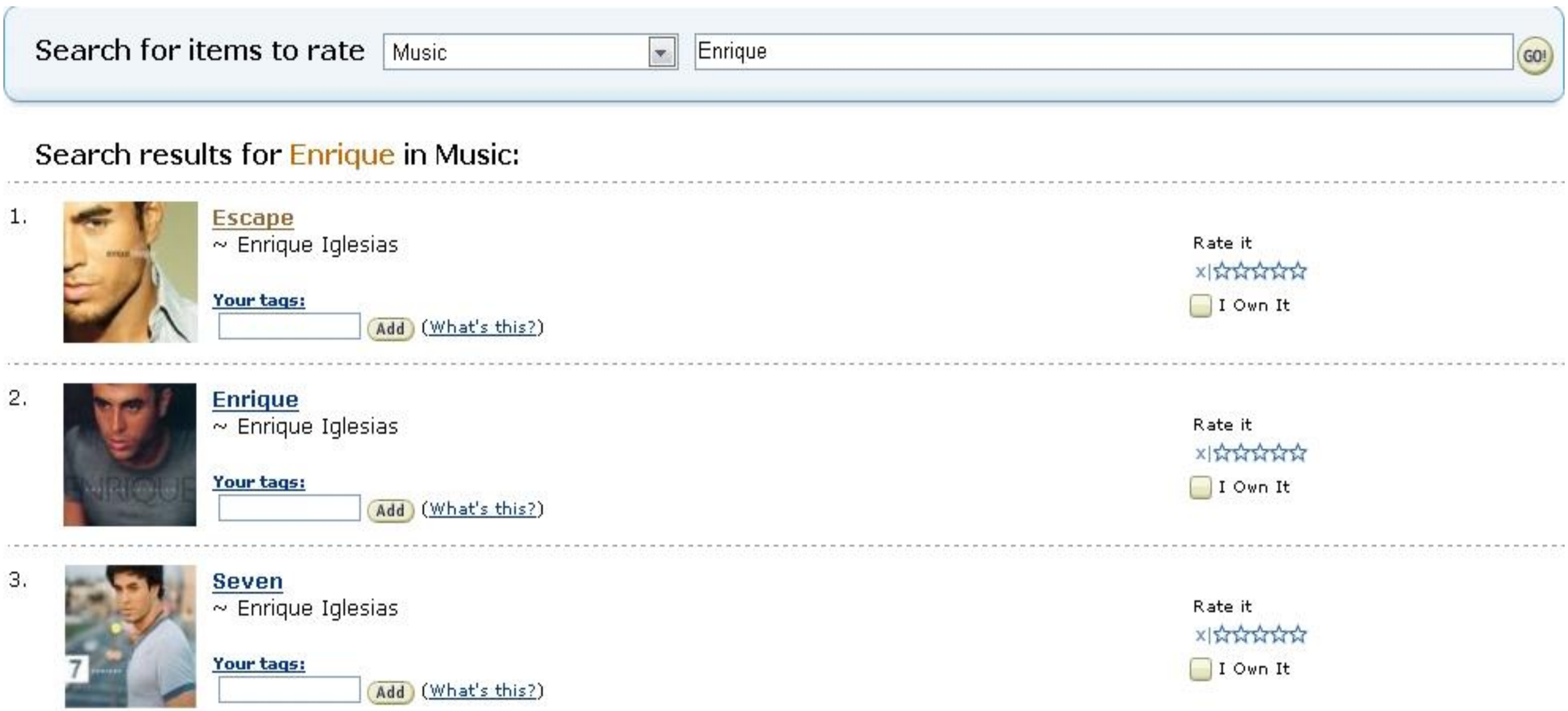
Violence by The Templars
New! Traditional Byrd69

View all | **invite more friends**

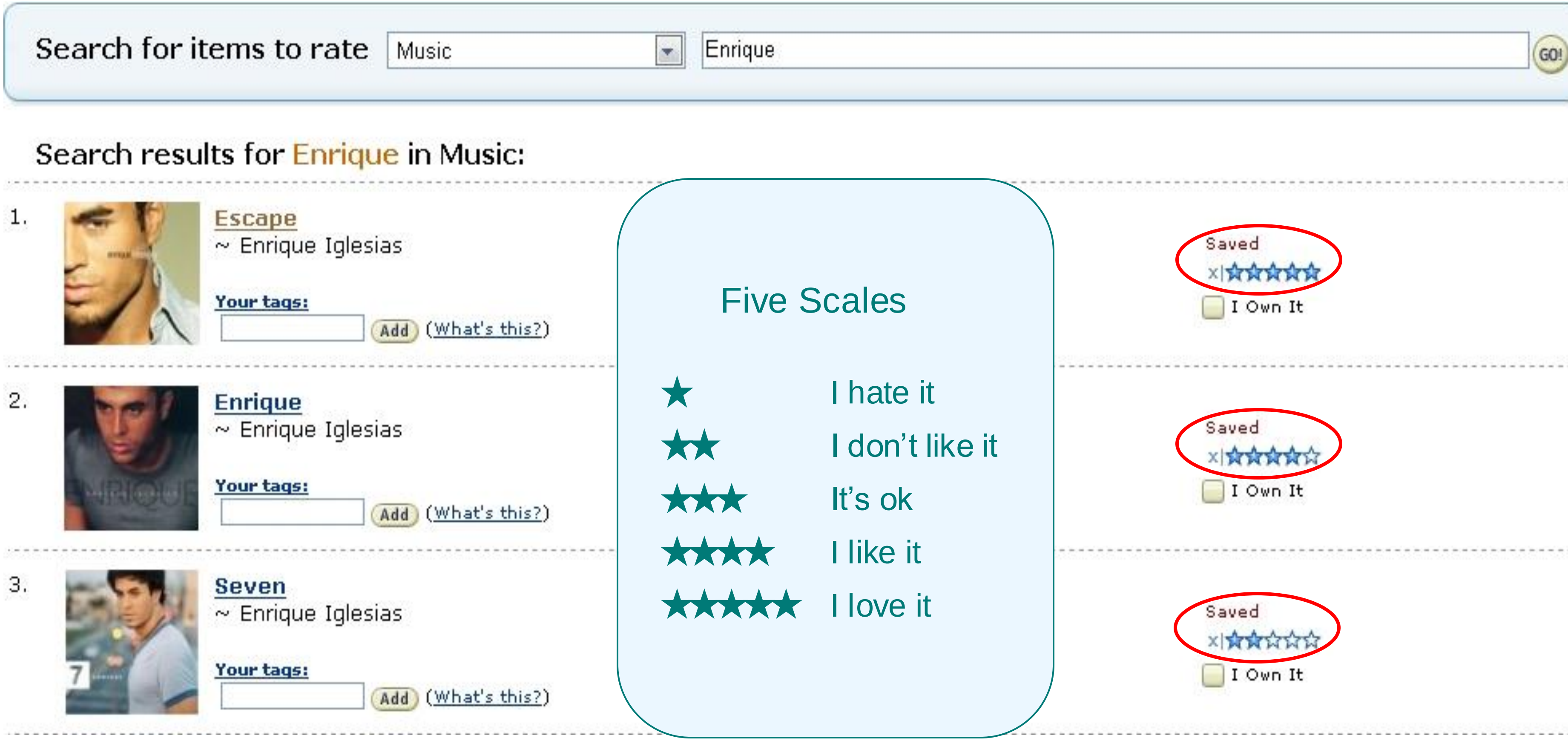
5-scale Ratings



5-scale Ratings



5-scale Ratings



Five Scales

★	I hate it
★★	I don't like it
★★★	It's ok
★★★★	I like it
★★★★★	I love it

Traditional Methods

- Memory-based Methods (Neighborhood-based Method)
  - Pearson Correlation Coefficient
  - User-based, Item-based
  - Etc.
- Model-based Method
  - Matrix Factorizations
  - Bayesian Models
  - Etc.

	v_1	v_2	v_3	v_4	v_5	v_6
u_1		5	2		3	
u_2	4			3		4
u_3			2			2
u_4	5			3		
u_5		5	5			3

User-based Method

Items

Users

u_1													
u_2	1	3		4		2		5			3	4	
u_3													
u_4		3		4			3	4		3	4		4
u_5													
u_6	1			3	5	2		4	1			3	

Matrix Factorization

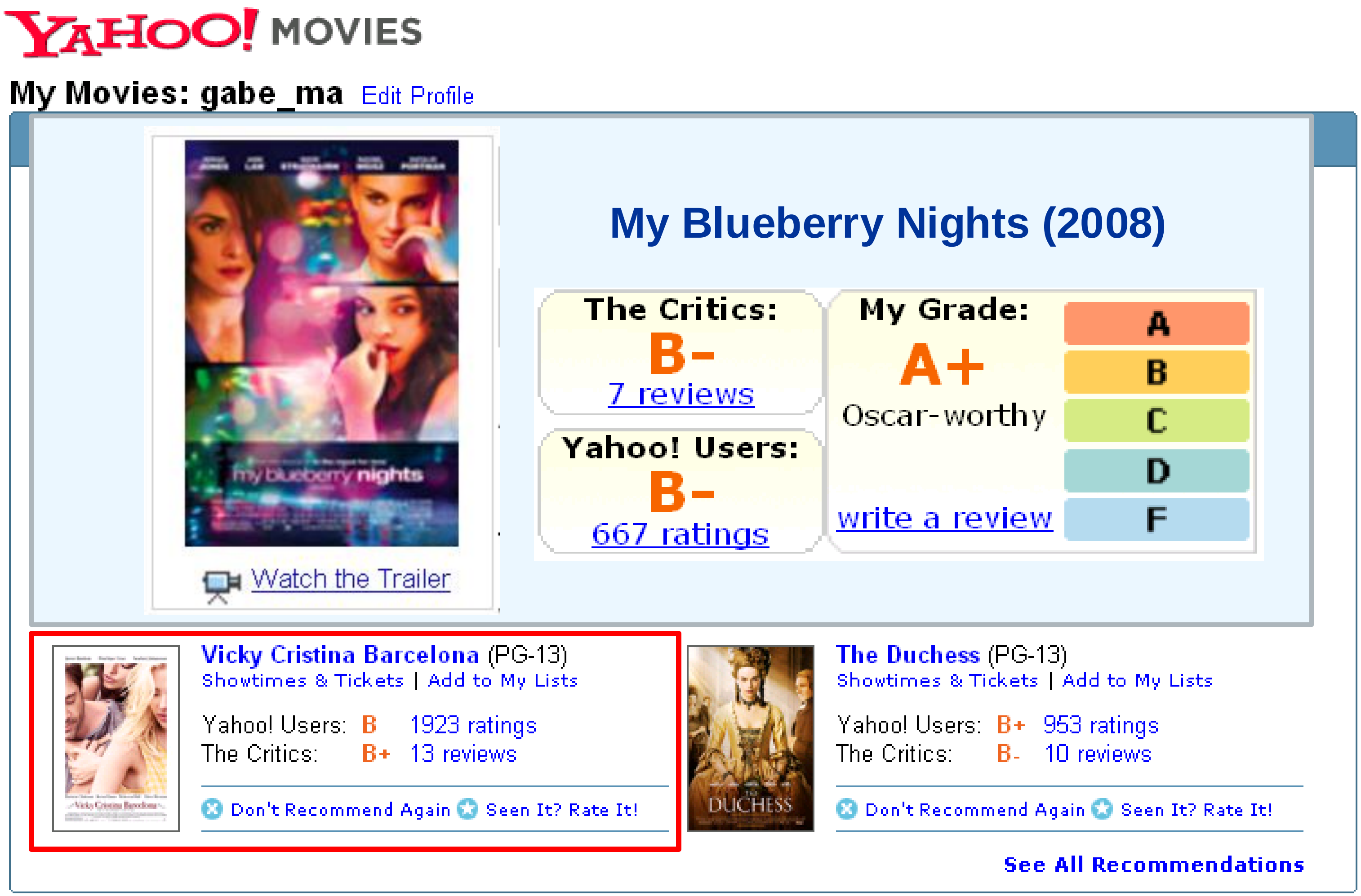
	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	5	2		3		4		
u_2	4	3			5			
u_3	4		2				2	4
u_4								
u_5	5	1	2		4	3		
u_6	4	3		2	4		3	5

	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8
u_1	5	2	**2.5**	3	**4.8**	4	**2.2**	**4.8**
u_2	4	3	**2.4**	**2.9**	5	**4.1**	**2.6**	**4.7**
u_3	4	**1.7**	2	**3.2**	**3.9**	**3.0**	2	4
u_4	**4.8**	**2.1**	**2.7**	**2.6**	**4.7**	**3.8**	**2.4**	**4.9**
u_5	5	1	2	**3.4**	4	3	**1.5**	**4.6**
u_6	4	3	**2.9**	2	4	**3.4**	3	5

$$U = \begin{bmatrix} 1.55 & 1.22 & 0.37 & 0.81 & 0.62 & -0.01 \\ 0.36 & 0.91 & 1.21 & 0.39 & 1.10 & 0.25 \\ 0.59 & 0.20 & 0.14 & 0.83 & 0.27 & 1.51 \\ 0.39 & 1.33 & -0.43 & 0.70 & -0.90 & 0.68 \\ 1.05 & 0.11 & 0.17 & 1.18 & 1.81 & 0.40 \end{bmatrix} \quad V = \begin{bmatrix} 1.00 & -0.05 & -0.24 & 0.26 & 1.28 & 0.54 & -0.31 & 0.52 \\ 0.19 & -0.86 & -0.72 & 0.05 & 0.68 & 0.02 & -0.61 & 0.70 \\ 0.49 & 0.09 & -0.05 & -0.62 & 0.12 & 0.08 & 0.02 & 1.60 \\ -0.40 & 0.70 & 0.27 & -0.27 & 0.99 & 0.44 & 0.39 & 0.74 \\ 1.49 & -1.00 & 0.06 & 0.05 & 0.23 & 0.01 & -0.36 & 0.80 \end{bmatrix}$$

Challenges

- Data sparsity problem



Challenges

- Data sparsity problem

My Movie Ratings

The Pursuit of Happyness (PG-13, 1 hr. 57 min.)
Buy DVD | Add to My Lists

Yahoo! Users: B+ 38992 ratings
The Critics: B- 13 reviews

My Rating: A+

Finding Nemo (G, 1 hr. 40 min.)
Buy DVD | Add to My Lists

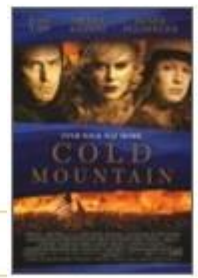
Yahoo! Users: B+ 137394 ratings
The Critics: A- 14 reviews

My Rating: A

My Blueberry Nights (PG-13, 1 hr. 30 min.)
Buy DVD | Add to My Lists

Yahoo! Users: B- 756 ratings
The Critics: B- 7 reviews

My Rating: A+



Cold Mountain (R, 2 hrs. 35 min.)
Buy DVD | Add to My Lists

Yahoo! Users: B 38986 ratings
The Critics: B+ 10 reviews

My Rating: B+

The Lord of the Rings: The Fellowship of the Ring
Buy DVD | Add to My Lists

Yahoo! Users: A- 110957 ratings
The Critics: A 15 reviews

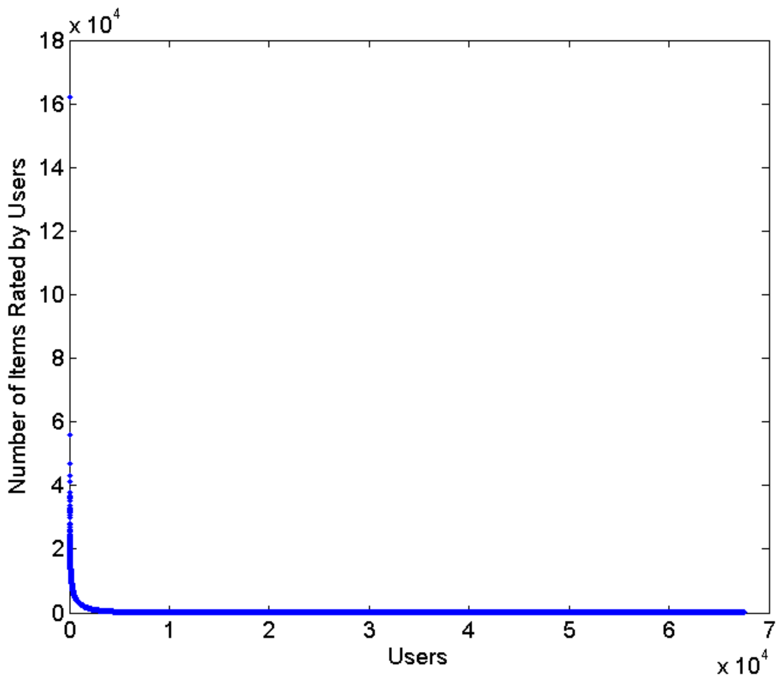
My Rating: A

Shrek 2 (PG, 1 hr. 32 min.)
Buy DVD | Add to My Lists

Yahoo! Users: B+ 150368 ratings
The Critics: B 15 reviews

My Rating: B

Number of Ratings per User



Data Extracted From Epinions.com

Challenges

- Traditional recommender systems ignore the social connections between users

Recommendations from friends

Contents

Traditional Methods

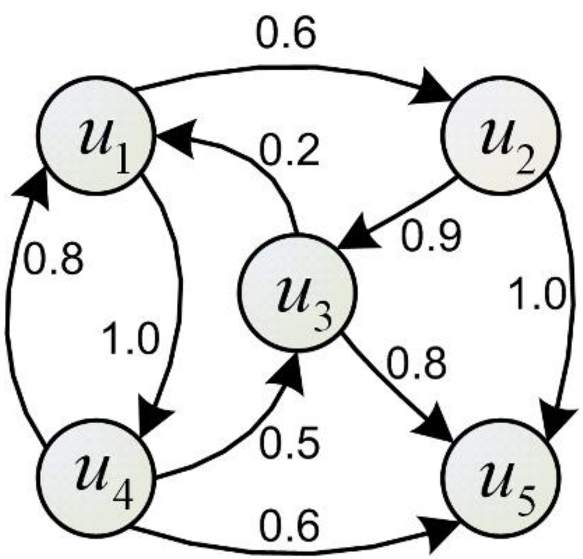
- Chapter 3: Effective Missing Data Prediction
- Chapter 4: Recommend with Global Consistency
- Chapter 5: Social Recommendation
- Chapter 6: Recommend with Social Trust Ensemble
- Chapter 7: Recommend with Social Distrust

Social Recommendation

Chapter 5

Social Recommendation

Problem Definition



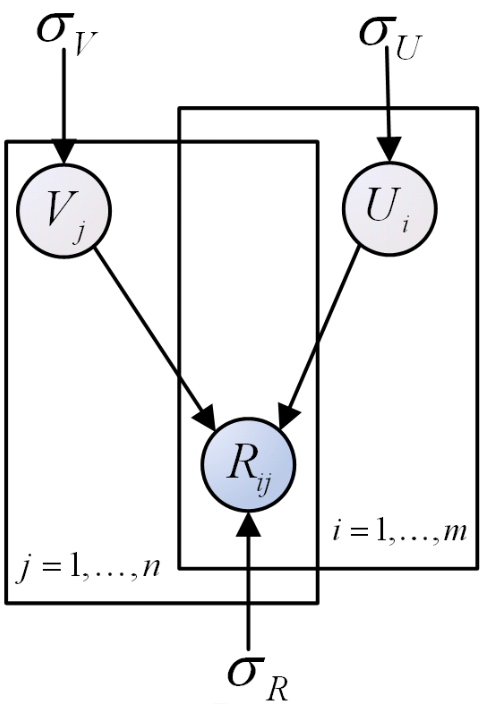
Social Trust Graph

	v_1	v_2	v_3	v_4	v_5	v_6
u_1		5	2		3	
u_2	4			3		4
u_3			2			2
u_4	5			3		
u_5		5	5			3

User-Item Rating Matrix

User-Item Matrix Factorization

	v_1	v_2	v_3	v_4	v_5	v_6
u_1		5	2		3	
u_2	4			3		4
u_3			2			2
u_4	5			3		
u_5		5	5			3



$$p(R|U,V,\sigma_R^2) = \prod_{i=1}^{m} \prod_{j=1}^{n} \left[\mathcal{N}\left(R_{ij}|g(U_i^T V_j), \sigma_R^2\right)\right]^{I_{ij}^R}$$

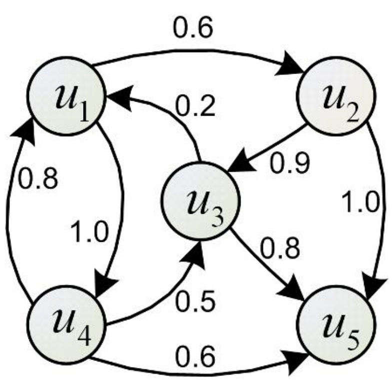
$$p(U|\sigma_U^2) = \prod_{i=1}^{m} \mathcal{N}(U_i|0, \sigma_U^2 \mathbf{I}) \qquad p(V|\sigma_V^2) = \prod_{j=1}^{n} \mathcal{N}(V_j|0, \sigma_V^2 \mathbf{I})$$

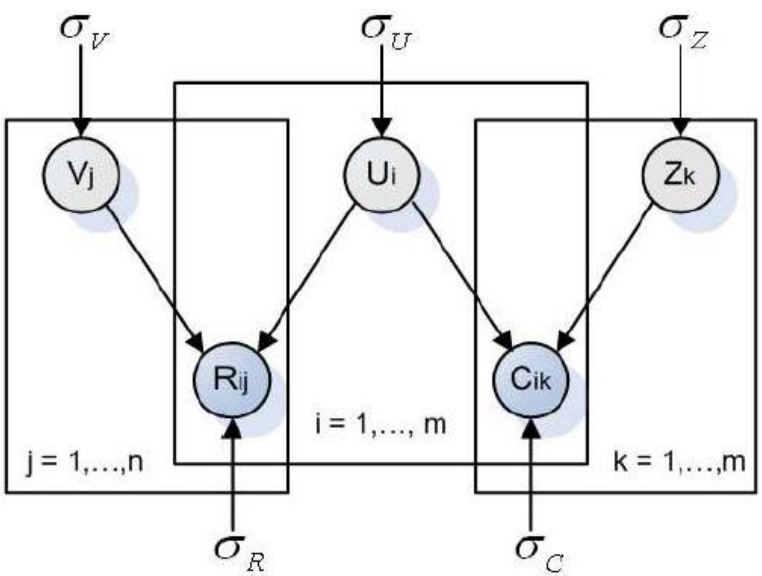
R. Salakhutdinov and A. Mnih (NIPS'08)

SoRec

- Social Recommendation (SoRec)

	v_1	v_2	v_3	v_4	v_5	v_6
u_1		5	2		3	
u_2	4			3		4
u_3			2			2
u_4	5			3		
u_5		5	5			3

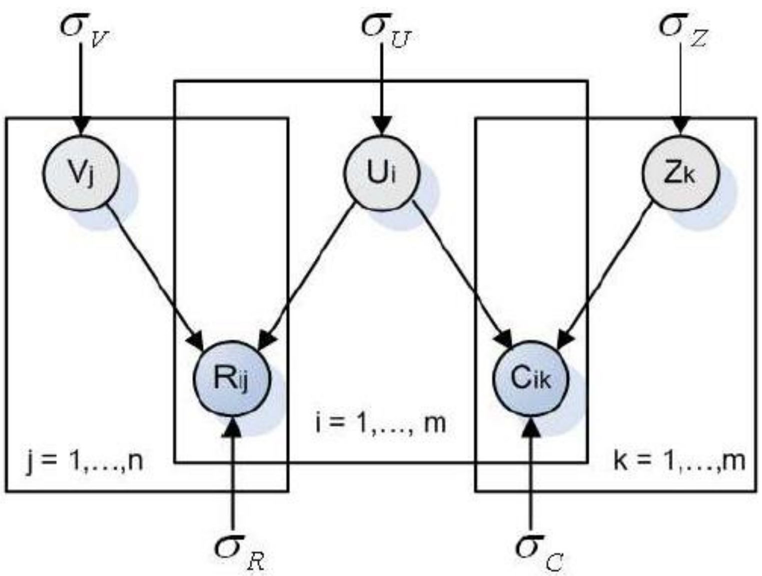




SoRec

SoRec

- Social Recommendation (SoRec)



$$p(R|U,V,\sigma_R^2) = \prod_{i=1}^{m}\prod_{j=1}^{n} \mathcal{N}\left[\left(r_{ij}|g(U_i^T V_j),\sigma_R^2\right)\right]^{I_{ij}^R}$$

$$p(C|U,Z,\sigma_C^2) = \prod_{i=1}^{m}\prod_{k=1}^{m} \mathcal{N}\left[\left(c_{ik}|g(U_i^T Z_k),\sigma_C^2\right)\right]^{I_{ik}^C}$$

$$p(U|\sigma_U^2) = \prod_{i=1}^{m} \mathcal{N}(U_i|0,\sigma_U^2\mathbf{I}) \quad p(V|\sigma_V^2) = \prod_{j=1}^{n} \mathcal{N}(V_j|0,\sigma_V^2\mathbf{I})$$

$$p(Z|\sigma_Z^2) = \prod_{k=1}^{m} \mathcal{N}(Z_k|0,\sigma_Z^2\mathbf{I})$$

$$\mathcal{L}(R,C,U,V,Z) =$$
$$\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}^R (r_{ij} - g(U_i^T V_j))^2 + \frac{\lambda_C}{2}\sum_{i=1}^{m}\sum_{k=1}^{m} I_{ik}^C (c_{ik}^* - g(U_i^T Z_k))^2$$
$$+ \frac{\lambda_U}{2}\|U\|_F^2 + \frac{\lambda_V}{2}\|V\|_F^2 + \frac{\lambda_Z}{2}\|Z\|_F^2,$$

SoRec

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial U_i} &= \sum_{j=1}^{n} I_{ij}^R g'(U_i^T V_j)(g(U_i^T V_j) - r_{ij})V_j \\
&+ \lambda_C \sum_{j=1}^{m} I_{ik}^C g'(U_i^T Z_k)(g(U_i^T Z_k) - c_{ik}^*)Z_k + \lambda_U U_i, \\
\frac{\partial \mathcal{L}}{\partial V_j} &= \sum_{i=1}^{m} I_{ij}^R g'(U_i^T V_j)(g(U_i^T V_j) - r_{ij})U_i + \lambda_V V_j, \\
\frac{\partial \mathcal{L}}{\partial Z_k} &= \lambda_C \sum_{i=1}^{m} I_{ik}^C g'(U_i^T Z_k)(g(U_i^T Z_k) - c_{ik}^*)U_i + \lambda_Z Z_k,
\end{aligned}
$$

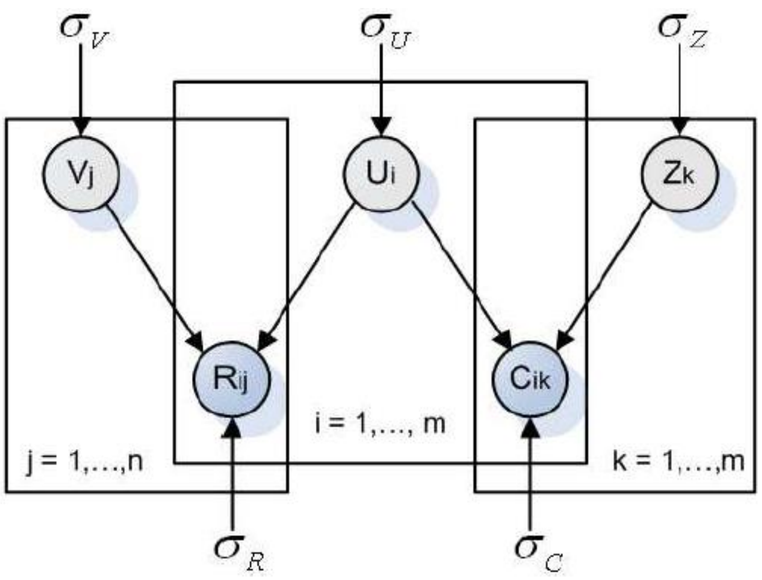
Complexity Analysis

- For the Objective Function $O(\rho_R l + \rho_C l)$
- For $\frac{\partial \mathcal{L}}{\partial U}$, the complexity is $O(\rho_R l + \rho_C l)$
- For $\frac{\partial \mathcal{L}}{\partial V}$, the complexity is $O(\rho_R l)$
- For $\frac{\partial \mathcal{L}}{\partial Z}$, the complexity is $O(\rho_C l)$

- In general, the complexity of our method is linear with the observations in these two matrices

Disadvantages of SoRec

- Lack of interpretability
- Does not reflect the real-world recommendation process

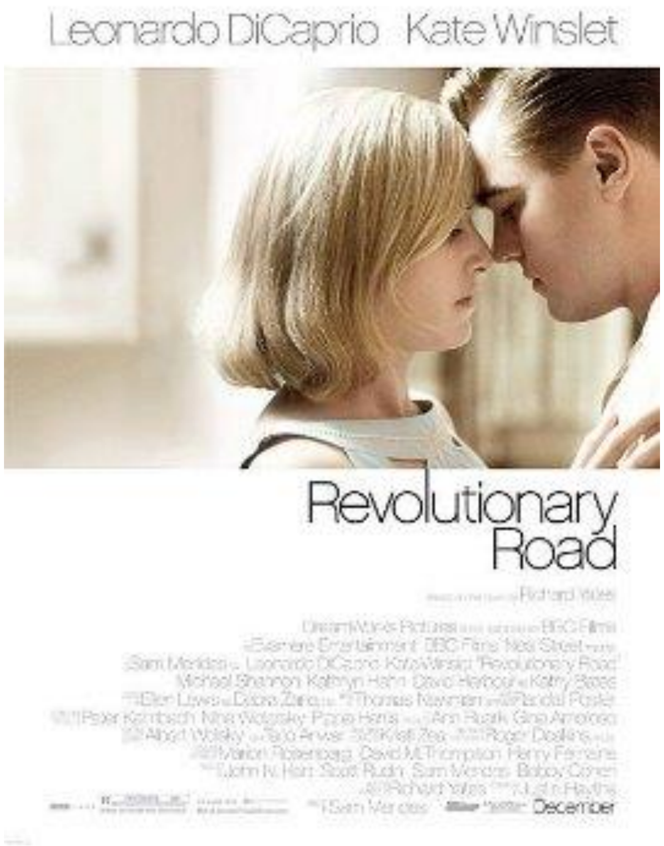


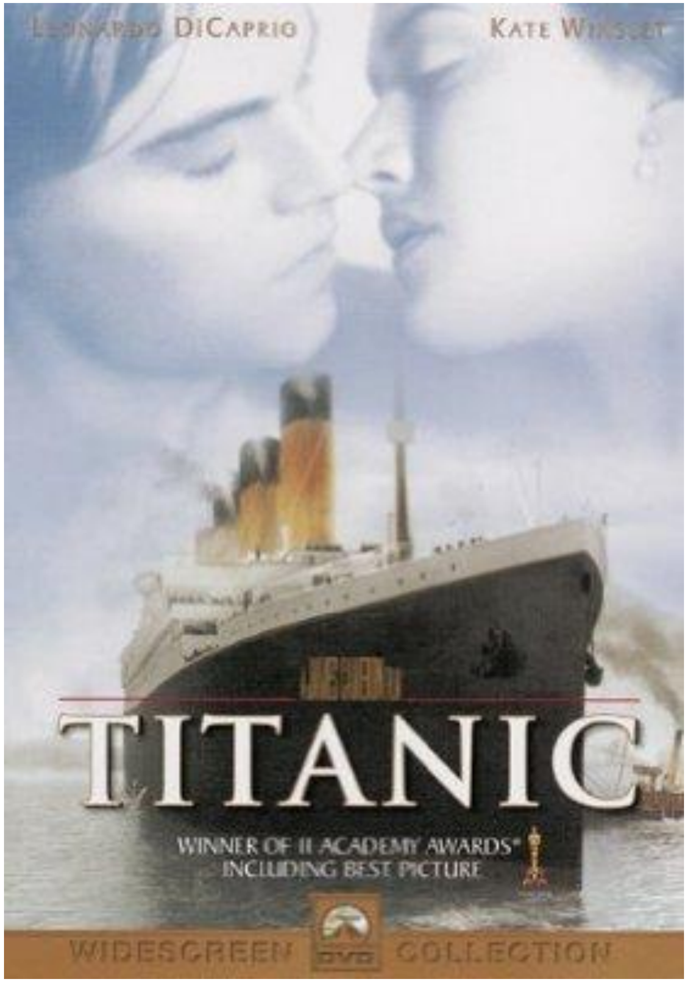
SoRec

Chapter 6

Recommend with Social Trust Ensemble

1st Motivation



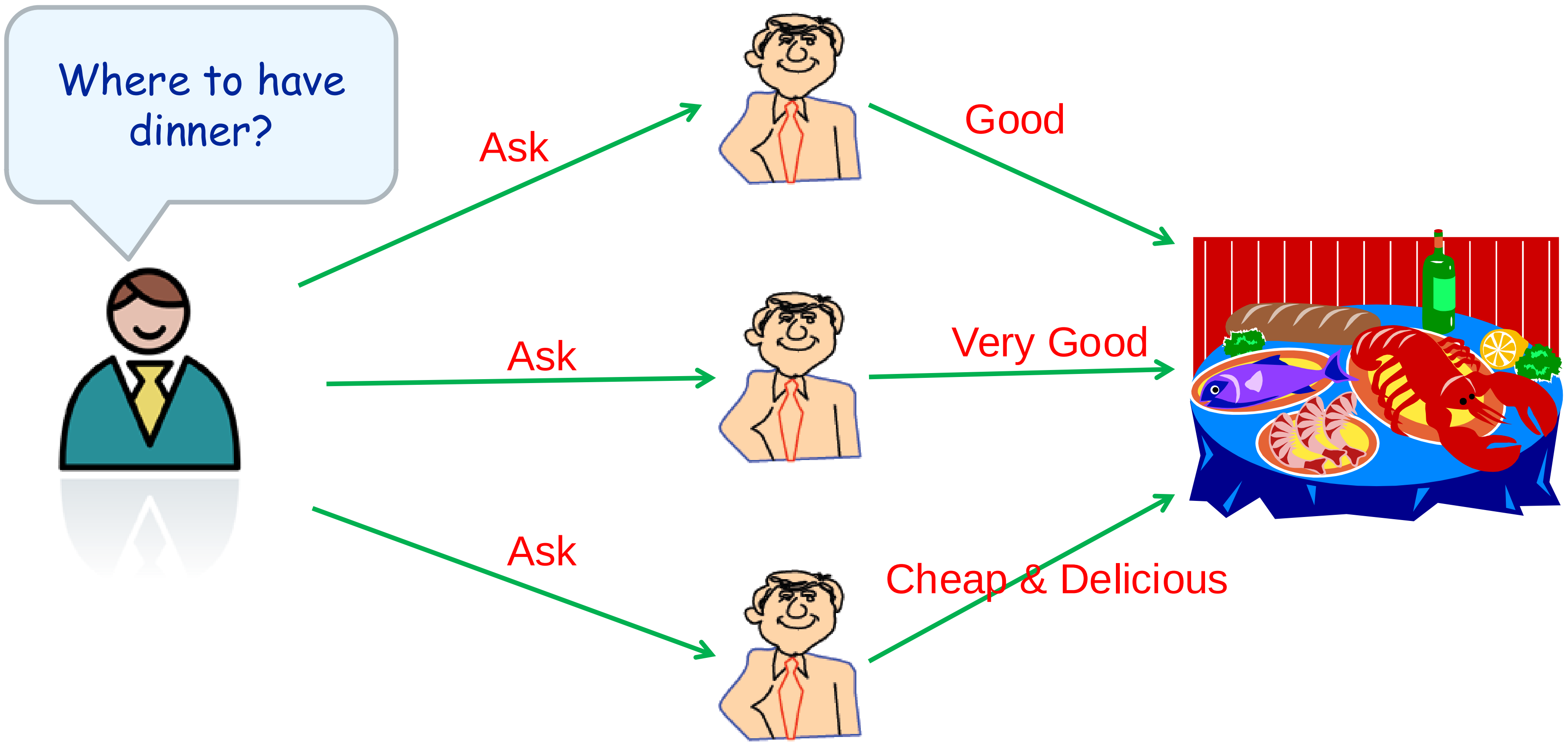


1st Motivation

1st Motivation

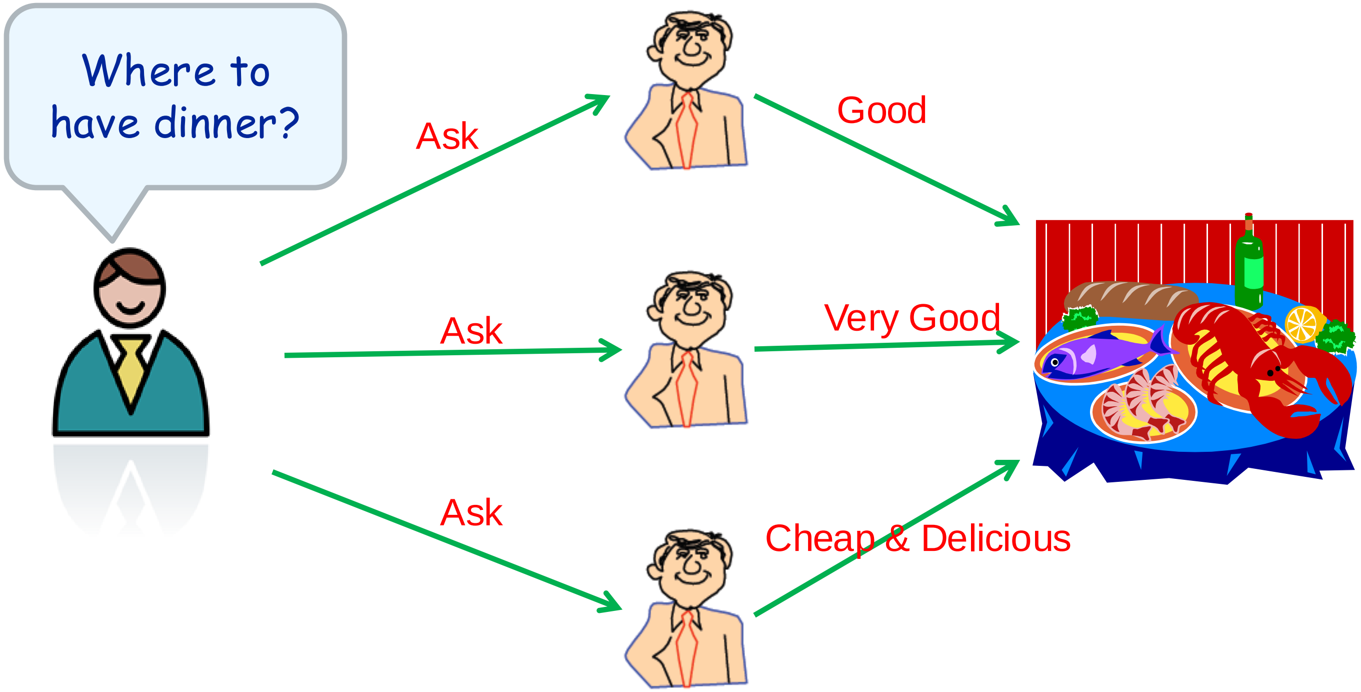
- Users have their own characteristics, and they have different tastes on different items, such as movies, books, music, articles, food, etc.

2nd Motivation



2nd Motivation

- Users can be easily influenced by the friends they trust, and prefer their friends' recommendations.



Motivations

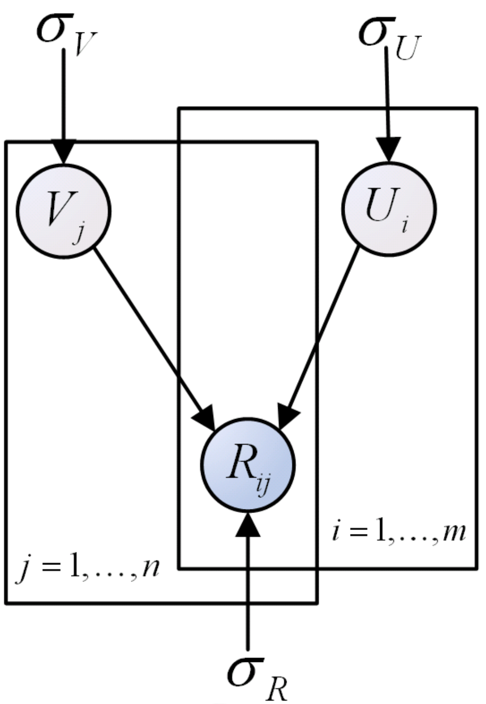
- Users have their own characteristics, and they have different tastes on different items, such as movies, books, music, articles, food, etc.

- Users can be easily influenced by the friends they trust, and prefer their friends' recommendations.

- One user's final decision is the balance between his/her own taste and his/her trusted friends' favors.

User-Item Matrix Factorization

	v_1	v_2	v_3	v_4	v_5	v_6
u_1		5	2		3	
u_2	4			3		4
u_3			2			2
u_4	5			3		
u_5		5	5			3



$$p(R|U,V,\sigma_R^2) = \prod_{i=1}^{m}\prod_{j=1}^{n}\left[\mathcal{N}\left(R_{ij}|g(U_i^T V_j),\sigma_R^2\right)\right]^{I_{ij}^R}$$

$$p(U|\sigma_U^2) = \prod_{i=1}^{m}\mathcal{N}(U_i|0,\sigma_U^2\mathbf{I}) \qquad p(V|\sigma_V^2) = \prod_{j=1}^{n}\mathcal{N}(V_j|0,\sigma_V^2\mathbf{I})$$

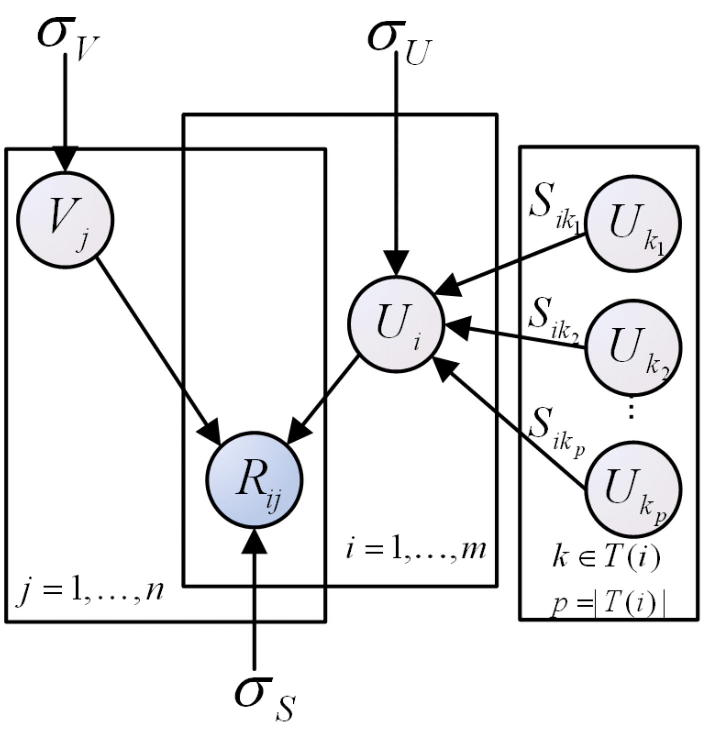
R. Salakhutdinov and A. Mnih (NIPS'08)

Recommendations by Trusted Friends

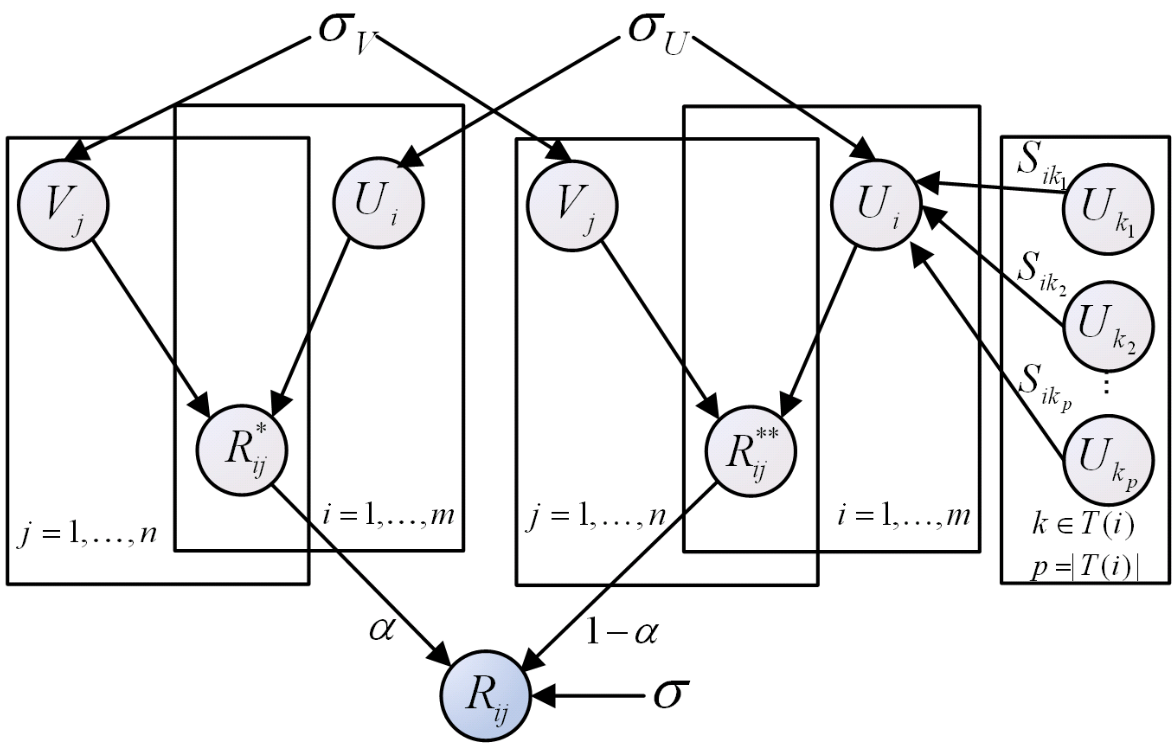
$$\widehat{R}_{ik} = \frac{\sum_{j \in \mathcal{T}(i)} R_{jk} S_{ij}}{|\mathcal{T}(i)|}$$

$$\widehat{R}_{ik} = \sum_{j \in \mathcal{T}(i)} R_{jk} S_{ij}$$

$$p(R|S,U,V,\sigma_R^2) = \prod_{i=1}^{m} \prod_{j=1}^{n} \left[\mathcal{N}\left(R_{ij} | g\left(\sum_{k \in \mathcal{T}(i)} S_{ik} U_k^T V_j \right), \sigma_S^2 \right) \right]^{I_{ij}^R}$$



Recommendation with Social Trust Ensemble



$$\prod_{i=1}^{m}\prod_{j=1}^{n}\left[\mathcal{N}\left(R_{ij}|g(\alpha U_i^T V_j + (1-\alpha)\sum_{k\in\mathcal{T}(i)} S_{ik}U_k^T V_j), \sigma^2\right)\right]^{I_{ij}^R}$$

Recommendation with Social Trust Ensemble

$$
\begin{aligned}
&\mathcal{L}(R,S,U,V)\\
&=\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}^{R}(R_{ij}-g(\alpha U_i^T V_j+(1-\alpha)\sum_{k\in\mathcal{T}(i)} S_{ik}U_k^T V_j))^2\\
&+\frac{\lambda_U}{2}\|U\|_F^2+\frac{\lambda_V}{2}\|V\|_F^2, \qquad (15)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial\mathcal{L}}{\partial U_i} =& \alpha\sum_{j=1}^{n} I_{ij}^{R} g'(\alpha U_i^T V_j+(1-\alpha)\sum_{k\in\mathcal{T}(i)} S_{ik}U_k^T V_j)V_j\\
&\times(g(\alpha U_i^T V_j+(1-\alpha)\sum_{k\in\mathcal{T}(i)} S_{ik}U_k^T V_j)-R_{ij})\\
&+(1-\alpha)\sum_{p\in\mathcal{B}(i)}\sum_{j=1}^{n} I_{pj}^{R} g'(\alpha U_p^T V_j+(1-\alpha)\sum_{k\in\mathcal{T}(p)} S_{pk}U_k^T V_j)\\
&\times(g(\alpha U_p^T V_j+(1-\alpha)\sum_{k\in\mathcal{T}(p)} S_{pk}U_k^T V_j)-R_{pj})S_{pi}V_j\\
&+\lambda_U U_i,
\end{aligned}
$$

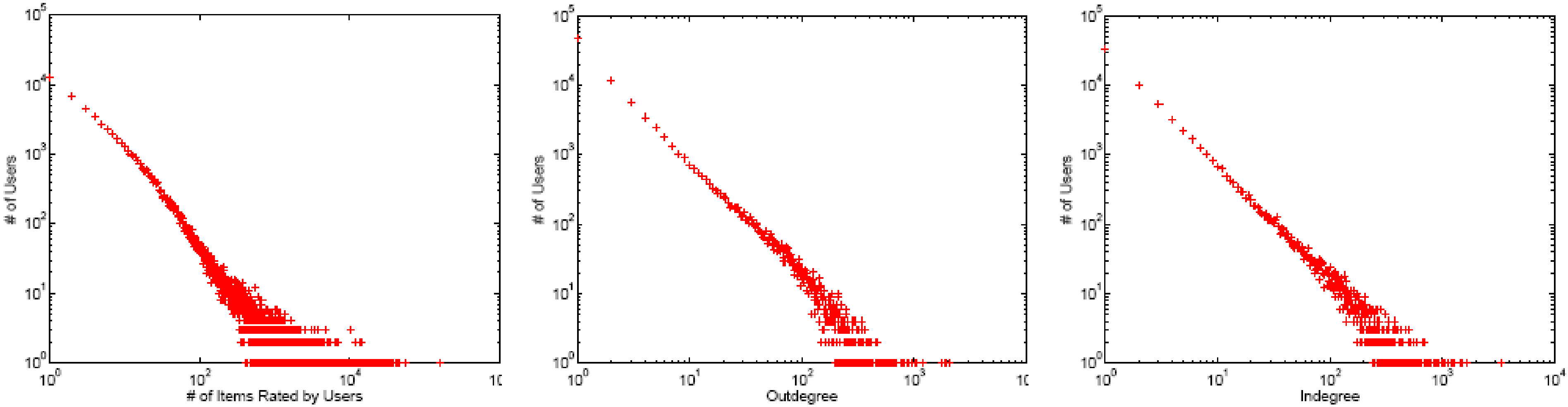
$$
\begin{aligned}
\frac{\partial\mathcal{L}}{\partial V_j} =& \sum_{i=1}^{m} I_{ij}^{R} g'(\alpha U_i^T V_j+(1-\alpha)\sum_{k\in\mathcal{T}(i)} S_{ik}U_k^T V_j)\\
&\times(g(\alpha U_i^T V_j+(1-\alpha)\sum_{k\in\mathcal{T}(i)} S_{ik}U_k^T V_j)-R_{ij})\\
&\times(\alpha U_i+(1-\alpha)\sum_{k\in\mathcal{T}(i)} S_{ik}U_k^T)+\lambda_V V_j,
\end{aligned}
$$

Complexity

- In general, the complexity of this method is linear with the observations the user-item matrix

Epinions Dataset

- 51,670 users who rated 83,509 items with totally 631,064 ratings
- Rating Density 0.015%
- The total number of issued trust statements is 511,799



(a) Items per User Distribution (b) Trust Graph Outdegree Distribution (c) Trust Graph Indegree Distribution

Figure 3: Power-Law Distributions of the Epinions Dataset

Metrics

- Mean Absolute Error and Root Mean Square Error

$$MAE = \frac{\sum_{i,j} |r_{i,j} - \widehat{r}_{i,j}|}{N}$$

$$RMSE = \sqrt{\frac{\sum_{i,j} (r_{i,j} - \widehat{r}_{i,j})^2}{N}}$$

Comparisons

Table 3: Performance Comparisons (A Smaller MAE or RMSE Value Means a Better Performance)

Training Data	Metrics	Dimensionality = 5				Dimensionality = 10			
		Trust	PMF	SoRec	RSTE	Trust	PMF	SoRec	RSTE
90%	MAE	0.9054	0.8676	0.8484	**0.8377**	0.9039	0.8651	0.8426	**0.8367**
	RMSE	1.1959	1.1575	1.1418	**1.1109**	1.1917	1.1544	1.1365	**1.1094**
80%	MAE	0.9221	0.8951	0.8654	**0.8594**	0.9215	0.8886	0.8605	**0.8537**
	RMSE	1.2140	1.1826	1.1517	**1.1346**	1.2132	1.1760	1.1586	**1.1256**

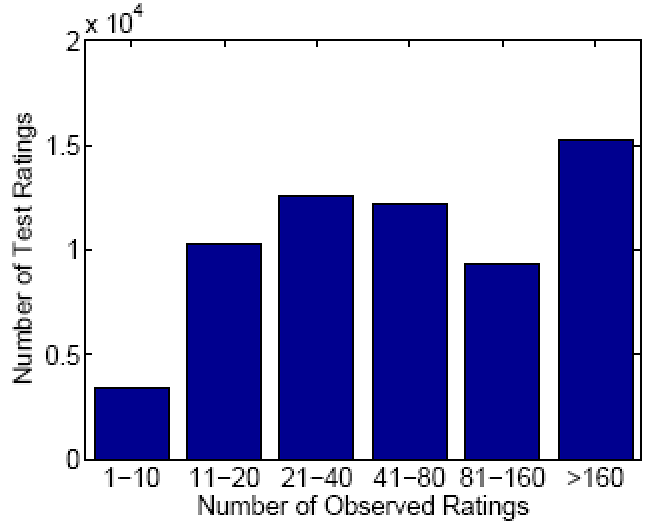
PMF --- R. Salakhutdinov and A. Mnih (NIPS 2008)

SoRec --- H. Ma, H. Yang, M. R. Lyu and I. King (CIKM 2008)

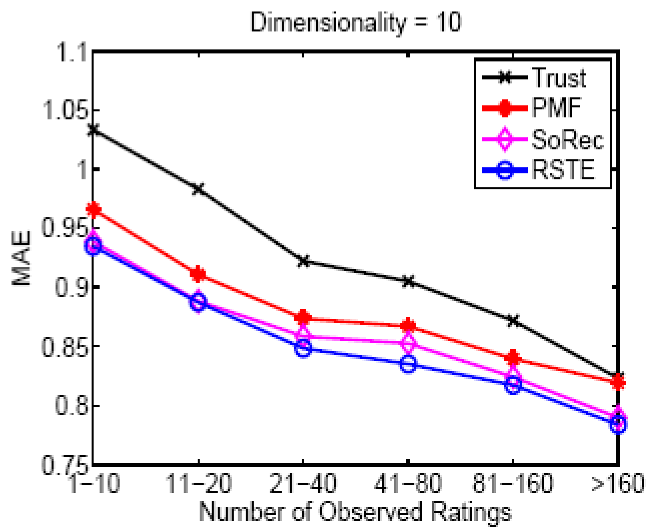
Trust, RSTE --- H. Ma, I. King and M. R. Lyu (SIGIR 2009)

Performance on Different Users

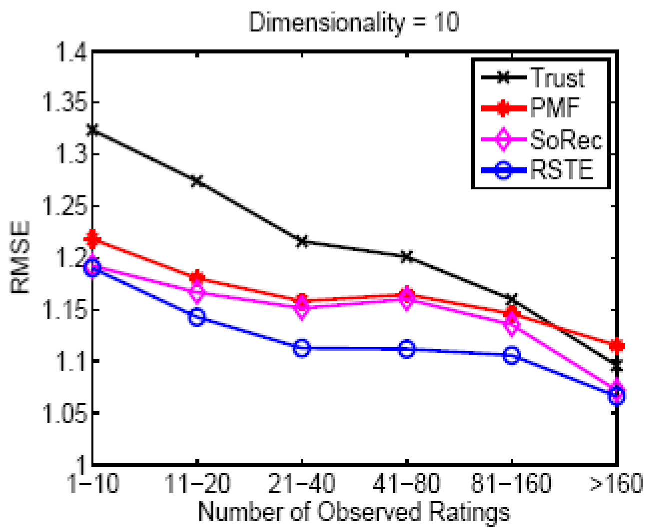
- Group all the users based on the number of observed ratings in the training data
- 6 classes: "1 – 10", "11 – 20", "21 – 40", "41 – 80", "81 – 160", "> 160",



(a) Distribution of Testing Data (90% as Training Data)

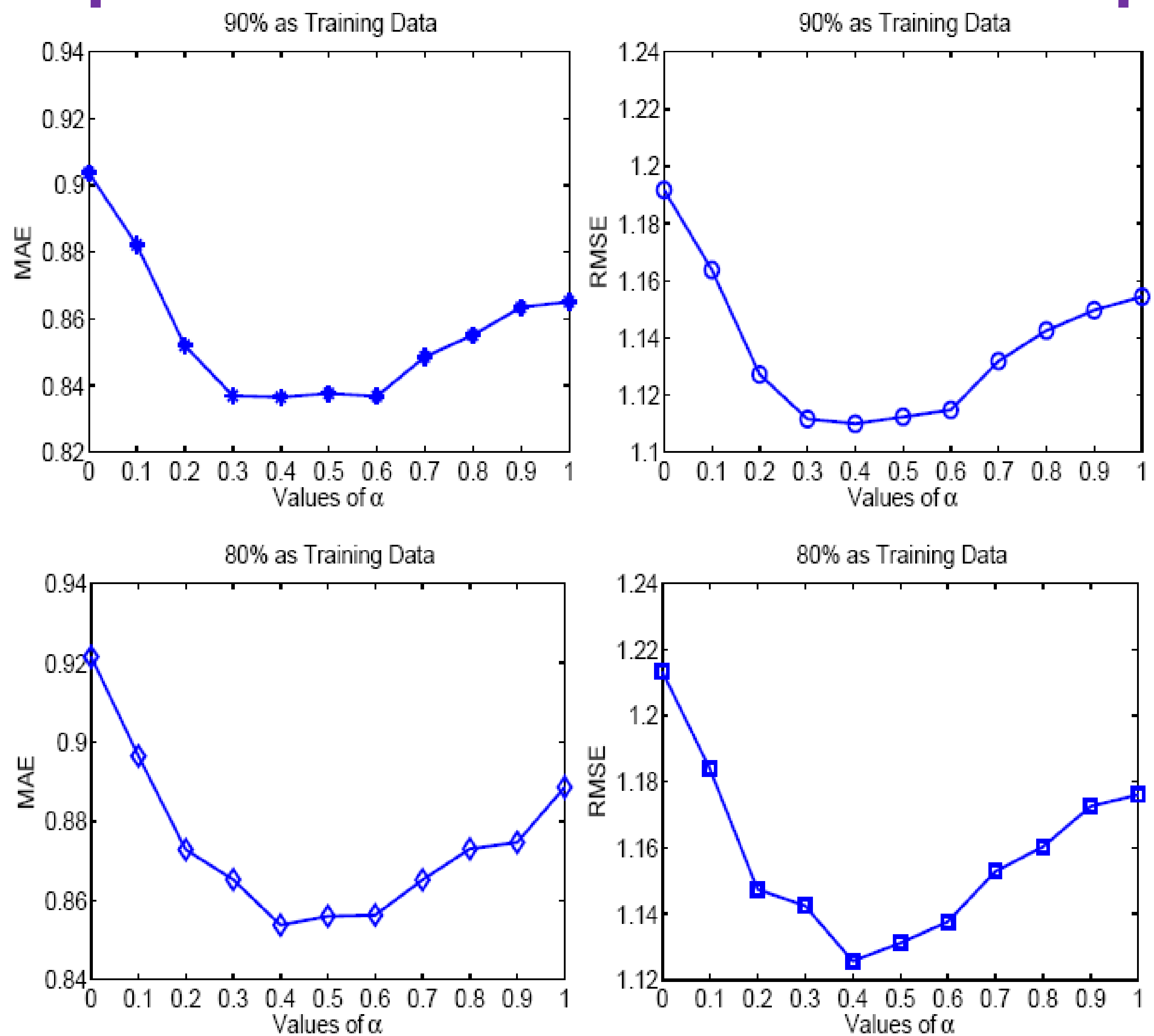


(b) MAE Comparison on Different User Rating Scales (90% as Training Data)



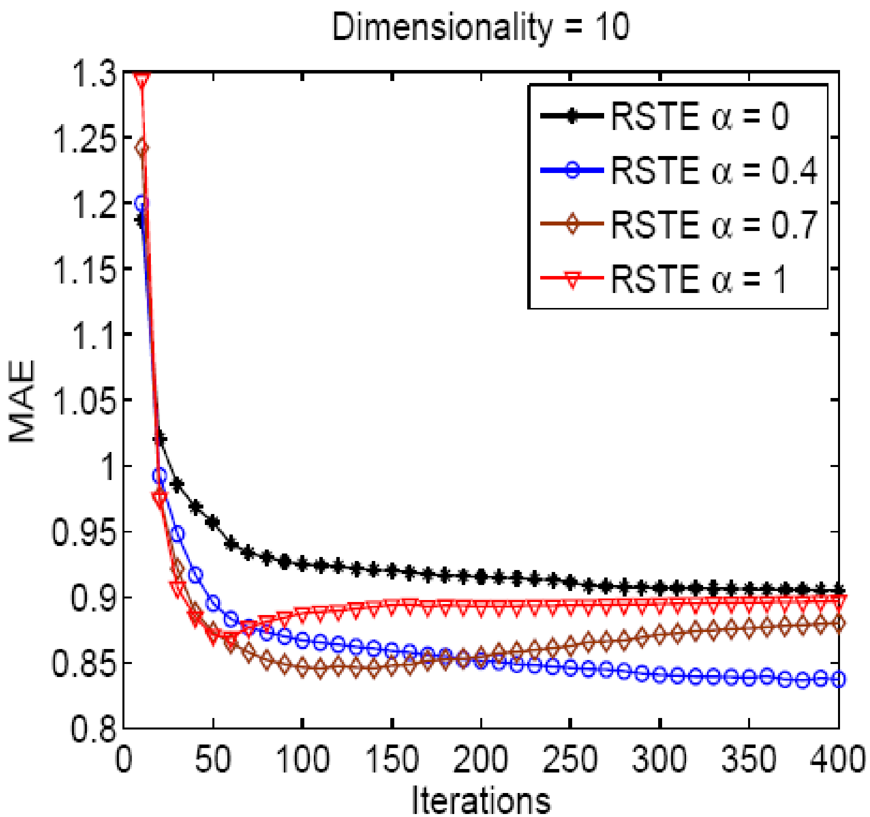
(c) RMSE Comparison on Different User Rating Scales (90% as Training Data)

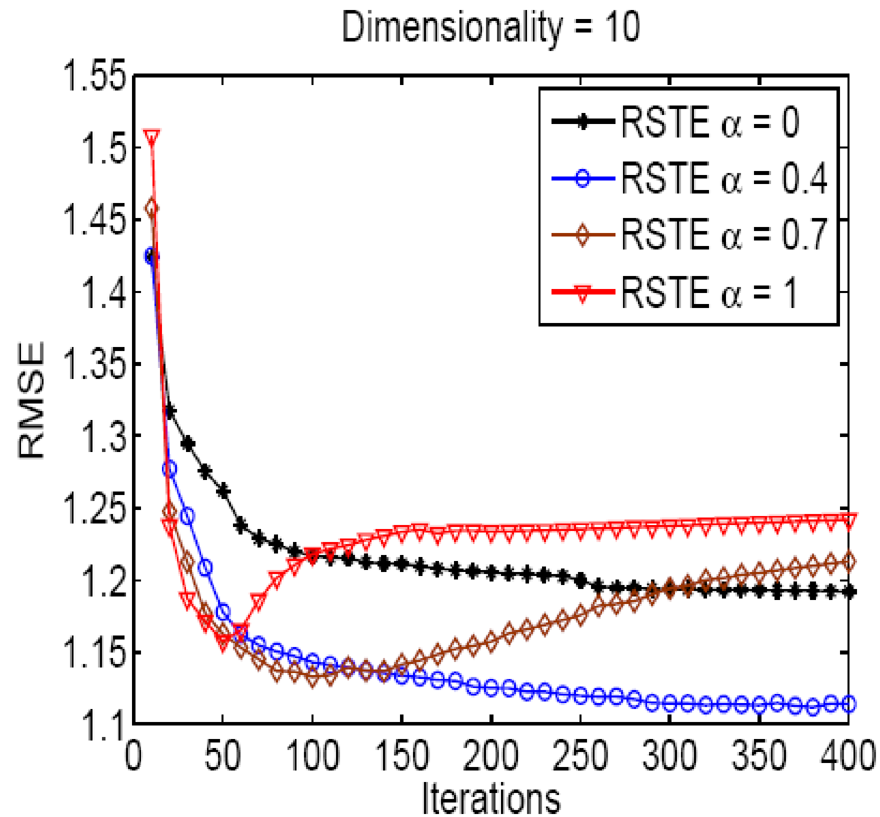
Impact of Parameter Alpha



Impact of Parameter α (Dimensionality = 10)

MAE and RMSE Changes with Iterations





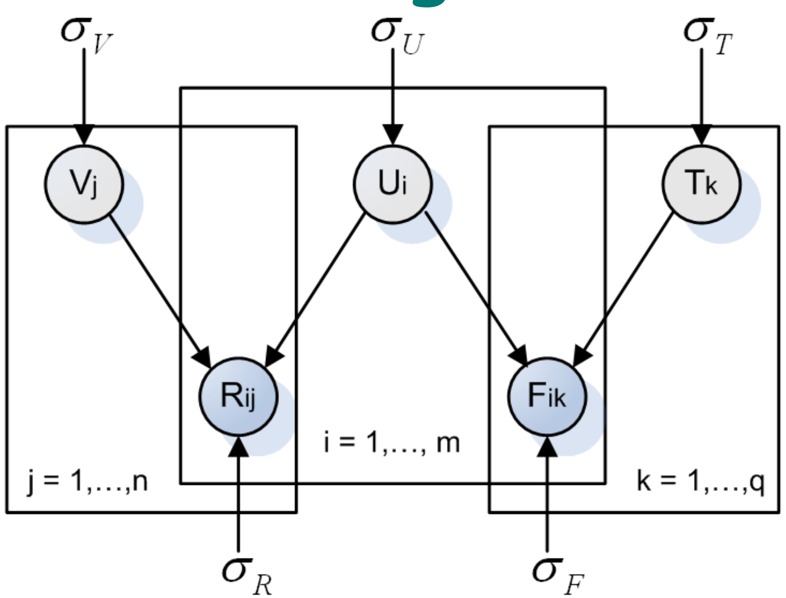
90% as Training Data

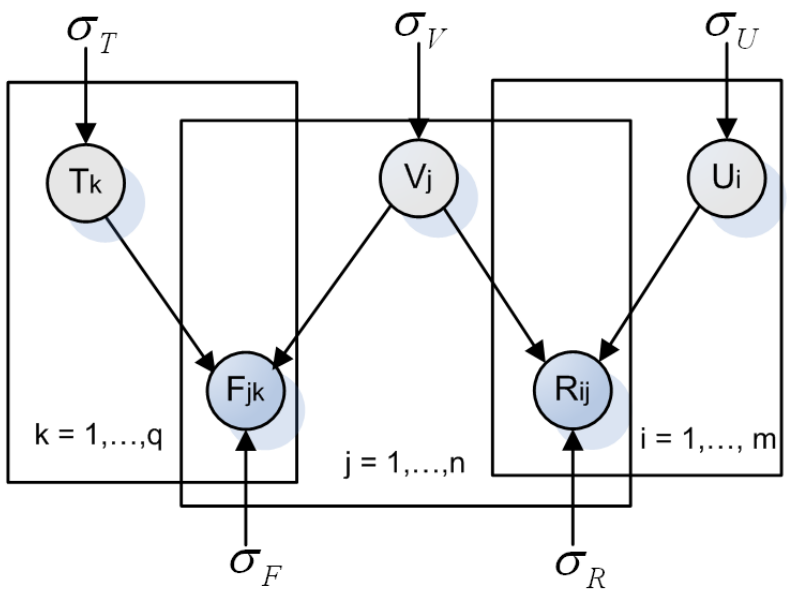
Conclusions of SoRec and RSTE

- Propose two novel Social Trust-based Recommendation methods
- Perform well
- Scalable to very large datasets
- Show the promising future of social-based techniques

Further Discussion of SoRec

- Improving Recommender Systems Using Social Tags





MovieLens Dataset
71,567 users, 10,681 movies,
10,000,054 ratings, 95,580 tags

Further Discussion of SoRec

- MAE

Table V: MAE comparison with other approaches on MovieLens dataset (A smaller MAE value means a better performance)

Methods		80% Training	50% Training	30% Training	10% Training
User Mean		0.7686	0.7710	0.7742	0.8234
Item Mean		0.7379	0.7389	0.7399	0.7484
5D	SVD	0.6390	0.6547	0.6707	0.7448
	PMF	0.6325	0.6542	0.6698	0.7430
	SoRecUser	0.6209	0.6419	0.6607	0.7040
	SoRecItem	**0.6199**	**0.6407**	**0.6395**	**0.7026**
10D	SVD	0.6386	0.6534	0.6693	0.7431
	PMF	0.6312	0.6530	0.6683	0.7417
	SoRecUser	0.6197	0.6408	0.6595	0.7028
	SoRecItem	**0.6187**	**0.6395**	**0.6584**	**0.7016**

Further Discussion of SoRec

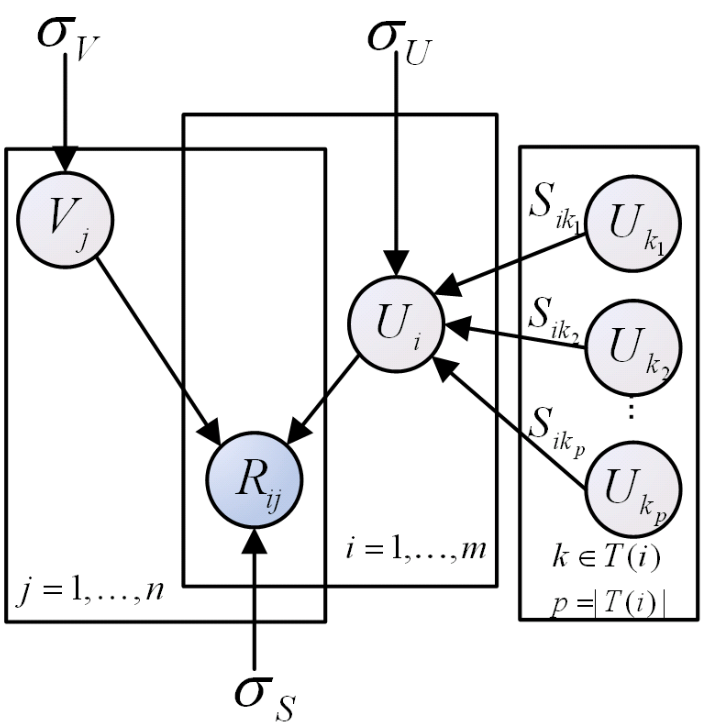
- RMSE

Table VI: RMSE comparison with other approaches on MovieLens dataset (A smaller RMSE value means a better performance)

Methods		80% Training	50% Training	30% Training	10% Training
User Mean		0.9779	0.9816	0.9869	1.1587
Item Mean		0.9440	0.9463	0.9505	0.9851
5D	SVD	0.8327	0.8524	0.8743	0.9892
	PMF	0.8310	0.8582	0.8758	0.9698
	SoRecUser	0.8121	0.8384	0.8604	0.9042
	SoRecItem	**0.8112**	**0.8370**	**0.8591**	**0.9033**
10D	SVD	0.8312	0.8509	0.8728	0.9878
	PMF	0.8295	0.8569	0.8743	0.9681
	SoRecUser	0.8110	0.8372	0.8593	0.9034
	SoRecItem	**0.8097**	**0.8359**	**0.8578**	**0.9019**

Further Discussion of RSTE

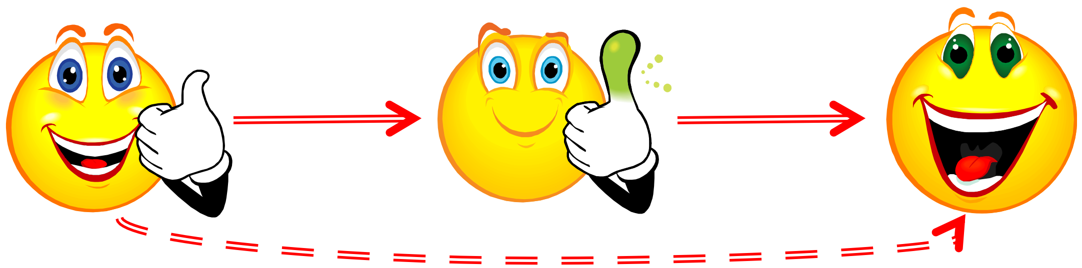
- Relationship with Neighborhood-based methods



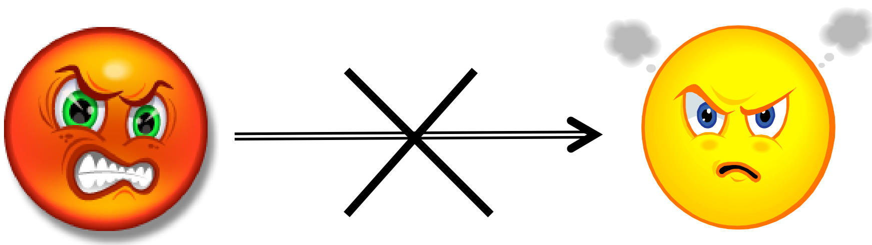
- The trusted friends are actually the explicit neighbors
- We can easily apply this method to include implicit neighbors
- Using PCC to calculate similar users for every user

What We Cannot Model Using SoRec and RSTE?

- Propagation of trust



- Distrust



Chapter 7

Recommend with Social Distrust

Distrust

- Users' distrust relations can be interpreted as the "dissimilar" relations
  - On the web, user Ui distrusts user Ud indicates that user Ui disagrees with most of the opinions issued by user Ud.

Distrust

$$\max_{U} \frac{1}{2} \sum_{i=1}^{m} \sum_{d \in \mathcal{D}+(i)} S_{id}^{\mathcal{D}} ||U_i - U_d||_F^2$$

$$\begin{aligned}
\min_{U,V} \mathcal{L}_{\mathcal{D}}(R, S^{\mathcal{D}}, U, V) \quad &= \quad \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij}^{R} (R_{ij} - g(U_i^T V_j))^2 \\
&+ \quad \frac{\beta}{2} \sum_{i=1}^{m} \sum_{d \in \mathcal{D}+(i)} (-S_{id}^{\mathcal{D}} ||U_i - U_d||_F^2) \\
&+ \quad \frac{\lambda_U}{2} ||U||_F^2 + \frac{\lambda_V}{2} ||V||_F^2.
\end{aligned}$$

Trust

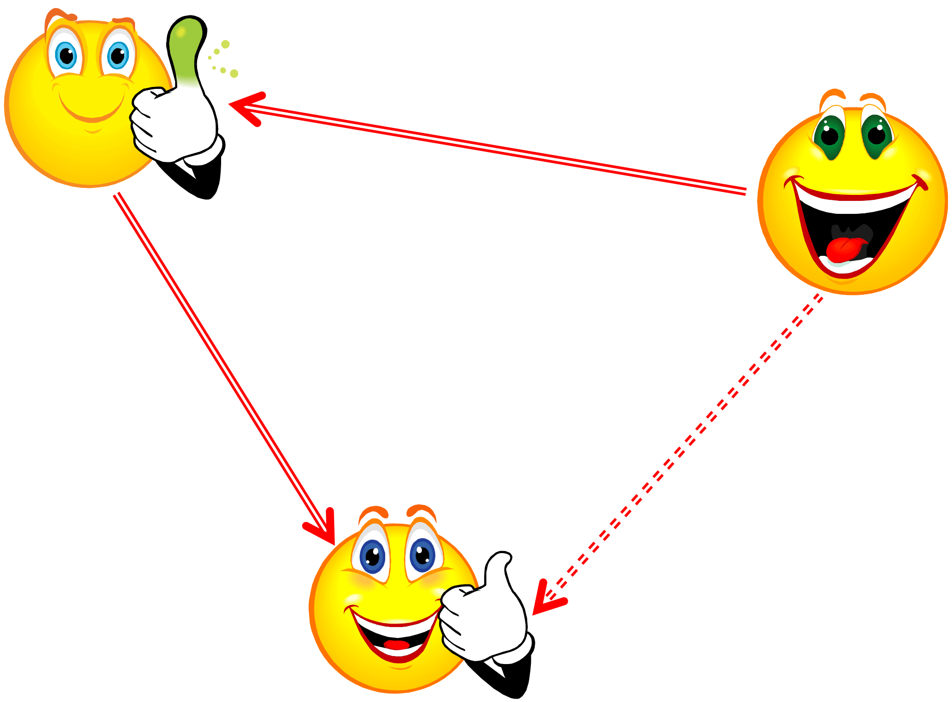
- Users' trust relations can be interpreted as the "similar" relations
  - On the web, user Ui trusts user Ut indicates that user Ui agrees with most of the opinions issued by user Ut.

Trust

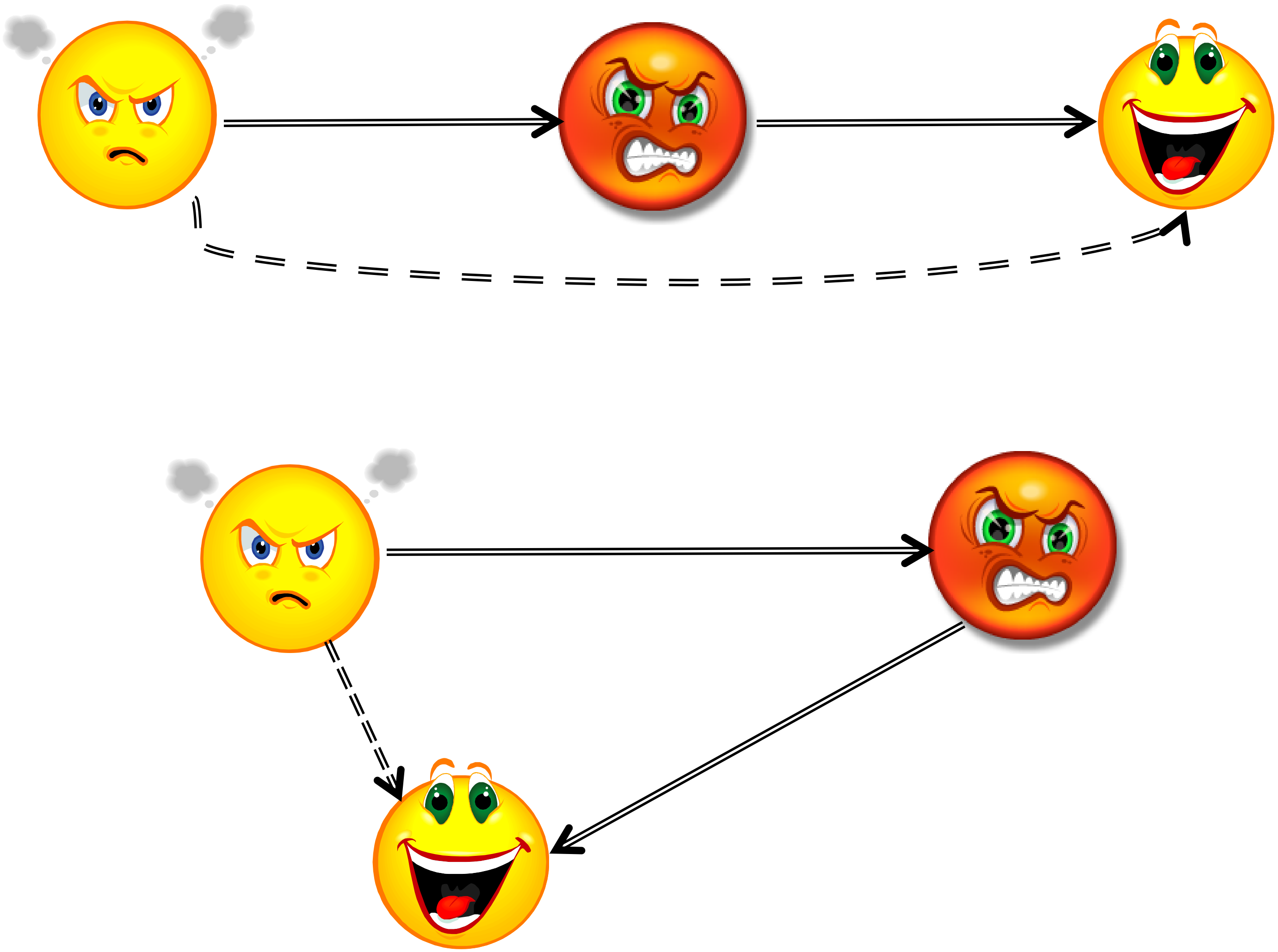
$$\min_{U} \frac{1}{2} \sum_{i=1}^{m} \sum_{t \in \mathcal{T}^{+}(i)} S_{it}^{\mathcal{T}} \|U_i - U_t\|_F^2$$

$$\begin{aligned}
\min_{U,V} \mathcal{L}_{\mathcal{T}}(R, S^{\mathcal{T}}, U, V) &= \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij}^{R} (R_{ij} - g(U_i^T V_j))^2 \\
&+ \frac{\alpha}{2} \sum_{i=1}^{m} \sum_{t \in \mathcal{T}^{+}(i)} (S_{it}^{\mathcal{T}} \|U_i - U_t\|_F^2) \\
&+ \frac{\lambda_U}{2} \|U\|_F^2 + \frac{\lambda_V}{2} \|V\|_F^2 .
\end{aligned}$$

Trust Propagation



Distrust Propagation?



Experiments

- Dataset - Epinions
- 131,580 users, 755,137 items, 13,430,209 ratings
- 717,129 trust relations, 123,670 distrust relations

Data Statistics

Table 1: Statistics of User-Item Rating Matrix of Epinions

Statistics	User	Item
Min. Num. of Ratings	1	1
Max. Num. of Ratings	162169	1179
Avg. Num. of Ratings	102.07	17.79

Table 2: Statistics of Trust Network of Epinions

Statistics	Trust per User	Be Trusted per User
Max. Num.	2070	3338
Avg. Num.	5.45	5.45

Table 3: Statistics of Distrust Network of Epinions

Statistics	Distrust per User	Be Distrusted per User
Max. Num.	1562	540
Avg. Num.	0.94	0.94

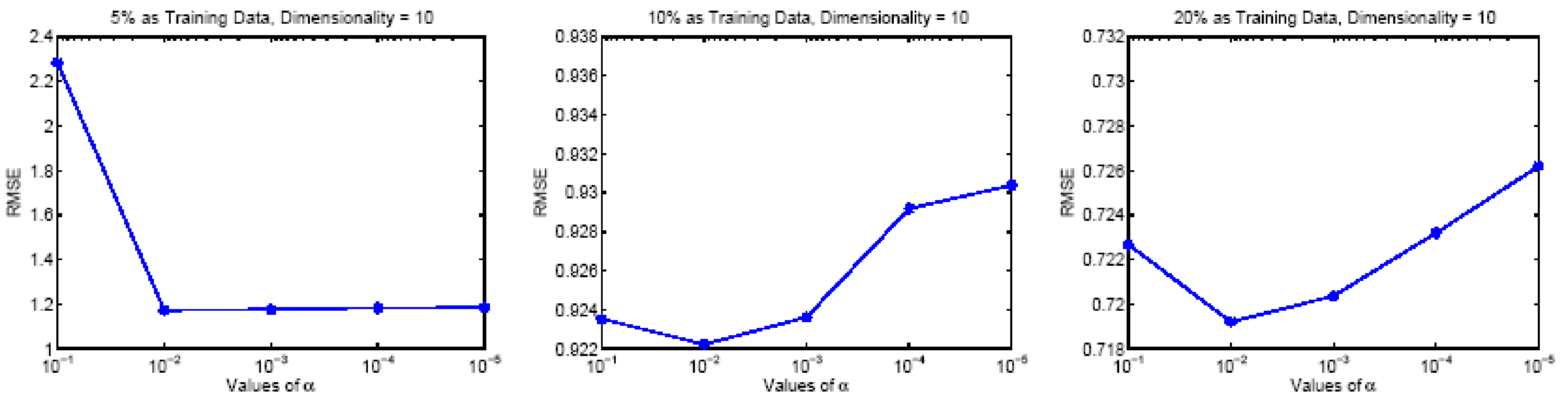
Experiments

RMSE

Dataset	Traning Data	Dimensionality	PMF	SoRec	RWD	RWT
Epinions	5%	5D	1.228	1.199	1.186	**1.177**
		10D	1.214	1.198	1.185	**1.176**
	10%	5D	0.990	0.944	0.932	**0.924**
		10D	0.977	0.941	0.931	**0.923**
	20%	5D	0.819	0.788	0.723	**0.721**
		10D	0.818	0.787	0.723	**0.720**

131,580 users, 755,137 items, 13,430,209 ratings
717,129 trust relations, 123,670 distrust relations

Impact of Parameters



(a) 5% as Training Data (b) 10% as Training Data (c) 20% as Training Data

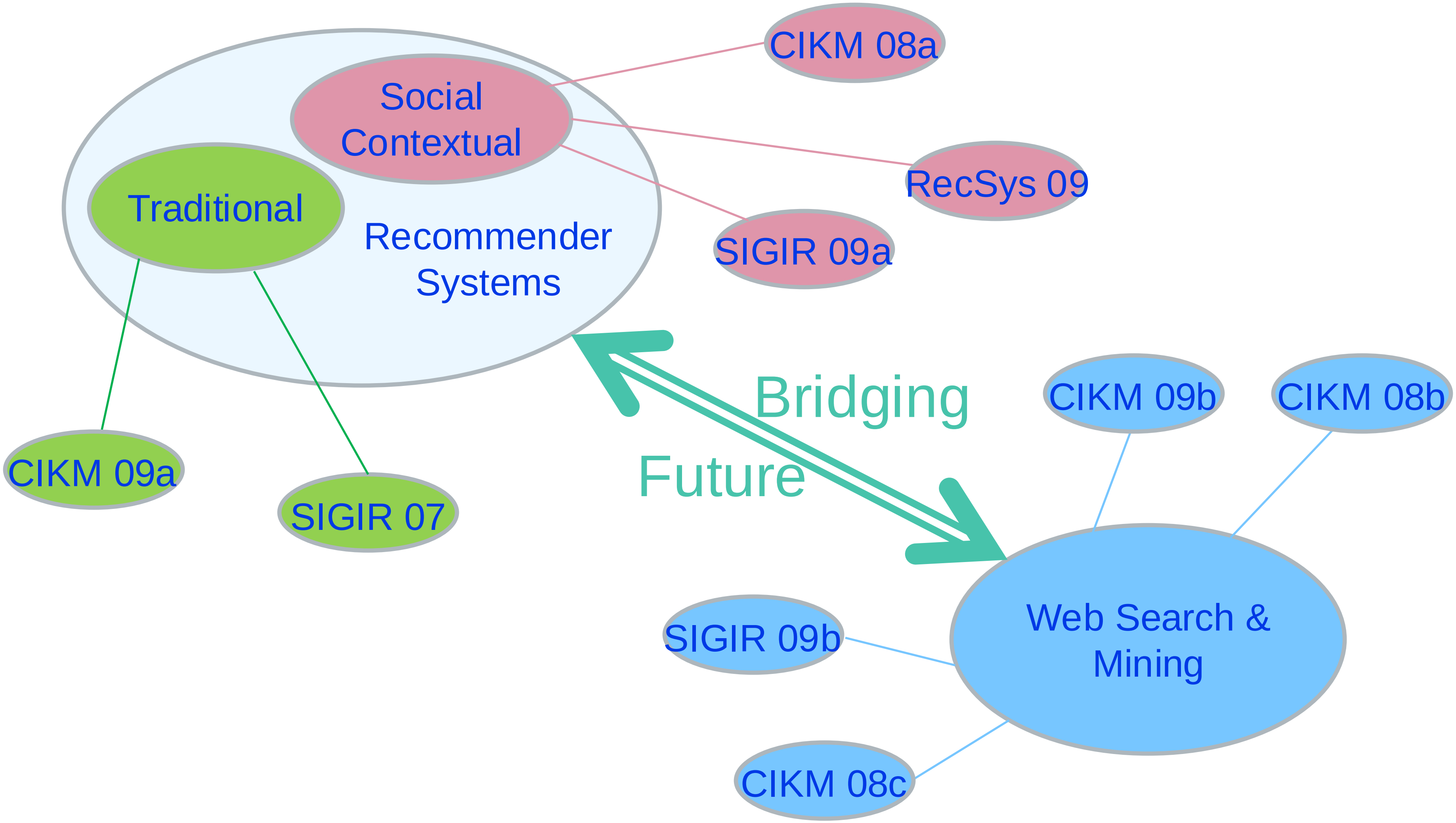
Figure 6: Impact of Parameter α

Alpha = 0.01 will get the best performance!
Parameter beta basically shares the same trend!

Summary

- 5 methods for Improving Recommender
  - 2 traditional recommendation methods
  - 3 social recommendation approaches
- Effective and efficient
- Very general, and can be applied to different applications, including search-related problems

A Roadmap of My Work



Search and Recommendation

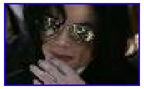
Google

michael jackson

Google Search | I'm Feeling Lucky

Advanced Search
Language Tools

News results for **michael jackson**


guardian.co.uk

Propofol dosage reported in **Michael Jackson** case is low, experts say - 1 hour ago
By Kimi Yoshino If **Michael Jackson** died from lethal levels of the powerful anesthetic propofol, then he must have been injected with much more of the drug **...**
Los Angeles Times - 4334 related articles »
The continuing fantasy of **Michael Jackson's** future in Vegas -
Los Angeles Times - 164 related articles »
People:A&E channel sets reality show starring **Michael Jackson's ...** -
San Jose Mercury News - 38 related articles »

Michael Jackson | **Michael Jackson** This Is It | **Michael Jackson ...**
The Official **Michael Jackson** site including info on This Is It, the **Michael Jackson** Movie, **Michael Jackson** Music, Videos and Lyrics from hits like; **...**
www.**michaeljackson**.com/ - Cached - Similar -

Michael Jackson - Wikipedia, the free encyclopedia
Michael Joseph **Jackson** (August 29, 1958 – June 25, 2009), known as the "King of Pop", was an American musician and one of the most commercially successful **...**
en.wikipedia.org/wiki/**Michael_Jackson** - Cached - Similar -

Michael Jackson (I)
American superstar **Michael Jackson** was born in Gary, Indiana, on August 29... Visit IMDb for Photos, Filmography, Discussions, Bio, News, Awards, Agent, **...**
www.imdb.com/name/nm0001391/ - Cached - Similar -

Image results for **michael jackson** - Report images

Passive Recommender System

Search and Recommendation

- We need a more active and intelligent search engine to understand users' interests

- Recommendation technology represents the new paradigm of search

Search and Recommendation

CNNMoney.com
A Service of CNN, Fortune & Money

Jeffrey M. O'Brien

FORTUNE

- The Web
  - Is leaving the era of search
  - Is entering one of discovery
- What's the difference?
  - Search is what you do when you're looking for something.
  - Discovery is when something wonderful that you didn't know existed, or didn't know how to ask for, finds you. Recommendation!!!

Search and Recommendation

- By mining user browsing graph or clickthrough data using the proposed methods in this thesis, we can:
  - Build personalized web site recommendations
  - Improve the ranking
  - Learn more accurate features of URLs or Queries
  - ……

Publications

1. **Hao Ma**, Haixuan Yang, Irwin King, Michael R. Lyu. Semi-Nonnegative Matrix Factorization with Global Statistical Consistency in Collaborative Filtering. ***ACM CIKM'09***, Hong Kong, China, November 2-6, 2009.
2. **Hao Ma**, Raman Chandrasekar, Chris Quirk, Abhishek Gupta. Improving Search Engines Using Human Computation Games. ***ACM CIKM'09***, Hong Kong, China, November 2-6, 2009.
3. **Hao Ma**, Michael R. Lyu, Irwin King. Learning to Recommend with Trust and Distrust Relationships. ***ACM RecSys'09***, New York City, NY, USA, October 22-25, 2009.
4. **Hao Ma**, Irwin King, Michael R. Lyu. Learning to Recommend with Social Trust Ensemble. ***ACM SIGIR'09***, Boston, MA, USA, July 19-23, 2009.
5. **Hao Ma**, Raman Chandrasekar, Chris Quirk, Abhishek Gupta. Page Hunt: Improving Search Engines Using Human Computation Games. ***ACM SIGIR'09***, Boston, MA, USA, July 19-23, 2009.

Publications

6. **Hao Ma**, Haixuan Yang, Michael R. Lyu, Irwin King. SoRec: Social Recommendation Using Probabilistic Matrix Factorization. **ACM CIKM'08**, pages 931-940, Napa Valley, California USA, October 26-30, 2008.

7. **Hao Ma**, Haixuan Yang, Irwin King, Michael R. Lyu. Learning Latent Semantic Relations from Clickthrough Data for Query Suggestion. **ACM CIKM'08**, pages 709-718, Napa Valley, California USA, October 26-30, 2008.

8. **Hao Ma**, Haixuan Yang, Michael R. Lyu, Irwin King. Mining Social Networks Using Heat Diffusion Processes for Marketing Candidates Selection. **ACM CIKM'08**, pages 233-242, Napa Valley, California USA, October 26-30, 2008.

9. **Hao Ma**, Irwin King, Michael R. Lyu. Effective Missing Data Prediction for Collaborative Filtering. **ACM SIGIR'07**, pages 39-46, Amsterdam, the Netherlands, July 23-27, 2007.

Thank You!

Q & A

Hao Ma
hma@cse.cuhk.edu.hk