Scalable Machine Learning

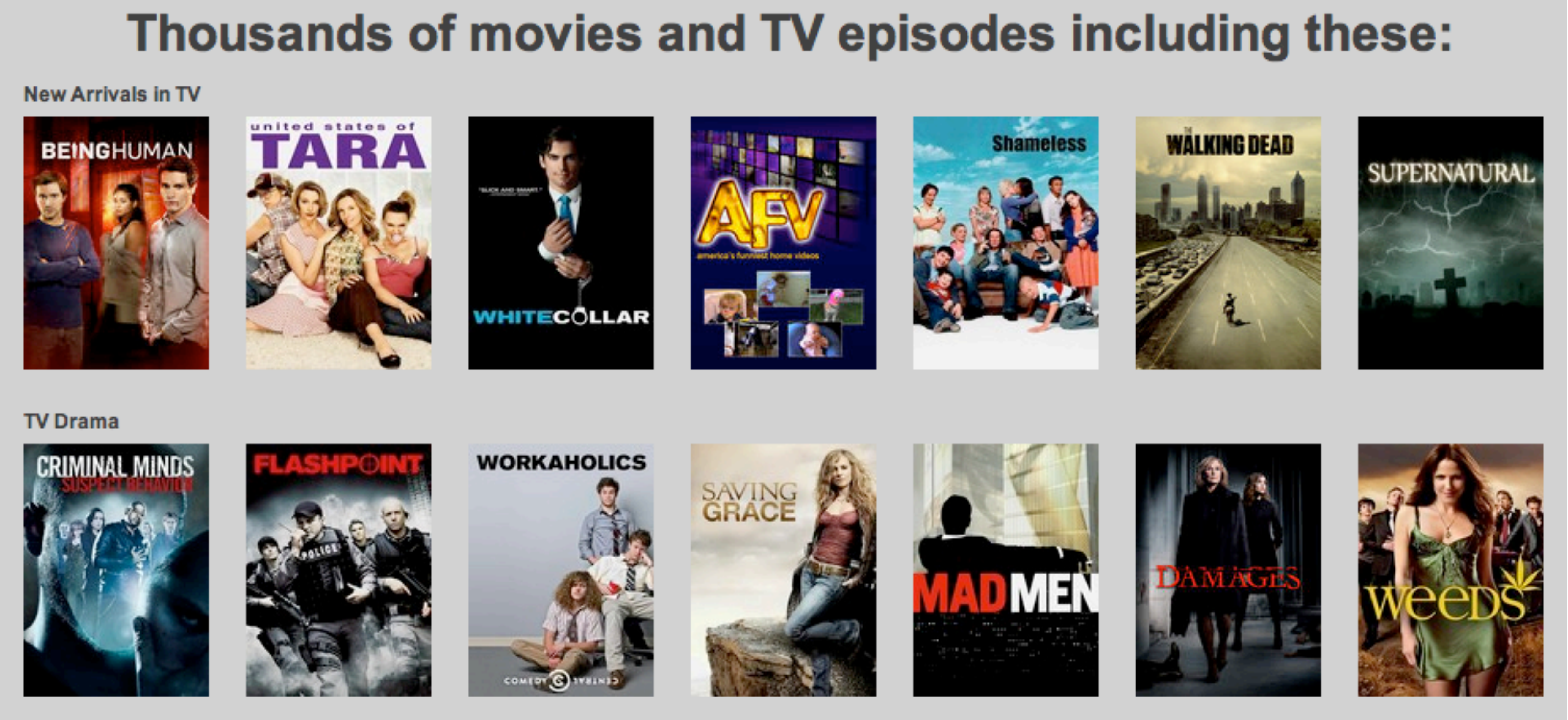
8. Recommender Systems

Alex Smola
Yahoo! Research and ANU

http://alex.smola.org/teaching/berkeley2012
Stat 260 SP 12

Significant content courtesy of Yehuda Koren

8. Recommender Systems



Much content courtesy of (Mr Netflix) Yehuda Koren

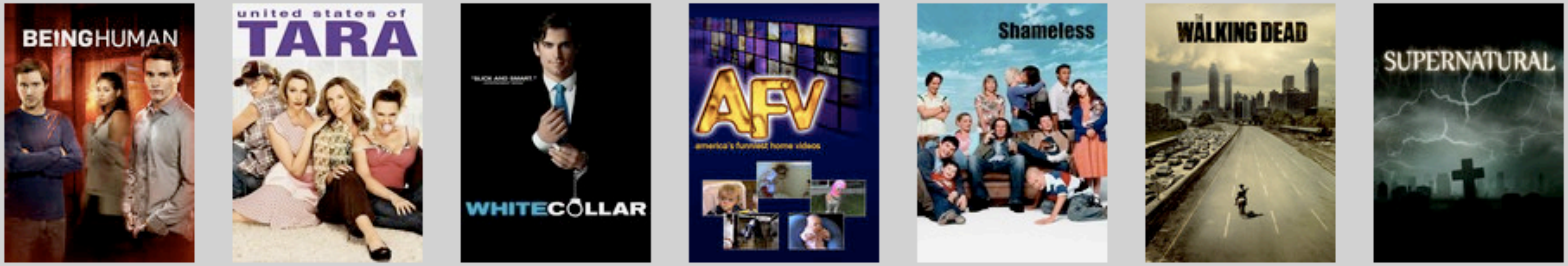
Outline

- Neighborhood methods
  - User / movie similarity
  - Iteration on graph
- Matrix Factorization
  - Singular value decomposition
  - Convex reformulation
- Ranking and Session Modeling
  - Ordinal regression
  - Session models
- Features
  - Latent dense (Bayesian Probabilistic Matrix Factorization)
  - Latent sparse (Dirichlet process factorization)
  - Coldstart problem (inferring features)
- Hashing

Why

Thousands of movies and TV episodes including these:

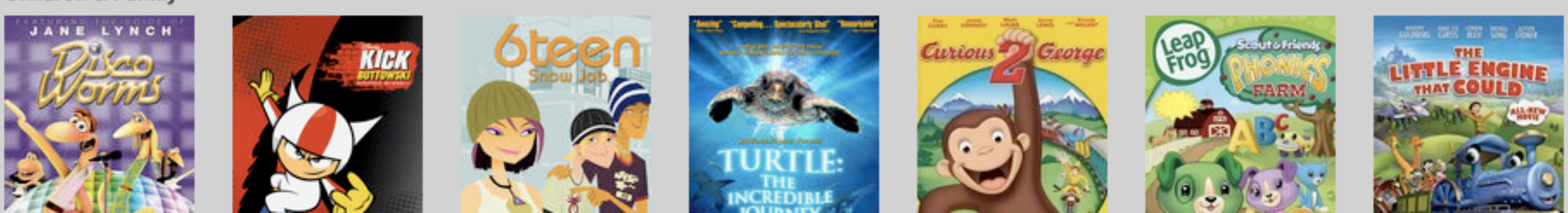
New Arrivals in TV



TV Drama

TV Comedy

Children & Family




Thousands of movies and TV episodes including these:
New Arrivals in TV
BEINGHUMAN
united states of TARA
WHITECOLLAR
AFV
america's funniest home videos
Shameless
WALKING DEAD
SUPERNATURAL
TV Drama
CRIMINAL MINDS
SUSPECT BEHAVIOR
FLASHPOINT
WORKAHOLICS
SAVING GRACE
MAD MEN
DAMAGES
weeds
TV Comedy
Parenthood
Raising HOPE
HOT in CLEVELAND
TV LAND
my boys
psych
archer
MONK
Children & Family
JANE LYNCH
Disco Worms
KICK BUTTOWSKI
6teen
Snow Job
TURTLE: THE INCREDIBLE JOURNEY
Curious George 2
Leap Frog
Scout & Friends
PHONICS FARM
THE LITTLE ENGINE THAT COULD

Netflix

amazon.com Prime

Hello, Alexander Smola. We have recommendations for you. (Not Alexander?)

Alexander's Amazon.com | Today's Deals | Gifts & Wish Lists | Gift Cards

Easter Clothes, Candy, Décor and More

Your Digital Items | Your Account | Help

Shop All Departments | Search | Amazon Instant Video | GO | Cart | Wish List

Amazon Instant Video | Most Popular | Getting Started | Watch Anywhere | Prime Instant Videos | Your Video Library | Passes and Pre-orders | Get Help | DVD & Blu-ray

Three Kings R

(398 customer reviews)

George Clooney, Mark Wahlberg, Ice Cube conspire to steal a huge cache of gold hidden near their desert base.

Starring: George Clooney, Mark Wahlberg
Directed by: David O. Russell
Runtime: 1 hour 56 minutes
Release year: 1999
Studio: Warner Bros.

Also available in **HD** with **Amazon Instant Video on Your TV**

Prime

Your Amazon Prime membership now includes unlimited, commercial-free, instant streaming of thousands of movies and TV shows at no additional cost.

amazon instant video

Prime instant videos

Watch now unlimited streaming $0.00

48 hour rental

1-Click® $2.99

Buy movie

1-Click® $9.99

Learn more about renting and buying

Add to Wish List

[Send us Feedback]

Have a promotion code? Redeem a gift card or promotion code View Balance

Customers Who Bought This Item Also Bought

Page 1 of 19

Tower Heist Amazon Instant Video ~ Eddie Murphy (42) $3.99

Syriana Amazon Instant Video ~ George Clooney (355) $2.99

Five Minutes of Heaven Amazon Instant Video ~ Liam Neeson (70) $2.99

Foolproof Amazon Instant Video ~ Ryan Reynolds (81) $2.99

The Recruit Amazon Instant Video ~ Al Pacino (161) $1.99

Alexander's Amazon.com › **Recommended for You**
(If you're not Alexander Smola, click here.)

Just For Today

Browse Recommended

Recommendations

- Amazon Instant Video
- Appstore for Android
- Baby
- Beauty
- Books
- Books on Kindle
- Camera & Photo
- Clothing & Accessories
- Computers & Accessories
- Electronics
- Grocery & Gourmet Food
- Health & Personal Care
- Home Improvement
- Industrial & Scientific
- Jewelry
- Kitchen & Dining
- MP3 Downloads
- Magazine Subscriptions
- Movies & TV
- Music
- Musical Instruments
- Office & School Supplies
- Patio, Lawn & Garden
- Shoes
- Software
- Sports & Outdoors
- Toys & Games
- Video Games
- Watches

These recommendations are based on items you own and more.

view: **All** | New Releases | Coming Soon   More results

1. 

Convex Optimization
by Stephen Boyd (March 8, 2004)
Average Customer Review: (13)
In Stock
List Price: ~~$84.00~~
Price: $68.13
44 used & new from $61.32
Add to Cart | Add to Wish List
I own it | Not interested | Rate this item
Recommended because you purchased **Nonlinear Programming** and more (Fix this)

2. 

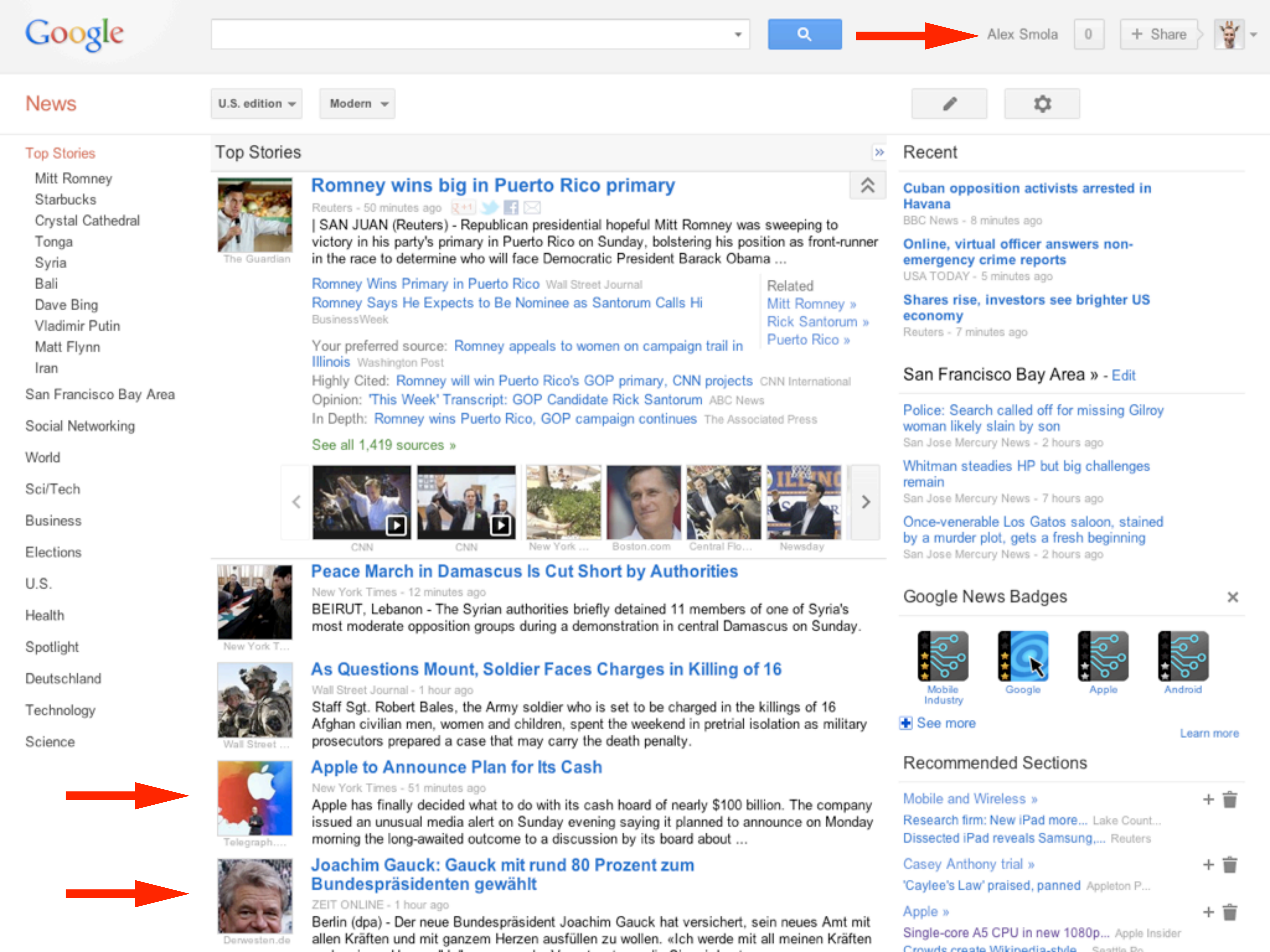
Probabilistic Graphical Models: Principles and Techniques (Adaptive Computation and Machine Learning series)
by Nir Friedman (July 31, 2009)
Average Customer Review: (11)
In Stock
List Price: ~~$95.00~~
Price: $93.55
47 used & new from $91.33
Add to Cart | Add to Wish List
I own it | Not interested | Rate this item
Recommended because you purchased **Nonlinear Programming** and more (Fix this)

3. 

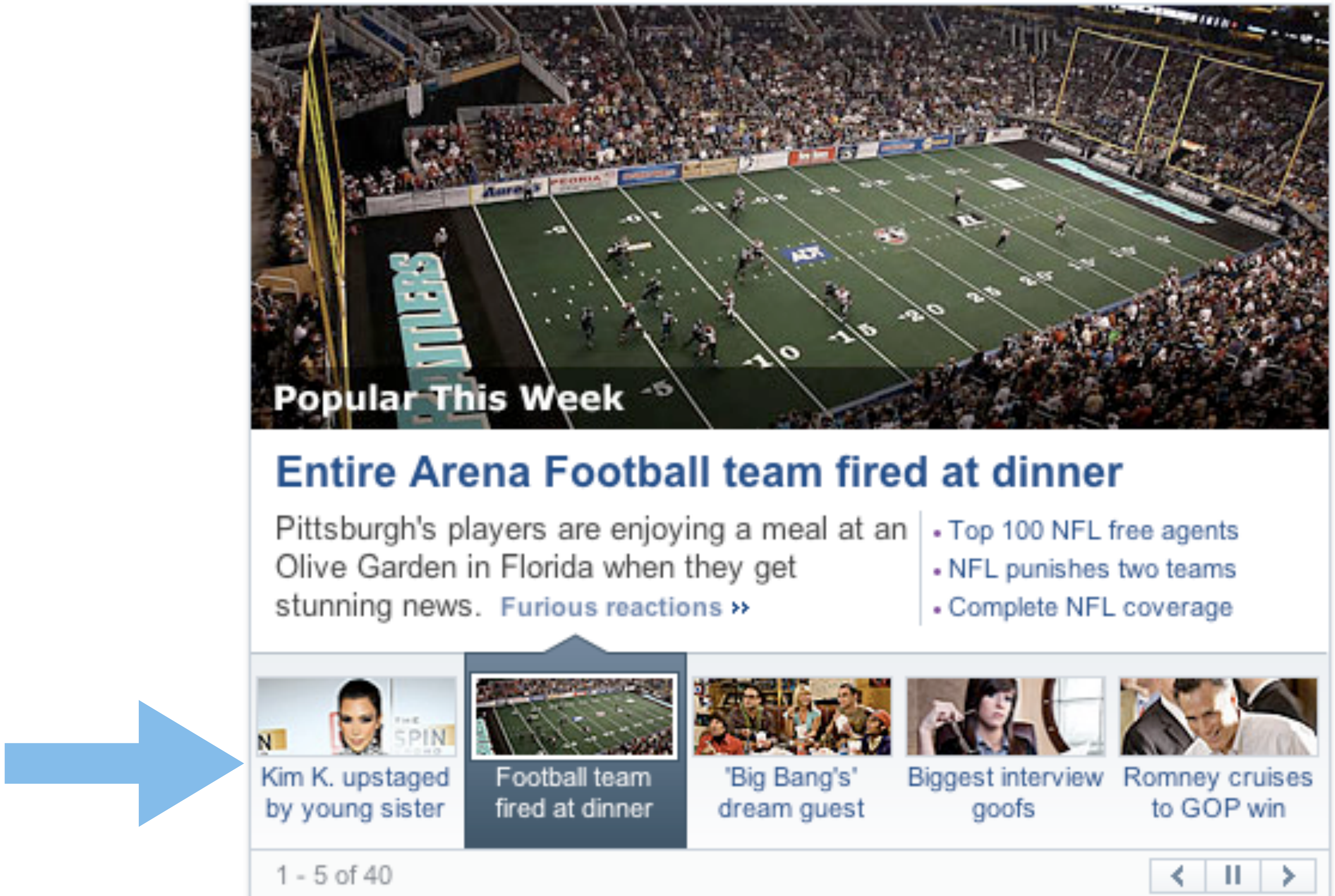
Doing Bayesian Data Analysis: A Tutorial with R and BUGS
by John K. Kruschke (November 10, 2010)
Average Customer Review: (15)
In Stock
List Price: ~~$89.95~~
Price: $77.98
44 used & new from $68.40
Add to Cart | Add to Wish List
I own it | Not interested | Rate this item
Recommended because you purchased **Bayesian Nonparametrics** and more (Fix this)

4. 

Parallel and Distributed Computation: Numerical Methods (Optimization and Neural Computation)
by Dimitri P. Bertsekas (January 1, 1997)
Average Customer Review: (1)
In Stock
Price: $49.50
15 used & new from $45.49
Add to Cart | Add to Wish List
I own it | Not interested | Rate this item
Recommended because you purchased **Nonlinear Programming** (Fix this)


Google
Alex Smola
0
Share
News
U.S. edition
Modern
Top Stories
Mitt Romney
Starbucks
Crystal Cathedral
Tonga
Syria
Bali
Dave Bing
Vladimir Putin
Matt Flynn
Iran
San Francisco Bay Area
Social Networking
World
Sci/Tech
Business
Elections
U.S.
Health
Spotlight
Deutschland
Technology
Science
Top Stories
Romney wins big in Puerto Rico primary
Reuters - 50 minutes ago
The Guardian
| SAN JUAN (Reuters) - Republican presidential hopeful Mitt Romney was sweeping to victory in his party's primary in Puerto Rico on Sunday, bolstering his position as front-runner in the race to determine who will face Democratic President Barack Obama ...
Romney Wins Primary in Puerto Rico Wall Street Journal
Romney Says He Expects to Be Nominee as Santorum Calls Hi BusinessWeek
Related
Mitt Romney »
Rick Santorum »
Puerto Rico »
Your preferred source: Romney appeals to women on campaign trail in Illinois Washington Post
Highly Cited: Romney will win Puerto Rico's GOP primary, CNN projects CNN International
Opinion: 'This Week' Transcript: GOP Candidate Rick Santorum ABC News
In Depth: Romney wins Puerto Rico, GOP campaign continues The Associated Press
See all 1,419 sources »
CNN
CNN
New York ...
Boston.com
Central Flo...
Newsday
Peace March in Damascus Is Cut Short by Authorities
New York Times - 12 minutes ago
New York T...
BEIRUT, Lebanon - The Syrian authorities briefly detained 11 members of one of Syria's most moderate opposition groups during a demonstration in central Damascus on Sunday.
As Questions Mount, Soldier Faces Charges in Killing of 16
Wall Street Journal - 1 hour ago
Wall Street ...
Staff Sgt. Robert Bales, the Army soldier who is set to be charged in the killings of 16 Afghan civilian men, women and children, spent the weekend in pretrial isolation as military prosecutors prepared a case that may carry the death penalty.
Apple to Announce Plan for Its Cash
New York Times - 51 minutes ago
Telegraph....
Apple has finally decided what to do with its cash hoard of nearly $100 billion. The company issued an unusual media alert on Sunday evening saying it planned to announce on Monday morning the long-awaited outcome to a discussion by its board about ...
Joachim Gauck: Gauck mit rund 80 Prozent zum Bundespräsidenten gewählt
ZEIT ONLINE - 1 hour ago
Derwesten.de
Berlin (dpa) - Der neue Bundespräsident Joachim Gauck hat versichert, sein neues Amt mit allen Kräften und mit ganzem Herzen ausfüllen zu wollen. «Ich werde mit all meinen Kräften
Recent
Cuban opposition activists arrested in Havana
BBC News - 8 minutes ago
Online, virtual officer answers non-emergency crime reports
USA TODAY - 5 minutes ago
Shares rise, investors see brighter US economy
Reuters - 7 minutes ago
San Francisco Bay Area » - Edit
Police: Search called off for missing Gilroy woman likely slain by son
San Jose Mercury News - 2 hours ago
Whitman steadies HP but big challenges remain
San Jose Mercury News - 7 hours ago
Once-venerable Los Gatos saloon, stained by a murder plot, gets a fresh beginning
San Jose Mercury News - 2 hours ago
Google News Badges
Mobile Industry
Google
Apple
Android
See more
Learn more
Recommended Sections
Mobile and Wireless »
Research firm: New iPad more... Lake Count...
Dissected iPad reveals Samsung,... Reuters
Casey Anthony trial »
'Caylee's Law' praised, panned Appleton P...
Apple »
Single-core A5 CPU in new 1080p... Apple Insider

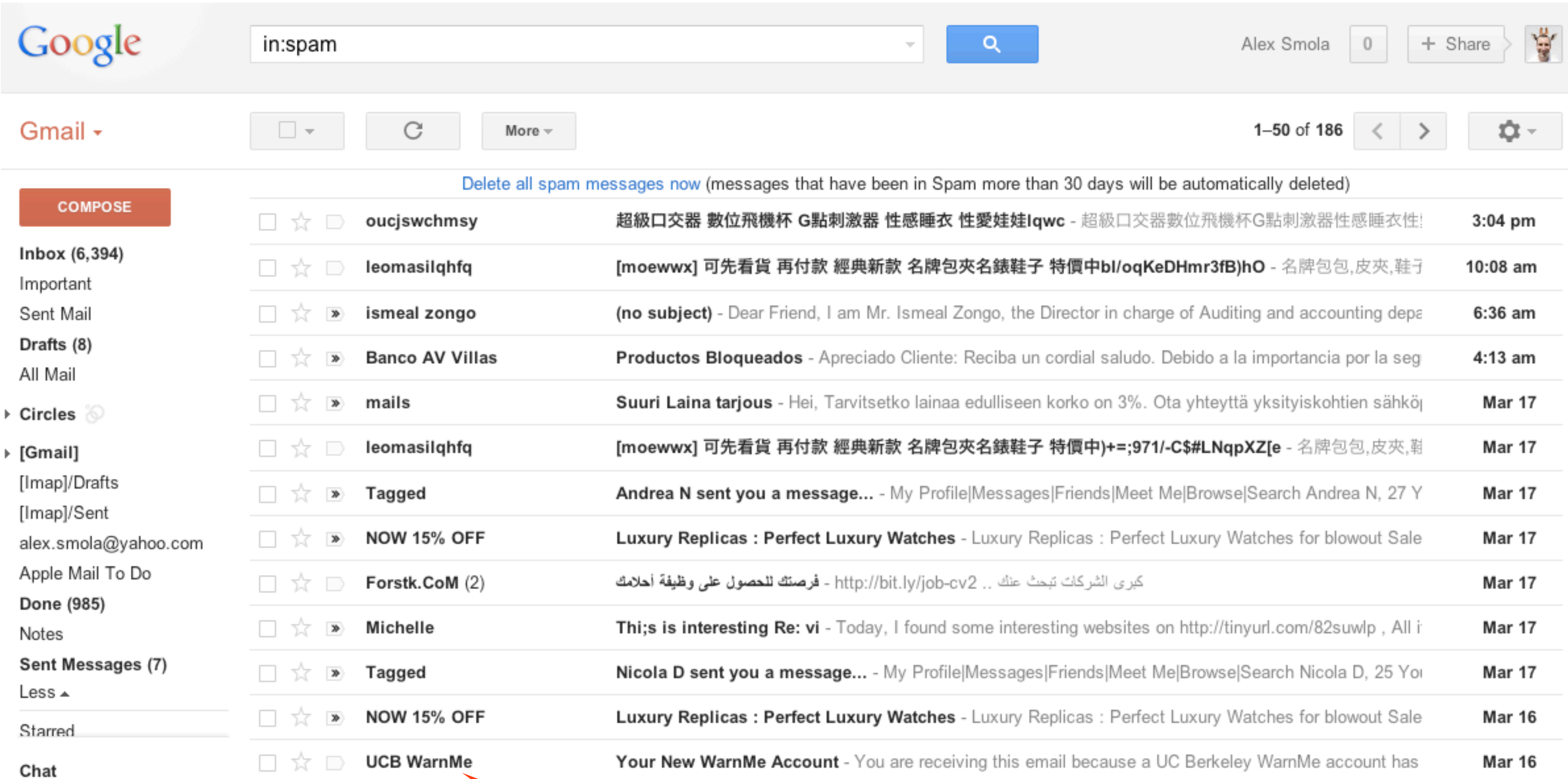
Personalized Content



adapt to general popularity

pick based on user preferences

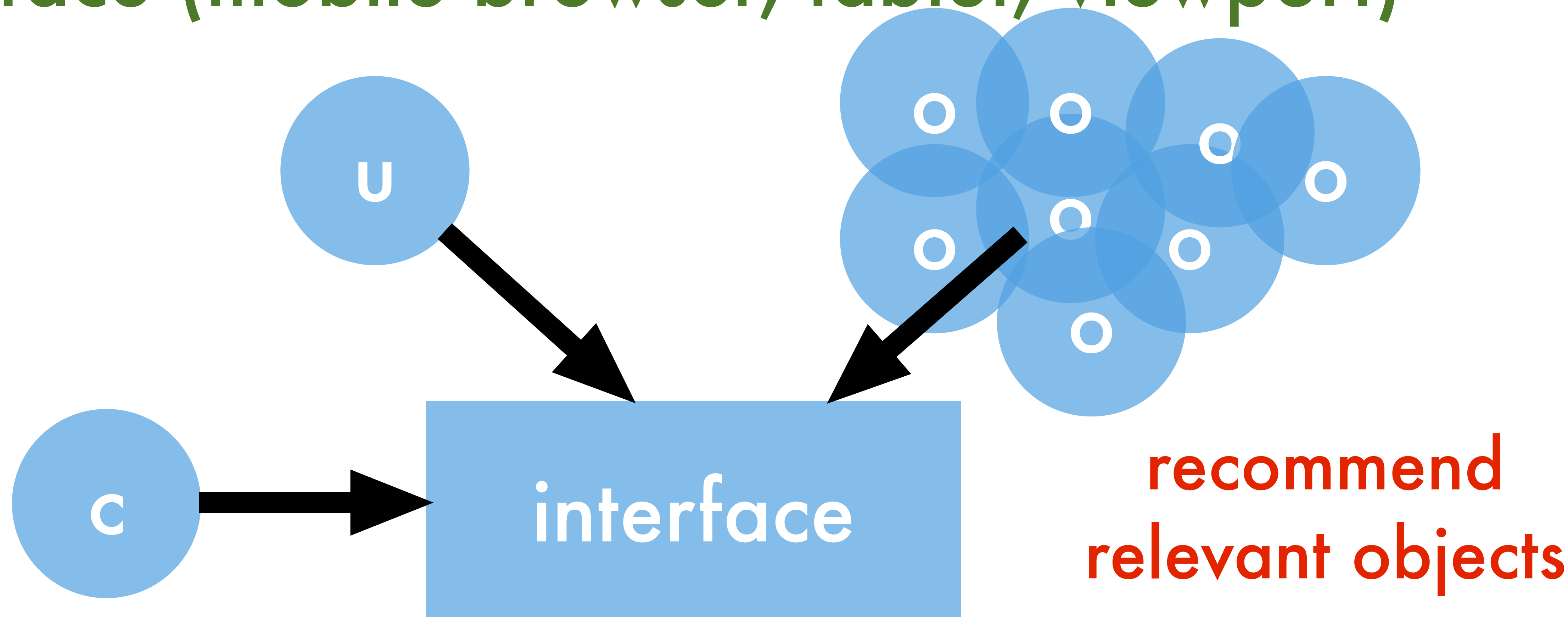
Spam Filtering



Something went wrong!

A more formal view

- User (requests content)
- Objects (that can be displayed)
- Context (device, location, time)
- Interface (mobile browser, tablet, viewport)



Examples

- Movie recommendation (Netflix)
- Related product recommendation (Amazon)
- Web page ranking (Google)
- Social recommendation (Facebook)
- News content recommendation (Yahoo)
- Priority inbox & spam filtering (Google)
- Online dating (OK Cupid)
- Computational Advertising (Yahoo)

Running Example
Netflix Movie Recommendation

Training data

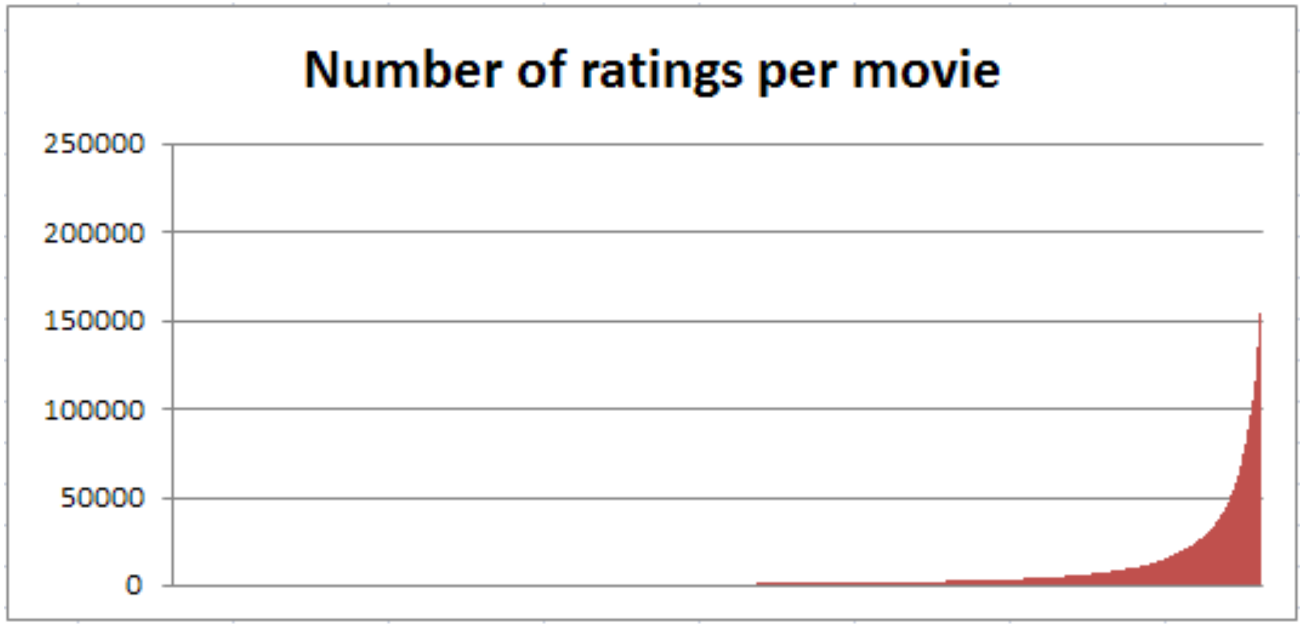
user	movie	date	score
1	21	5/7/02	1
1	213	8/2/04	5
2	345	3/6/01	4
2	123	5/1/05	4
2	768	7/15/02	3
3	76	1/22/01	5
4	45	8/3/00	4
5	568	9/10/05	1
5	342	3/5/03	2
5	234	12/28/00	2
6	76	8/11/02	5
6	56	6/15/03	4

Test data

user	movie	date	score
1	62	1/6/05	?
1	96	9/13/04	?
2	7	8/18/05	?
2	3	11/22/05	?
3	47	6/13/02	?
3	15	8/12/01	?
4	41	9/1/00	?
4	28	8/27/05	?
5	93	4/4/05	?
5	74	7/16/03	?
6	69	2/14/04	?
6	83	10/3/03	?

Challenges

- Scalability
  - Millions of objects
  - 100s of millions of users
- Cold start
  - Changing user base
  - Changing inventory (movies, stories, goods)
  - Attributes
- Imbalanced dataset
  User activity / item reviews
  are power law distributed



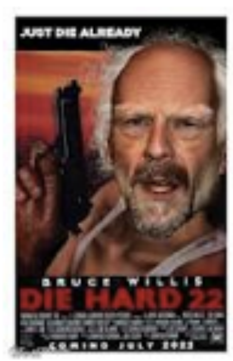
http://www.igvita.com/2006/10/29/dissecting-the-netflix-dataset/

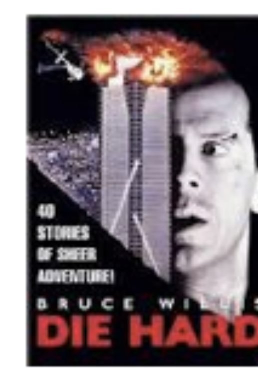
Netflix competition yardstick

- Least mean squares prediction error
  - Easy to define

$$\mathrm{rmse}(S) = \sqrt{|S|^{-1} \sum_{(i,u)\in S} (\hat{r}_{ui} - r_{ui})^2}$$

  - Wrong measure for composing sessions!

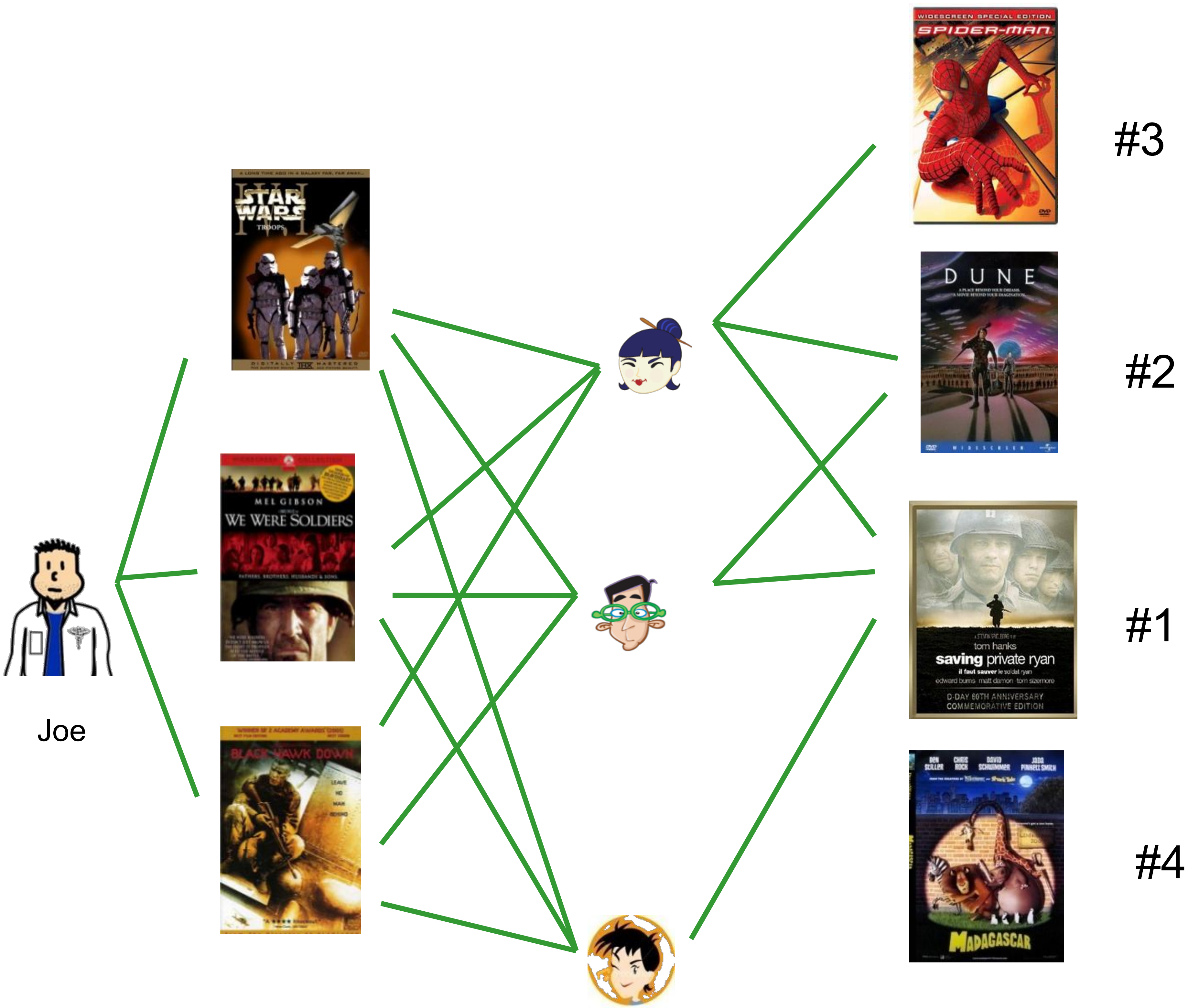


  - Consistent (in large sample size limit this will converge to minimizer)

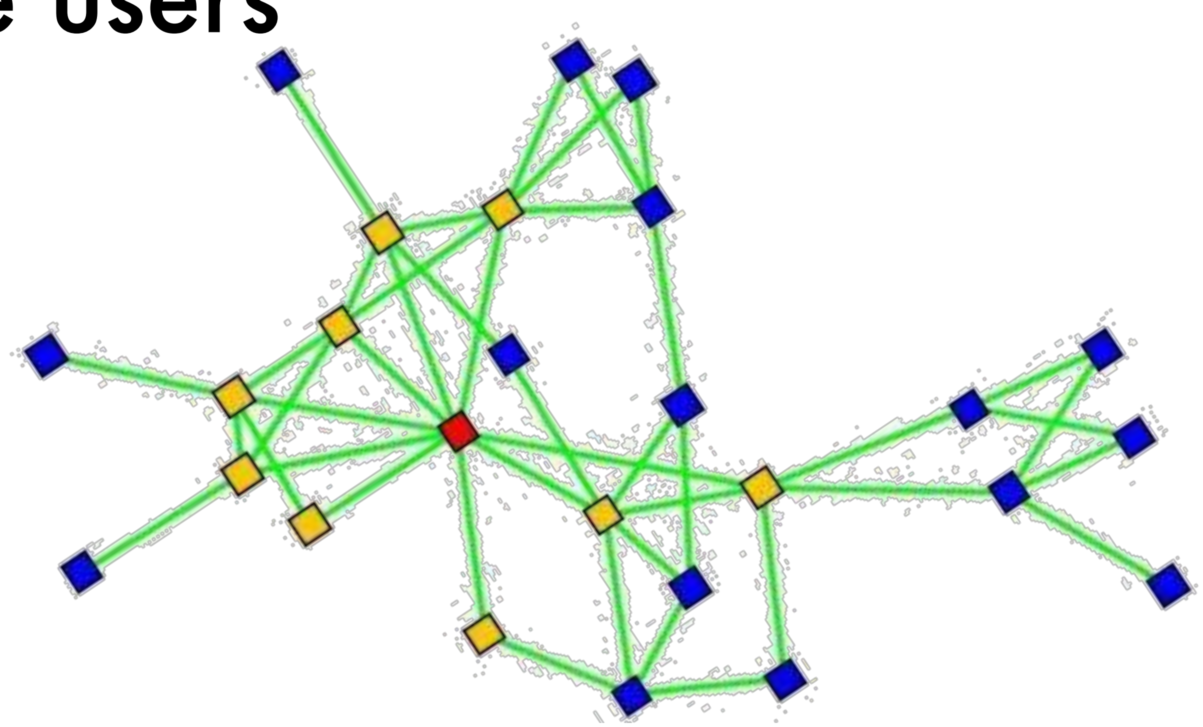
1 Neighborhood Methods

Basic Idea



Basic Idea

- (user,user) similarity to recommend items
  - good if item base is smaller than user base
  - good if item base changes rapidly
  - traverse bipartite similarity graph
- (item,item) similarity to recommend new items that were also liked by the same users
  - good if the user base is small is small
- Oldest known CF method



Neighborhood based CF

users

items	1	2	3	4	5	6	7	8	9	10	11	12
1	1		3			5			5		4	
2			5	4			4			2	1	3
3	2	4		1	2		3		4	3	5	
4		2	4		5			4			2	
5			4	3	4	2					2	5
6	1		3		3			2			4	

☐ - unknown rating ■ - rating between 1 to 5

Neighborhood based CF

users

items	1	2	3	4	5	6	7	8	9	10	11	12
1	1		3		?	5			5		4	
2			5	4			4			2	1	3
3	2	4		1	2		3		4	3	5	
4		2	4		5			4			2	
5			4	3	4	2					2	5
6	1		3		3			2			4	

- unknown rating (white cell)

- rating between 1 to 5 (yellow cell)

Neighborhood based CF

users

items	1	2	3	4	5	6	7	8	9	10	11	12
1	1		3		?	5			5		4	
2			5	4			4			2	1	3
3	2	4		1	2		3		4	3	5	
4		2	4		5			4			2	
5			4	3	4	2					2	5
6	1		3		3			2			4	

- unknown rating (white cell)

- rating between 1 to 5 (yellow cell)

Neighborhood based CF

users

items	1	2	3	4	5	6	7	8	9	10	11	12
1	1		3		?	5			5		4	
2			5	4			4			2	1	3
3	2	4		1	2		3		4	3	5	
4		2	4		5			4			2	
5			4	3	4	2					2	5
6	1		3		3			2			4	

similarity

$s_{13} = 0.2$

$s_{16} = 0.3$

- unknown rating

- rating between 1 to 5

Neighborhood based CF

users

items	1	2	3	4	5	6	7	8	9	10	11	12
1	1		3		2.6	5			5		4	
2			5	4			4			2	1	3
$\boxed{3}$	2	4		1	$\boxed{2}$		3		4	3	5	
4		2	4		5			4			2	
5			4	3	4	2					2	5
$\boxed{6}$	1		3		$\boxed{3}$			2			4	

similarity

$s_{13} = 0.2$

$s_{16} = 0.3$

weighted average

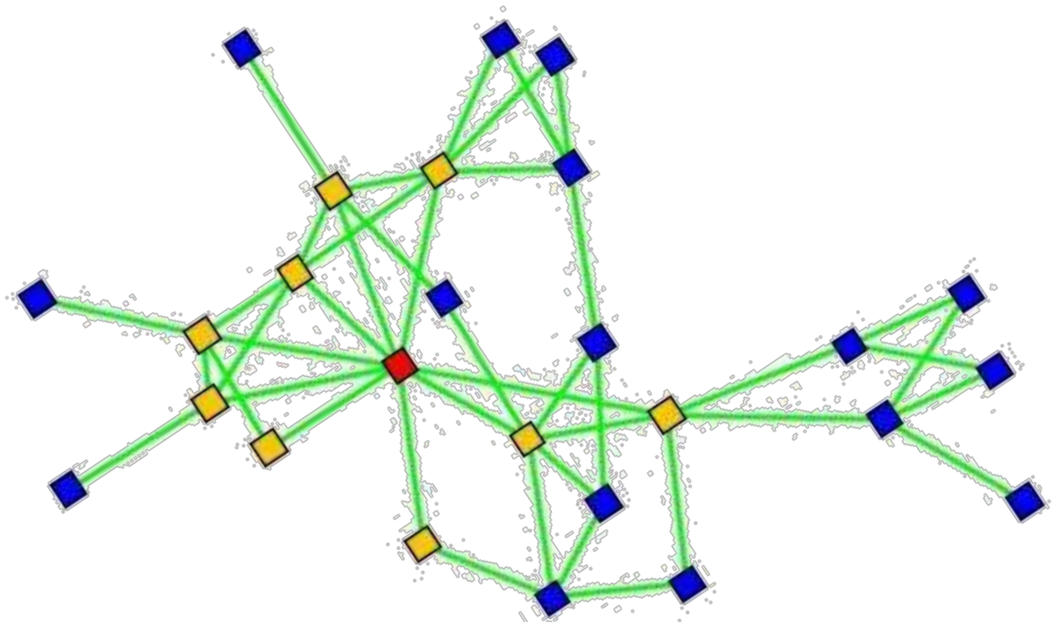
$$\frac{0.2 \cdot 2 + 0.3 \cdot 3}{0.2 + 0.3} = 2.6$$

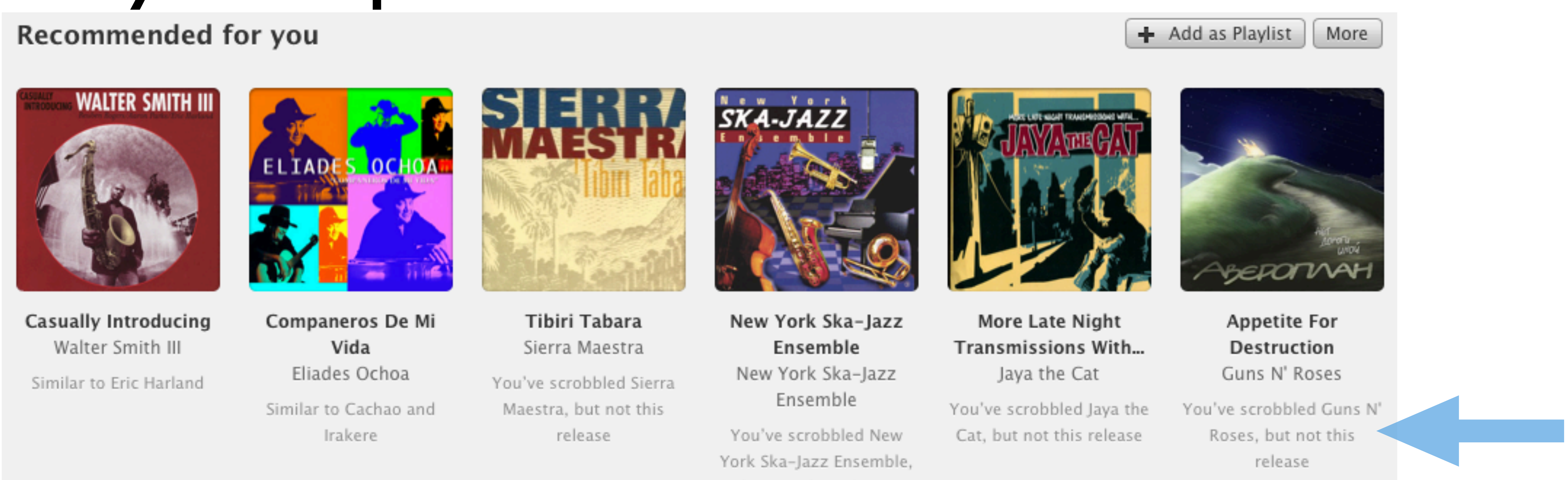
- unknown rating

- rating between 1 to 5

Properties

- Intuitive
- No (substantial) training
- Handles new users / items
- Easy to explain to user





- Accuracy & scalability questionable

Normalization / Bias

- Problem
  - Some items are significantly higher rated
  - Some users rate substantially lower
  - Ratings change over time
- Bias correction is crucial for nearest neighborhood recommender algorithm
  - Offset per user
  - Offset per movie
  - Time effects
  - Global bias

$$b_{ui} = \underbrace{\mu}_{\text{global}} + \underbrace{b_u}_{\text{user}} + \underbrace{b_i}_{\text{item}}$$

Bell & Koren ICDM 2007
http://public.research.att.com/~volinsky/netflix/BellKorICDM07.pdf

Baseline estimation

- Mean rating is 3.7
- Troll Hunter is 0.7 above mean
- User rates 0.2 below mean
- Baseline is 4.2 stars
- Least mean squares problem

$$\underset{b}{\text{minimize}} \sum_{(u,i)} (r_{ui} - \mu - b_u - b_i)^2 + \lambda \left[\sum_u b_u^2 + \sum_i b_i^2 \right]$$

- Jointly convex. Alternatively remove mean & iterate

$$b_i = \frac{\sum_{u \in R(i)} (r_{ui} - \mu - b_u)}{\lambda + |R(i)|} \text{ and } b_u = \frac{\sum_{i \in R(u)} (r_{ui} - \mu - b_i)}{\lambda + |R(u)|}$$

Parzen Windows style CF

- Similarity measure s_{ij} between items
- Find set $s_k(i,u)$ of k-nearest neighbors to i that were rated by user u
- Weighted average over the set

$$\hat{r}_{ui} = b_{ui} + \frac{\sum_{j \in s_k(i,u)} s_{ij}(r_{uj} - b_{uj})}{\sum_{j \in s_k(i,u)} s_{ij}} \text{ where } b_{ui} = \mu + b_u + b_i$$

- How to compute s_{ij}?

(item,item) similarity measures

User ratings for item i:

1	?	?	5	5	3	?	?	?	4	2	?	?	?	?	4	?	5	4	1	?

User ratings for item j:

?	?	4	2	5	?	?	1	2	5	?	?	2	?	?	3	?	?	?	5	4

- Pearson correlation coefficient
  - nonuniform support
  - compute only over shared support
  - shrinkage towards 0 to address problem of small support (typically few items in common)

$$s_{ij} = \frac{\mathrm{Cov}[r_{ui}, r_{uj}]}{\mathrm{Std}[r_{ui}]\mathrm{Std}[r_{uj}]}$$

(item,item) similarity

- Empirical Pearson correlation coefficient

$$\hat{\rho}_{ij} = \frac{\sum_{u \in U(i,j)} (r_{ui} - b_{ui})(r_{uj} - b_{uj})}{\sqrt{\sum_{u \in U(i,j)} (r_{ui} - b_{ui})^2 \sum_{u \in U(i,j)} (r_{uj} - b_{uj})^2}}$$

- Smoothing towards 0 for small support

$$s_{ij} = \frac{|U(i,j)| - 1}{|U(i,j)| - 1 + \lambda} \hat{\rho}_{ij}$$

- Make neighborhood more peaked $s_{ij} \to s_{ij}^2$
- Shrink towards baseline for small neighborhood

$$\hat{r}_{ui} = b_{ui} + \frac{\sum_{j \in s_k(i,u)} s_{ij} (r_{uj} - b_{uj})}{\lambda + \sum_{j \in s_k(i,u)} s_{ij}}$$

Similarity for binary data

- Pearson correlation meaningless
  - Views
  - Purchase behavior
  - Clicks
- Jaccard similarity (intersection vs. joint)
- Observed/expected ratio Improve by counting per user (many users better than heavy users)

m_i users acting on i

m_{ij} users acting on both i and j

m total number of users

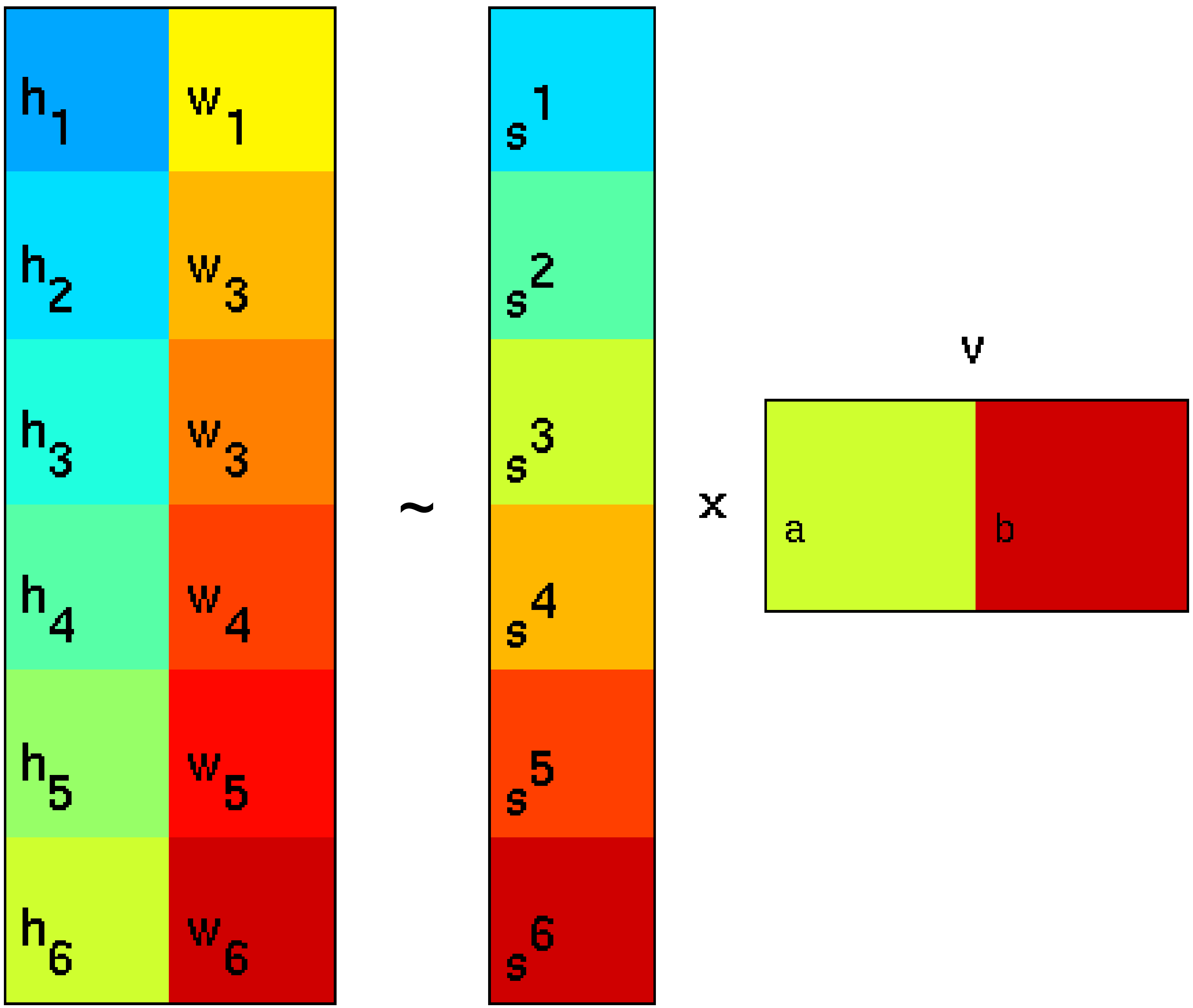
$$s_{ij} = \frac{m_{ij}}{\alpha + m_i + m_j - m_{ij}}$$

$$s_{ij} = \frac{\text{observed}}{\text{expected}} \approx \frac{m_{ij}}{\alpha + m_i m_j / m}$$

2 Matrix Factorization

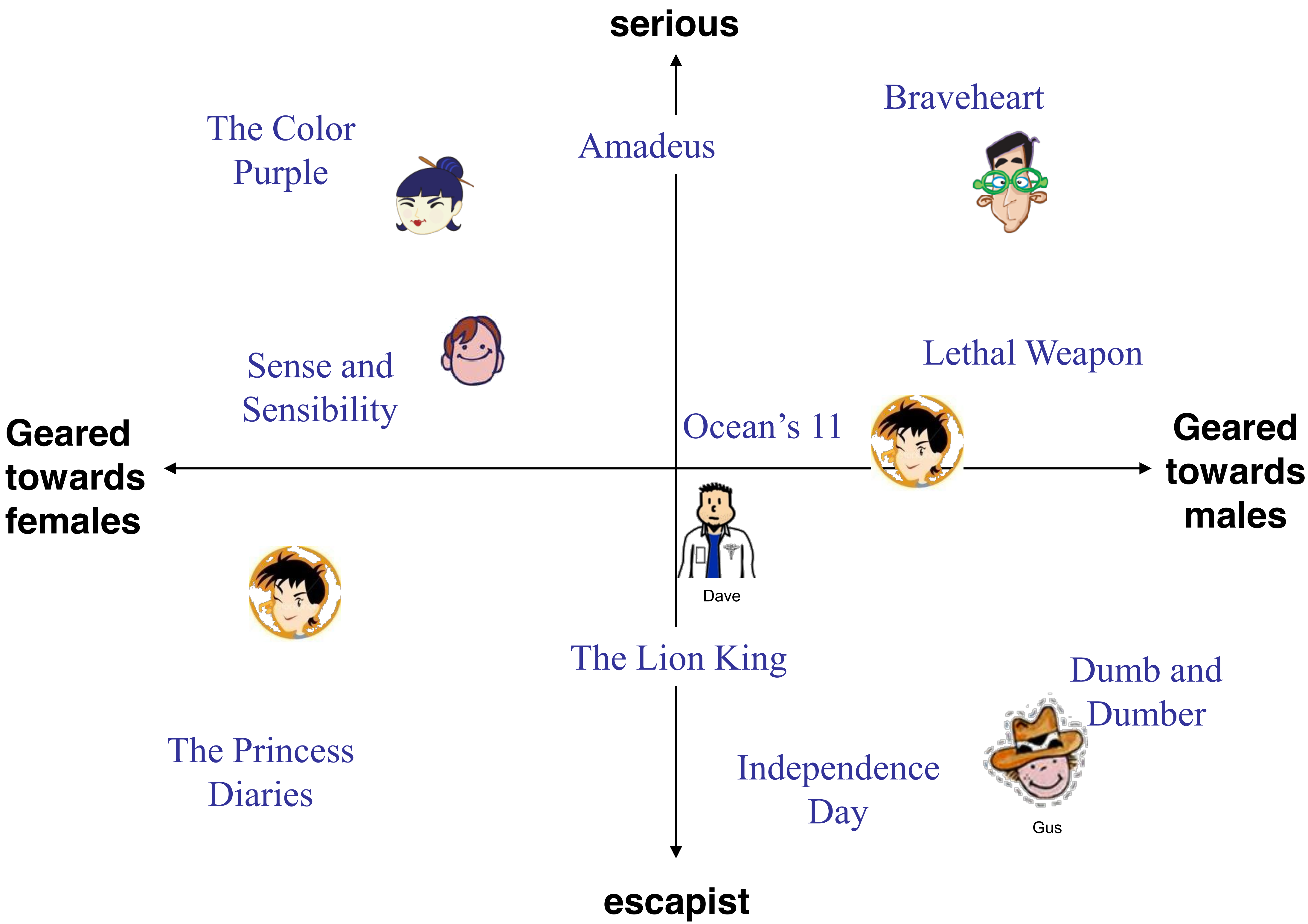
Basics

Basic Idea



$$M \approx U \cdot V$$

Latent variable view



Basic matrix factorization

users

items

1		3			5			5		4	
		5	4			4			2	1	3
2	4		1	2		3		4	3	5	
	2	4		5			4			2	
		4	3	4	2					2	5
1		3		3			2			4	

~

~ items

.1	-.4	.2
-.5	.6	.5
-.2	.3	.5
1.1	2.1	.3
-.7	2.1	-2
-1	.7	.3

•

users

1.1	-.2	.3	.5	-2	-.5	.8	-.4	.3	1.4	2.4	-.9
-.8	.7	.5	1.4	.3	-1	1.4	2.9	-.7	1.2	-.1	1.3
2.1	-.4	.6	1.7	2.4	.9	-.3	.4	.8	.7	-.6	.1

A rank-3 SVD approximation

Estimate unknown ratings as inner products of latent factors

users

items

1		3			5			5		4	
		5		?		4			2	1	3
2	4		1	2		3		4	3	5	
	2	4		5			4			2	
		4	3	4	2					2	5
1		3		3			2			4	

~

~ items

.1	-.4	.2
-.5	.6	.5
-.2	.3	.5
1.1	2.1	.3
-.7	2.1	-2
-1	.7	.3

●

users

1.1	-.2	.3	.5	-2	-.5	.8	-.4	.3	1.4	2.4	-.9
-.8	.7	.5	1.4	.3	-1	1.4	2.9	-.7	1.2	-.1	1.3
2.1	-.4	.6	1.7	2.4	.9	-.3	.4	.8	.7	-.6	.1

A rank-3 SVD approximation

Estimate unknown ratings as inner products of latent factors

users

items

1		3			5			5		4	
		5		?		4			2	1	3
2	4		1	2		3		4	3	5	
	2	4		5			4			2	
		4	3	4	2					2	5
1		3		3			2			4	

~

~ items

.1	-.4	.2
-.5	.6	.5
-.2	.3	.5
1.1	2.1	.3
-.7	2.1	-2
-1	.7	.3

•

users

1.1	-.2	.3	.5	-2	-.5	.8	-.4	.3	1.4	2.4	-.9
-.8	.7	.5	1.4	.3	-1	1.4	2.9	-.7	1.2	-.1	1.3
2.1	-.4	.6	1.7	2.4	.9	-.3	.4	.8	.7	-.6	.1

A rank-3 SVD approximation

Estimate unknown ratings as inner products of latent factors

users

items

1		3			5			5		4	
		5		2.4		4			2	1	3
2	4		1	2		3		4	3	5	
	2	4		5			4			2	
		4	3	4	2					2	5
1		3		3			2			4	

~

~

items

.1	-.4	.2
-.5	.6	.5
-.2	.3	.5
1.1	2.1	.3
-.7	2.1	-2
-1	.7	.3

●

users

1.1	-.2	.3	.5	-2	-.5	.8	-.4	.3	1.4	2.4	-.9
-.8	.7	.5	1.4	.3	-1	1.4	2.9	-.7	1.2	-.1	1.3
2.1	-.4	.6	1.7	2.4	.9	-.3	.4	.8	.7	-.6	.1

A rank-3 SVD approximation

Properties

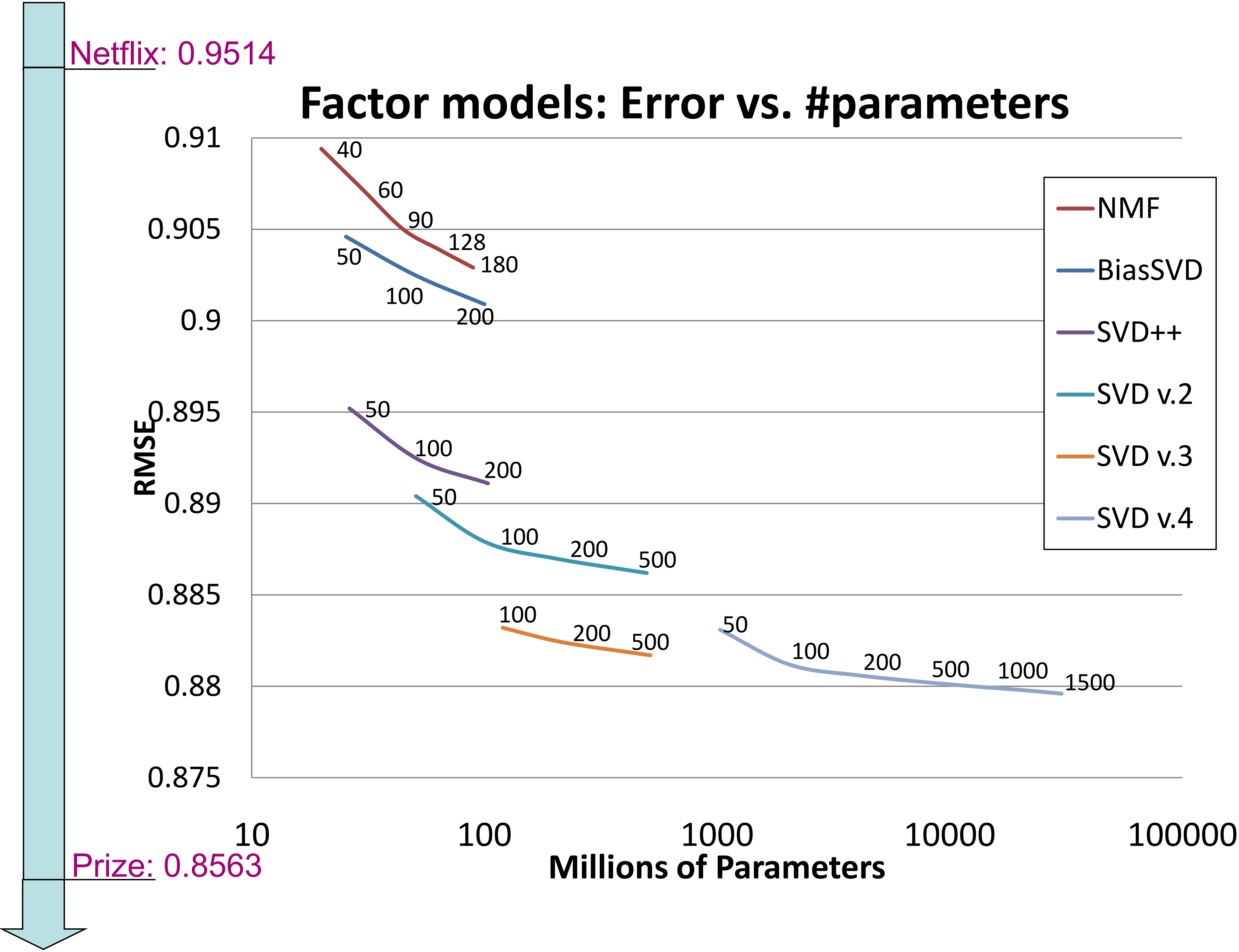
1		3			5			5		4	
		5	4			4			2	1	3
2	4		1	2		3		4	3	5	
	2	4		5			4			2	
		4	3	4	2					2	5
1		3		3			2			4	

~

.1	-.4	.2
-.5	.6	.5
-.2	.3	.5
1.1	2.1	.3
-.7	2.1	-2
-1	.7	.3

1.1	-.2	.3	.5	-2	-.5	.8	-.4	.3	1.4	2.4	-.9
-.8	.7	.5	1.4	.3	-1	1.4	2.9	-.7	1.2	-.1	1.3
2.1	-.4	.6	1.7	2.4	.9	-.3	.4	.8	.7	-.6	.1

- SVD is undefined for missing entries
    - stochastic gradient descent (faster)
    - alternating optimization
- Overfitting without regularization particularly if fewer reviews than dimensions
- Very popular on Netflix


Netflix: 0.9514
Factor models: Error vs. #parameters
RMSE
Millions of Parameters
NMF
BiasSVD
SVD++
SVD v.2
SVD v.3
SVD v.4
0.91
0.905
0.9
0.895
0.89
0.885
0.88
0.875
10
100
1000
10000
100000
40
60
90
128
180
50
100
200
50
100
200
50
100
200
500
100
200
500
50
100
200
500
1000
1500
Prize: 0.8563

Risk Minimization View

- **Objective Function**

$$\underset{p,q}{\text{minimize}} \sum_{(u,i)\in S} \left(r_{ui} - \langle p_u, q_i \rangle\right)^2 + \lambda \left[\|p\|^2_{\text{Frob}} + \|q\|^2_{\text{Frob}}\right]$$

- **Alternating least squares**

$$p_u \leftarrow \left[\lambda \mathbf{1} + \sum_{i|(u,i)\in S} q_i q_i^\top\right]^{-1} \sum_i q_i r_{ui}$$

$$q_i \leftarrow \left[\lambda \mathbf{1} + \sum_{u|(u,i)\in S} p_u p_u^\top\right]^{-1} \sum_i p_u r_{ui}$$

good for MapReduce

Risk Minimization View

- Objective Function

$$\underset{p,q}{\text{minimize}} \sum_{(u,i)\in S} (r_{ui} - \langle p_u, q_i \rangle)^2 + \lambda \left[\|p\|^2_{\text{Frob}} + \|q\|^2_{\text{Frob}} \right]$$

- Stochastic gradient descent

$$p_u \leftarrow (1 - \lambda\eta_t)p_u - \eta_t q_i(r_{ui} - \langle p_u, q_i \rangle)$$
$$q_i \leftarrow (1 - \lambda\eta_t)q_i - \eta_t p_u(r_{ui} - \langle p_u, q_i \rangle)$$

much faster

  - No need for locking
  - Multicore updates asynchronously (Recht, Re, Wright, 2012 - Hogwild)

Theoretical Motivation

deFinetti Theorem

- Independent random variables

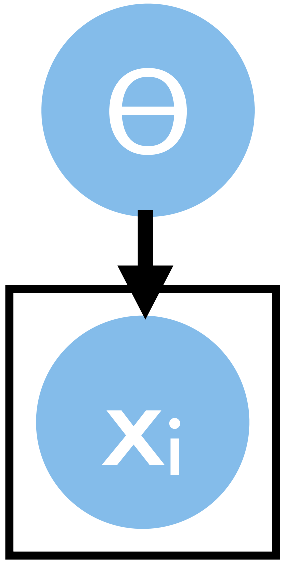
$$p(X) = \prod_{i=1}^{m} p(x_i)$$

- Exchangeable random variables

$$p(X) = p(x_1, \ldots, x_m) = p(x_{\pi(1)}, \ldots, x_{\pi(m)})$$

- There exists a conditionally independent representation of exchangeable r.v.

$$p(X) = \int dp(\theta) \prod_{i=1}^{m} p(x_i|\theta)$$



This motivates latent variable models

Aldous Hoover Factorization

- Matrix-valued set of random variable
  Example - Erdos Renyi graph model

$$p(E) = \prod_{i,j} p(V_{ij})$$

- Independently exchangeable on matrix

$$p(E) = p(E_{11}, E_{12}, \ldots, E_{mn}) = p(E_{\pi(1)\rho(1)}, E_{\pi(1)\rho(2)}, \ldots, E_{\pi(m)\rho(n)})$$

- Aldous Hoover Theorem

$$p(E) = \int dp(\theta) \int \prod_{i=1}^{m} dp(u_i) \prod_{j=1}^{n} dp(v_j) \prod_{i,j} p(E_{ij}|u_i, v_j, \theta)$$

Aldous Hoover Factorization

	u_1	u_2	u_3	u_4	u_5	u_6
v_1	e_{11}	e_{12}			e_{15}	e_{16}
v_2				e_{24}		
v_3		e_{32}				
v_4			e_{43}			e_{46}
v_5					e_{55}	

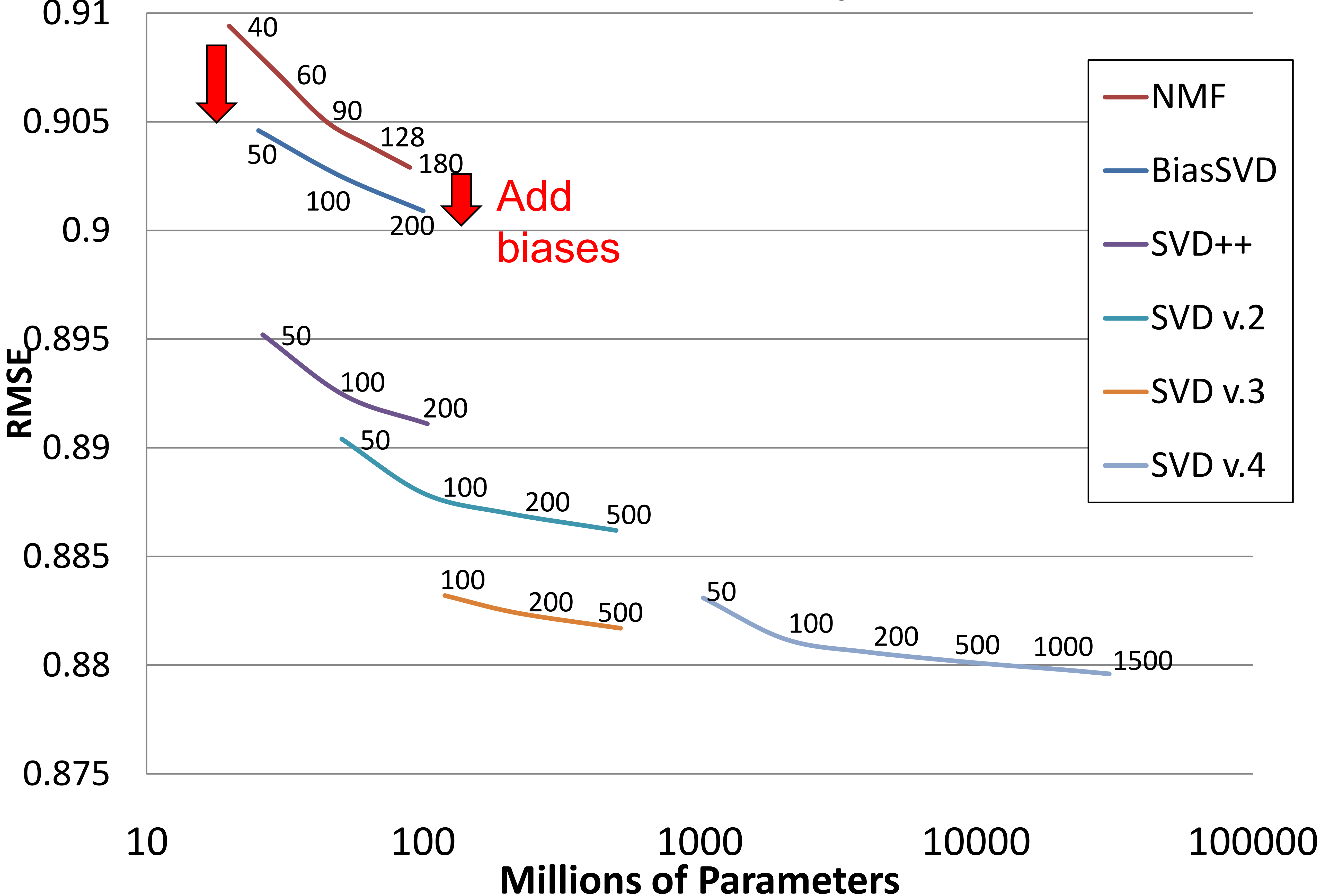
- Rating matrix is (row, column) exchangeable
- Draw latent variables per row and column
- Draw matrix entries independently given pairs
- Absence / presence of rating is a signal
- Can be extended to graphs with vertex attributes

Aldous Hoover variants

- Jointly exchangeable matrix
  - Social network graphs
  - Draw vertex attributes first, then edges
- Cold start problem
  - New user appears
  - Attributes (age, location, browser)
  - Can estimate latent variables from that
- User and item factors in matrix factorization problem can be viewed as AH-factors

Improvements

Factor models: Error vs. #parameters



Bias

- **Objective Function**

$$\underset{p,q}{\text{minimize}} \sum_{(u,i)\in S} \left(r_{ui} - (\mu + b_u + b_i + \langle p_u, q_i \rangle)\right)^2 +$$

$$\lambda \left[\|p\|_{\text{Frob}}^2 + \|q\|_{\text{Frob}}^2 + \|b_{\text{users}}\|^2 + \|b_{\text{items}}\|^2 \right]$$

- **Stochastic gradient descent**

$$p_u \leftarrow (1 - \lambda\eta_t) p_u - \eta_t q_i \rho_{ui}$$

$$q_i \leftarrow (1 - \lambda\eta_t) q_i - \eta_t p_u \rho_{ui}$$

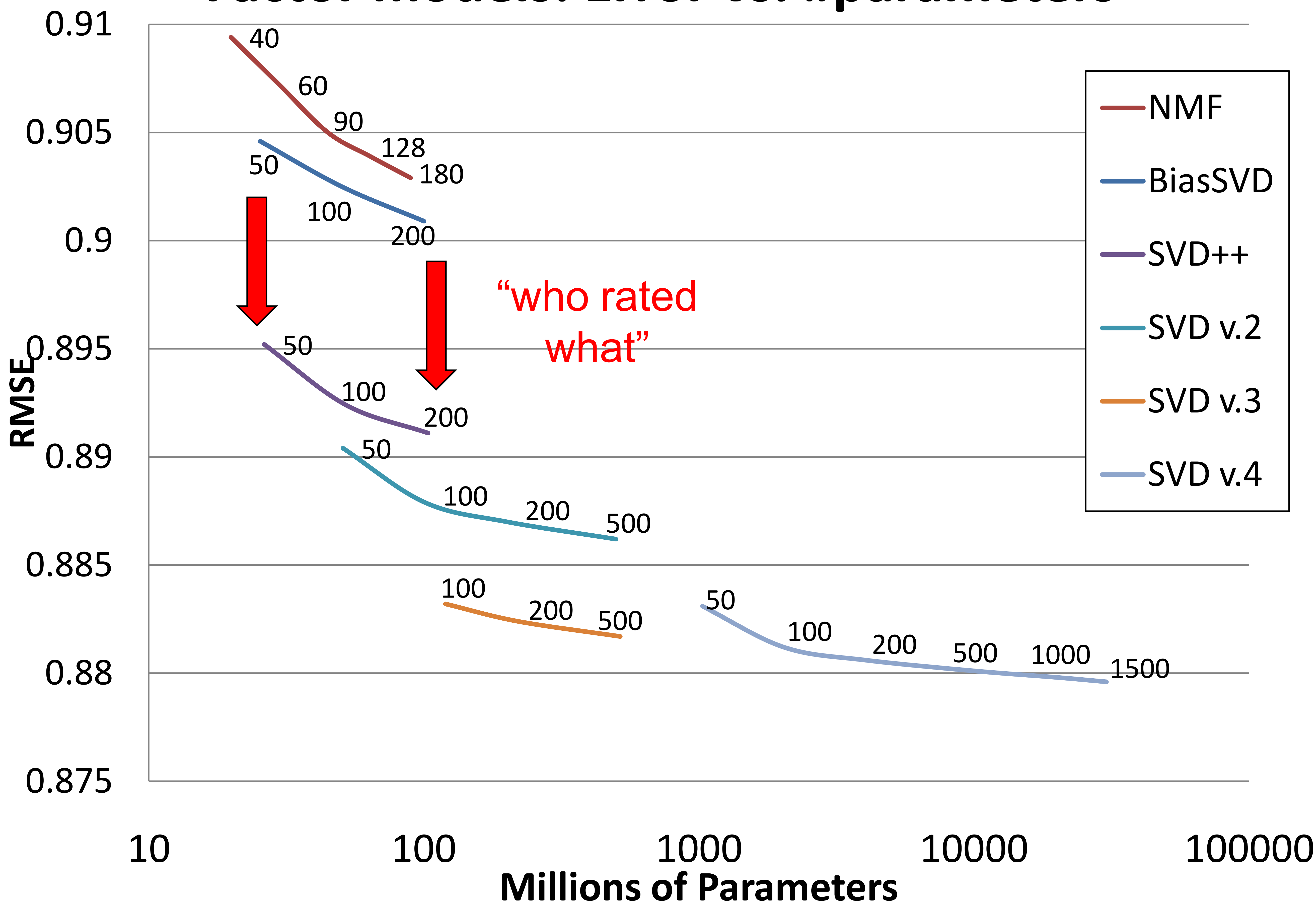
$$b_u \leftarrow (1 - \lambda\eta_t) b_u - \eta_t \rho_{ui}$$

$$b_i \leftarrow (1 - \lambda\eta_t) b_i - \eta_t \rho_{ui}$$

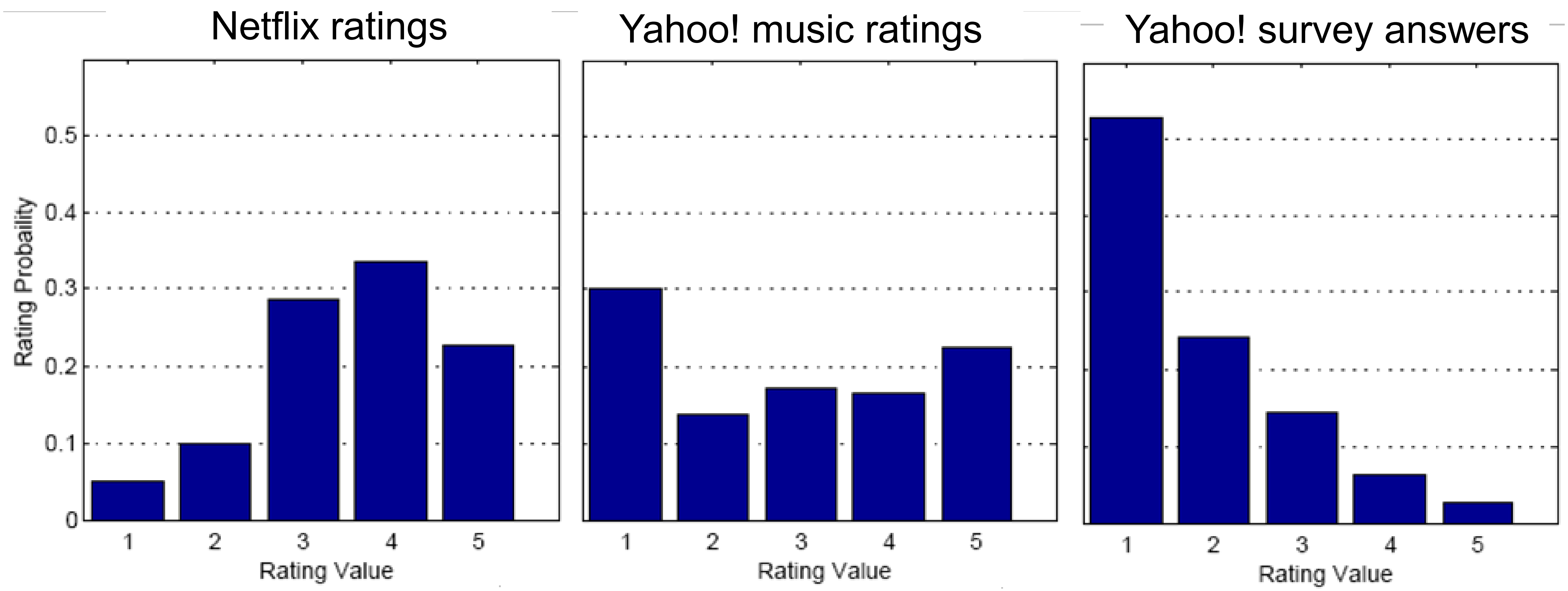
$$\mu \leftarrow (1 - \lambda\eta_t) \mu - \eta_t \rho_{ui}$$

$$\text{where } \rho_{ui} = \left(r_{ui} - (\mu + b_i + b_u + \langle p_u, q_i \rangle)\right)$$

Factor models: Error vs. #parameters



Ratings are not given at random



- Marlin et al. "Collaborative Filtering and the Missing at Random Assumption" UAI 2007

Movie rating matrix

users

movies

1		3			5			5		4	
		5	4			4			2	1	3
2	4		1	2		3		4	3	5	
	2	4		5			4			2	
		4	3	4	2					2	5
1		3		3			2			4	

r_{ui}

users

movies

1	0	1	0	0	1	0	0	1	0	1	0
0	0	1	1	0	0	1	0	0	1	1	1
1	1	0	1	1	0	1	0	1	1	1	0
0	1	1	0	1	0	0	1	0	0	1	0
0	0	1	1	1	1	0	0	0	0	1	1
1	0	1	0	1	0	0	1	0	0	1	0

c_{ui}

- Characterize users by **which** movies they rated
  Edge attributes (observed, rating)
- Adding features to recommender system

$$r_{ui} = \mu + b_u + b_i + \langle p_u, q_i \rangle + \langle c_u, x_i \rangle$$

regression

Alternative integration

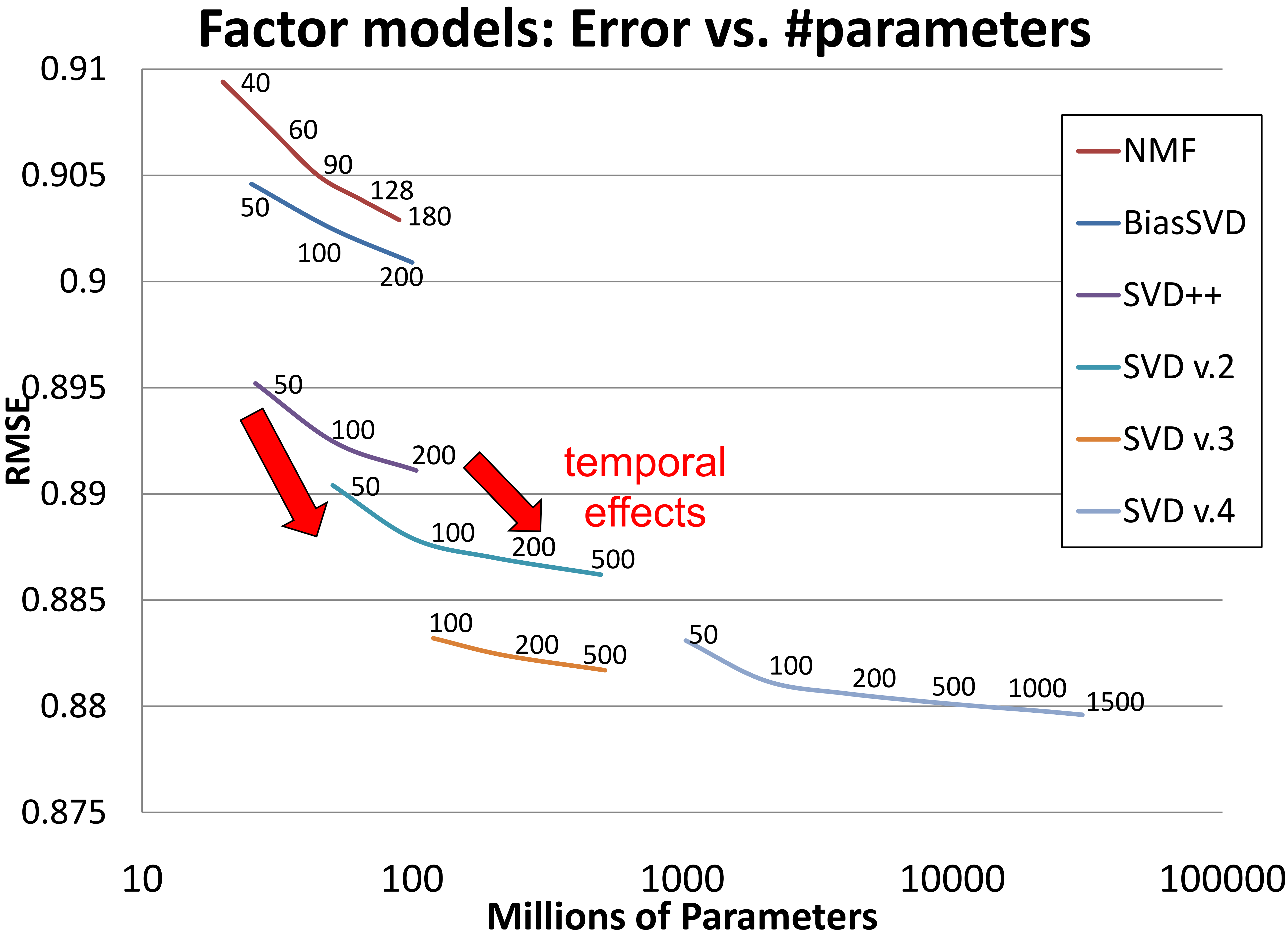
- Key idea - use related ratings to average
- Salakhudtinov & Mnih, 2007

$$q_i \leftarrow q_i + \sum_u c_{ui} p_u$$

- Koren et al., 2008

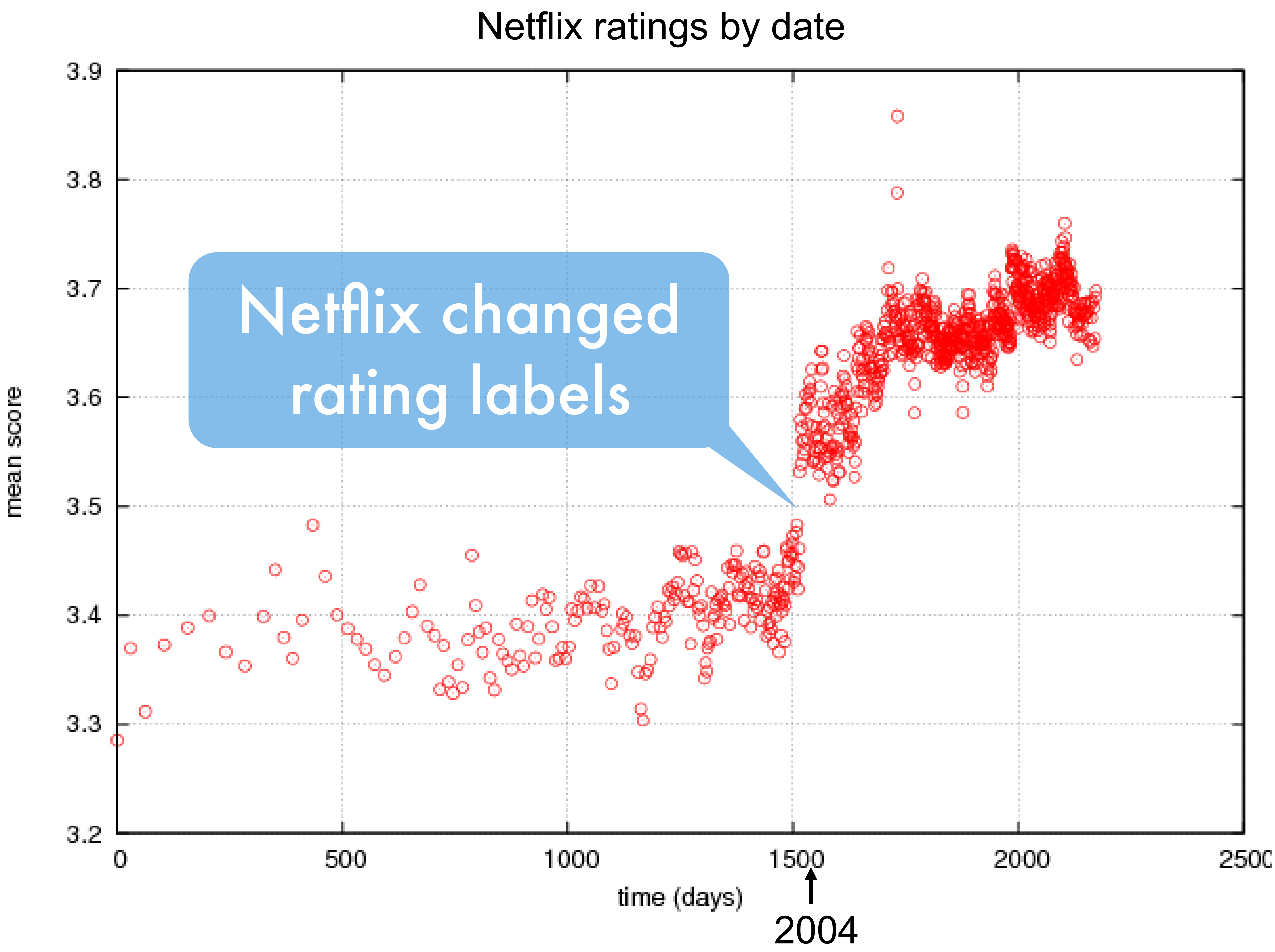
$$q_i \leftarrow q_i + \sum_u c_{ui} x_j$$

Overparametrize items by q and x

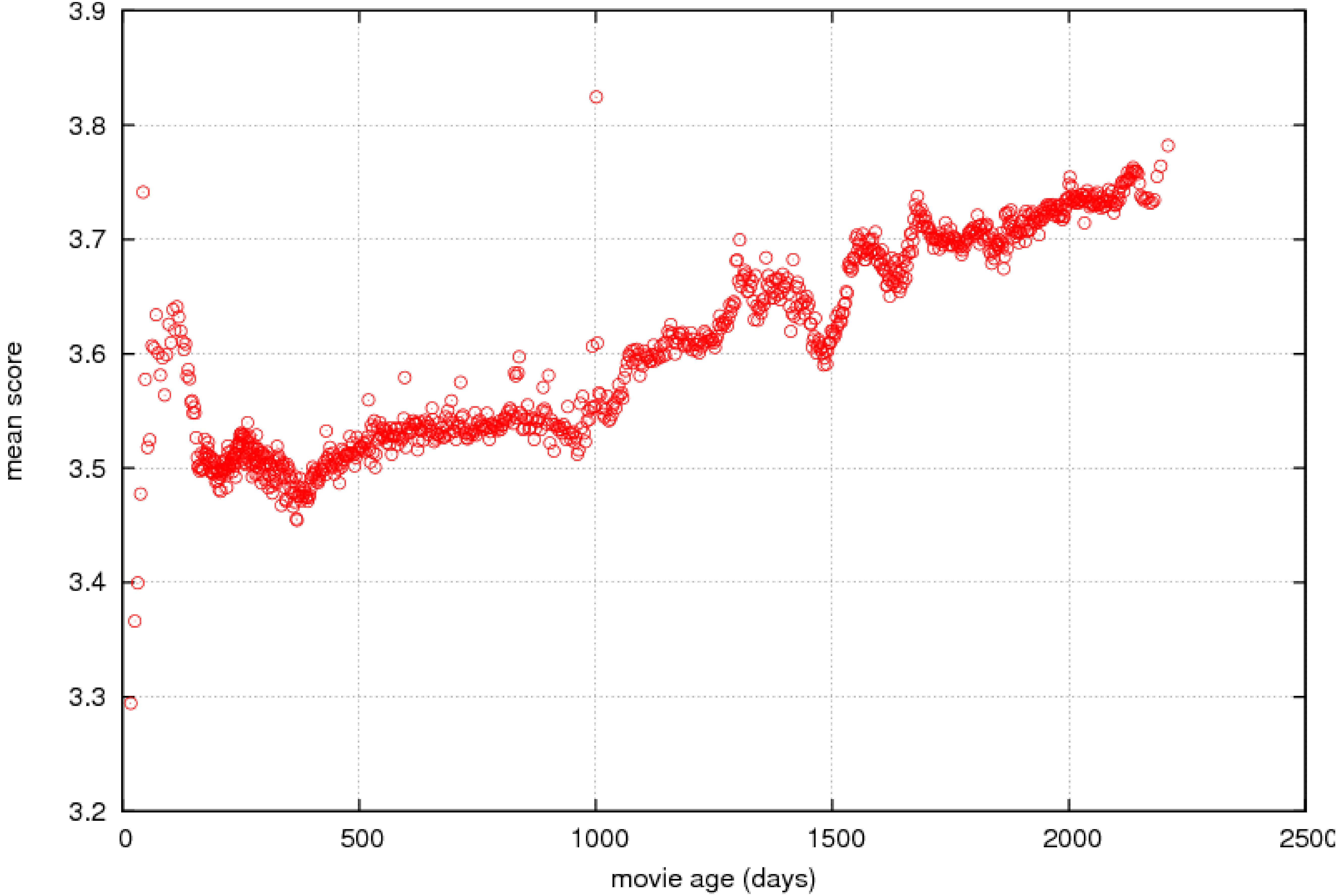


Factor models: Error vs. #parameters
0.91
0.905
0.9
0.895
0.89
0.885
0.88
0.875
RMSE
10
100
1000
10000
100000
Millions of Parameters
NMF
BiasSVD
SVD++
SVD v.2
SVD v.3
SVD v.4
40
60
90
128
180
50
100
200
50
100
200
50
100
200
500
100
200
500
50
100
200
500
1000
1500
temporal
effects

Something Happened in Early 2004…



Are movies getting better with time?



Sources of temporal change

- Items
  - Seasonal effects
    (Christmas, Valentine's day, Holiday movies)
  - Public perception of movies (Oscar etc.)
- Users
  - Changed labeling of reviews
  - Anchoring (relative to previous movie)
  - Change of rater in household
  - Selection bias for time of viewing

Modeling temporal change

- Time-dependent bias
- Time-dependent user preferences

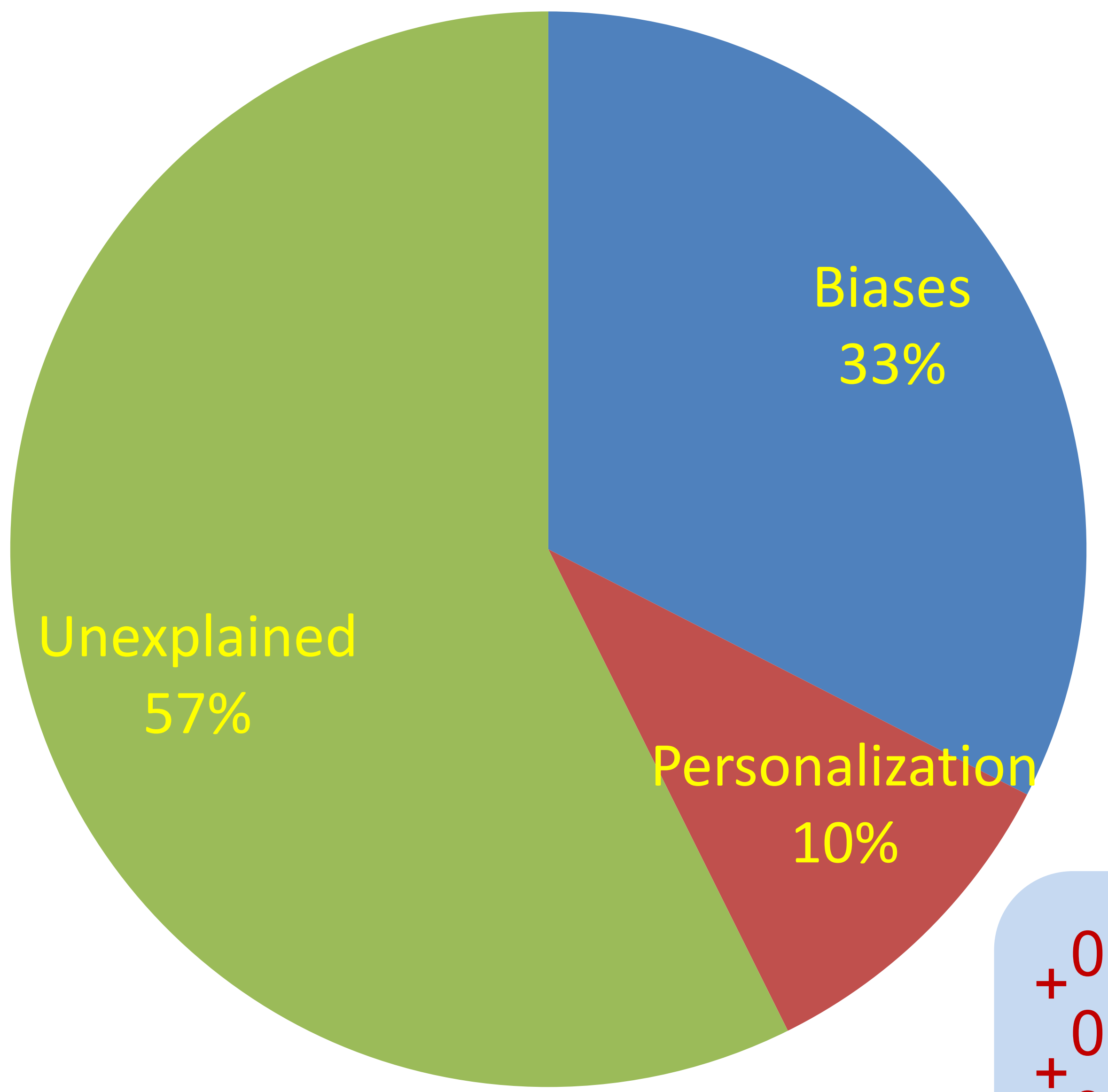
$$r_{ui}(t) = \mu + b_u(t) + b_i(t) + \langle q_i, p_u(t) \rangle$$

- Parameterize functions b and p
  - Slow changes for items
  - Fast sudden changes for users
  - Good parametrization is key

Koren et al., KDD 2009 (CF with temporal dynamics)

Bias matters

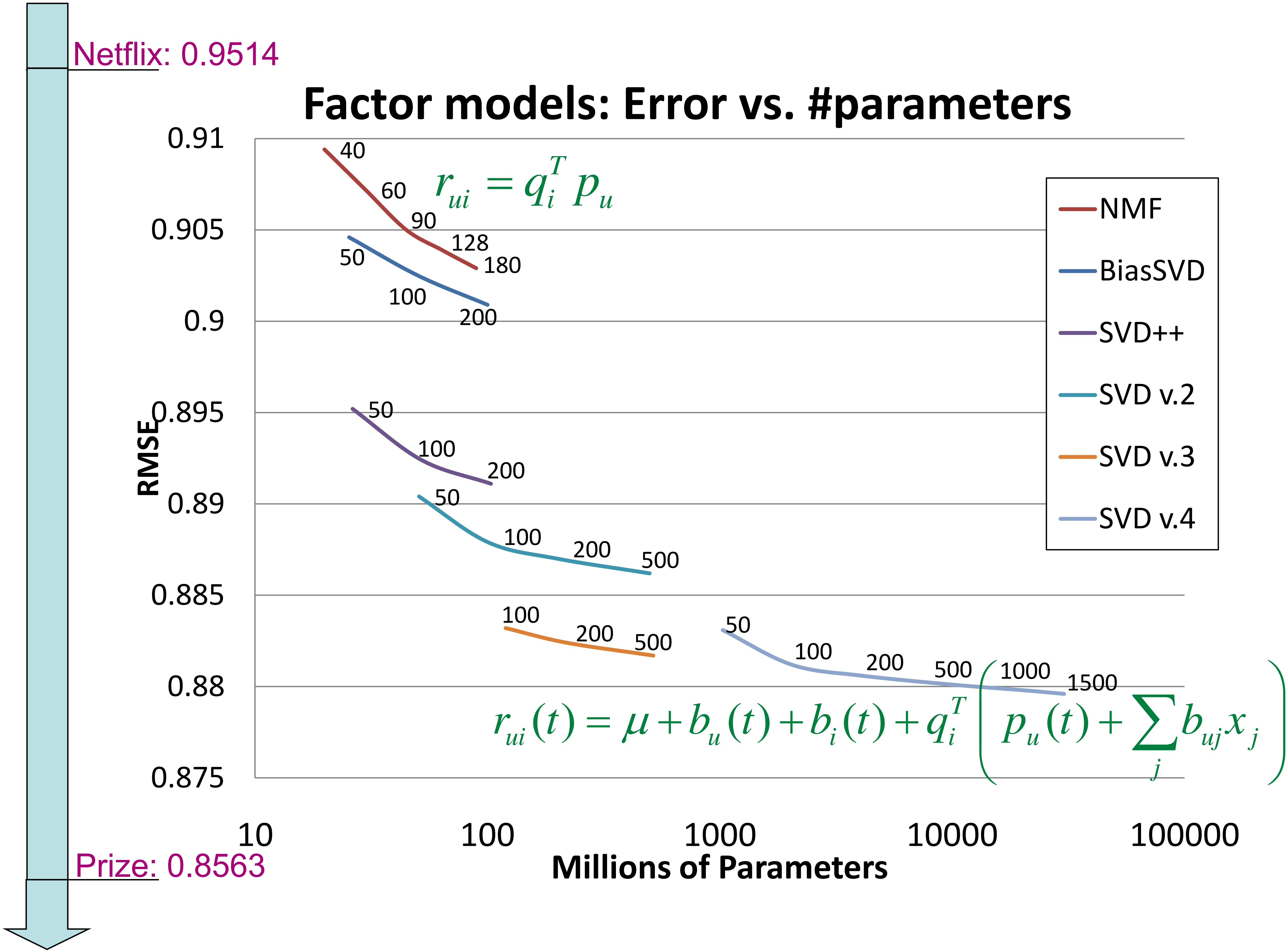
Sources of Variance in Netflix data



0.732 (unexplained)
\+ 0.415 (biases)
\+ 0.129 (personalization)

1.276 (total variance)

Factor models: Error vs. #parameters



More ideas

- Explain factorizations
- Cold start (new users)
- Different regularization for different parameter groups / different users
- Sharing of statistical strength between users
- Hierarchical matrix co-clustering / factorization (write a paper on that)

3 Session Modeling

Motivation

User interaction

- Explicit search query
  - Search engine
  - Genre selection on movie site
- Implicit search query
  - News site
  - Priority inbox
  - Comments on article
  - Viewing specific movie (see also ...)
  - Sponsored search (advertising)

Space, users' time and attention are limited.

+Alex Search Images Maps YouTube News Gmail Documents Calendar More

session modeling

Search

4 personal results. 40,000,000 other results (0.29 seconds)

- Everything
- Images
- Maps
- Videos
- News
- Shopping
- More

Mountain View, CA
Change location

Show search tools

Sessions Modeling Studio
www.**sessionsmodeling**.com/
Sessions modeling studio is a licensed agency. We offer print modeling, promotions, fashion shows, movies, tv commercials, and more. Locates in Jacksonville ...
+ Show map of 12627 San Jose Blvd # 401, Jacksonville, FL 32223

Super Models
The Campbell Soup Company hired **Sessions models** for a ...

Contact Us
Contact Us. **Sessions Modeling** Studio 12627 San Jose Blvd ...

Model Search
Contact Us. **Sessions Modeling** Studio 12627 San Jose Blvd ...

Why Sessions
For 27 years **SESSIONS MODELING** STUDIO has ...

Child Sessions
Contact Us. **Sessions Modeling** Studio. 12627 San Jose Blvd ...

Teen Sessions
Contact Us. **Sessions Modeling** Studio. 12627 San Jose Blvd ...

More results from sessionsmodeling.com »

Sessions Modeling Studio - Local Business - Jacksonville, FL ...
www.facebook.com/pages/**Sessions-Modeling**-Studio/99577445805
Sessions Modeling Studio - **Sessions Modeling** Studio is a licensed agency. ... To interact with **Sessions Modeling** Studio you need to sign up for Facebook first.
+ Show map of 12627 San Jose Blvd # 401, Jacksonville, FL 32223

Rethinking Modeling Sessions
www.agile**modeling**.com/essays/**modelingSessions**.htm
Recently reviewed, A **modeling session** is an activity where one or more people focus on the development of one or more **models**. **Modeling sessions** are an ...

Session W25: Focus Session: Modeling of Rare Events
meetings.aps.org/Meeting/MAR12/**Session**Index2/?**Session**EventID...
Mar 2, 2012 – Session W25: Focus **Session**: **Modeling** of Rare Events: Methods and Applications I. Show Abstracts. Sponsoring Units: DCOMP Chair: Weinan ...

+Alex Search Images Maps YouTube News Gmail Documents Calendar More

session modeling

Search

4 personal results. 40,000,000 other results (0.29 seconds)

session? models?

Everything
Images
Maps
Videos
News
Shopping
More

Mountain View, CA
Change location

Show search tools

Sessions Modeling Studio
www.**sessionsmodeling**.com/
Sessions modeling studio is a licensed agency. We offer print modeling, promotions, fashion shows, movies, tv commercials, and more. Locates in Jacksonville ...
+ Show map of 12627 San Jose Blvd # 401, Jacksonville, FL 32223

Super Models
The Campbell Soup Company hired **Sessions models** for a ...

Contact Us
Contact Us. **Sessions Modeling** Studio 12627 San Jose Blvd ...

Model Search
Contact Us. **Sessions Modeling** Studio 12627 San Jose Blvd ...

Why Sessions
For 27 years **SESSIONS MODELING** STUDIO has ...

Child Sessions
Contact Us. **Sessions Modeling** Studio. 12627 San Jose Blvd ...

Teen Sessions
Contact Us. **Sessions Modeling** Studio. 12627 San Jose Blvd ...

More results from sessionsmodeling.com »

Sessions Modeling Studio - Local Business - Jacksonville, FL ...
www.facebook.com/pages/**Sessions-Modeling**-Studio/99577445805
Sessions Modeling Studio - **Sessions Modeling** Studio is a licensed agency. ... To interact with **Sessions Modeling** Studio you need to sign up for Facebook first.
+ Show map of 12627 San Jose Blvd # 401, Jacksonville, FL 32223

Rethinking **Modeling Sessions**
www.agile**modeling**.com/essays/**modelingSessions**.htm
Recently reviewed, A **modeling session** is an activity where one or more people focus on the development of one or more **models**. **Modeling sessions** are an ...

Session W25: Focus **Session**: **Modeling** of Rare Events
meetings.aps.org/Meeting/MAR12/**Session**Index2/?**Session**EventID...
Mar 2, 2012 – Session W25: Focus **Session**: **Modeling** of Rare Events: Methods and Applications I. Show Abstracts. Sponsoring Units: DCOMP Chair: Weinan ...

[PDF] Sagan Workshop Hands-on **Sessions (Modeling)** At present ...
nexsci.caltech.edu/workshop/2011/Tues_HandsOn.pdf
File Format: PDF/Adobe Acrobat - Quick View
Sagan Workshop Hands-on **Sessions** (**Modeling**). At present, searching for planets with microlensing requires selecting a few targets out of hundreds discovered ...

GIS and Agent-Based **Modelling**: AAG SPECIAL **SESSION** ...
gisagents.blogspot.com/.../aag-special-**session-modeling**-geographic....
Sep 3, 2009 – AAG SPECIAL **SESSION**: **Modeling** Geographic Complexity. For those interested we are organizing a special session(s) at the forthcoming ...

Technical **Session** 31: **Modeling** & Control for Renewable Energy
www.apec-conf.org/2011/conference-at-a-glance/337?task=view
Title. Author(s). Fault Impacts on Solar Power Unit Reliability. Ali Bazzi, Katherine Kim, Brian Johnson, Philip Krein, Alejandro Do... Analysis of Boundary Control ...

Plenary **Session**: **Modeling** Social Behavior with Aggregated ...
video.mit.edu/.../plenary-**session-modeling**-social-behavior-with-aggr...
Ted Morgan, CEO, Skyhook Wireless; Kipp Jones, Chief Architect, Skyhook Wireless. 10/12/2009.

[PDF] Case Based **Session Modeling** and Personalization in a Travel ...
www.inf.unibz.it/~ricci/papers/07-arslan.pdf
File Format: PDF/Adobe Acrobat - Quick View
by B Arslan - Cited by 5 - Related articles
Knowledge intensive **session modeling** and mixed initiative recommendation are introduced in the CBR framework. The advantages of this approach, with ...

Sessions modeling studio - YouTube
www.youtube.com/watch?v=eD1KJHwLxVY
Mar 30, 2011 – Trainer Davey at Fitness America Weekend 2010 Las Vegasby TrainerDavey177 views; Studio **Modeling session** swimsuit **model** & ...

Goooooooooogle ›

1 2 3 4 5 6 7 8 9 10 Next

Advanced search   Search Help   Give us feedback

Google Home   Advertising Programs   Business Solutions   Privacy
About Google

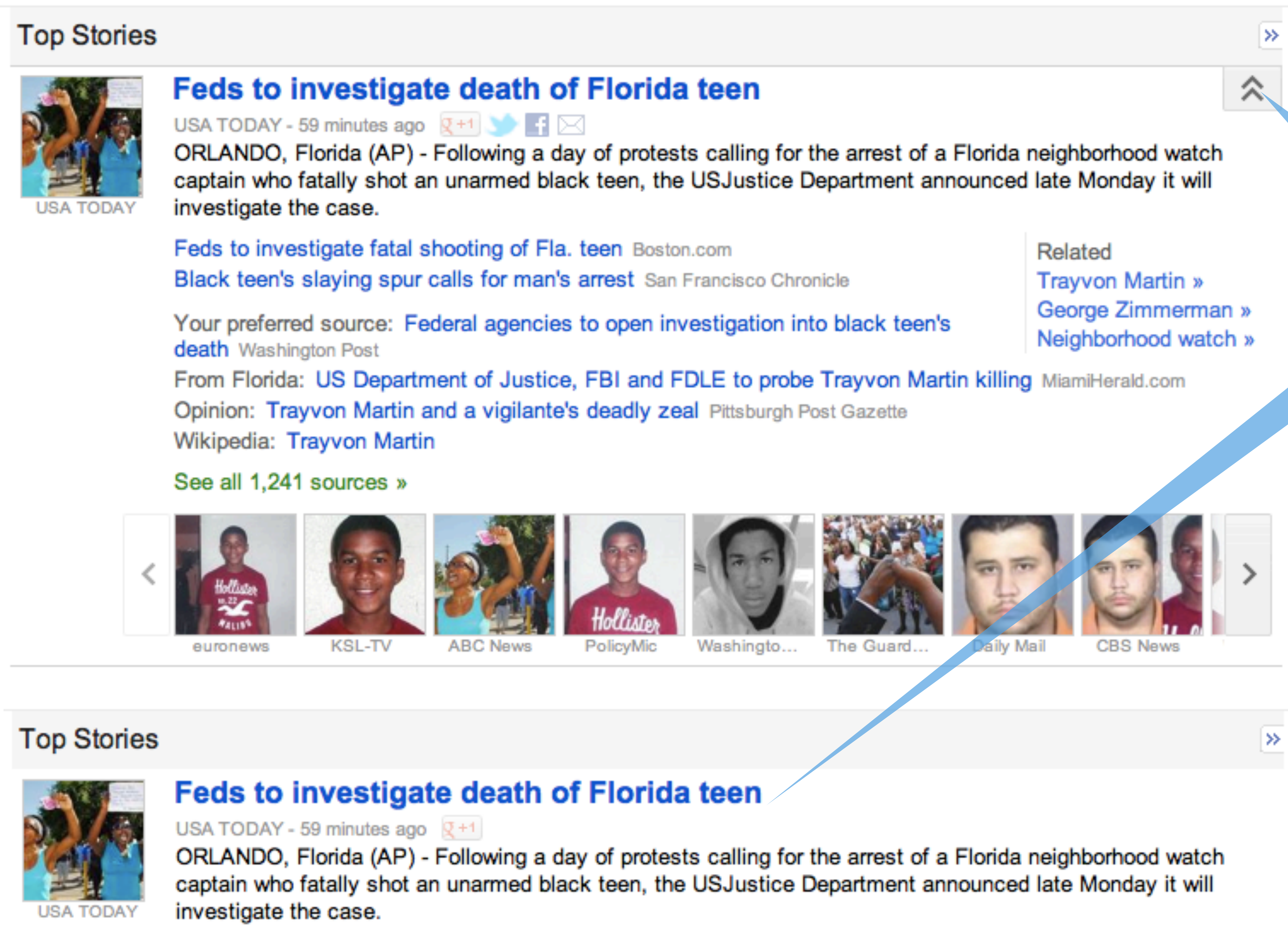
Did the user SCROLL DOWN?

Bad ideas ...

- Show items based on relevance

  - Yes, this user likes Die Hard.
  - But he likes other movies, too
- Show items only for majority of users
  ‘apple’ vs. ‘Apple’

User response



collapse

implicit
user interest

log it!

Web Images Videos Shopping News Maps More | MSN Hotmail

bing

bieber

Web

Web News Music Images Videos More▾

RELATED SEARCHES

- Justin Bieber
- Justin Bieber **Games**
- Justin Bieber **Cuts His Hair**
- Justin Bieber **New Haircut**
- Justin Bieber **Nail Polish**
- Justin Bieber **Kissing**
- Justin Bieber **Hair**
- Justin Bieber **Selena Gomez Kiss**

SEARCH HISTORY

- user
- user response
- See all
- Clear all · Turn off

NARROW BY DATE

- **All results**
- Past 24 hours
- Past week
- Past month

ALL RESULTS

1-10 of 209,000,000 results · Advanced

Justin Bieber - Wikipedia, the free encyclopedia
Life and career · Image · Discography · Tours
Justin Drew **Bieber** is a Canadian pop/ R&B singer, songwriter and actor. **Bieber** was discovered in 2008 by Scooter Braun, who came across **Bieber**'s videos on YouTube and ...
en.wikipedia.org/wiki/Justin_**Bieber**

Justin Bieber
Official site of Justin **Bieber**. Includes news and blog, webshop and online video.
www.justin**bieber**music.com

bieber - Bing News

Justin **Bieber** gets beaten bloody in the boxing ring for Complex, talks about his 'feminine qualities'
Justin **Bieber** plans on being very open about his love for girlfriend Selena Gomez - but he won't let her get in the way of his music. In an interview with...
New York Daily News · 11 hours ago
Justin **Bieber** gets bloody for 'Complex' magazine AZCentral.com
Justin **Bieber** Takes A Few Punches For Complex MTV

Bieber Tours - Home
Fuel Surcharge - diesel fuel prices continue to climb, no increase for March ticket prices: Read More...
www.**bieber**tourways.com

Bieber, California - Wikipedia, the free encyclopedia
History · Demographics · Politics
Bieber (formerly, Chalk Ford) is a census-designated place (CDP) in Lassen County, California. It is located on the Pit River 55 miles (89 km) north-northwest of ...
en.wikipedia.org/wiki/**Bieber**,_California

Bieber by Adam - Bing Music

Bieber 3:05
Album: Bieber - Single

hover on link

PAGE SECTIONS

1. History
The settlement sprang up at the F
1877. [3] The first post office at
2. Demographics
The 2010 United States Census [
More on this page had a population of 31
3. Politics
In the state legislature Bieber is l
1st Senate District, represented b
4. References

Search within wikipedia.org

Search

Cached page

www.bing.com/search?q=bieber&go=&qs=n&form=QBRE&pq=bieber&sc=8-5&sp=-1&sk=#

Response is conditioned on available options

- User search for ‘chocolate’

user picks this

- What the user really would have wanted
  - User can only pick from available items
  - Preferences are often relative

Models

Independent click model

- Each object has click probability
- Object is viewed independently
  - Used in computational advertising (with some position correction)
  - Horribly wrong assumption
  - OK if probability is very small (OK in ads)

$$p(x|s) = \prod_{i=1}^{n} \frac{1}{1 + e^{-x_i s_i}}$$

Logistic click model

no click

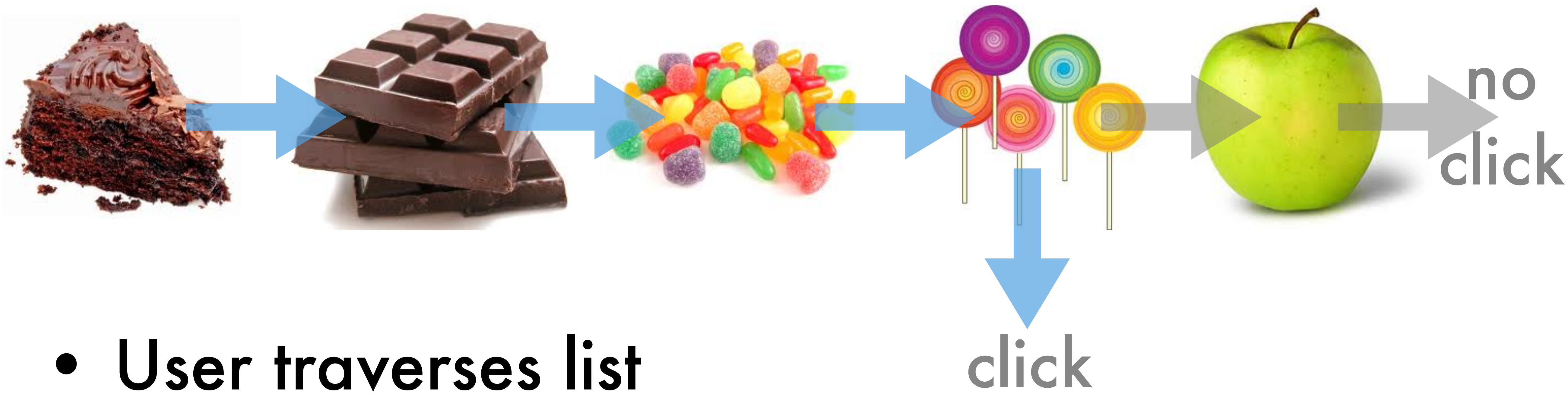
- User picks at most one object
- Exponential family model for click

$$p(x|s) = \frac{e^{s_x}}{e^{s_0} + \sum_{x'} e^{s_{x'}}} = \exp\left(s_x - g(s)\right)$$

no click

- Ignores order of objects
- Assumes that the user looks at all before taking action

Sequential click model

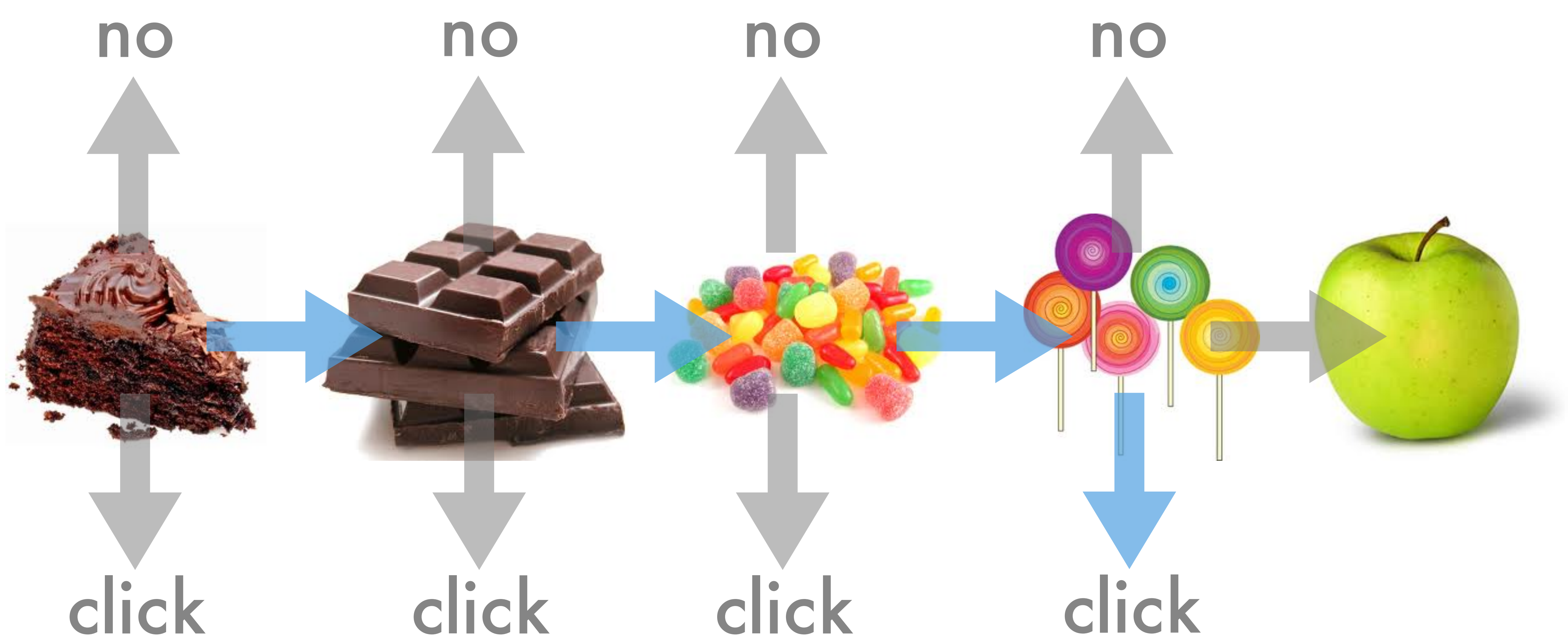


- User traverses list
- At each position some probability of clicking
- When user reaches end of the list he aborts

$$p(x = j|s) = \left[\prod_{i=1}^{j-1} \frac{1}{1 + e^{s_i}}\right] \frac{1}{1 + e^{-s_j}}$$

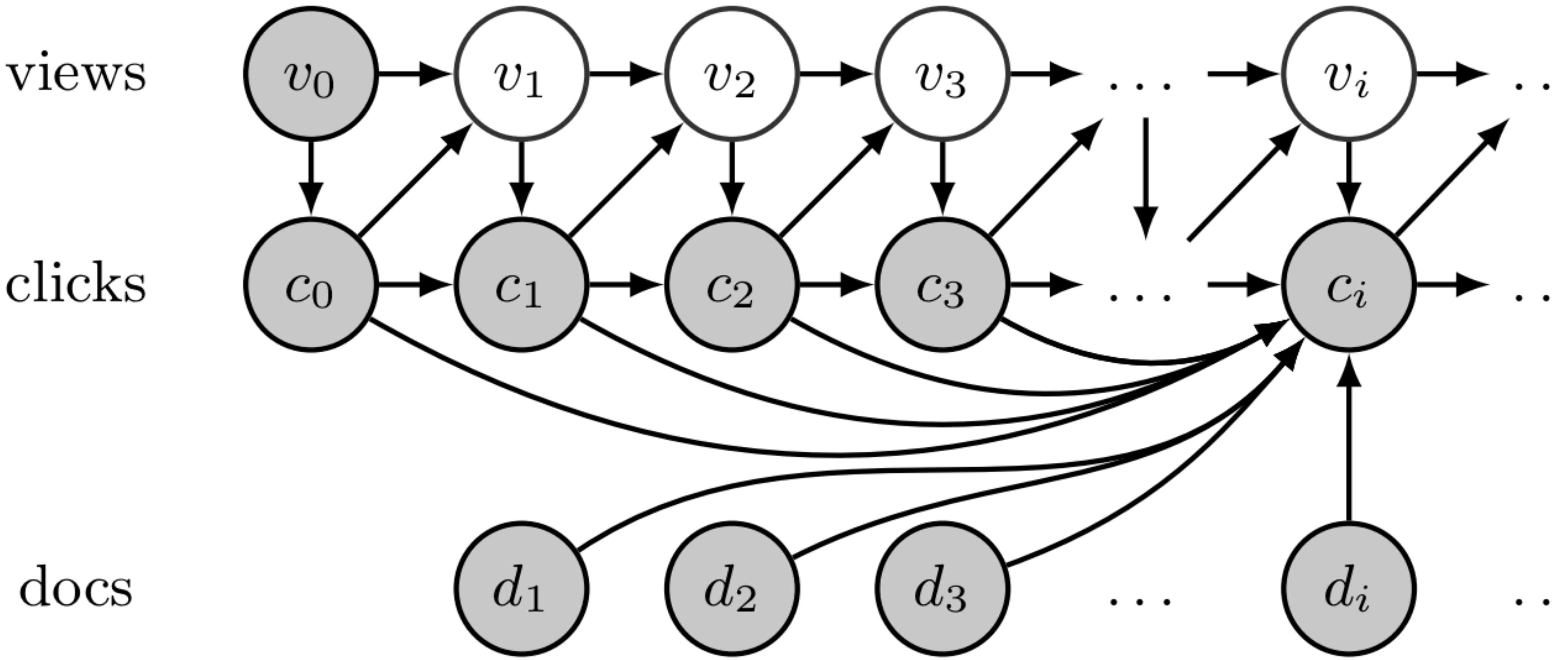
- This assumes that a patient user viewed all items

Skip click model



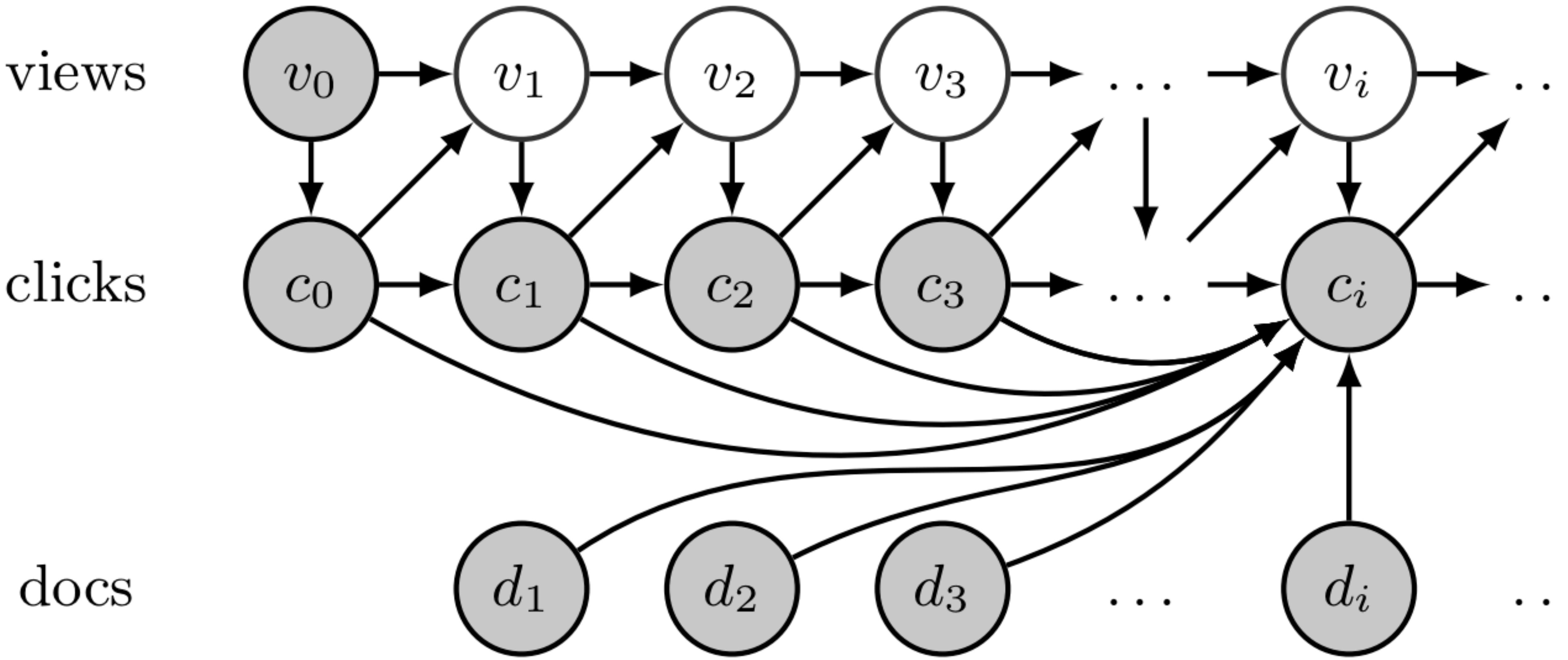
- User traverses list
- At each position some probability of clicking
- At each position the user may abandon the process
- This assumes that user traverses list sequentially

Context skip click model



- User traverses list
- At each position some probability of clicking which depends on previous content
- At each position the user may abandon the process
- User may click more than once

Context skip click model



$$p(v, c|d) = \prod_{i=1}^{n} \left[p(v_i|v_{i-1}, c_{i-1}) p(c_i|v_i, c^{i-1}, d^i) \right]$$

Context skip click model

- **Viewing probability**

user is gone

$$p(v_i = 1 | v_{i-1} = 0) = 0$$

$$p(v_i = 1 | v_{i-1} = 1, c_{i-1} = 0) = \frac{1}{1 + e^{-\alpha_i}}$$

user returns

$$p(v_i = 1 | v_{i-1} = 1, c_{i-1} = 1) = \frac{1}{1 + e^{-\beta_i}}$$

- **Click probability (only if viewed)**

prior context

$$p(c_i = 1 | v_i = 1, c^{i-1}, d^i) = \frac{1}{1 + e^{-f(|c^{i-1}|, d_i, d^{i-1})}}$$

$$p(v, c | d) = \prod_{i=1}^{n} \left[p(v_i | v_{i-1}, c_{i-1}) p(c_i | v_i, c^{i-1}, d^i) \right]$$

Incremental gains score

$$
\begin{aligned}
& f(|c^{i-1}|, d_i, d^{i-1}) \\
:= & \rho(S, d^i | a, b) - \rho(S, d^{i-1} | a, b) + \gamma_{|c^{i-1}|} + \delta_i \\
:= & \sum_{s \in S} \sum_j [s]_j \left(a_j \sum_{d \in d^i} [d]_j + b_j \left(\rho_j(d^i) - \rho_j(d^{i-1}) \right) \right) \\
& + \gamma_{|c^{i-1}|} + \delta_i
\end{aligned}
$$

- Submodular gain per additional document
- Relevance score per document
- Coverage over different aspects
- Position dependent score
- Score dependent on number of previous clicks

Optimization

- Latent variables

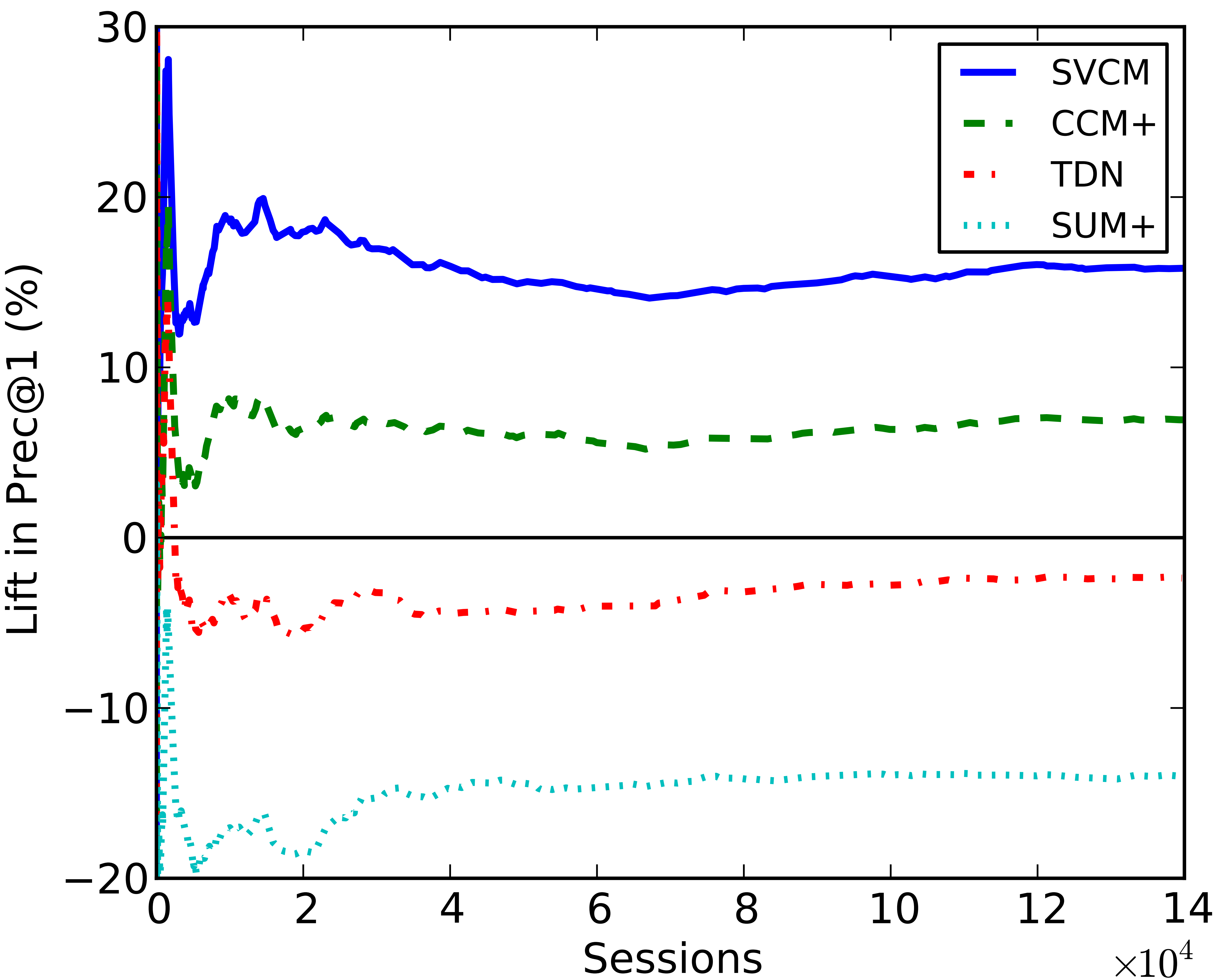
$$p(v, c|d) = \prod_{i=1}^{n} \left[p(v_i|v_{i-1}, c_{i-1}) p(c_i|v_i, c^{i-1}, d^i) \right]$$

  We don't know v whether user viewed result
- Use variational inference to integrate out v (more next week in graphical models)

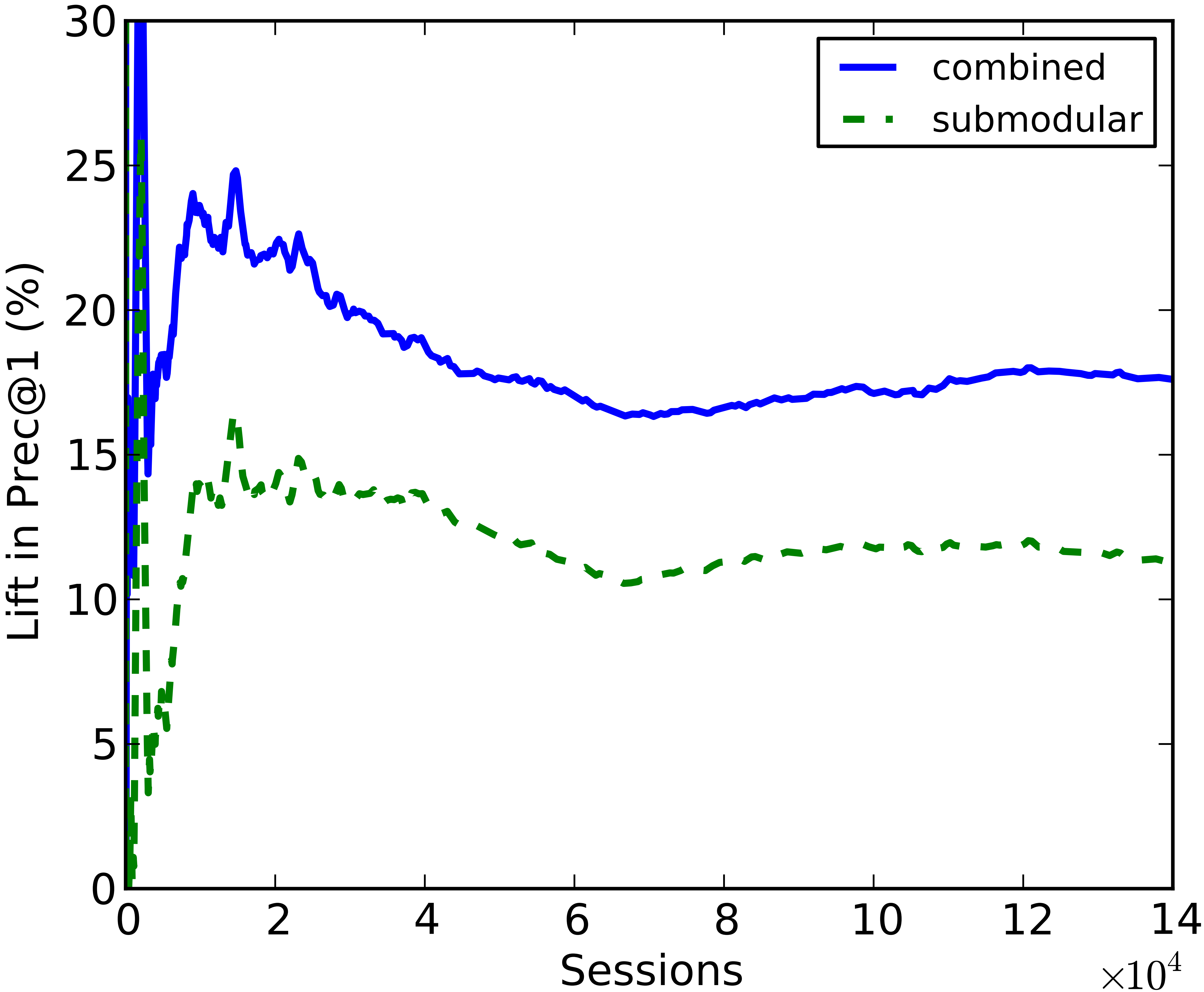
$$\begin{aligned} -\log p(c) &\leq -\log p(c) + D(q(v)\|p(v|c)) \\ &= \mathbf{E}_{v\sim q(v)}\left[-\log p(c) + \log q(v) - \log p(v|c)\right] \\ &= \mathbf{E}_{v\sim q(v)}\left[-\log p(c, v)\right] - H(q(v)). \end{aligned}$$

Optimization

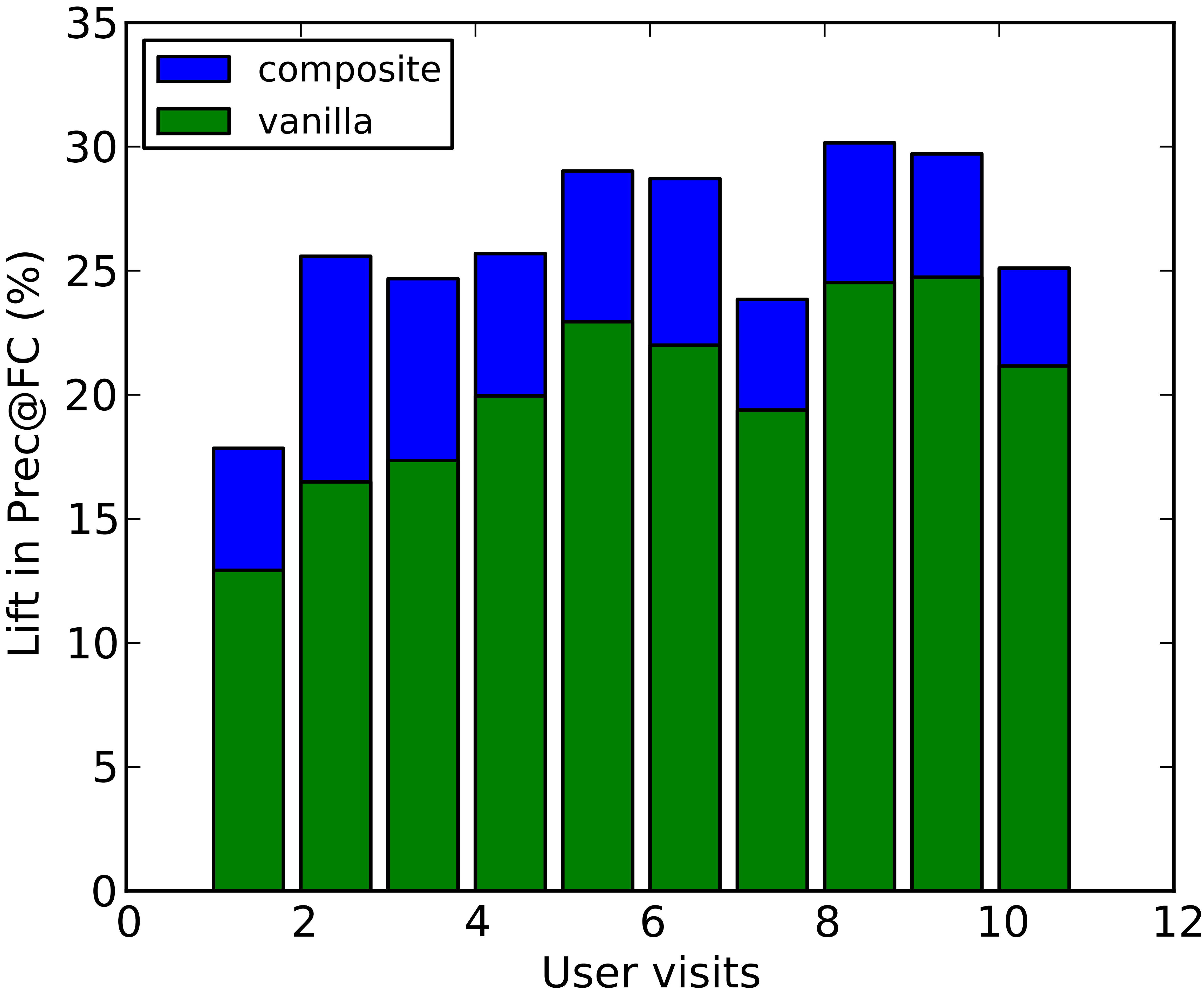
- Compute latent viewing probability given clicks
  - Easy since we only have one transition from views to no views (no DP needed)
  - Expected log-likelihood under viewing model
  - Convex expected log-likelihood
- Stochastic gradient descent
- Parametrization uses personalization, too (user, position, viewport, browser)



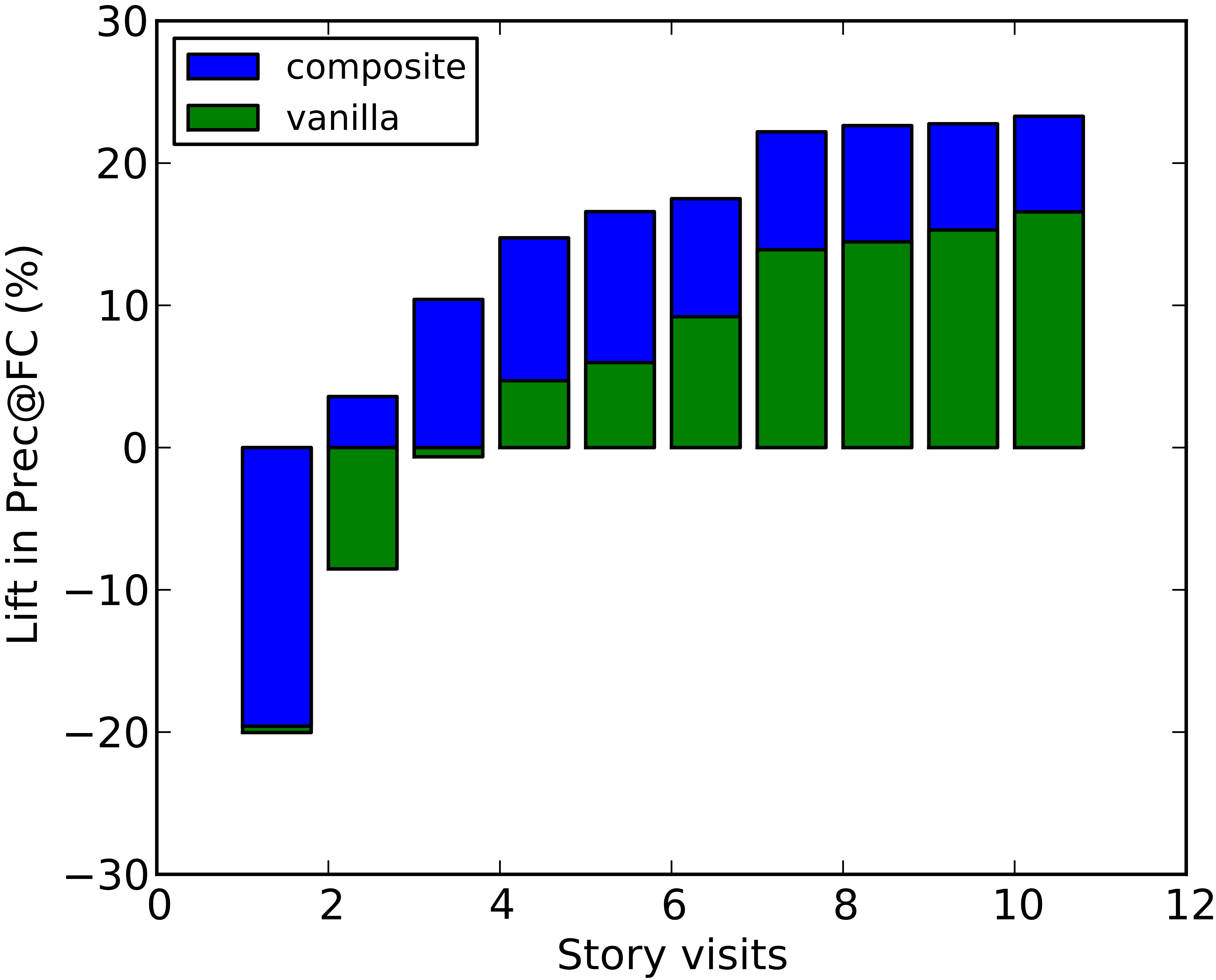
SVCM
CCM+
TDN
SUM+
30
20
10
0
−10
−20
Lift in Prec@1 (%)
0
2
4
6
8
10
12
14
Sessions
$\times 10^4$



combined
submodular
Lift in Prec@1 (%)
Sessions
$\times 10^4$
0
5
10
15
20
25
30
0
2
4
6
8
10
12
14



composite
vanilla
Lift in Prec@FC (%)
User visits
0
5
10
15
20
25
30
35
0
2
4
6
8
10
12

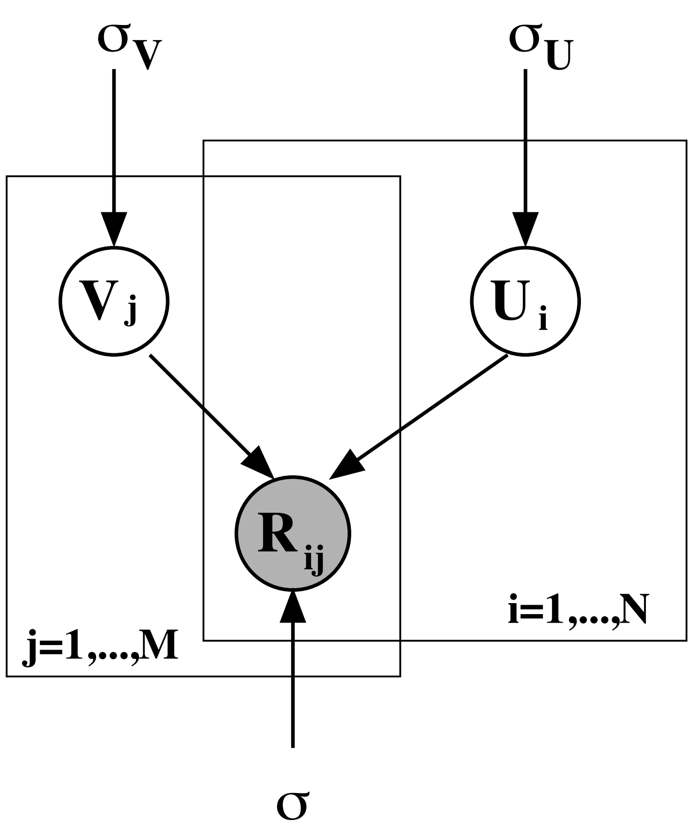

composite
vanilla
Lift in Prec@FC (%)
Story visits
30
20
10
0
−10
−20
−30
0
2
4
6
8
10
12

4 Feature Representation

Bayesian Probabilistic Matrix Factorization

Statistical Model

- Aldous-Hoover factorization
  - normal distribution for user and item attributes
  - rating given by inner product
- Ratings



$$p(R_{ij}|U_i, V_j, \sigma^2) = \mathcal{N}(R_{ij}|U_i^T V_j, \sigma^2)$$

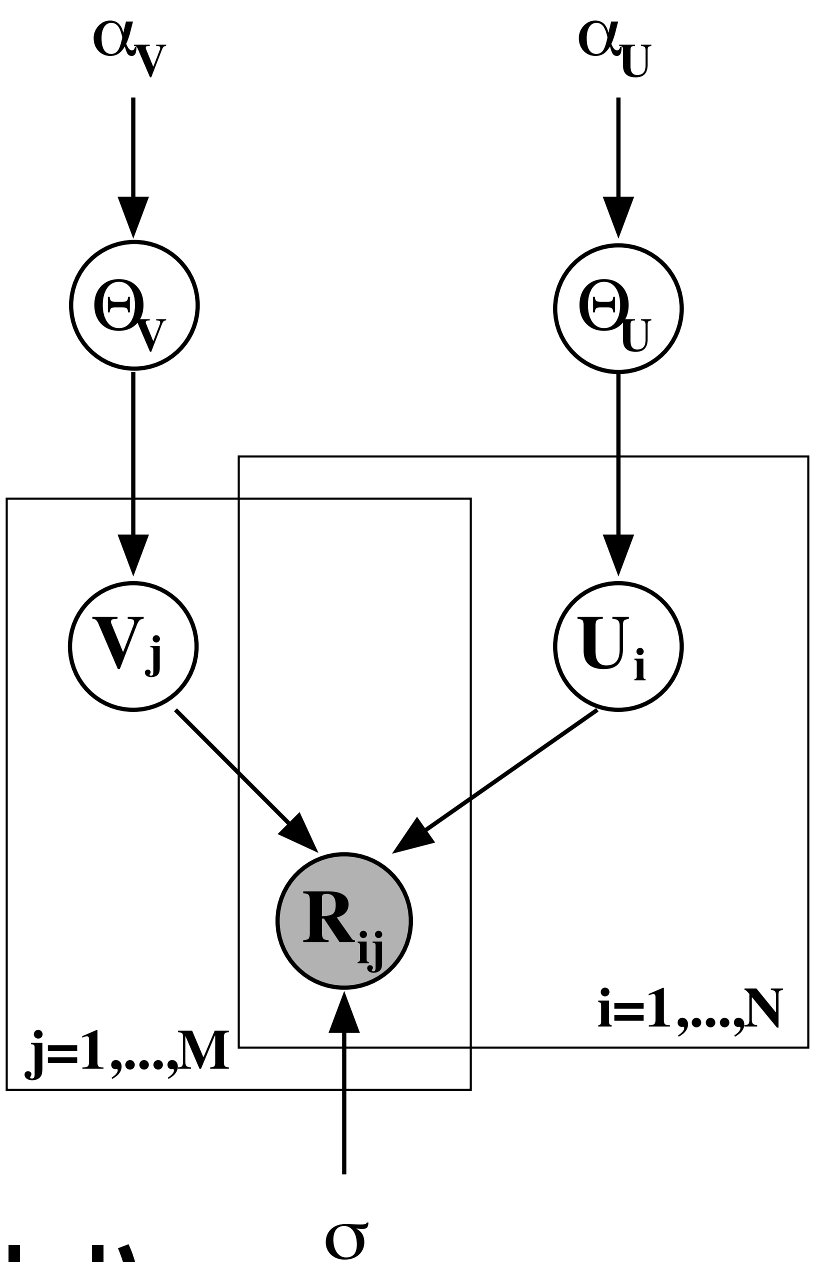
- Latent factors

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

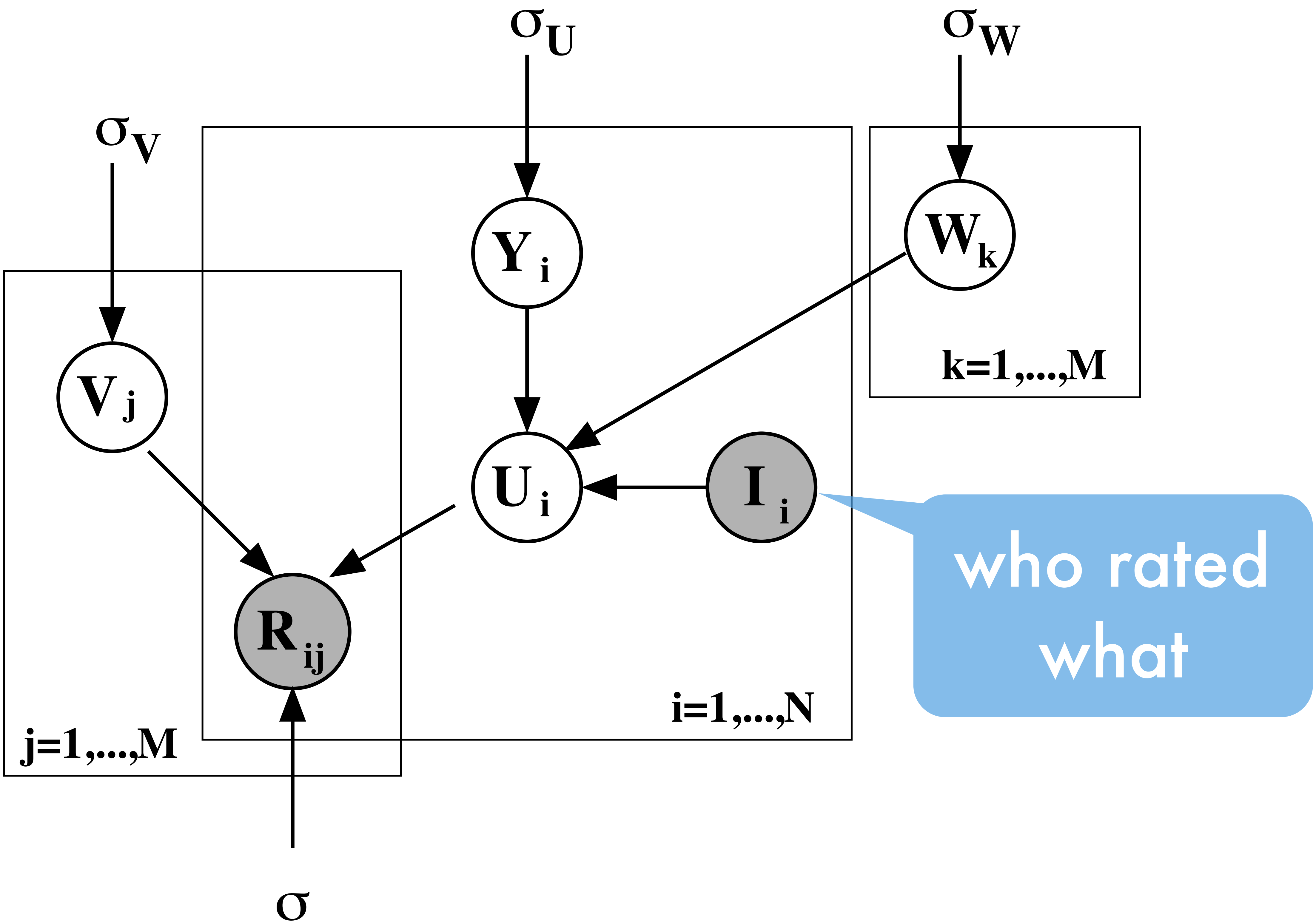
Salakhudtinov & Mnih, ICML 2008 BPMF

Details

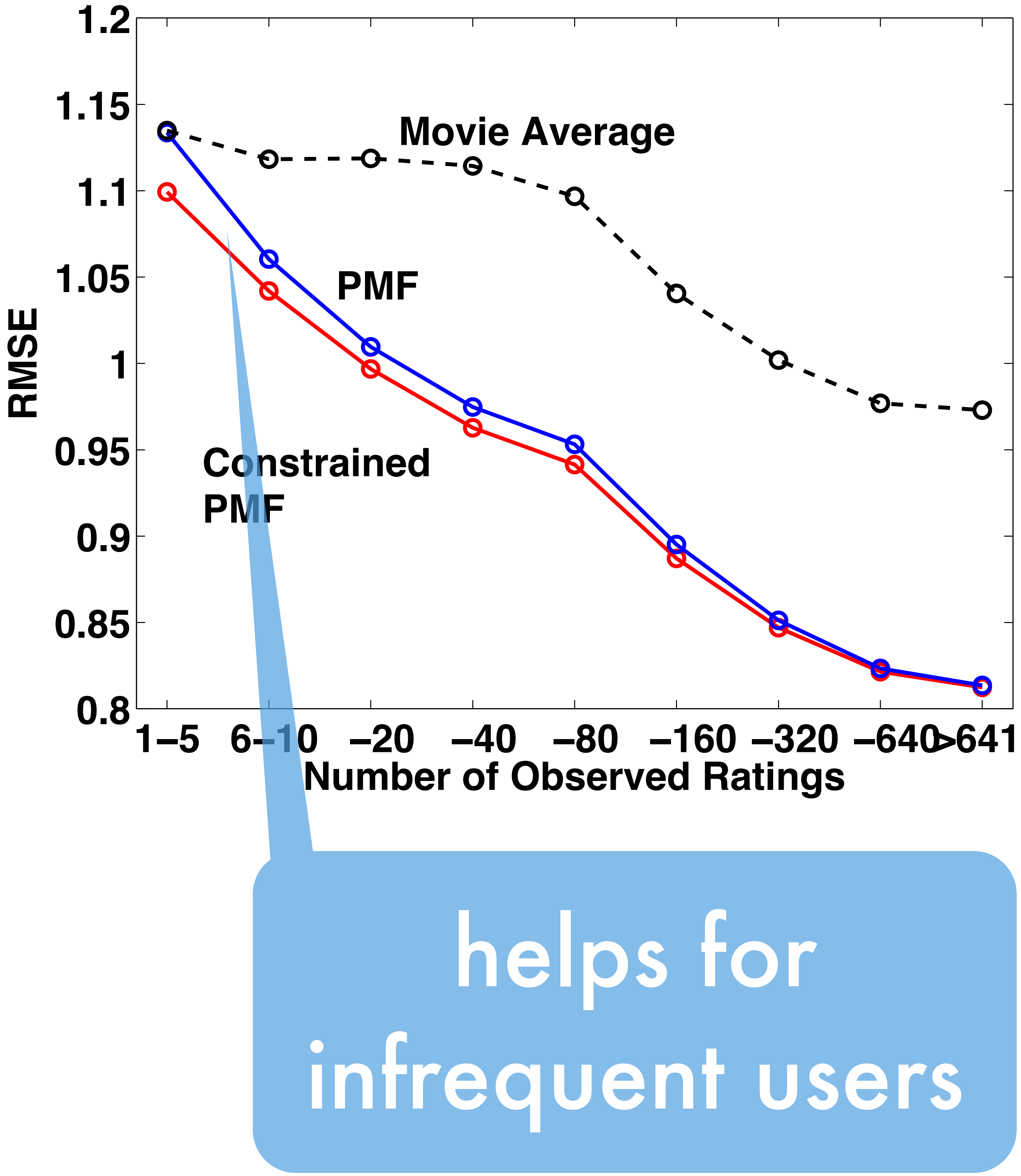
- Priors on all factors
- Wishart prior is conjugate to Gaussian, hence use it
- Allows us to adapt the variance automatically
- Inference (Gibbs sampler)
  - Sample user factors (parallel)
  - Sample movie factors (parallel)
  - Sample hyperparameters (parallel)

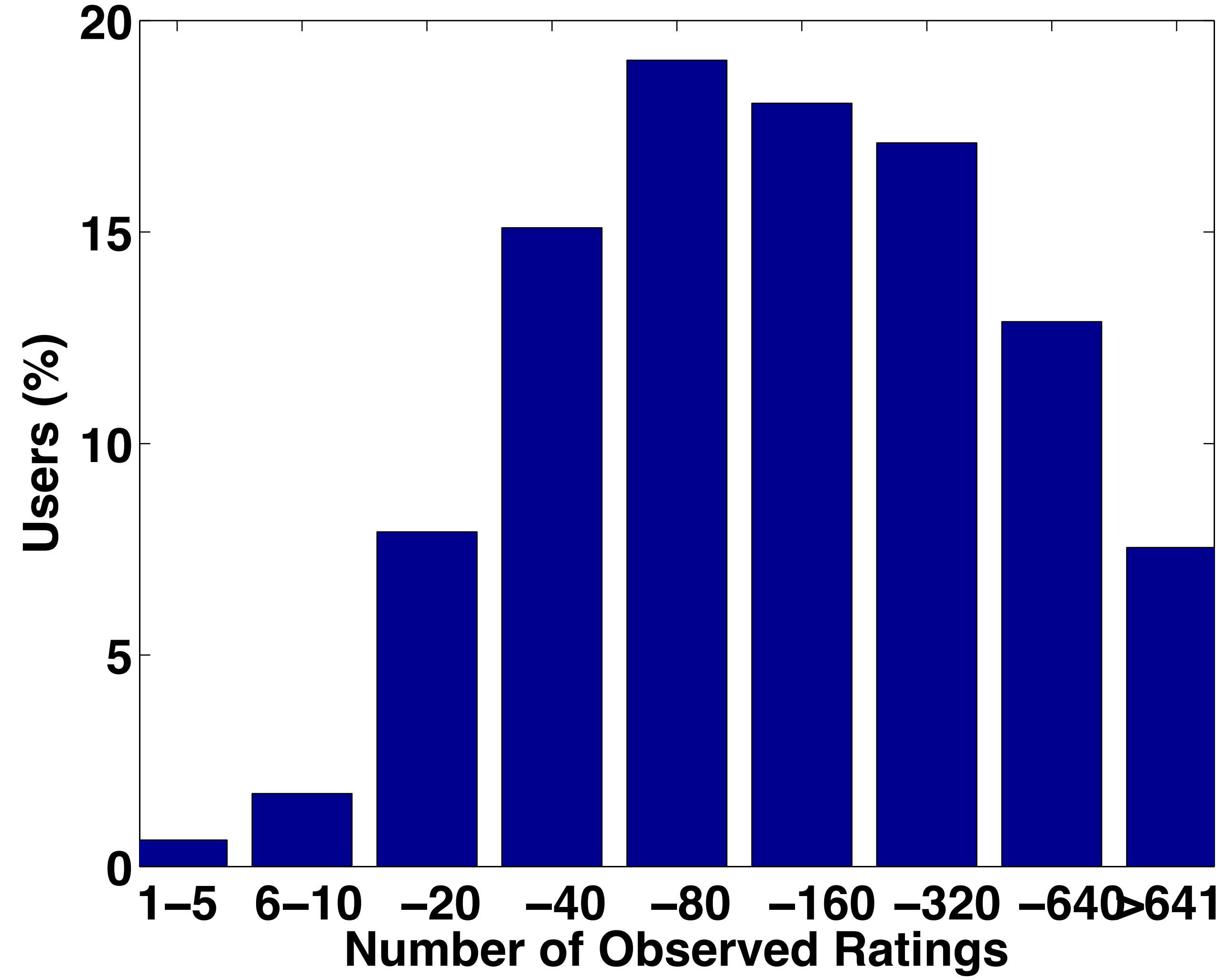


Making it fancier (constrained BPMF)



Results (Mnih & Salakthudtinov)



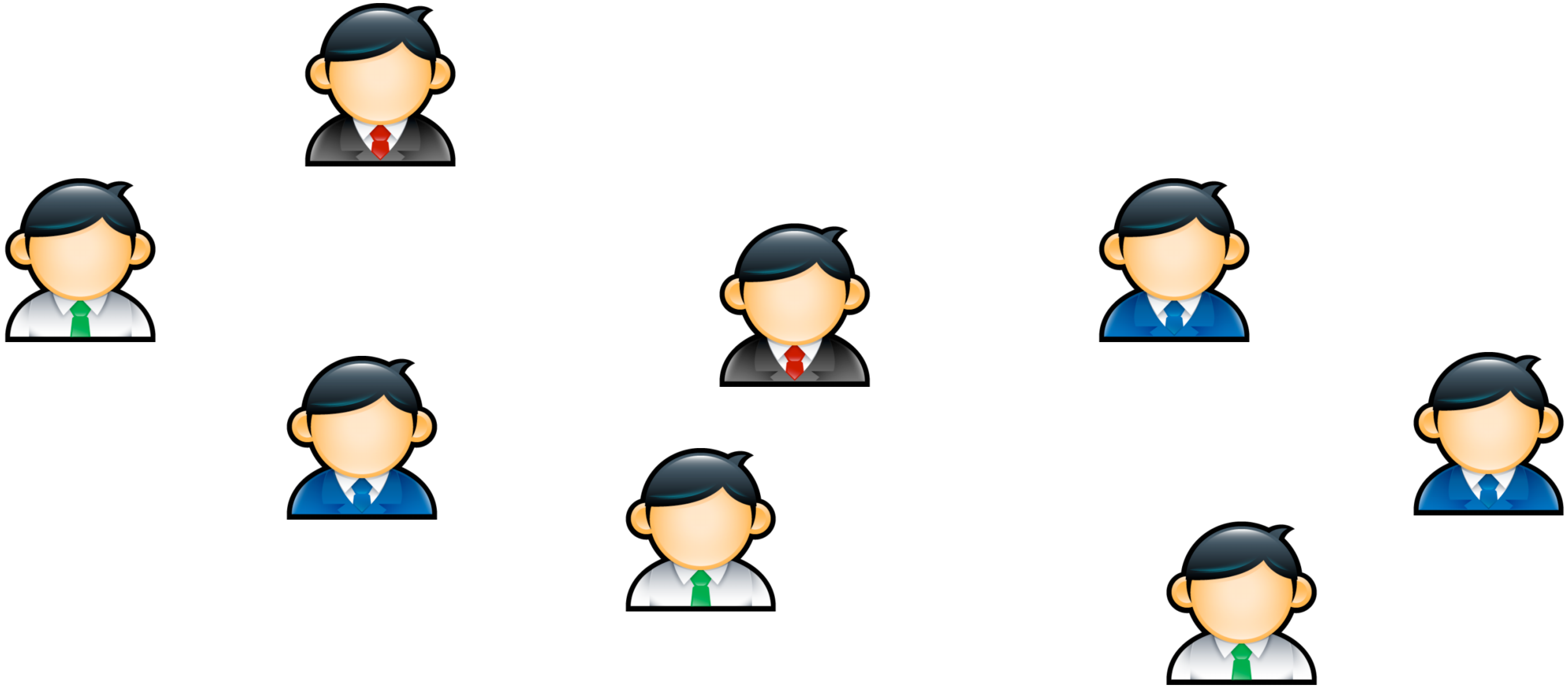


Multiple Sources

Social Network Data

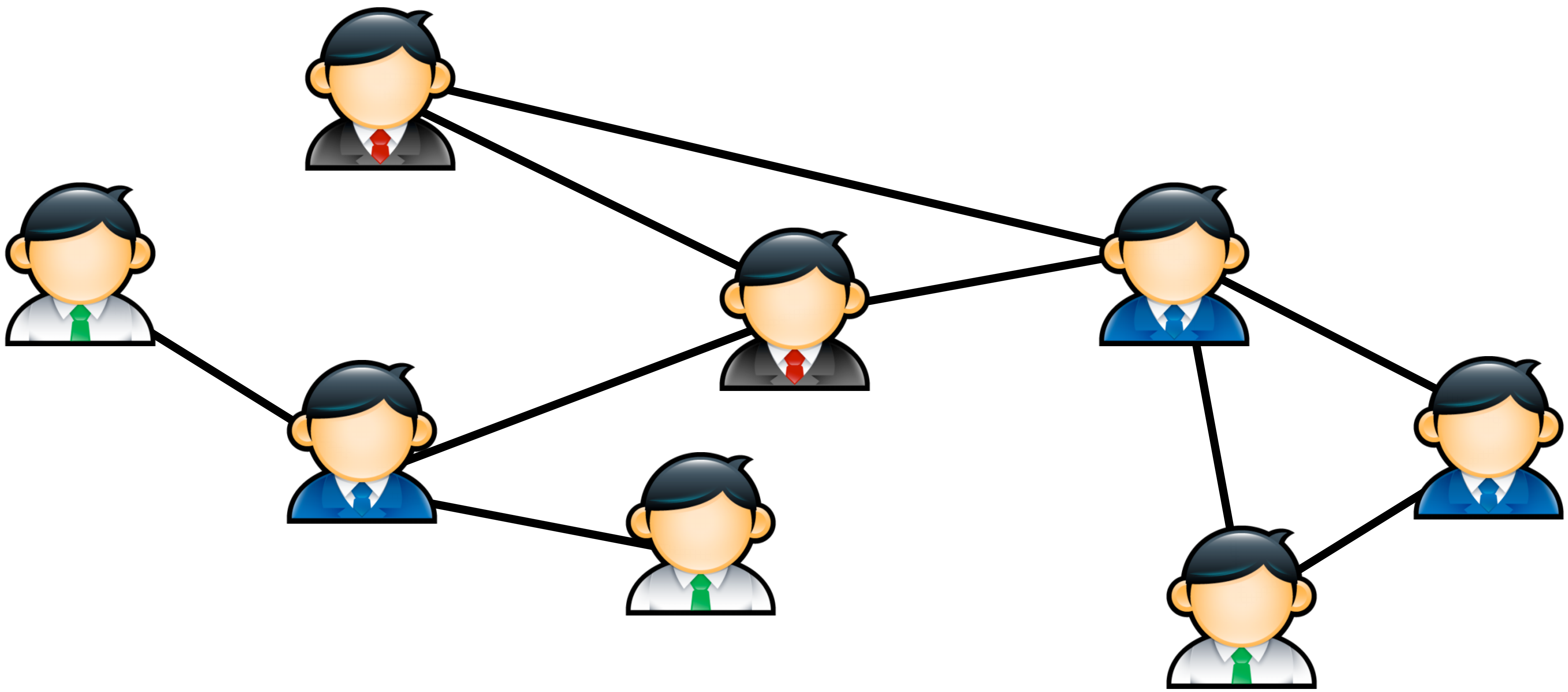
Data: users, connections, features

Goal: suggest connections



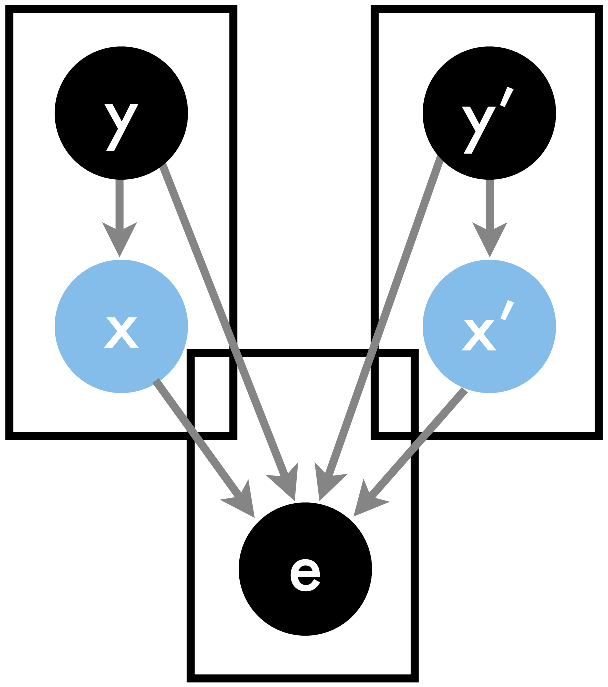
Social Network Data

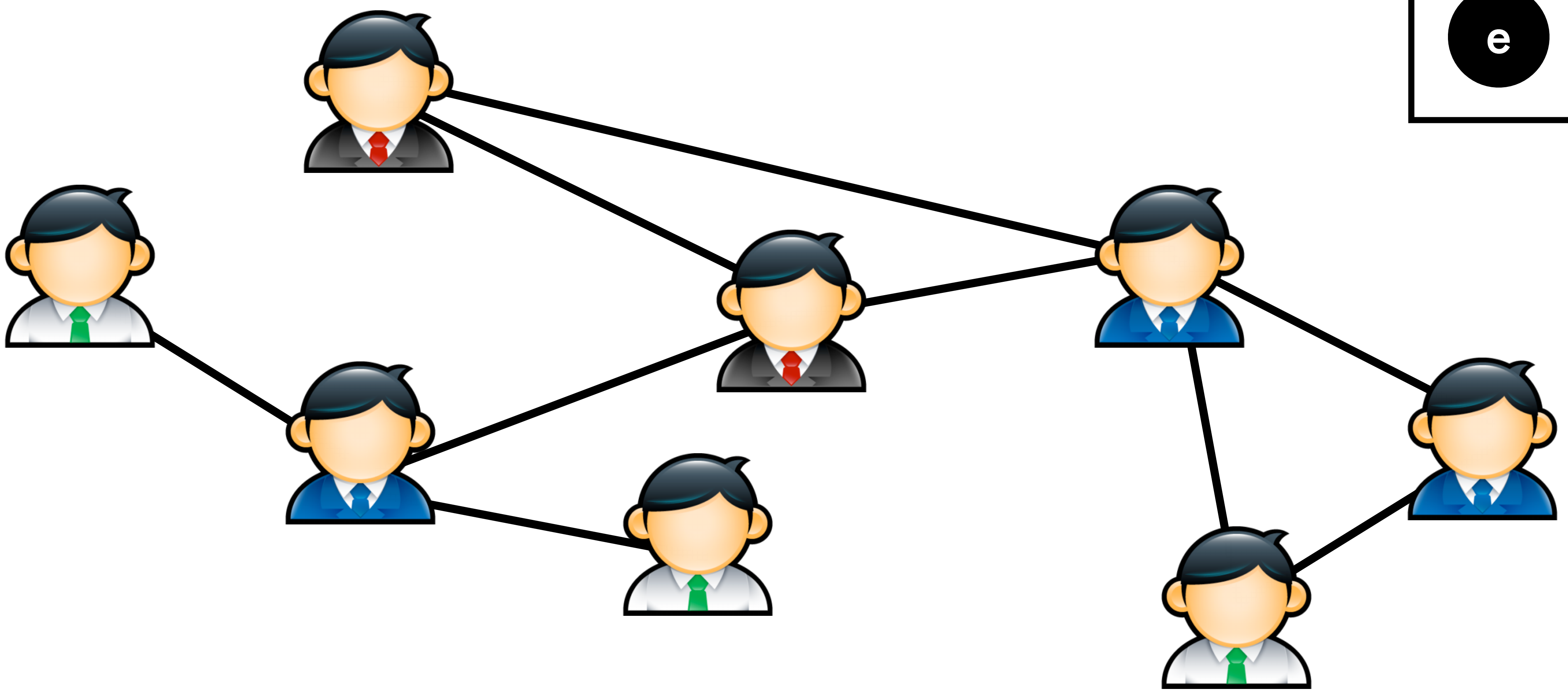
Data: users, connections, features
Goal: suggest connections



Social Network Data

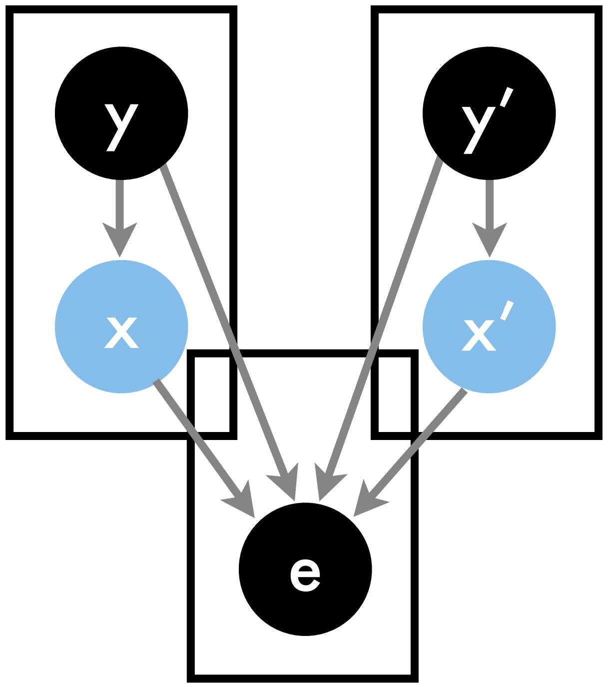
Data: users, connections, features
Goal: suggest connections





Social Network Data

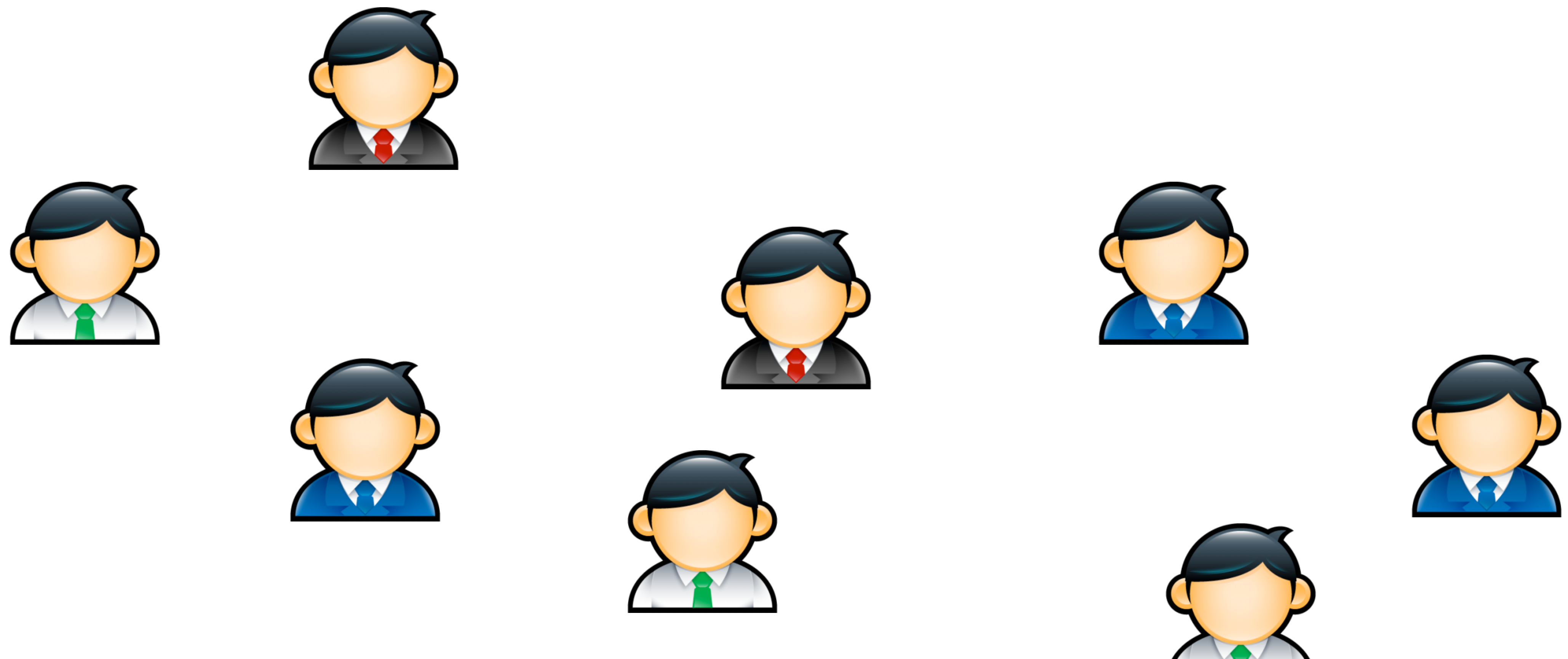
Data: users, connections, features
Goal: model/suggest connections



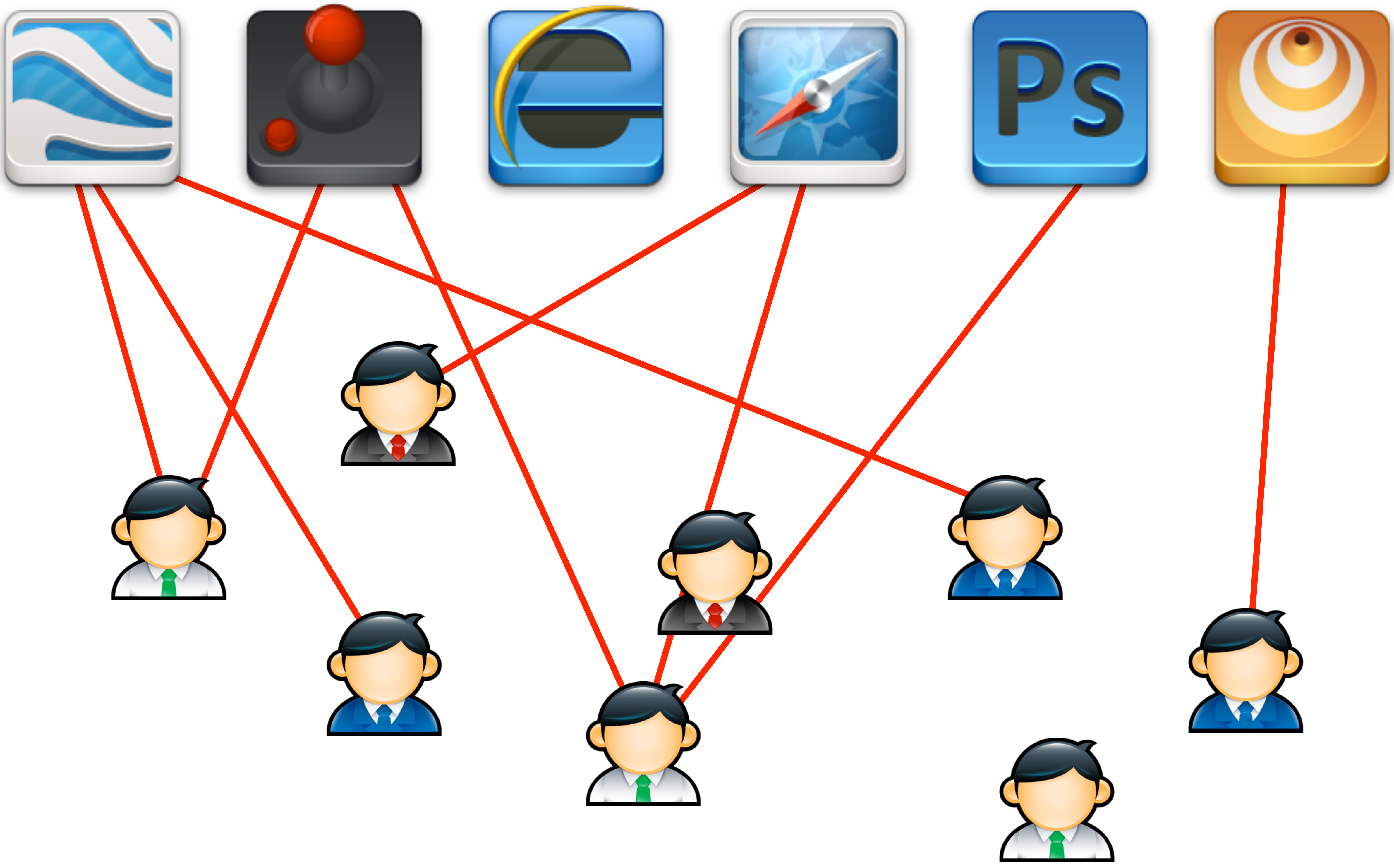
$$p(x, y, e) = \prod_{i \in \text{Users}} p(y_i) p(x_i | y_i) \prod_{i,j \in \text{Users}} p(e_{ij} | x_i, y_i, x_j, y_j)$$

Direct application of the Aldous-Hoover theorem.
Edges are conditionally independent.

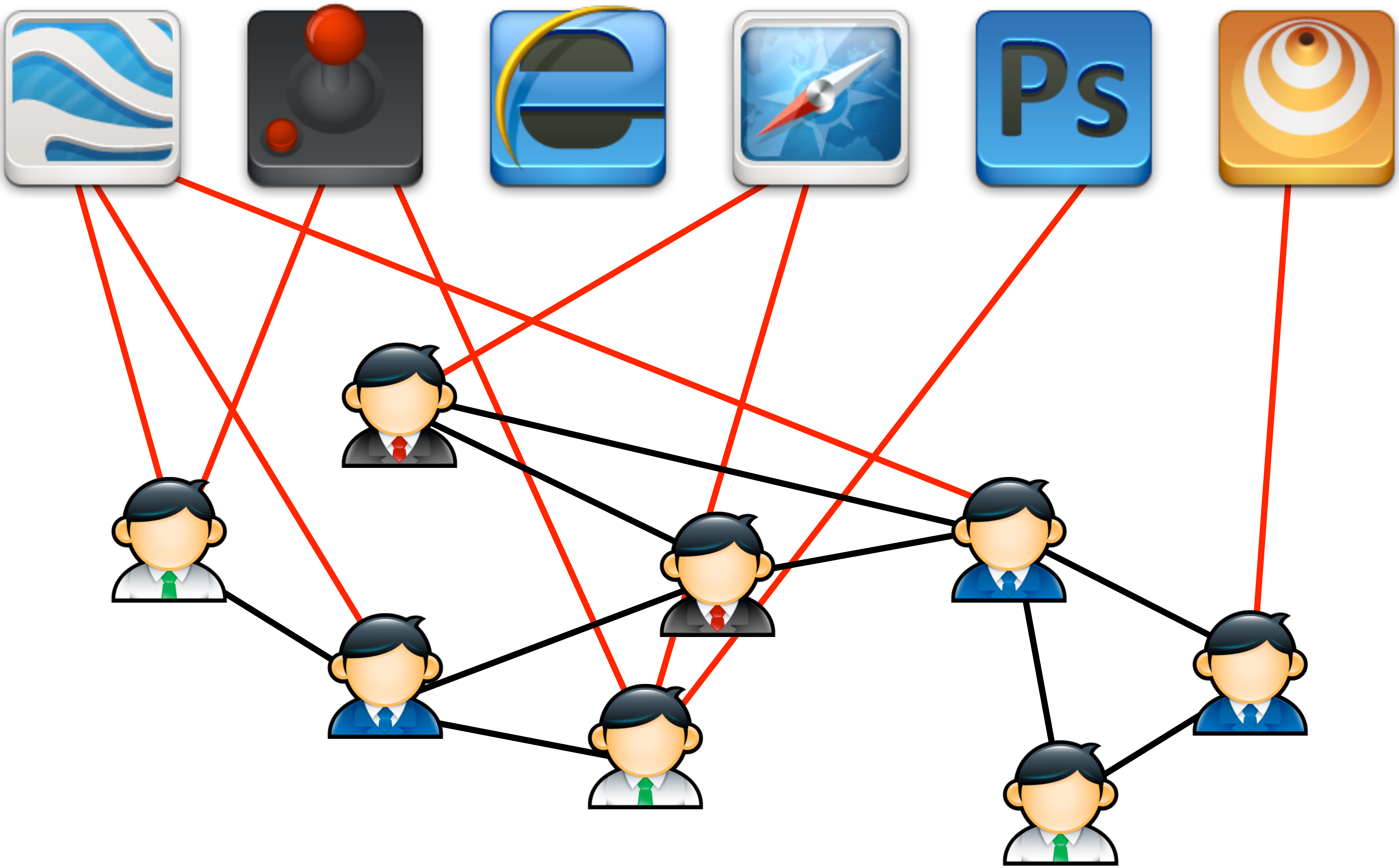
Applications



Applications

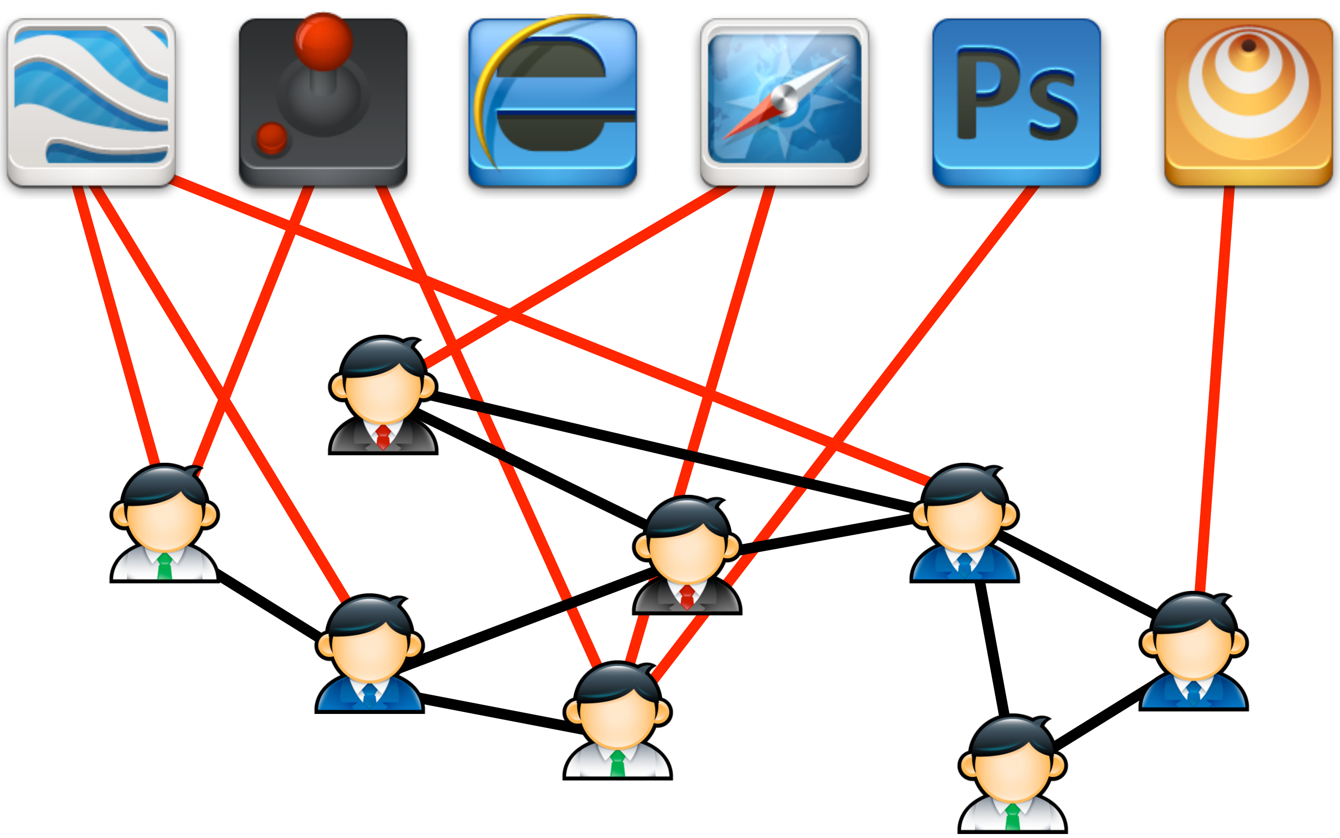


Applications



Applications

social network = friendship + interests

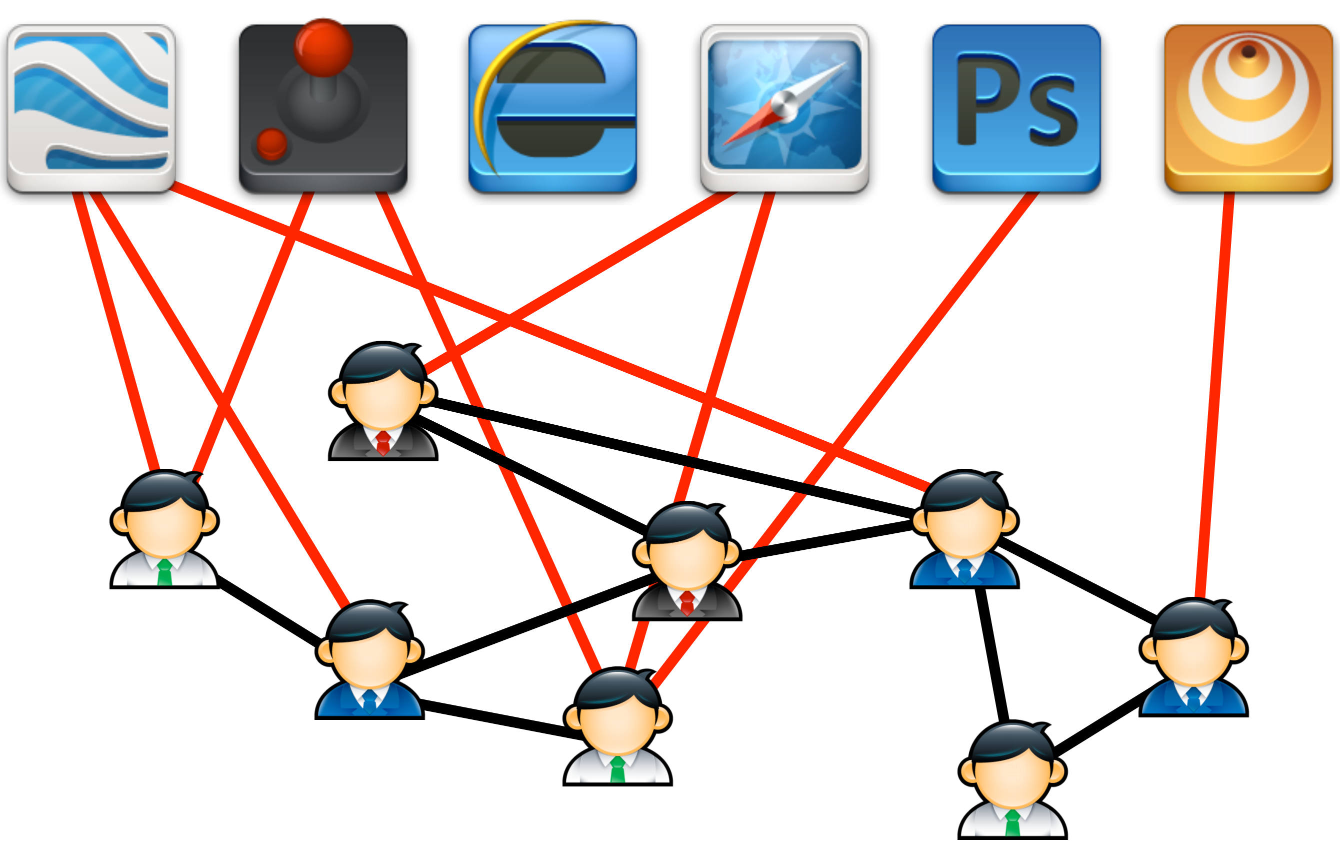


Applications

social network = friendship + interests

recommend users based on friendship & interests

recommend apps based on friendship & interests



Social Recommendation

recommend users based on friendship & interests	recommend apps based on friendship & interests
• boost traffic • make the user graph more dense • increase user population • stickiness	• boost traffic • increased revenue • increased user participation • make app graph more dense

... usually addressed by separate tools ...

Homophily

recommend users based on friendship & interests

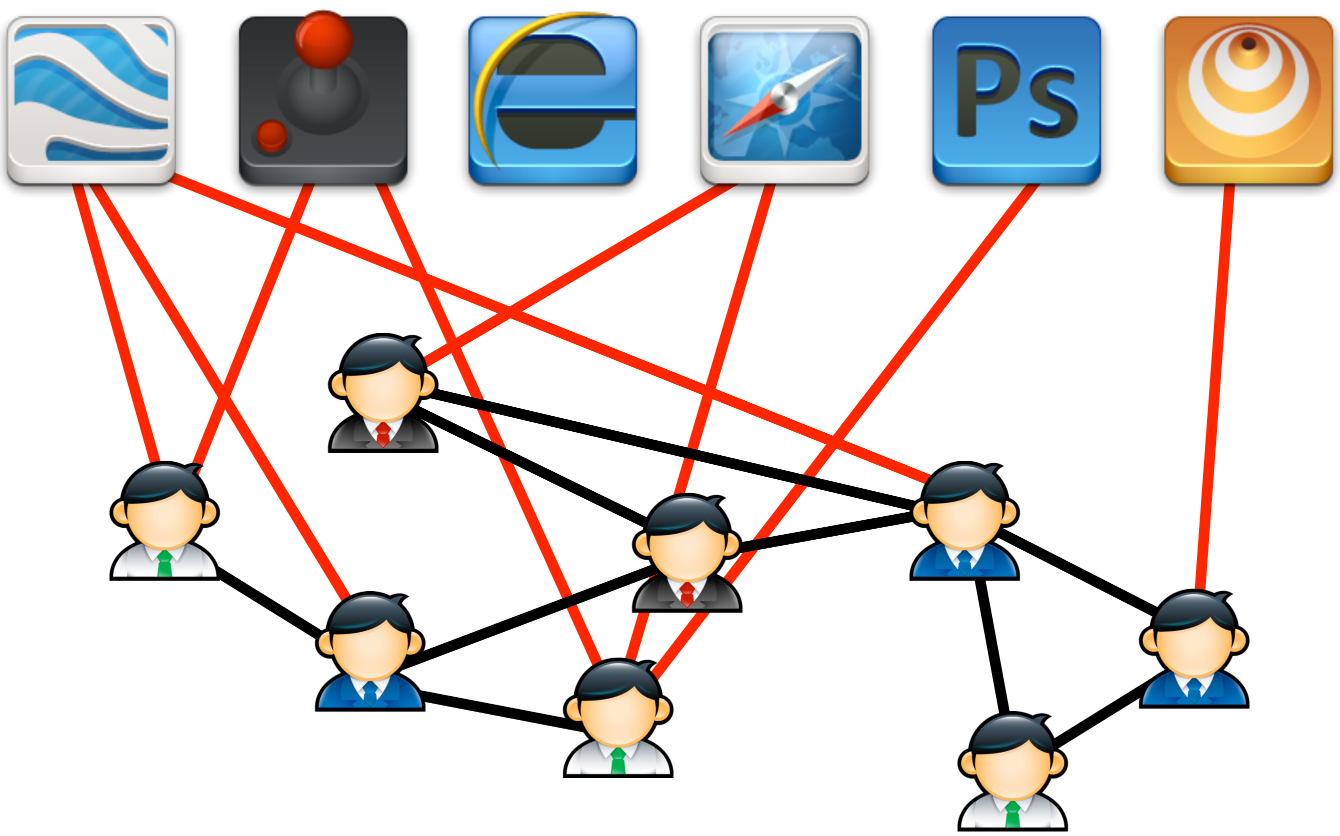
recommend apps based on friendship & interests

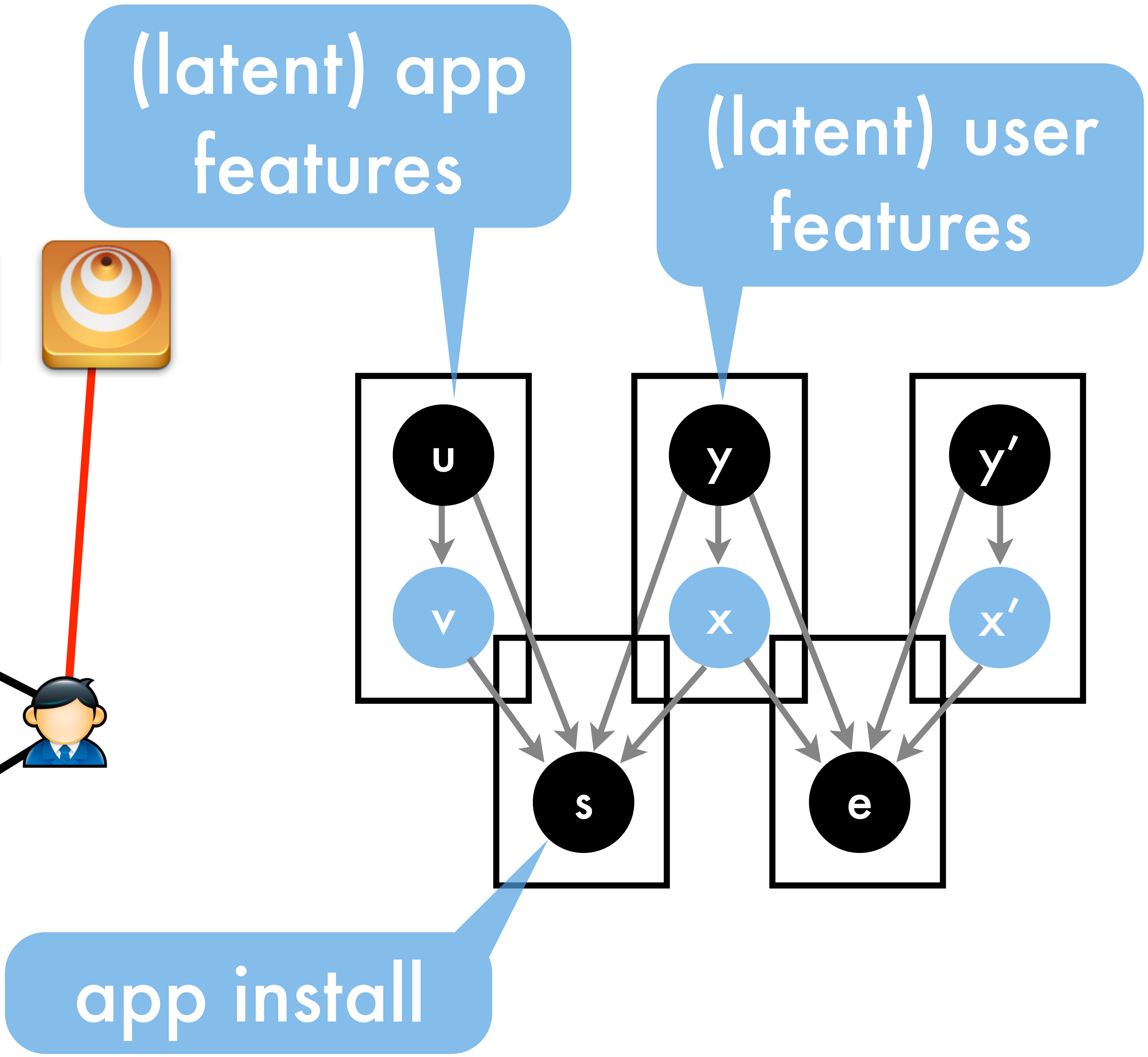
- users with similar interests are more likely to connect

- friends install similar applications

Highly correlated. Estimate both jointly

Model





(latent) app features
(latent) user features
u
v
y
x
y′
x′
s
e
app install

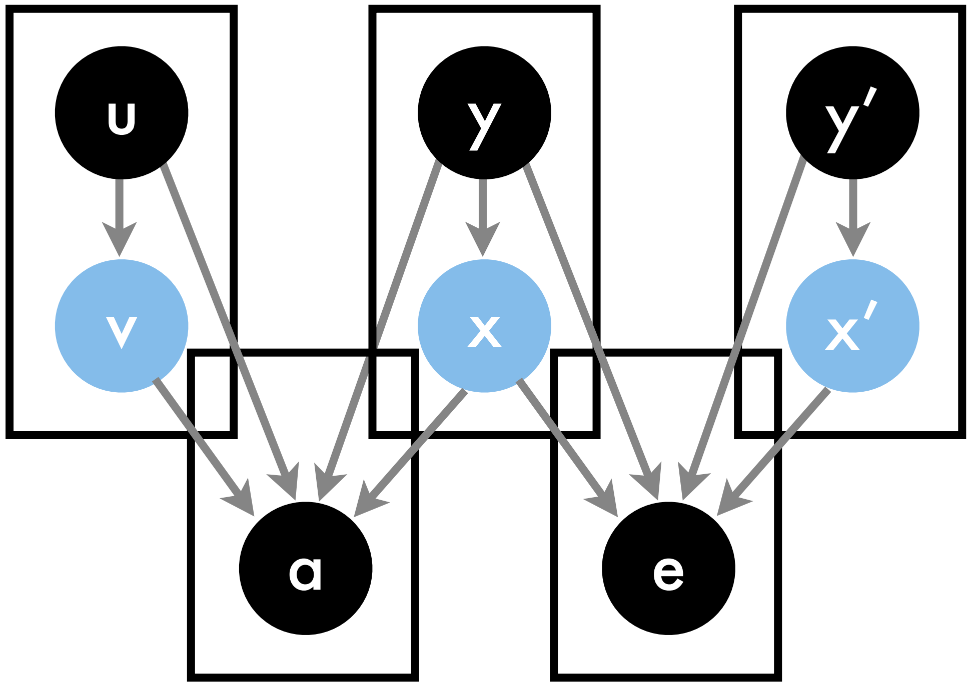
Model

- Social interaction

$$x_i \sim p(x|y_i)$$
$$x_j \sim p(x|y_j)$$
$$e_{ij} \sim p(e|x_i, y_i, x_j, y_j, \Phi)$$

- App install

$$x_i \sim p(x|y_i)$$
$$v_j \sim p(v|u_j)$$
$$a_{ij} \sim p(a|x_i, y_i, u_j, v_j, \Phi)$$



Model

- **Social interaction**

$$x_i \sim p(x|y_i)$$
$$x_j \sim p(x|y_j)$$
$$e_{ij} \sim p(e|x_i, y_i, x_j, y_j, \Phi)$$

cold start

latent features

$$x_i = Ay_i + \epsilon_i$$
$$v_j = Bu_j + \tilde{\epsilon}_j$$

- **App install**

$$x_i \sim p(x|y_i)$$
$$v_j \sim p(v|u_j)$$
$$a_{ij} \sim p(a|x_i, y_i, u_j, v_j, \Phi)$$

$$e_{ij} \sim p(e|x_i^\top x_j + y_i^\top W y_j)$$
$$a_{ij} \sim p(a|x_i^\top v_j + y_i^\top M u_j)$$

bilinear features

Optimization Problem

minimize $\lambda_e \sum_{(i,j)} l(e_{ij}, x_i^\top x_j + y_i^\top W y_j) +$

Optimization Problem

minimize $\lambda_e \sum_{(i,j)} l(e_{ij}, x_i^\top x_j + y_i^\top W y_j) +$

social

Optimization Problem

minimize $$\lambda_e \sum_{(i,j)} l(e_{ij}, x_i^\top x_j + y_i^\top W y_j) +$$

social

$$\lambda_a \sum_{(i,j)} l(a_{ij}, x_i^\top v_j + y_i^\top M u_j) +$$

app

Optimization Problem

minimize $$\lambda_e \sum_{(i,j)} l(e_{ij}, x_i^\top x_j + y_i^\top W y_j) +$$

social

$$\lambda_a \sum_{(i,j)} l(a_{ij}, x_i^\top v_j + y_i^\top M u_j) +$$

app

reconstruction

$$\lambda_x \sum \gamma(x_i | y_i) + \lambda_v \sum \gamma(v_i | u_i) +$$

Optimization Problem

minimize $$\lambda_e \sum_{(i,j)} l(e_{ij}, x_i^\top x_j + y_i^\top W y_j) +$$ social

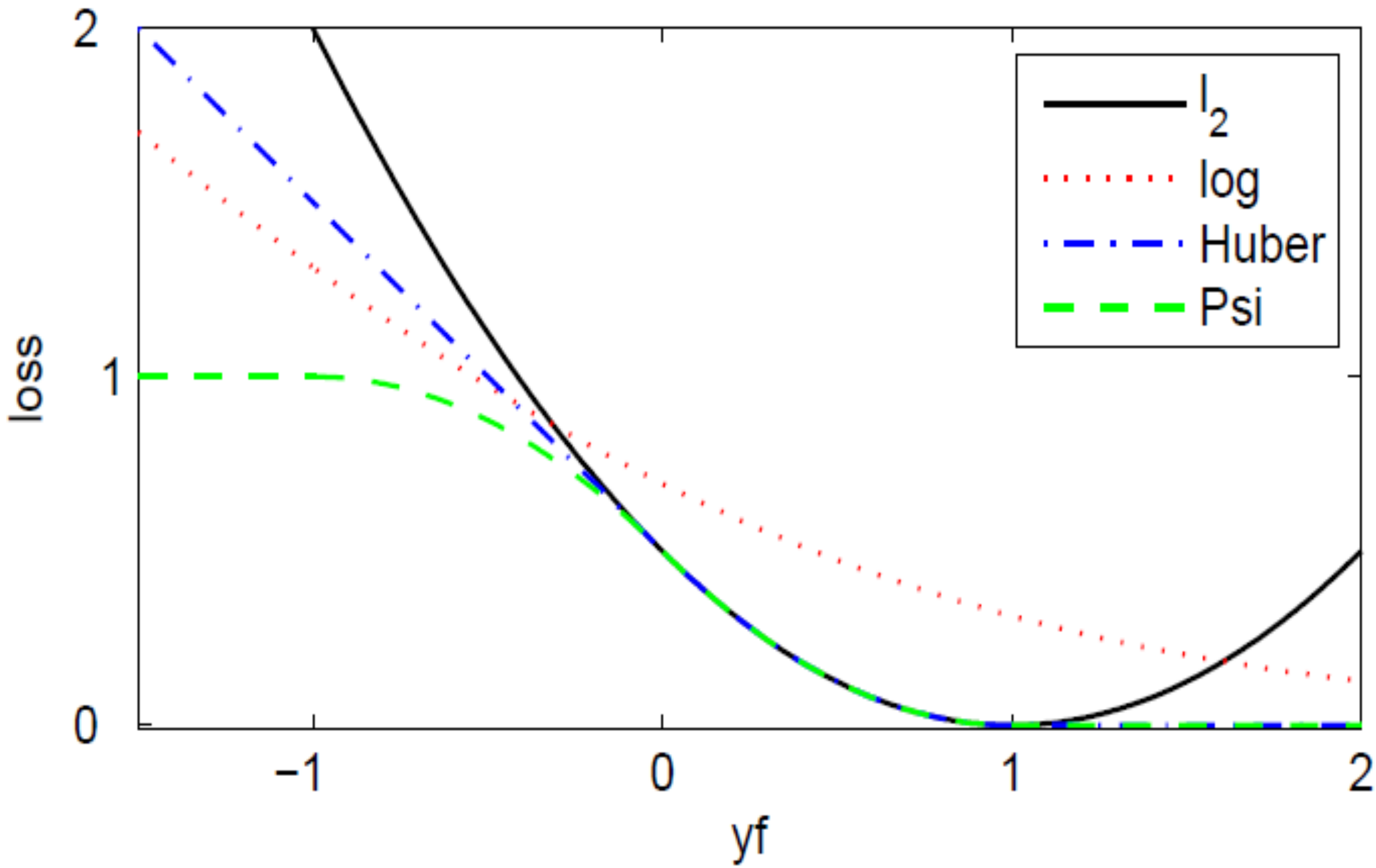
$$\lambda_a \sum_{(i,j)} l(a_{ij}, x_i^\top v_j + y_i^\top M u_j) +$$ app

reconstruction $$\lambda_x \sum_i \gamma(x_i | y_i) + \lambda_v \sum_i \gamma(v_i | u_i) +$$

$$\lambda_W \|W\|^2 + \lambda_M \|M\|^2 + \lambda_A \|A\|^2 + \lambda_B \|B\|^2$$

regularizer

Loss Function

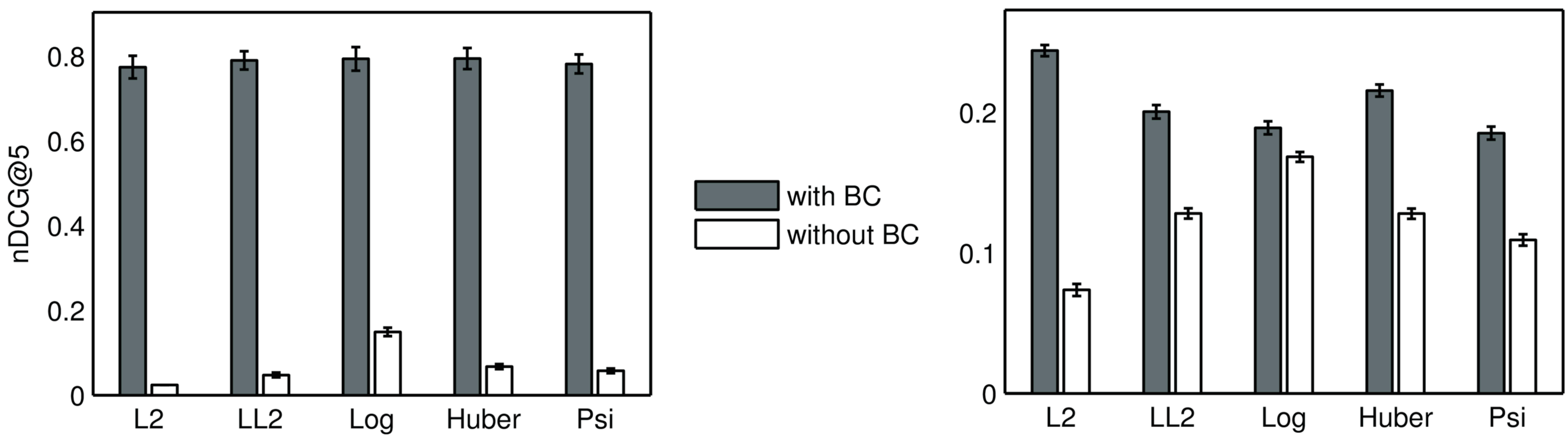


Loss

- Much more evidence of application non-install (i.e. many more negative examples)
- Few links between vertices in friendship network (even within short graph distance)

- Generate ranking problems (link, non-link) with non-links drawn from background set

Loss



application recommendation

social recommendation

Optimization

- Nonconvex optimization problem
- Large set of variables

$$x_i = Ay_i + \epsilon_i$$
$$v_j = Bu_j + \tilde{\epsilon}_j$$

- Stochastic gradient descent on x, v, ε for speed

$$e_{ij} \sim p(e|x_i^\top x_j + y_i^\top W y_j)$$
$$a_{ij} \sim p(a|x_i^\top v_j + y_i^\top M u_j)$$

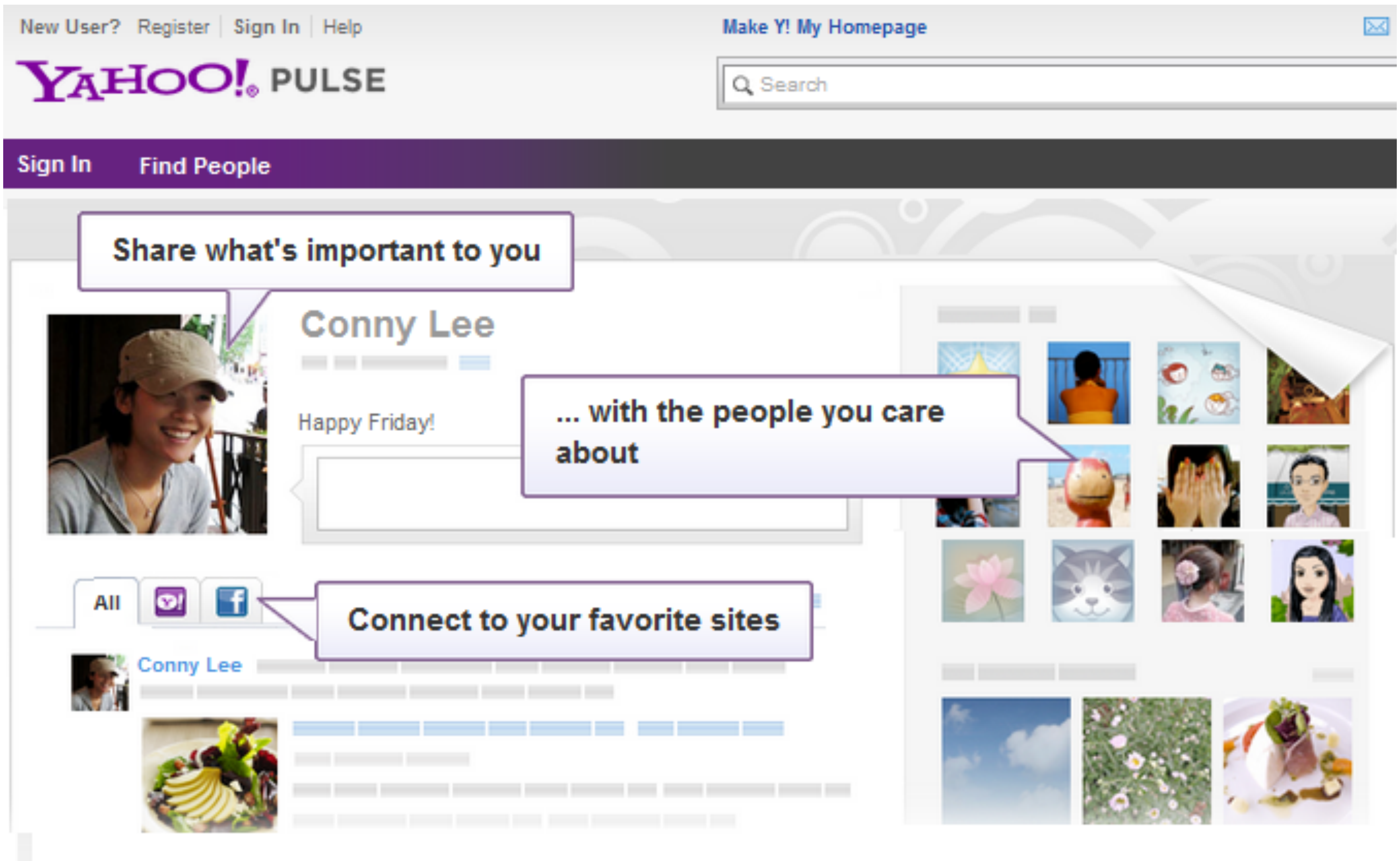
- Use hashing to reduce memory load, i.e.

$$x_{ij} = \sigma(i,j) X[h(i,j)]$$

binary hash

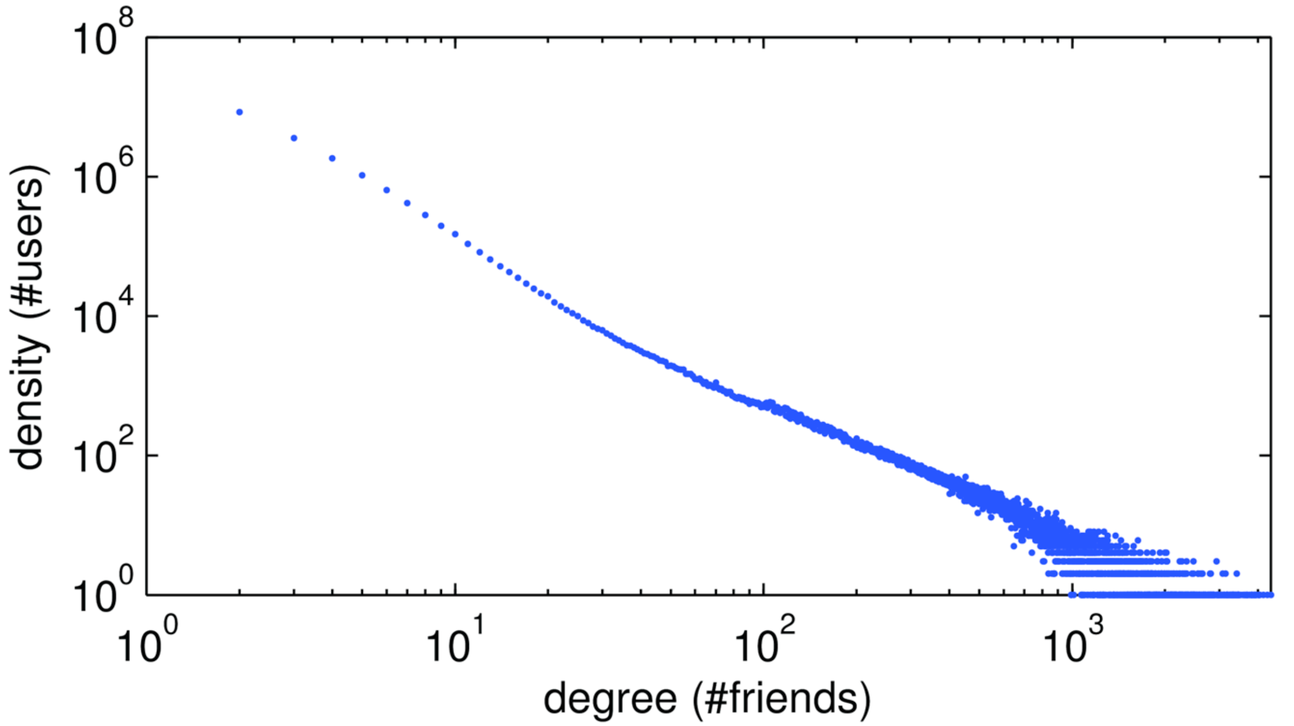
hash

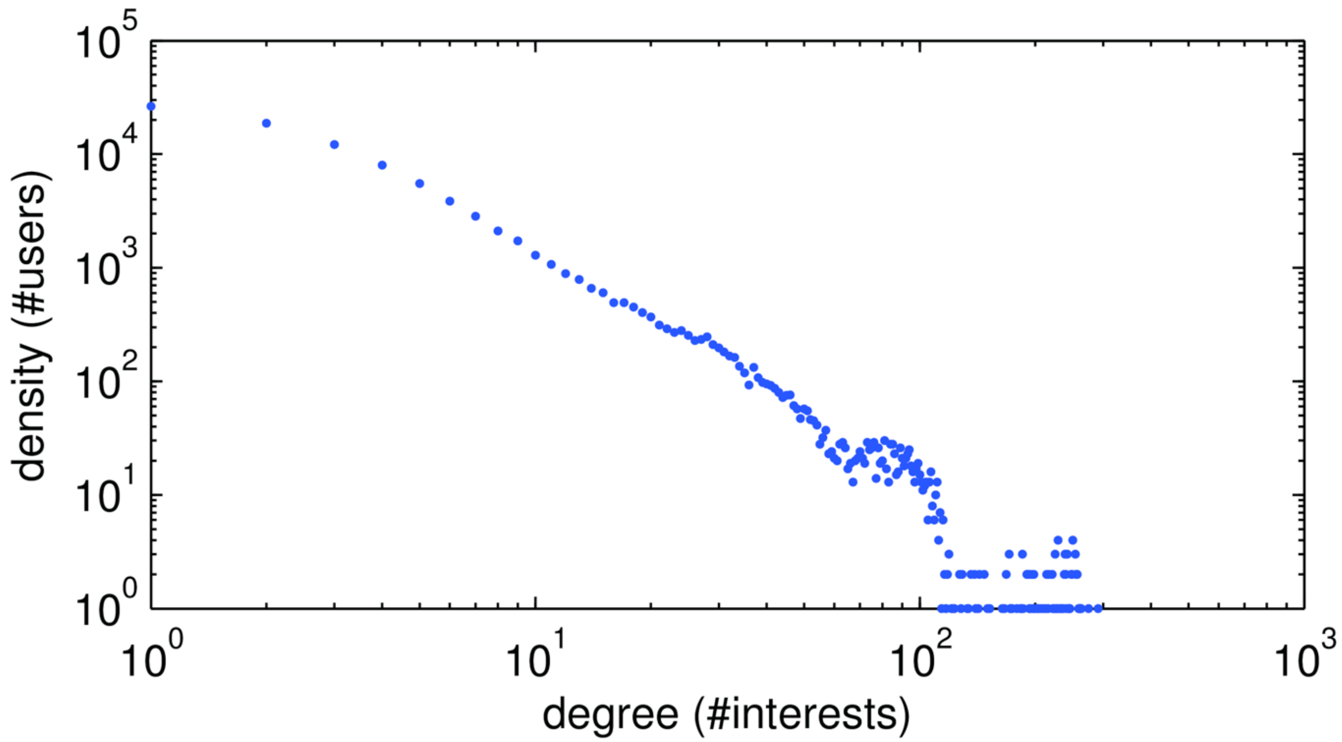
Y! Pulse



New User? Register | Sign In | Help
Make Y! My Homepage
YAHOO! PULSE
Search
Sign In
Find People
Share what's important to you
Conny Lee
Happy Friday!
... with the people you care about
All
Connect to your favorite sites
Conny Lee

Y! Pulse Data





1.2M users, 386 items
6.1M friend connections
29M interest indications

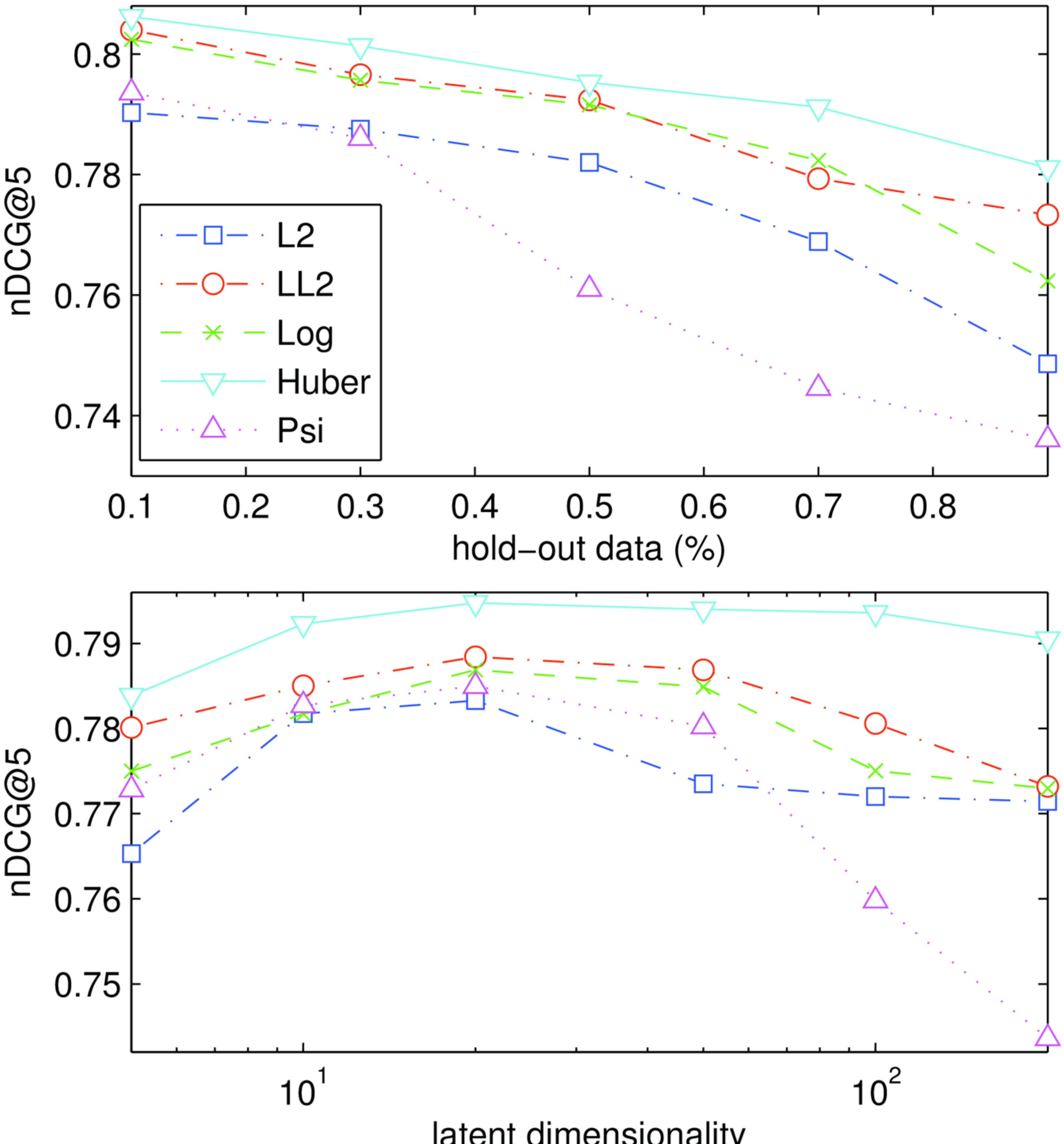
App Recommendation

Models	loss	$\Omega[\cdot]$	MAP@5	MAR@5	nDCG@5
SIM			0.630	0.186	0.698
RLFM			0.729	0.211	0.737
NLFM			0.748	0.222	0.761
FIP	ℓ_2	ℓ_2	0.768	0.228	0.774
FIP	lazy ℓ_2	ℓ_2	0.781	0.232	0.790
FIP	logistic	ℓ_2	0.781	0.232	0.793
FIP	Huber	ℓ_2	0.781	0.232	0.794
FIP	Ψ	ℓ_2	0.777	0.231	0.771
FIP	ℓ_2	ℓ_1	0.778	0.231	0.787
FIP	lazy ℓ_2	ℓ_1	0.780	0.231	0.791
FIP	logistic	ℓ_1	0.779	0.231	0.792
FIP	Huber	ℓ_1	**0.786**	**0.233**	**0.797**
FIP	Ψ	ℓ_1	0.765	0.215	0.772

SIM: similarity based model;
RLFM: regression based latent factor model (Chen&Agarwal); NLFM: SIM&RLFM

Social recommendation

Models	loss	$\Omega[\cdot]$	MAP@5	MAR@5	nDCG@5
RLFM			0.164	0.202	0.174
FIP	ℓ_2	ℓ_2	**0.359**	**0.284**	**0.244**
FIP	lazy ℓ_2	ℓ_2	0.193	0.269	0.200
FIP	logistic	ℓ_2	0.174	0.220	0.189
FIP	Huber	ℓ_2	0.210	0.234	0.215
FIP	Ψ	ℓ_2	0.187	0.255	0.185
FIP	ℓ_2	ℓ_1	0.186	0.230	0.214
FIP	lazy ℓ_2	ℓ_1	0.180	0.223	0.194
FIP	logistic	ℓ_1	0.183	0.217	0.189
FIP	Huber	ℓ_1	0.188	0.222	0.200
FIP	Ψ	ℓ_1	0.178	0.208	0.179

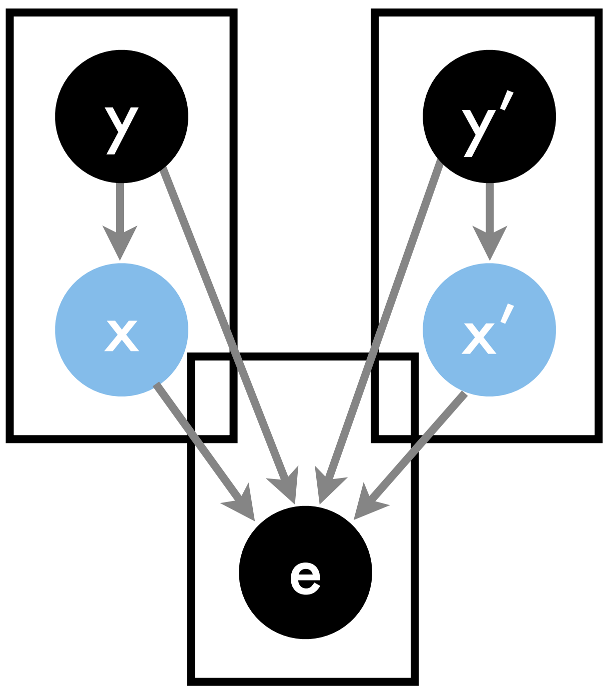


app recommendation
L2 penalty

Extensions

- Multiple relations

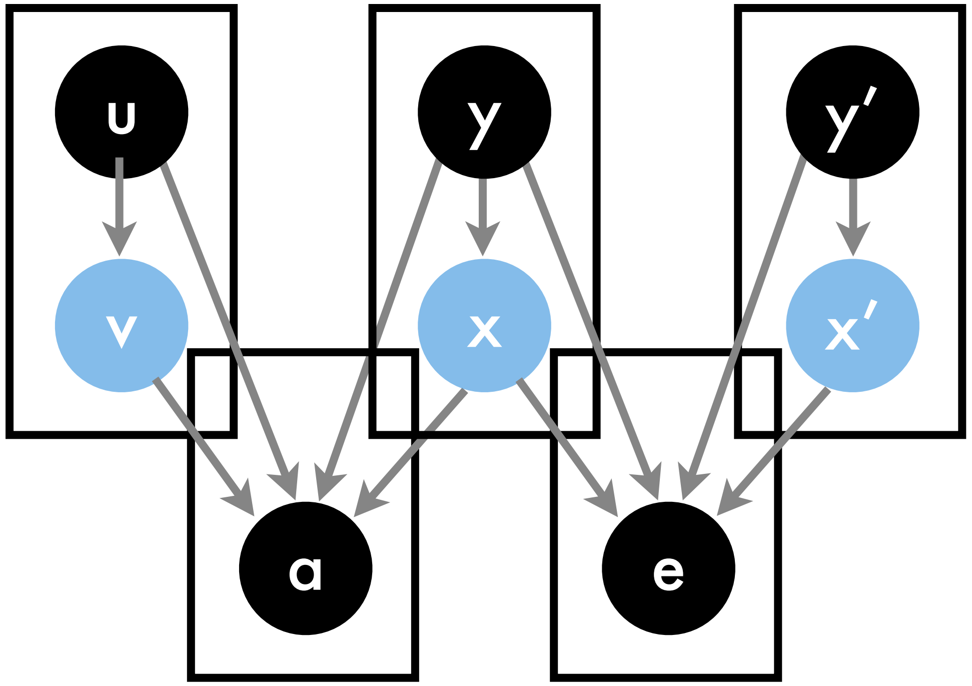
  (user, user)
  (user, app)
  (app, advertisement)



Extensions

- Multiple relations

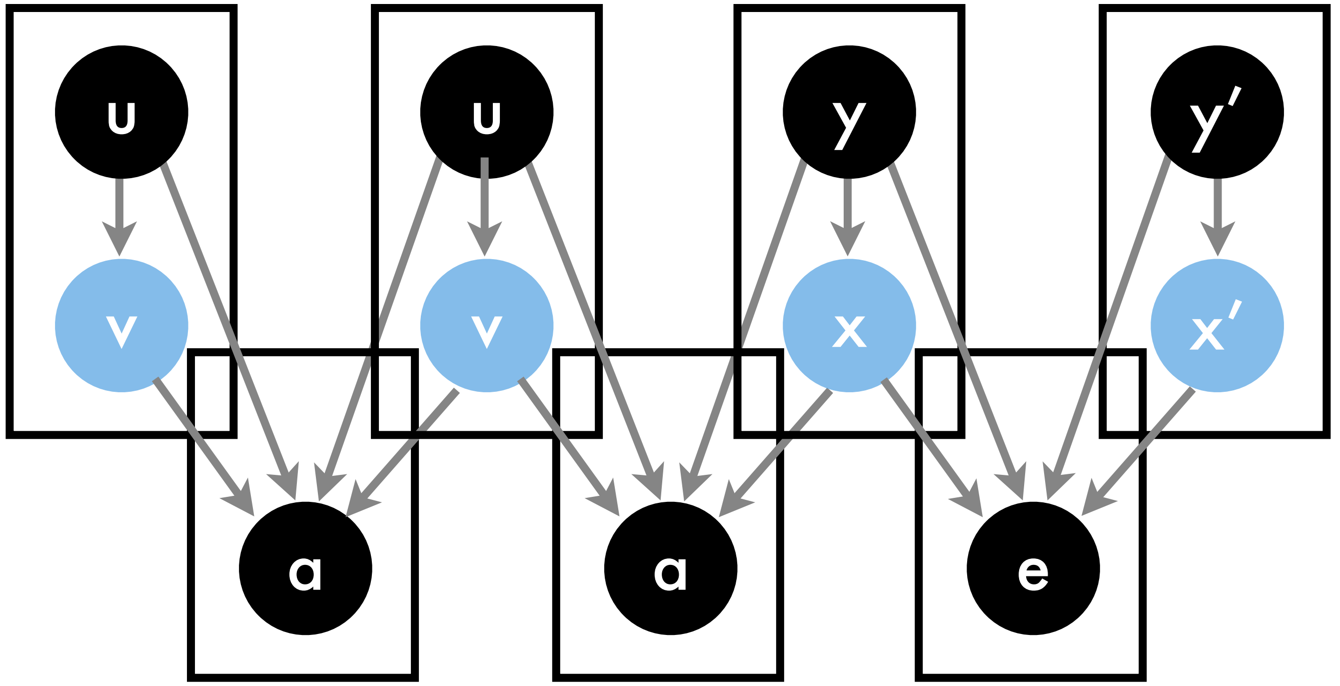
  (user, user)
  (user, app)
  (app, advertisement)



Extensions

- Multiple relations

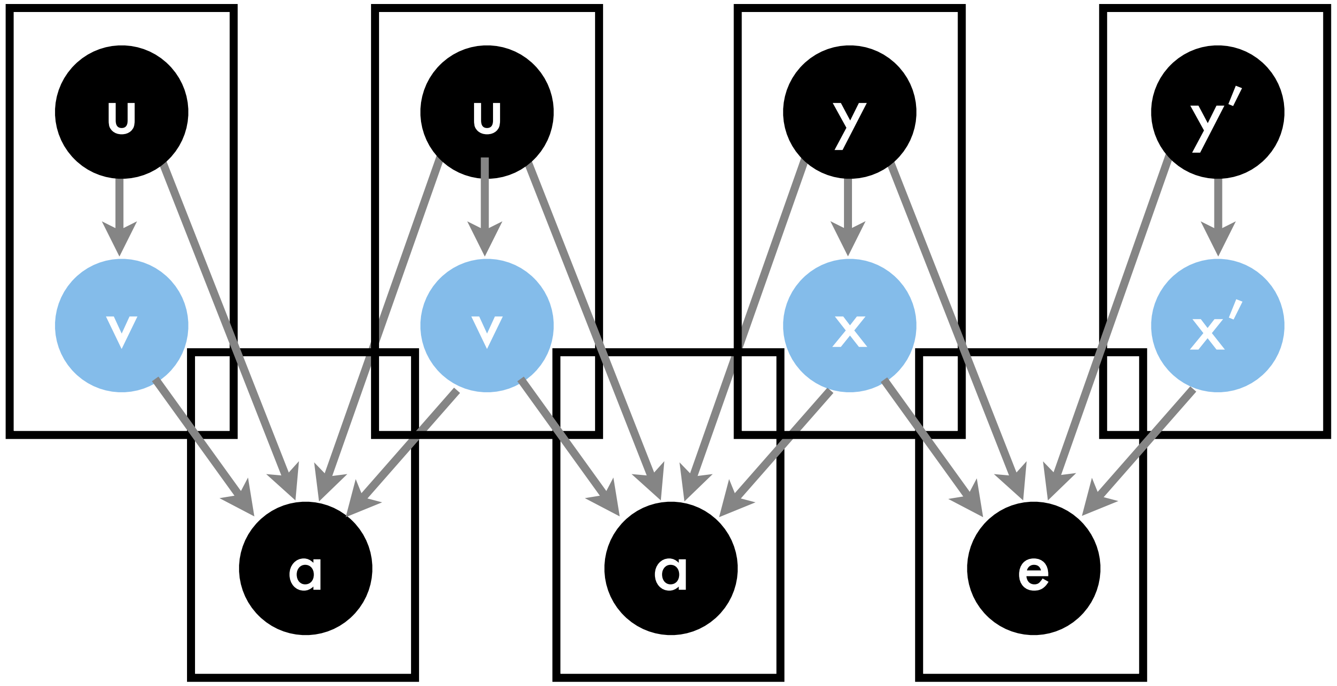
  (user, user)
  (user, app)
  (app, advertisement)



Extensions

- Multiple relations

  (user, user)
  (user, app)
  (app, advertisement)



- Users visiting several properties
  news, mail, frontpage, social network, etc.
- Different statistical models
  - Latent Dirichlet Allocation for latent factors
  - Indian Buffet Process

More strategies

Multiple factor LDA

- Discrete set of preferences (Porteous, Bart, Welling, 2008)
  - User picks one to assess movie
  - Movie represented by a discrete attribute
- Inference by Gibbs sampler
- Works fairly well

- Extension by Lester Mackey and coworkers to combine with BPMF model

More state representations

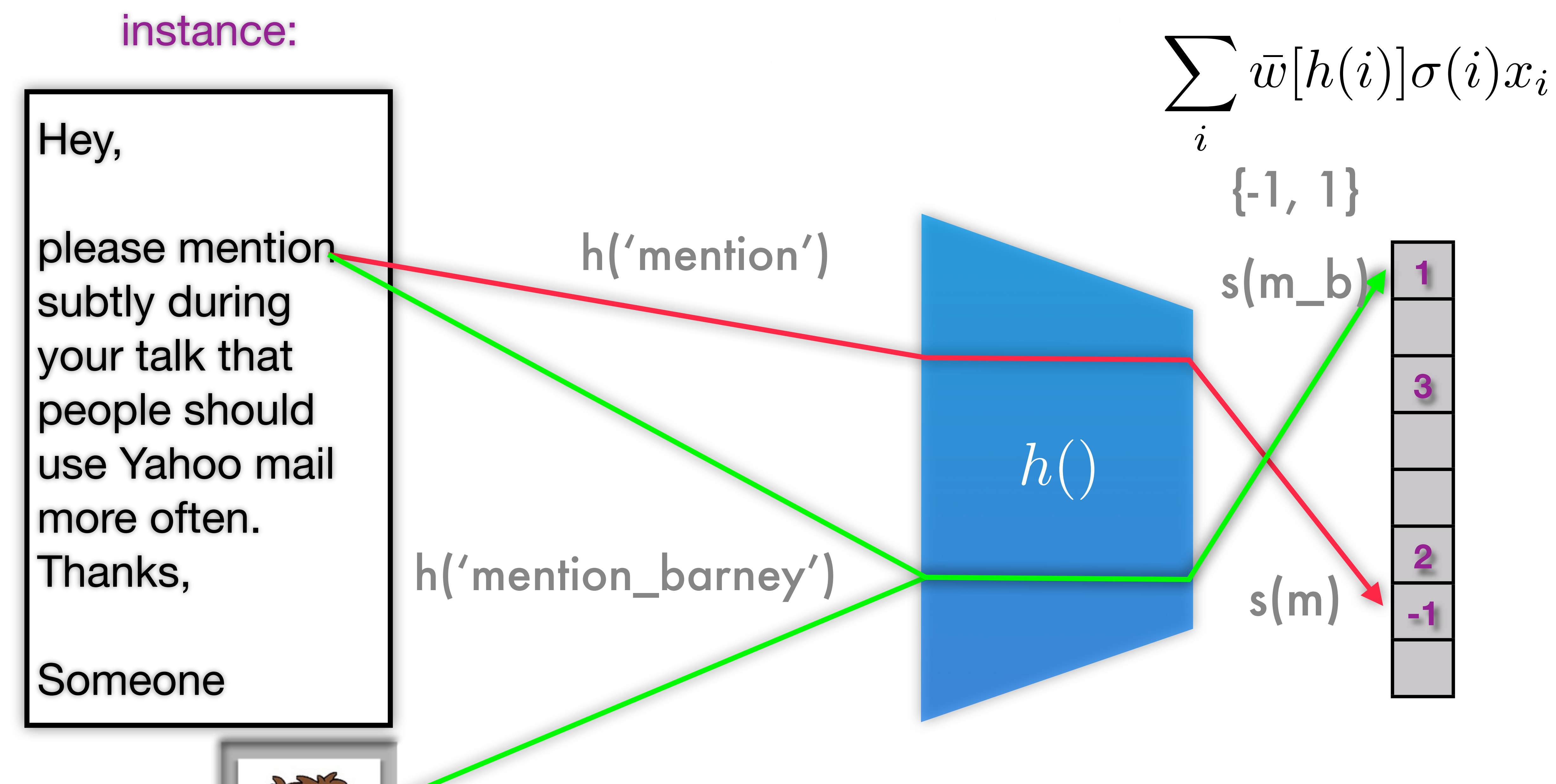
- Indian Buffet Process (Griffiths & Ghahramani, 2005)
  - Attribute vector is binary string
  - Models preferences naturally & very compact (Inference is costly)
- Hierarchical attribute representation and clustering over users ... TO DO

5 Hashing

Parameter Storage

- We have millions of users
- We have millions of products
- Storage - for 100 factors this requires
  $10^6 \times 10^6 \times 8 = 8\text{TB}$
- We want a model that can be kept in **RAM** (<16GB)
  - Instant response for each user
  - Disks have 20 IOP/s at best (SSDs much better)
- Privacy (what if parameter vector leaks)

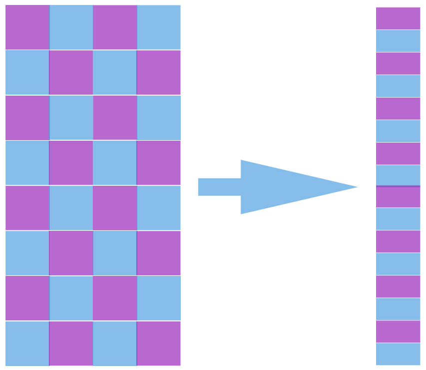
Recall - Hash Kernels



Similar to count hash
(Charikar, Chen, Farrach-Colton, 2003)

Collaborative Filtering

- **Hashing compression**



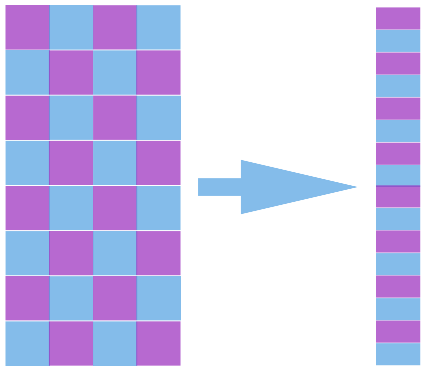
$$u_i = \sum_{j,k:h(j,k)=i} \xi(j,k) U_{jk} \text{ and } v_i = \sum_{j,k:h'(j,k)=i} \xi'(j,k) V_{jk}.$$

$$X_{ij} := \sum_k \xi(k,i)\xi'(k,j) u_{h(k,i)} v_{h'(k,j)}.$$

- **Approximation is O(1/n)**
  - To show that estimate is unbiased take expectation over Rademacher hash.

Collaborative Filtering

- **Hashing compression**



$$u_i = \sum_{j,k:h(k,j)=i} \xi(k,j) U_{kj} \text{ and } v_i = \sum_{j,k:h'(k,j)=i} \xi'(k,j) V_{kj}.$$

$$X_{ij} := \sum_k \xi(k,i)\xi'(k,j) u_{h(k,i)} v_{h'(k,j)}.$$

- **Expectation**

expectation vanishes

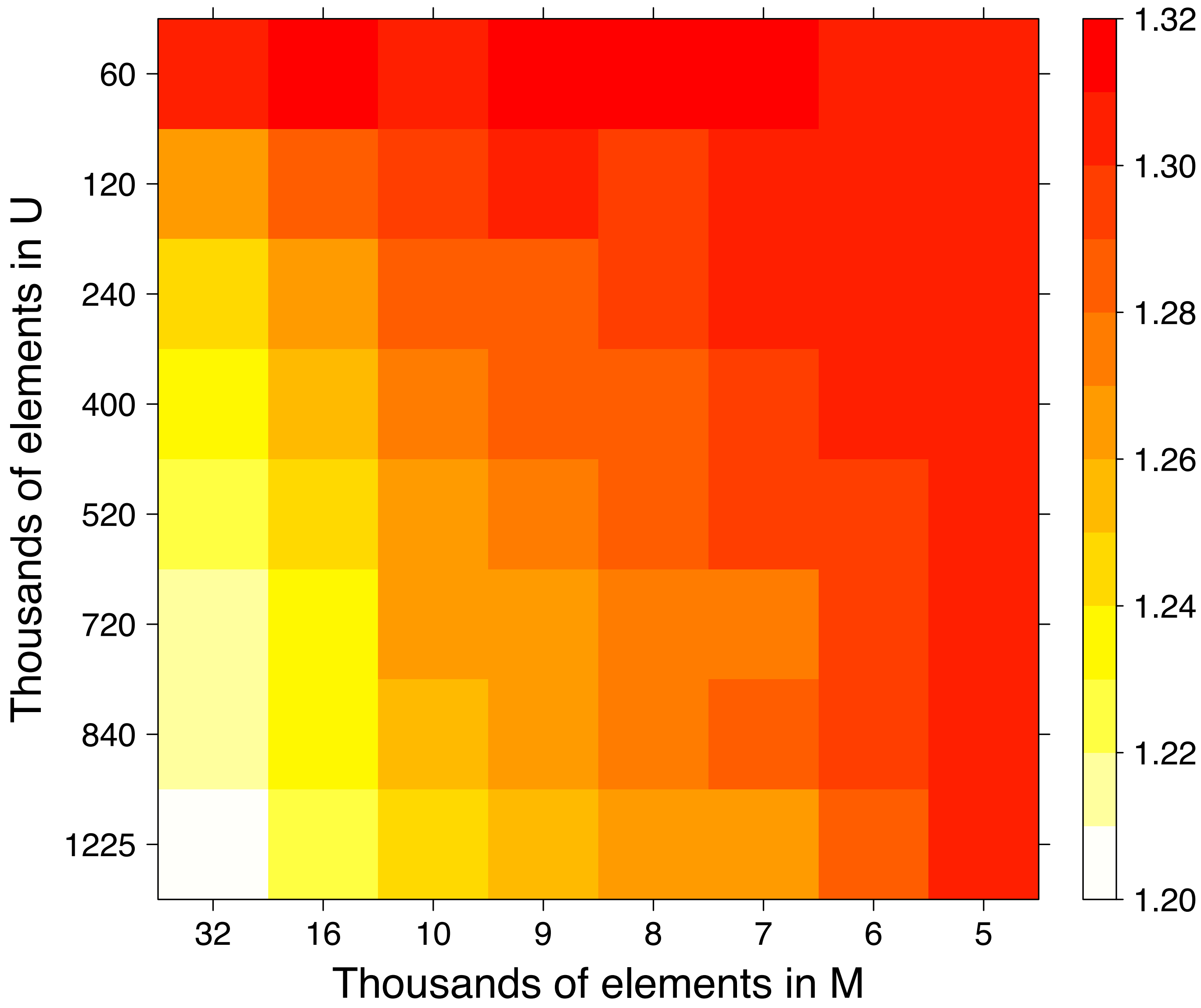
$$X_{ij} := \sum_k \xi(k,i)\xi'(k,j) \sum_{l,k:h(k,l)=h(k,i)} \sum_{o,k:h'(k,o)=h'(k,j)} \xi(k,l)\xi'(k,o) U_{kl} V_{ko}$$

Collaborative Hashing

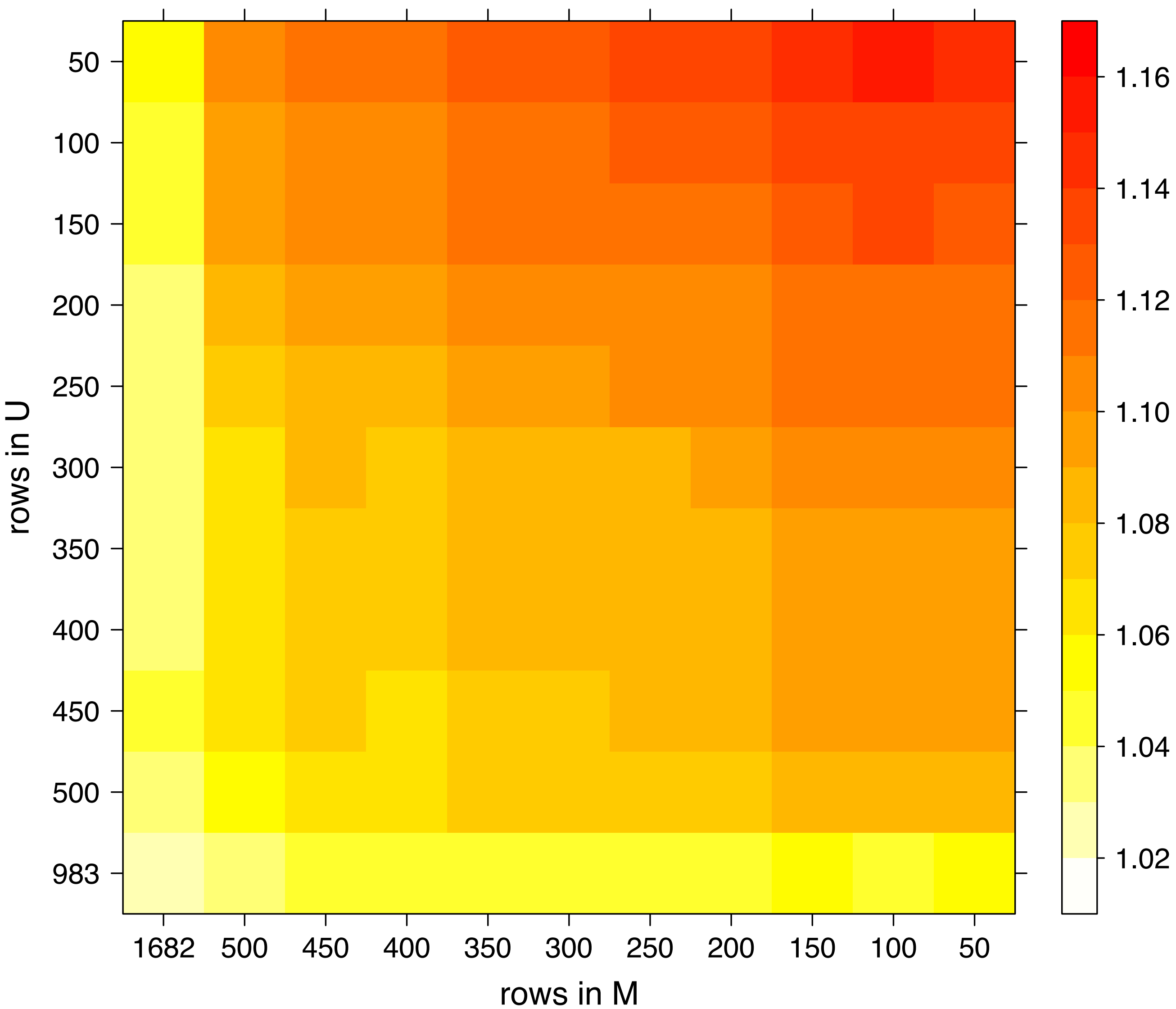
- Combine with stochastic gradient descent
- Random access in memory is expensive (we now have to do k lookups per pair)
- Feistel networks can accelerate this

- Distributed optimization without locking

Examples



Eachmovie



MovieLens

Summary

- Neighborhood methods
  - User / movie similarity
  - Iteration on graph
- Matrix Factorization
  - Singular value decomposition
  - Convex reformulation
- Ranking and Session Modeling
  - Ordinal regression
  - Session models
- Features
  - Latent dense (Bayesian Probabilistic Matrix Factorization)
  - Latent sparse (Dirichlet process factorization)
  - Coldstart problem (inferring features)
- Hashing

Further reading

- Collaborative Filtering with temporal dynamics
  http://research.yahoo.com/files/kdd-fp074-koren.pdf
- Neighborhood factorization
  http://research.yahoo.com/files/paper.pdf
- Matrix Factorization for recommender systems
  http://research.yahoo.com/files/ieeecomputer.pdf
- CoFi Rank (collaborative filtering & ranking)
  http://www.cofirank.org/
- Yehuda Koren's papers
  http://research.yahoo.com/Yehuda_Koren