

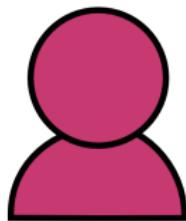
# Poisson Trust Factorization for Incorporating Social Networks into Personalized Item Recommendation

Allison J. B. Chaney

Princeton University  
David Blei (advisor)

October 15, 2013

# Problem



Till We Have Faces



East of Eden

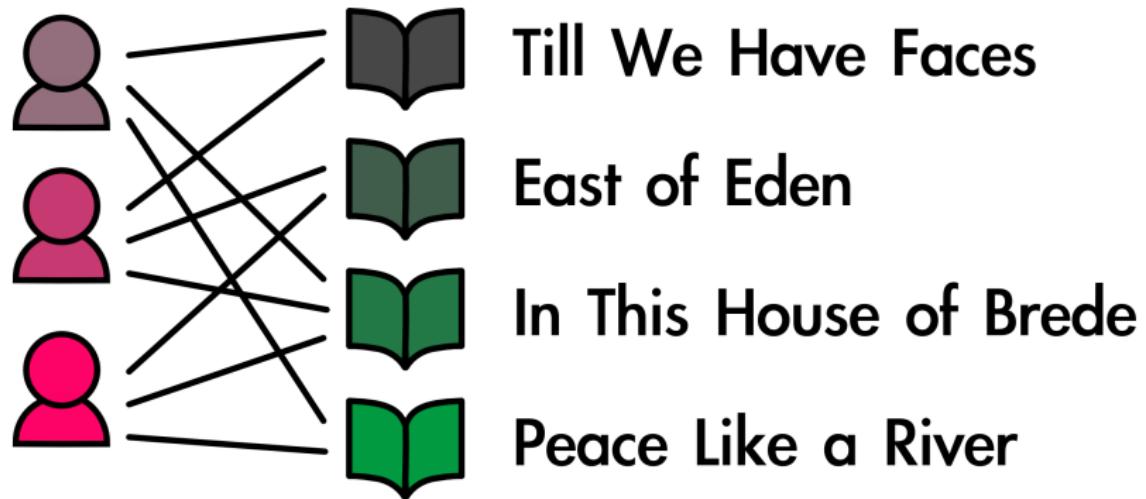


In This House of Brede



? ? ?

# Collaborative Filtering



# Collaborative Filtering: Matrix Factorization



# Collaborative Filtering: Matrix Factorization



# Collaborative Filtering: Matrix Factorization



# Matrix Factorization Extensions



written reviews

age & gender

social network

language

location

...

text description

author/director

date released

labeled genre

length

...

# Matrix Factorization Extensions



written reviews

age & gender

social network

language

location

...

text description

author/director

date released

labeled genre

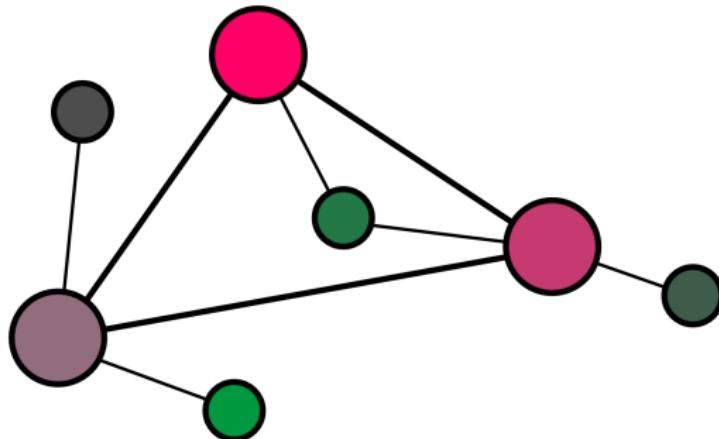
length

...

# Matrix Factorization Extensions



## Additional Data: Network Connections



# Existing Work on Incorporating Social Networks

**SoRec: social recommendation using probabilistic matrix factorization**, Ma, Yang, Lyu, and King, 2008.

**Learning to recommend with social trust ensemble**. Ma, King, and Lyu, 2009.

**A Matrix Factorization Technique with Trust Propagation for Recommendation in Social Networks**. Jamali and Ester, 2010.



## Jen &gt; Compare Books



- Jen's books: 1138 (1080 not in common)
- Books in common: 58 (21.17% of your library and 5.10% of her library)
- My books: 274 (216 not in common)

[back to Jen's profile »](#)  
[book compatibility test](#)  
[recommend a book to Jen](#)

common shelves

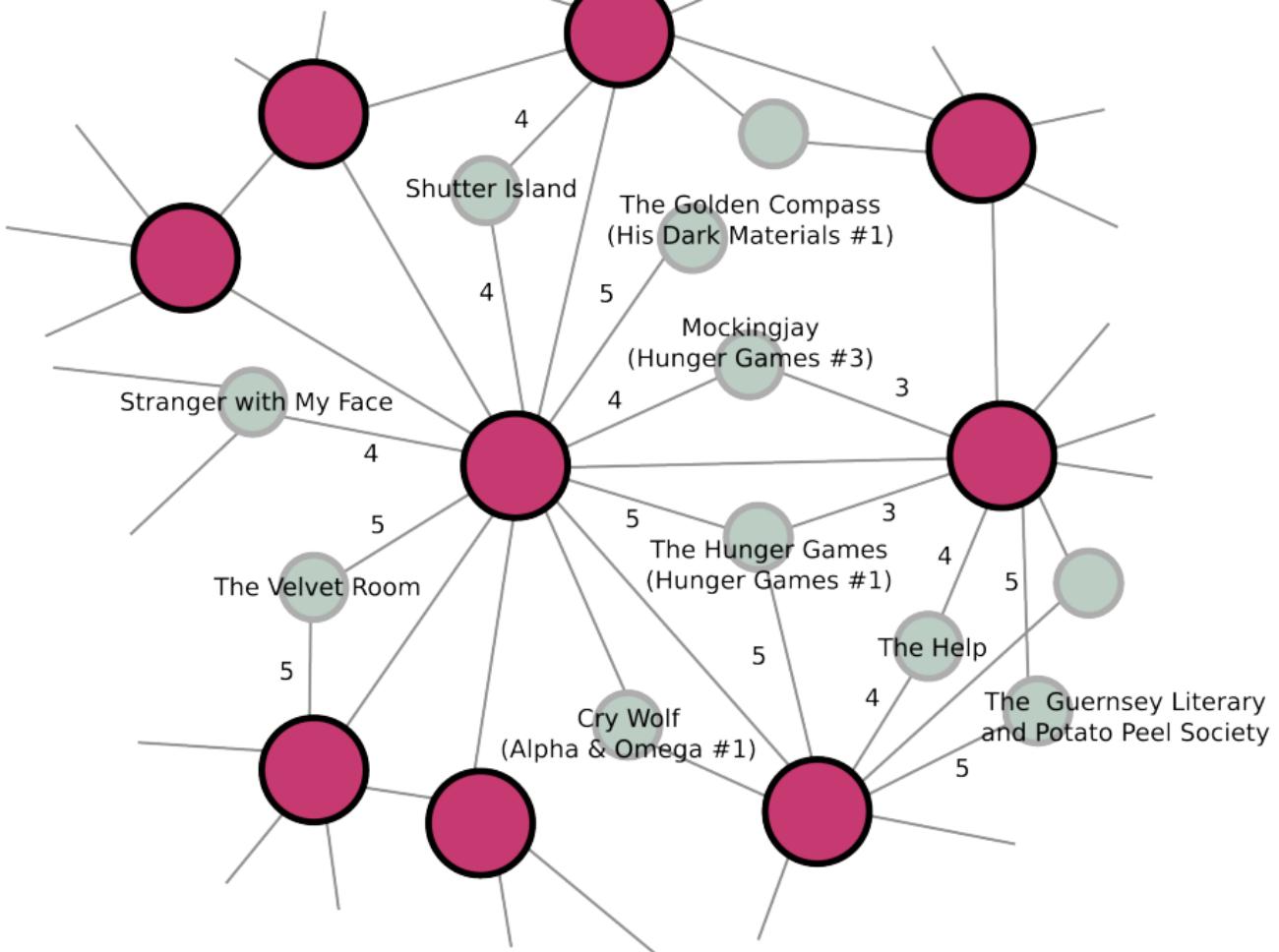
your shelves

their shelves

Your tastes are 71% similar for the books you both rated.

sort  (?)

book title	Jen's rating	my rating
My Mother She Killed Me, My Father He Ate Me: Forty New Fairy Tales by Kate Bernheimer (Editor)		to-read
The Bloody Chamber and Other Stories by Angela Carter		to-read
Haroun And The Sea Of Stories by Salman Rushdie	to-read	read
Neverwhere by Neil Gaiman		
How I Live Now by Meg Rosoff		to-read
The Sparrow by Mary Doria Russell	to-read	
Stardust by Neil Gaiman		
Swamplandia! by Karen Russell	to-read	to-read
A Wind in the Door by Madeleine L'Engle		to-read
Anthem by Ayn Rand		to-read



# Goodreads Scrape

## 1M Dataset

14,468 users

208,319 items

31,499 network  
connections

964,234 ratings (1 to 5)

## Ongoing

323,052 users

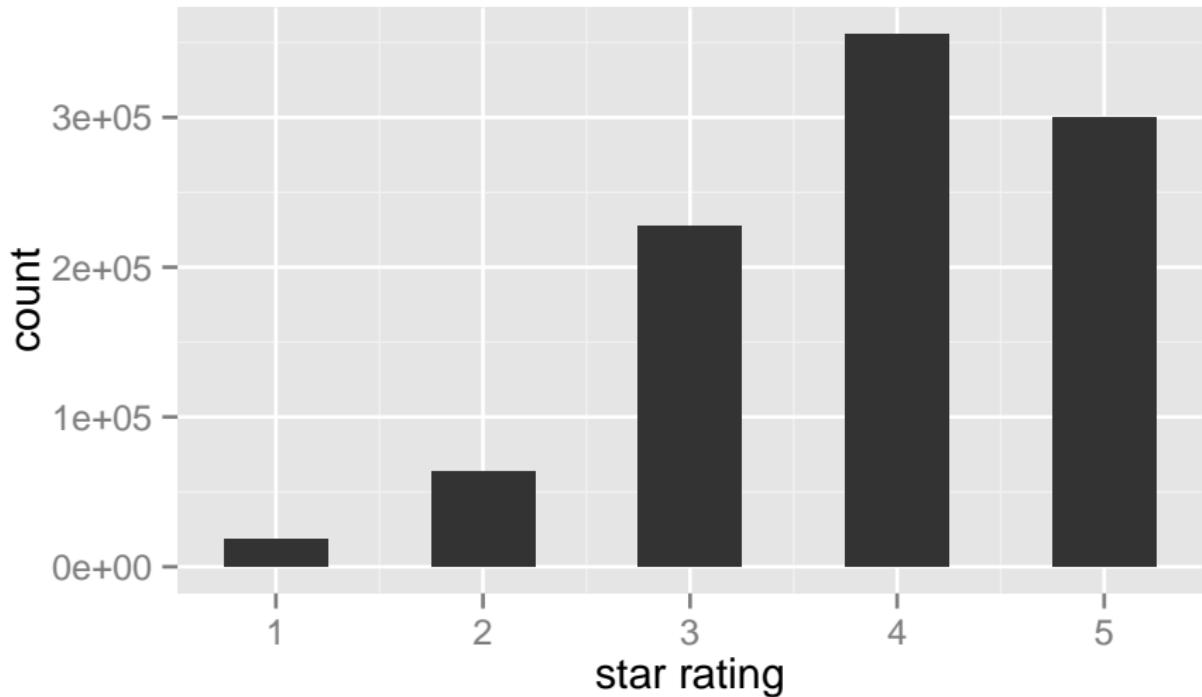
1,762,109 items

753,651 network  
connections

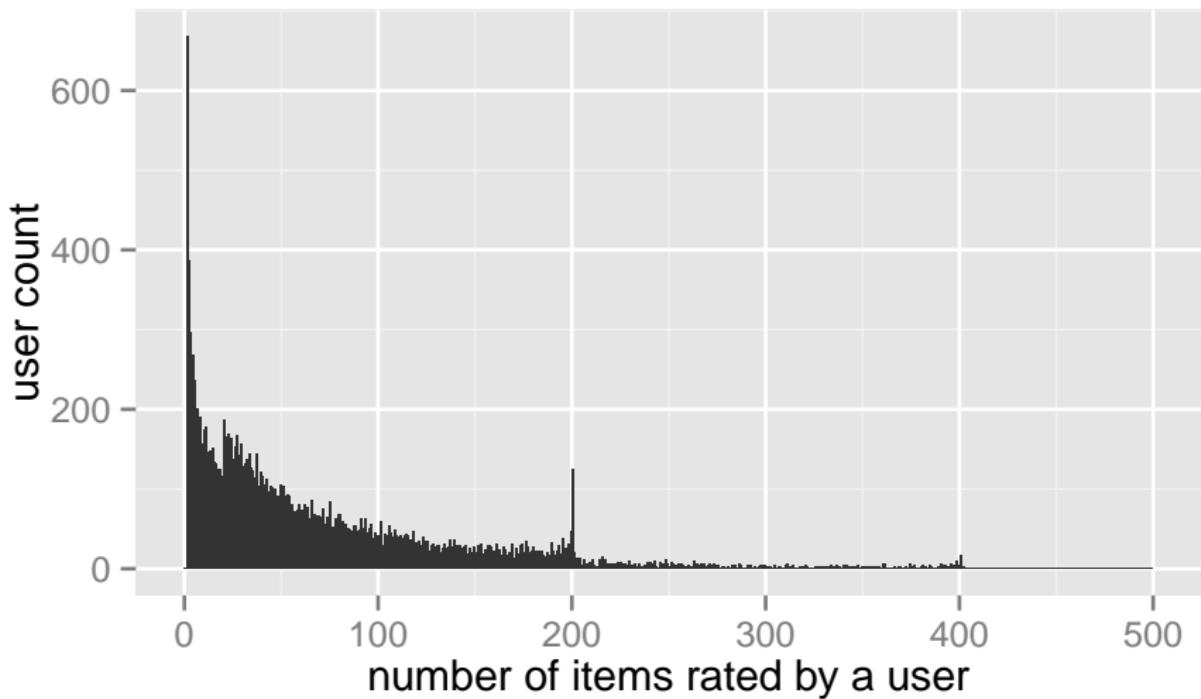
2,509,908 ratings

27,835,841 user-item  
connections

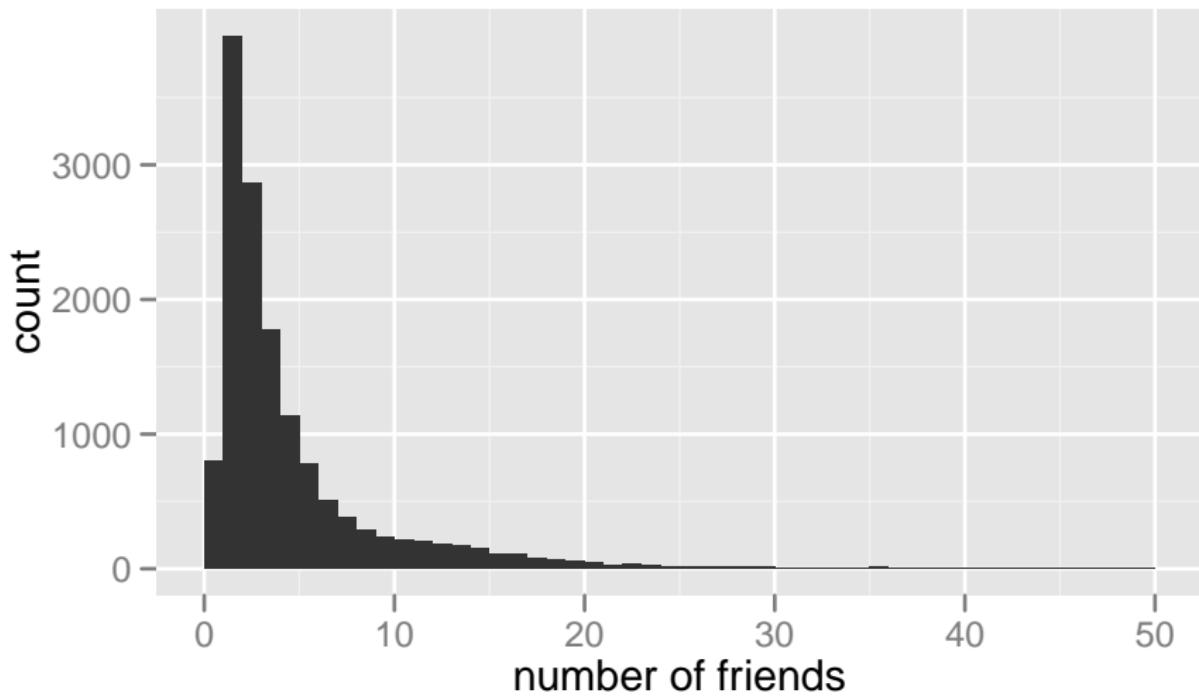
# Distribution of star ratings



## Distribution of user rating count



## Degree distribution



# Most popular books

The Hunger Games (The Hunger Games, #1)

Harry Potter and the Sorcerer's Stone (Harry Potter, #1)

To Kill a Mockingbird

The Great Gatsby

The Catcher in the Rye

Catching Fire (The Hunger Games, #2)

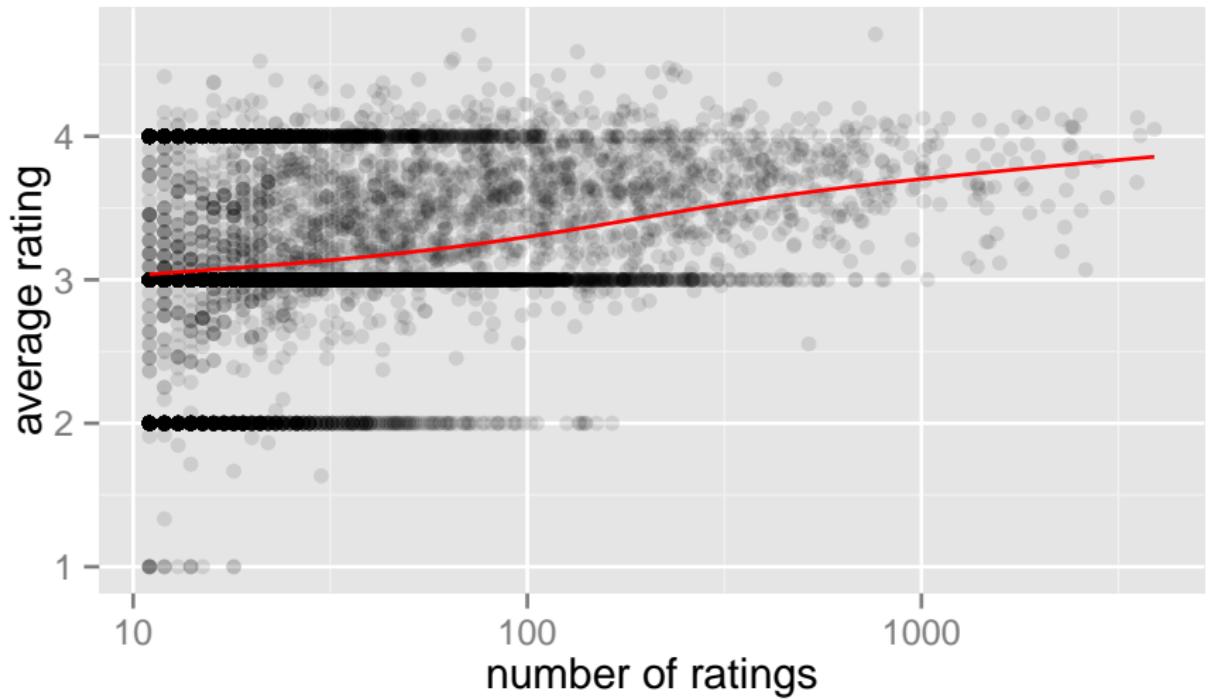
Twilight (Twilight, #1)

1984

Mockingjay (The Hunger Games, #3)

Harry Potter and the Deathly Hallows (Harry Potter, #7)

# Books by average rating



## Highest and lowest rated books

The Book of Mormon (4.71)

The Complete Calvin and Hobbes (4.70)

Calvin and Hobbes (4.59)

Chrysanthemum (4.54)

Jimmy Corrigan, the Smartest Kid on Earth (4.52)

...

Deception Point (Dan Brown novel)

Le Divorce (“worse than the movie”)

Splinter of the Mind’s Eye (Star Wars spinoff)

Star Wars: Outbound Flight

Fifty Shades of Grey (Fifty Shades, #1)

# LDA: Topics (Books as “words”)

## Classics

The Great Gatsby  
To Kill a Mockingbird  
The Catcher in the Rye  
1984  
Animal Farm  
Lord of the Flies

## Business Nonfiction

The Tipping Point  
Freakonomics  
Outliers: The Story of Success  
Steve Jobs  
The 7 Habits of Highly Effective People  
How to Win Friends and Influence People

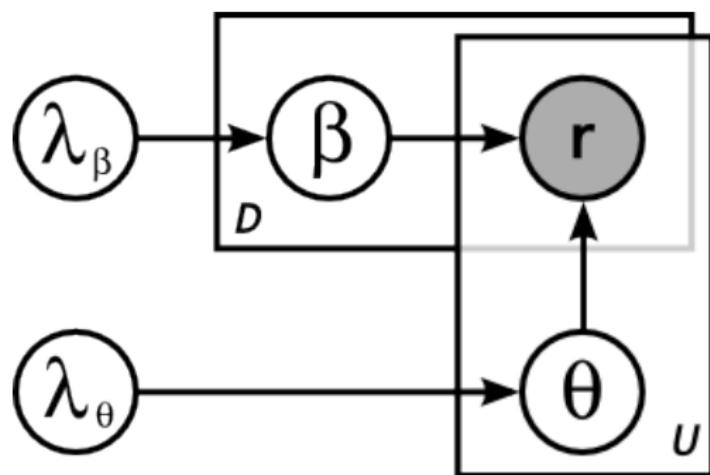
## Children's Books

Corduroy  
If You Give a Mouse a Cookie  
The Very Hungry Caterpillar  
Where the Wild Things Are  
How the Grinch Stole Christmas  
Goodnight Moon

## Religious Nonfiction

Overwhelmed by a Relentless God  
Mere Christianity  
The Reason for God  
The Ragamuffin Gospel  
Searching for God Knows What  
The Practice of the Presence of God

# Matrix Factorization: Graphical Model



# Poisson Matrix Factorization

User preferences:

$$\theta_u \sim \text{Gamma}(a, b) \quad (1)$$

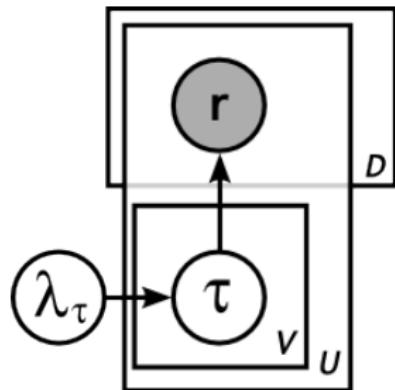
Item characteristics:

$$\beta_i \sim \text{Gamma}(c, d) \quad (2)$$

Ratings:

$$r_{ui} \sim \text{Poisson} \left( \theta_u^\top \beta_i \right) \quad (3)$$

# Basic Trust Poisson Model



# Basic Trust Poisson Model

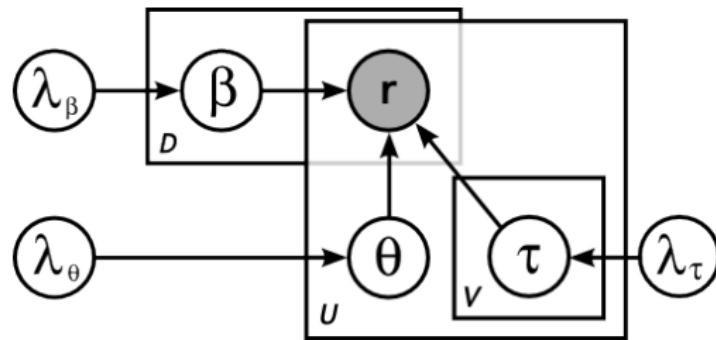
User trust (how much user  $u$  trusts neighboring user  $v$ ):

$$\tau_{uv} \sim \text{Gamma}(f, g) \quad (4)$$

Ratings:

$$r_{ui} \sim \text{Poisson} \left( \sum_{v \in N(u,i)} \tau_{uv} r_{vi} \right) \quad (5)$$

# Combining Matrix Factorization and Friend Trust



# Combining Matrix Factorization and Friend Trust

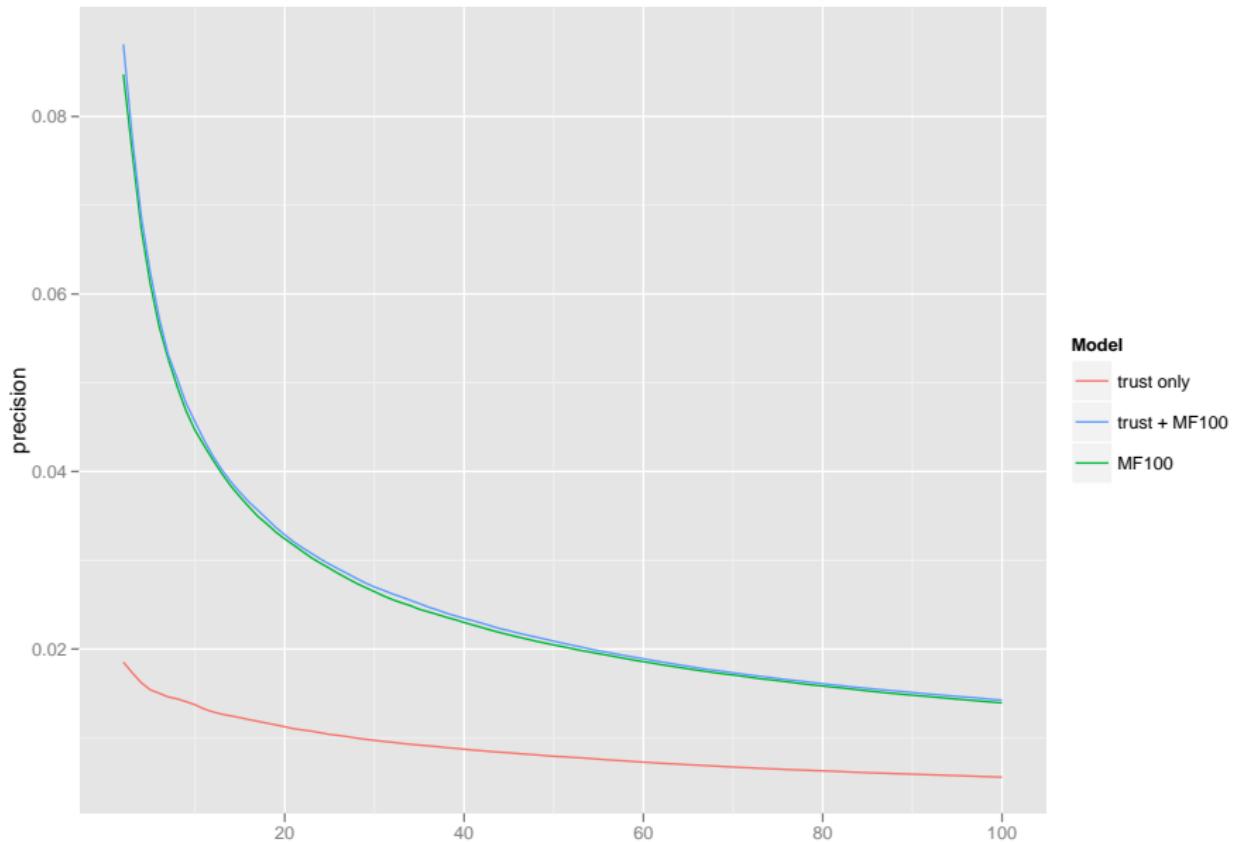
Ratings:

$$r_{ui} \sim \text{Poisson} \left( \theta_u^\top \beta_i + \sum_{v \in N(u,i)} \tau_{uv} r_{vi} \right) \quad (6)$$

## Measuring Performance with Precision

$$\text{precision} = \frac{\# \text{ of true positives}}{\# \text{ true positives} + \# \text{ false positives}}$$

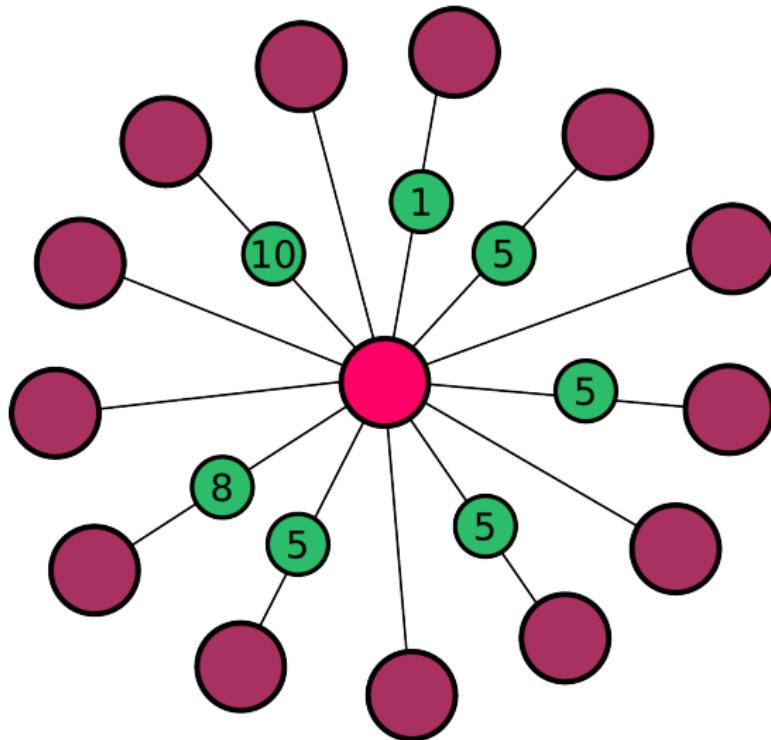
# Measuring Performance with Precision



# Measuring Performance with Precision

threshold	1	25	100
trust only	1.85%	1.03%	0.56%
MF100 only	8.47%	2.85%	1.39%
trust + MF100	8.80%	2.90%	1.42%

# Exploring Performance



# Exploring Performance

## 5★ Heldout Books

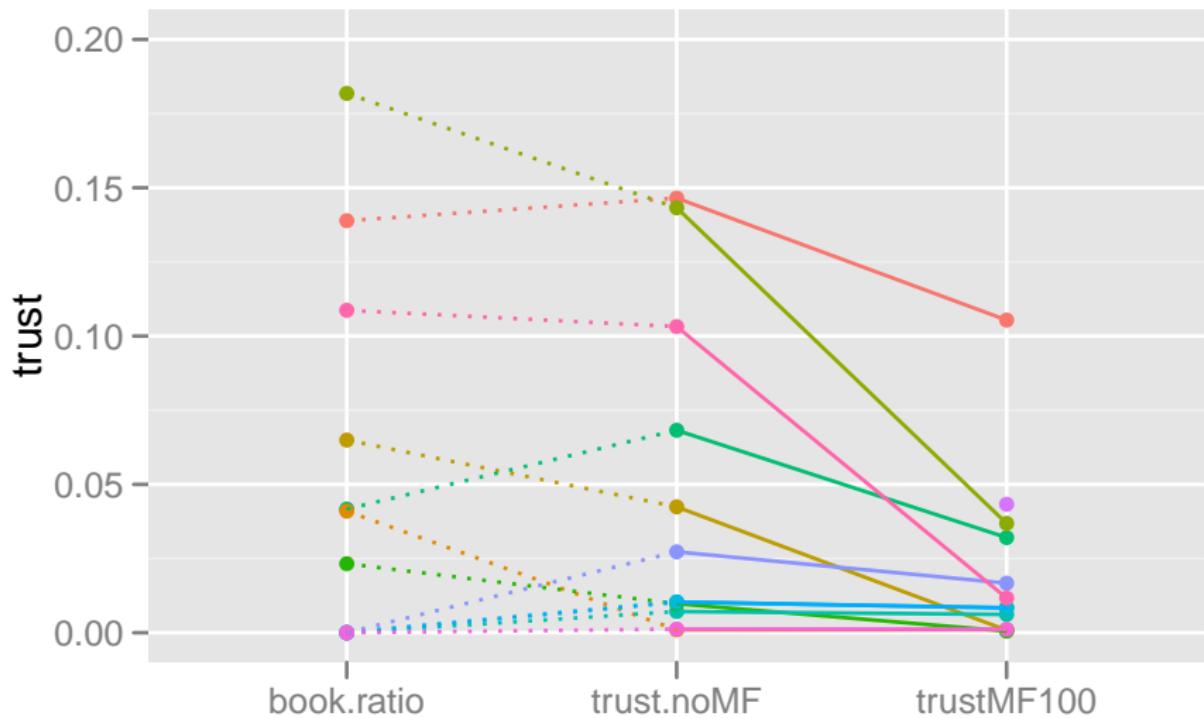
- Crime and Punishment
- Will You Please Be Quiet, Please?
- Mrs. Dalloway
- To the Lighthouse
- Powers Definitive Collection Vol. 2
- The Collected Poems of Frank...
- My Life in Heavy Metal: Stories
- Scalped, Vol. 3: Dead Mothers

## Recommendations

- Slaughterhouse-Five
- The Amazing Adventures of ...
- Breakfast of Champions
- The Wind-Up Bird Chronicle
- Lolita
- The Road
- Kafka on the Shore
- A Heartbreaking Work of St...

# Exploring Performance

user 108651's trust



# 5★ Books of Friends

## Most Trusted Friend

Disgrace  
Sleepless Nights  
Harry Potter #7  
The Catcher in the Rye  
Gilead

## Less Trusted Friend

Freakonomics  
Middlesex  
Slaughterhouse-Five  
Bluebook Uniform System Citation  
Cannery Row

# Redundancy

## Less Trusted Friend

Freakonomics

Middlesex

Slaughterhouse-Five

Bluebook Uniform System Citation

Cannery Row

## Topic 12

Slaughterhouse-Five

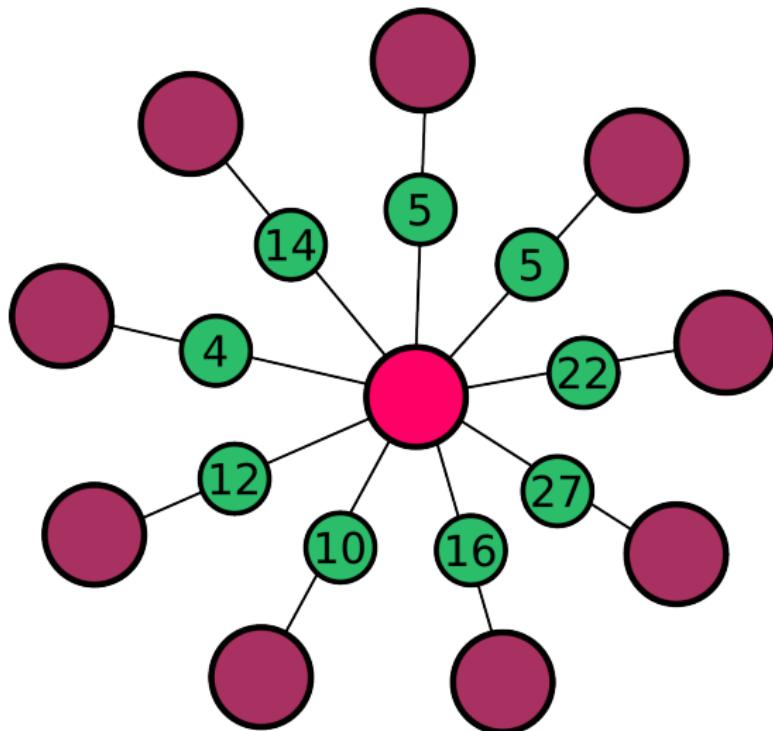
The Wind-Up Bird Chronicle

Lolita

The Road

Kafka on the Shore

# Exploring Performance



# Exploring Performance

## 5★ Training Books

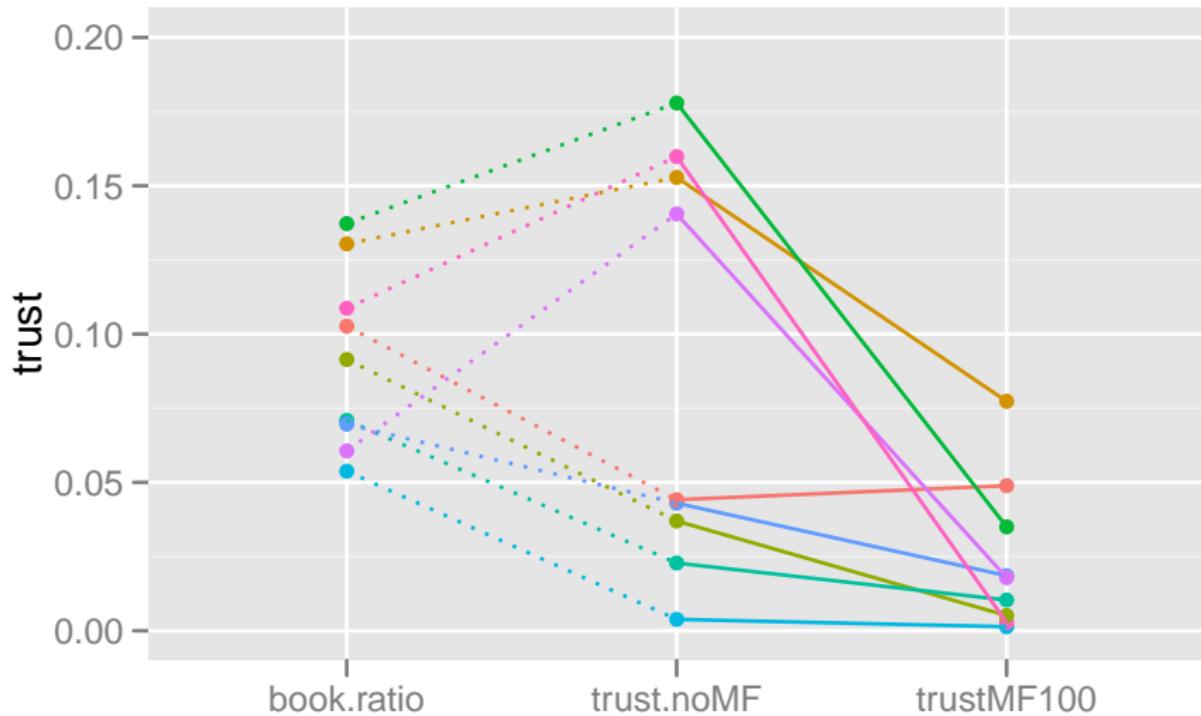
Othello  
Maus, Vol 1: My Father Bleeds...  
The Complete Calvin and Hobbes  
Catch-22  
The Adventures of Huckleberry...  
The Hitchhiker's Guide to the...  
Euclid's Elements  
The Principia: Mathematical...

## Recommendations

To Kill a Mockingbird  
The Great Gatsby  
1984  
Harry Potter and the Sorc...  
The Catcher in the Rye  
The Hobbit  
Animal Farm  
Lord of the Flies

# Exploring Performance

user 8496451's trust



# 5★ Books of Friends

## Most Trusted Friend

Stranger in a Strange Land  
To Kill a Mockingbird  
Persepolis 2: The Story of a Return  
The Bluest Eye  
I Know Why the Caged Bird Sings  
Savage Inequalities: Children in...

## Less Trusted Friend

Great Expectations  
To Kill a Mockingbird  
The Poetry of Pablo Neruda  
1984  
The Electric Kool-Aid Acid Test  
King Lear

# Redundancy

## Less Trusted Friend

Great Expectations  
To Kill a Mockingbird  
The Poetry of Pablo Neruda  
1984  
The Electric Kool-Aid Acid Test  
King Lear

## Topic 50

To Kill a Mockingbird  
The Great Gatsby  
The Catcher in the Rye  
1984  
Animal Farm  
Lord of the Flies

# Conclusions

## Contributions

- improves on basic matrix factorization

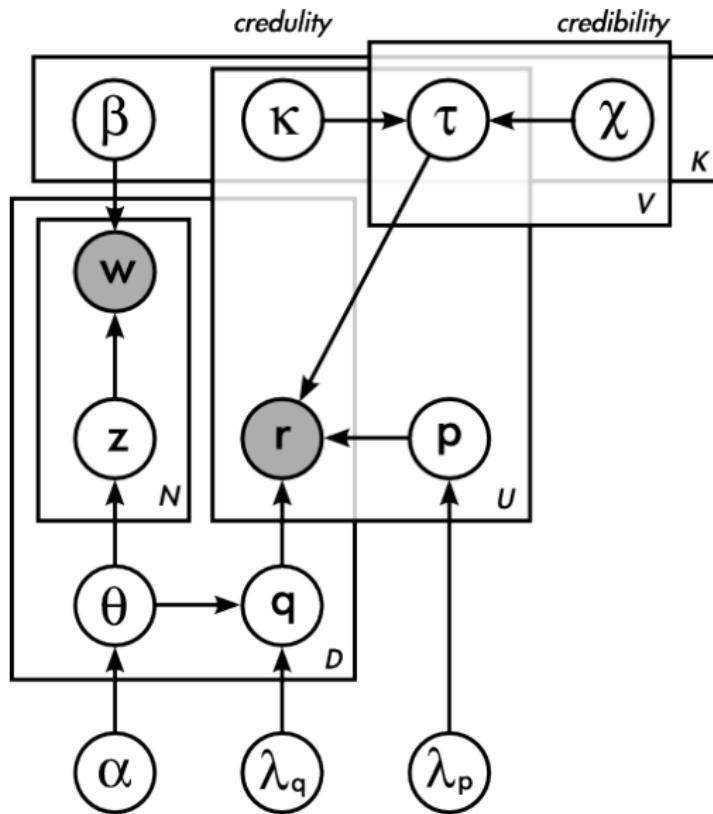
- model captures some notion of trust which can be used to explore data and contextualize recommendations

## Areas for Improvement

- weighting between MF and trust components

- not down-weighting trust for friends easily characterized by MF

# Hierarchical and Topical Trust



# Future Work

## User Studies

How do users respond to predictions? (model validation)

Understand the distributions of ratings (opt-in vs. must-rate-all)

Follow the feedback loop of ratings and recommendations

How to best present recommendations? ("because [friend] liked it")

## Misc

fine tuning: starred vs. binary data, parameter sweep

comparison to SoRec

security properties of these models

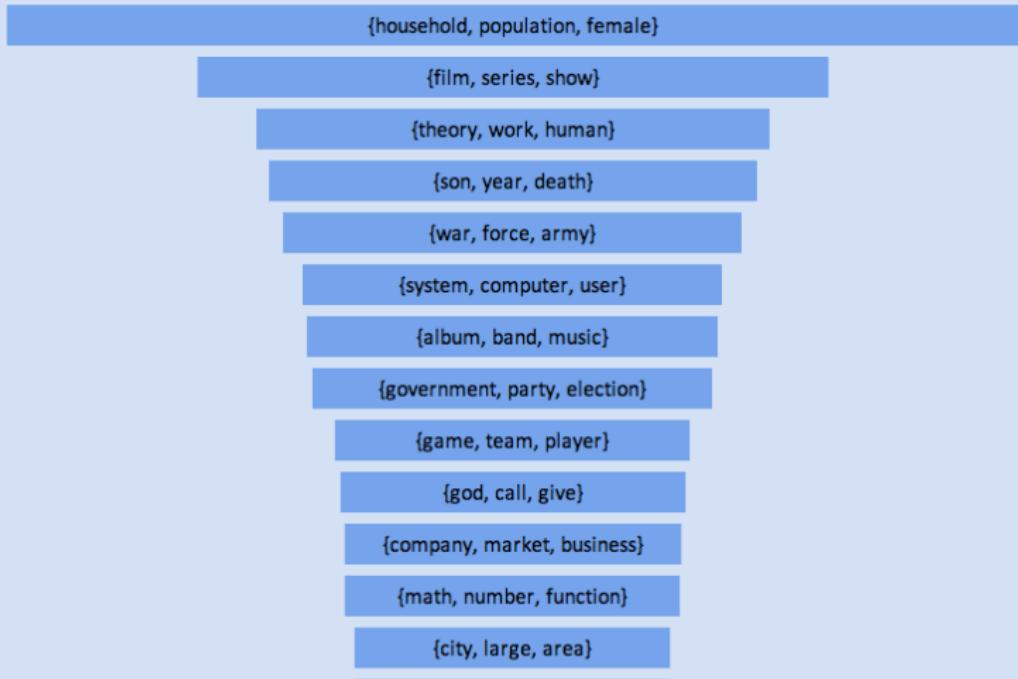
# Topic Model Visualization

Browse ▾ Topics Documents Terms About

Search ▾

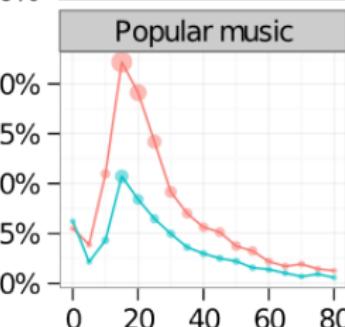
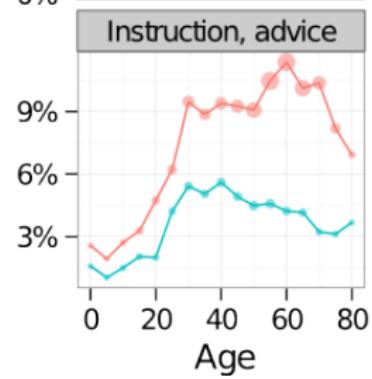
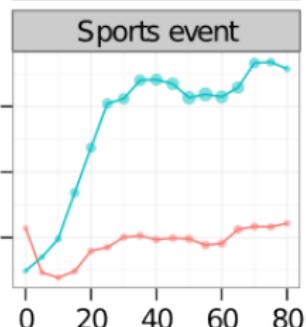
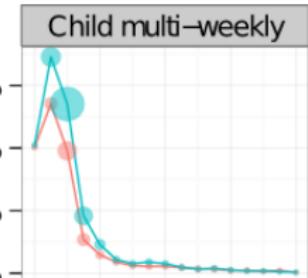
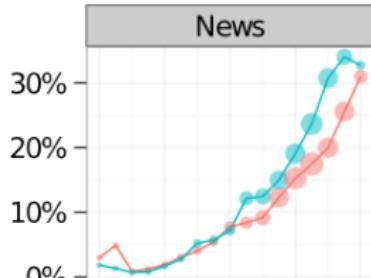
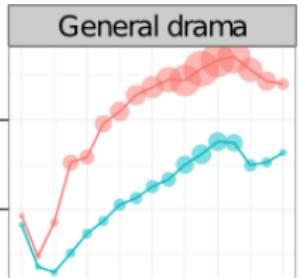
## Wikipedia Topics

*Relative Presence of Topics in all Documents*

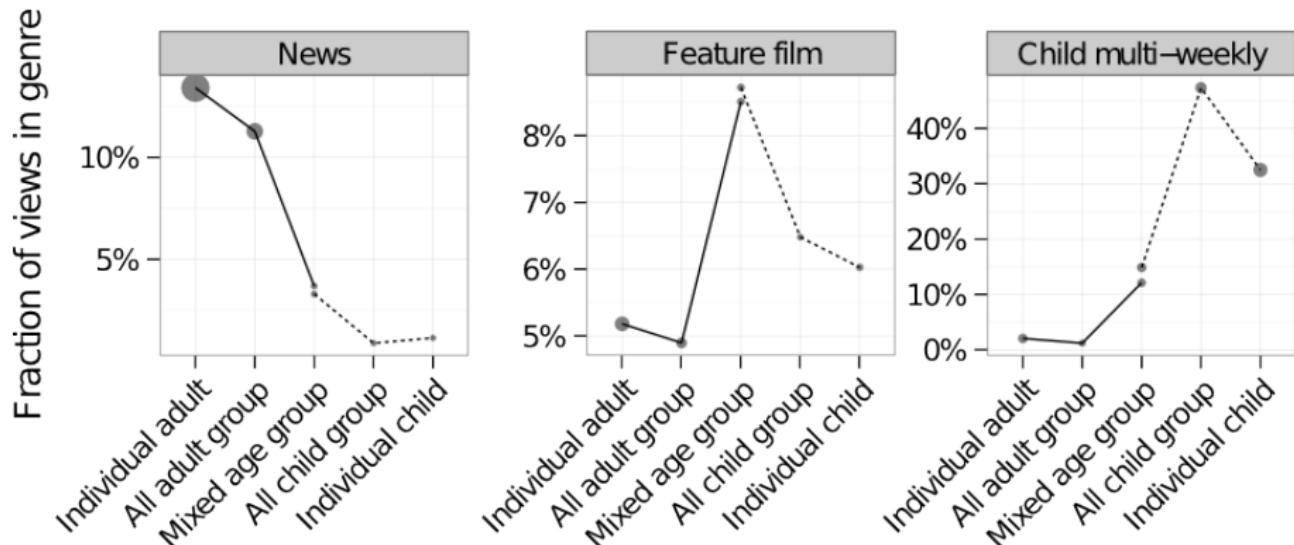


# A Large-scale Exploration of Group Viewing Patterns

Fraction of views in genre



# A Large-scale Exploration of Group Viewing Patterns



# Questions?

Thank you David Blei, Prem Gopalan, and Jake Hofman

# Generative model

User preferences:

$$\theta_u \sim \text{Gamma}(a, b) \quad (7)$$

Item characteristics:

$$\beta_i \sim \text{Gamma}(c, d) \quad (8)$$

User trust (how much user  $u$  trusts neighboring user  $v$ ):

$$\tau_{uv} \sim \text{Gamma}(f, g) \quad (9)$$

$$r_{ui} \sim \text{Poisson} \left( \theta_u^\top \beta_i + \sum_{v \in N(u,i)} \tau_{uv} r_{vi} \right) \quad (10)$$

Helper latent variables  $r_{ui} = \sum_n z_{uin}$  where  
 $n = \{n_1^{MF}, \dots, n_K^{MF}, n_1^{trust}, \dots, n_V^{trust}\}$ :

$$z_{uik}^{MF} \sim \text{Poisson}(\theta_{uk} \beta_{ik}) \quad (11)$$

$$z_{uiv}^{trust} \sim \text{Poisson}(\tau_{uv} r_{vi}) \quad (12)$$

# Complete Conditionals

$$\theta_{uk} \mid \beta, \tau, z, m, r \sim \text{Gamma}(a + \sum_i z_{uik}, b + \sum_i \beta_{ik}) \quad (13)$$

$$\beta_{ik} \mid \theta, \tau, z, m, r \sim \text{Gamma}(c + \sum_u z_{uik}, d + \sum_u \theta_{uk}) \quad (14)$$

$$\tau_{uv} \mid \theta, \beta, z, m, r \sim \text{Gamma}\left(f + \sum_{i \in D(u,v)} z_{uiv}, g + \sum_{i \in D(v)} r_{vi}\right) \quad (15)$$

$$z_{ui} \mid \theta, \beta, \tau, m, r \sim \text{Mult}(r_{ui}, \phi) \quad (16)$$

where

$$\phi \propto \langle \theta_{u1}\beta_{i1}, \dots, \theta_{uK}\beta_{iK}, \tau_{u1}r_{1i}, \dots, \tau_{uV}r_{Vi} \rangle \quad (17)$$

## Variational updates

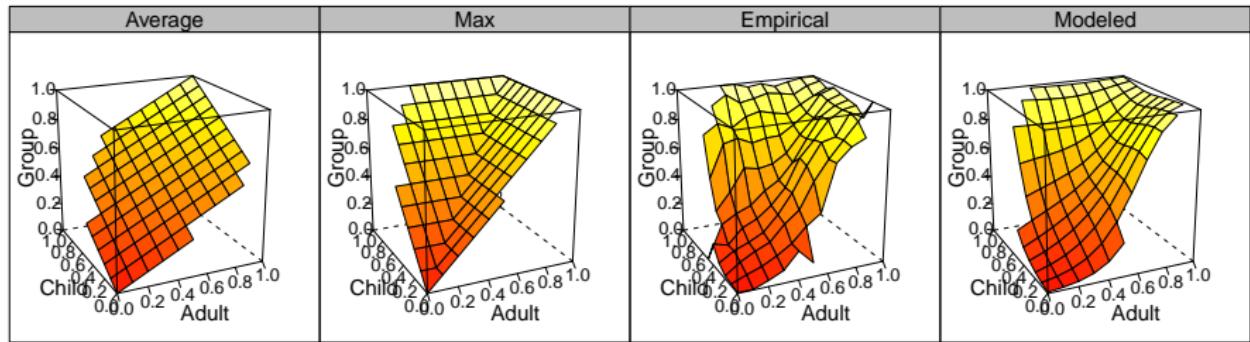
$$\gamma_{uk} = \left\langle a + \sum_i r_{ui} \phi_{uik}, \ b + \sum_i \lambda_{ik}^{\text{shp}} / \lambda_{ik}^{\text{ rte}} \right\rangle \quad (18)$$

$$\lambda_{ik} = \left\langle c + \sum_u r_{ui} \phi_{uik}, \ d + \sum_u \gamma_{uk}^{\text{shp}} / \gamma_{uk}^{\text{ rte}} \right\rangle \quad (19)$$

$$\nu_{uv} = \left\langle f + \sum_{i \in D(u,v)} r_{ui} \phi_{uiv}, \ g + \sum_{i \in D(v)} r_{vi} \right\rangle \quad (20)$$

$$\begin{aligned} \phi_{ui} \propto \exp \{ & \langle \Psi(\gamma_{u1}^{\text{shp}}) - \log \gamma_{u1}^{\text{ rte}} + \Psi(\lambda_{u1}^{\text{shp}}) - \log \lambda_{u1}^{\text{ rte}}, \\ & \dots \\ & \Psi(\gamma_{uK}^{\text{shp}}) - \log \gamma_{uK}^{\text{ rte}} + \Psi(\lambda_{uK}^{\text{shp}}) - \log \lambda_{uK}^{\text{ rte}}, \\ & \Psi(\nu_{u1}^{\text{shp}}) - \log \nu_{u1}^{\text{ rte}} + \log r_{1i}, \\ & \dots \\ & \Psi(\nu_{uV}^{\text{shp}}) - \log \nu_{uV}^{\text{ rte}} + \log r_{Vi} \rangle \} \end{aligned} \quad (21)$$

# A Large-scale Exploration of Group Viewing Patterns



# Friends and Books

