Exploiting content, structure, and communities in recommender systems

Julian McAuley, Stanford University

(with Jure Leskovec, Dan Jurafsky, and Hima Lakkaraju)

Overview

What are the aspects that contribute to users' ratings?	ICDM 2012 (w/ Leskovec & Jurafsky)
Can we discover these aspects automatically?	RecSys 2013 (w/ Leskovec)
How do users, ratings, and reviews evolve over time?	WWW 2013 (w/ Leskovec)

Overview...time permitting

Can we apply these ideas to social media submissions (e.g. on reddit.com)?	ICWSM 2013 (w/ Lakkaraju & Leskovec)
Can we recommend communities to users (e.g. circles on Google+)?	TKDD 2013 & NIPS 2012 (w/ Leskovec)

Item recommendation: recap

products/items

ratings

users

$$
\begin{bmatrix}
\cdot & 4 & \cdot & \cdot & 3 & \cdot \\
\cdot & \cdot & 2 & 3 & 1 & \cdot \\
\cdot & \cdot & \cdot & \cdot & 5 & \cdot \\
1 & 2 & \cdot & \cdot & \cdot & \cdot \\
3 & \cdot & \cdot & \cdot & 4 & 3 \\
\cdot & 2 & \cdot & 5 & 1 & \cdot
\end{bmatrix}
$$

Goal: infer missing ratings, and use them to make predictions/recommendations

Item recommendation: recap

social media submissions

votes

communities

$$\begin{bmatrix} \cdot & 4132 & \cdot & \cdot & 312 & \cdot \\ \cdot & \cdot & -152 & 613 & 11 & \cdot \\ \cdot & \cdot & \cdot & \cdot & 5 & \cdot \\ -101 & 22 & \cdot & \cdot & \cdot & \cdot \\ 315 & \cdot & \cdot & \cdot & 41 & 38 \\ \cdot & 762 & \cdot & -5 & 12 & \cdot \end{bmatrix}$$

Goal: infer missing ratings, and use them to make predictions/recommendations

Low-dimensional representations

Our goal in all of these settings is to identify low-dimensional representations of items, users, articles, communities, etc.

We do this to model the output variables, e.g.

$$\text{rating}(\text{julian}, \text{Harry Potter}) \simeq [0.1, 0.3, 0.8] \cdot [0.2, 0.5, 0.7]$$

My interest in special effects

Quality of HP's special effects

Low-dimensional representations

Our goal in all of these settings is to identify low-dimensional representations of items, users, articles, communities, etc.

We do this to model the output variables, e.g.

$$\text{rating}(\text{julian}, \text{Harry Potter}) \simeq [0.1, 0.3, 0.8] \cdot [0.2, 0.5, 0.7]$$

My interest in special effects

Quality of HP's special effects

But we also want our models to be interpretable, by using textual, temporal, and social information

What are the aspects that contribute to users' ratings?

Learning attitudes and attributes from multi-aspect reviews

McAuley, Jurafsky & Leskovec, ICDM 2012

(link)

"Aspects" on wikipedia

Jay-Z

From Wikipedia, the free encyclopedia

Shawn Corey Carter (born December 4, 1969),[1] better known by his stage name **Jay-Z**, is an American rapper, record producer, entrepreneur, and occasional actor. He is one of the most financially successful hip hop artists and entrepreneurs in America. In 2012, *Forbes* estimated Carter's net worth at nearly $500 million.[2][3] He has sold approximately 50 million albums worldwide, while receiving fourteen Grammy Awards for his musical work, and numerous additional nominations.[4][5] He is consistently ranked as one of the greatest rappers of all-time. He was ranked #1 by MTV in their list of *The Greatest MCs of All-Time* in 2006.[6] Two of his albums, *Reasonable Doubt* (1996) and *The Blueprint* (2001) are considered landmarks in the genre with both of them being ranked in *Rolling Stone* magazine's list of the 500 greatest albums of all time.[7][8] *Blender* included the former on their 500 CDs You Must Own Before You Die.[9]

Jay-Z

Rate this page

What's this?

View page ratings

Trustworthy	Objective	Complete	Well-written
★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★	★ ★ ★ ★ ★

☐ I am highly knowledgeable about this topic (optional)

Submit ratings

"Aspects" on wikipedia

Jay-Z

From Wikipedia, the free encyclopedia

Shawn Corey Carter (born December 4, 1969),[1] better known by his stage name **Jay-Z**, is an American rapper, record producer, entrepreneur, and occasional actor. He is one of the most financially successful hip hop artists and entrepreneurs in America. In 2012, *Forbes* estimated Carter's net worth at nearly $500 million.[2][3] He has sold approximately 50 million albums worldwide, while receiving fourteen Grammy Awards for his musical work, and numerous additional nominations.[4][5] He is consistently ranked as one of the greatest rappers of all-time. He was ranked #1 by MTV in their list of *The Greatest MCs of All-Time* in 2006.[6] Two of his albums, *Reasonable Doubt* (1996) and *The Blueprint* (2001) are considered landmarks in the genre with both of them being ranked in *Rolling Stone* magazine's list of the 500 greatest albums of all time.[7][8] *Blender* included the former on their 500 CDs You Must Own Before You Die.[9]

Jay-Z

Rate this page

What's this?

View page ratings

Trustworthy	Objective	Complete	Well-written
★★★★★	★★★★★	★★★★★	★★★★★

☐ I am highly knowledgeable about this topic (optional)

Your ratings have not been submitted yet

Submit ratings

"Aspects" on wikipedia

Jay-Z

From Wikipedia, the free encyclopedia

Shawn Corey Carter (born December 4, 1969),[1] better known by his stage name **Jay-Z**, is an American rapper, record producer, entrepreneur, and occasional actor. He is one of the most financially successful hip hop artists and entrepreneurs in America. In 2012, *Forbes* estimated Carter's net worth at nearly $500 million.[2][3] He has sold approximately 50 million albums worldwide, while receiving fourteen Grammy Awards for his musical work, and numerous additional nominations.[4][5] He is consistently ranked as one of the greatest rappers of all-time. He was ranked #1 by MTV in their list of *The Greatest MCs of All-Time* in 2006.[6] Two of his albums, *Reasonable Doubt* (1996) and *The Blueprint* (2001) are considered landmarks in the genre with both of them being ranked in *Rolling Stone* magazine's list of the 500 greatest albums of all time.[7][8] *Blender* included the former on their 500 CDs You Must Own Before You Die.[9]

Jay-Z

Rate this page

What's this?

View page ratings

? Trustworthy ? Objective ? Complete ? Well-written

☐ I am highly knowledgeable about this topic (optional)

Your ratings have not been submitted yet

Submit ratings

Aspects in online reviews

‘Partridge in a Pear Tree’, brewed by ‘The Bruery’

Dark brown with a light tan head, minimal lace and low retention. Excellent aroma of dark fruit, plum, raisin and red grape with light vanilla, oak, caramel and toffee. Medium thick body with low carbonation. Flavor has strong brown sugar and molasses from the start over bready yeast and a dark fruit and plum finish. Minimal alcohol presence. Actually, this is a nice quad.

Feel: 4.5 Look: 4 Smell: 4.5 Taste: 4 Overall: 4

Aspects in online reviews

‘Partridge in a Pear Tree’, brewed by ‘The Bruery’

Dark brown with a light tan head, minimal lace and low retention. Excellent aroma of dark fruit, plum, raisin and red grape with light vanilla, oak, caramel and toffee. Medium thick body with low carbonation. Flavor has strong brown sugar and molasses from the start over bready yeast and a dark fruit and plum finish. Minimal alcohol presence. Actually, this is a nice quad.

Feel: 4.5 Look: 4 Smell: 4.5 Taste: 4 Overall: 4

Dataset	Aspects	#Reviews
beer (beeradvocate)	feel, look, smell, taste, overall	1.6M
beer (ratebeer)	feel, look, smell, taste, overall	2.9M
pubs (beeradvocate)	food, price, quality, selection, service, vibe	18K
toys & games (amazon)	durability, educational, fun, overall	374K
audio books (audible)	author, narrator, overall	10K

Segmenting reviews into aspects

'Partridge in a Pear Tree', brewed by 'The Bruery'

Dark brown with a light tan head, minimal lace and low retention. Excellent aroma of dark fruit, plum, raisin and red grape with light vanilla, oak, caramel and toffee. Medium thick body with low carbonation.. Flavor has strong brown sugar and molasses from the start over bready yeast and a dark fruit and plum finish. Minimal alcohol presence. Actually, this is a nice quad.

Feel: 4.5 Look: 4 Smell: 4.5 Taste: 4 Overall: 4

Goal: to segment reviews using rating data, and use the segmented text to better summarize reviews and recommend products

Probabilistic model of aspects in reviews

$$P(\text{aspect}(s) = k | \text{sentence } s, \text{rating } v) \propto$$

$$\exp \sum_{w \in s} \left\{ \underbrace{\theta_{kw}}_{\text{aspect weights}} + \underbrace{\phi_{kv_k w}}_{\text{sentiment weights}} \right\}$$

The model separates the probability into:

1. Words that depend on the aspect, but **not** the sentiment
2. Words that depend on the aspect **and** the sentiment

Model fitting

Repeat steps (1) and (2) until convergence:

$$\arg\max_{\theta,\phi} \underbrace{l^{(\theta,\phi)}(\text{labels}|\text{sentences},\text{ratings})}_{\text{corpus likelihood}} - \underbrace{\Omega(\theta,\phi)}_{\text{regularizer}}$$

(solved via gradient ascent using L-BFGS)

Step 1:
fit the model parameters by maximum likelihood

$$\arg\max_{\text{labels}} \underbrace{l^{(\theta,\phi)}(\text{labels}|\text{sentences},\text{ratings})}_{\text{corpus likelihood}}$$

(solved via linear assignment)

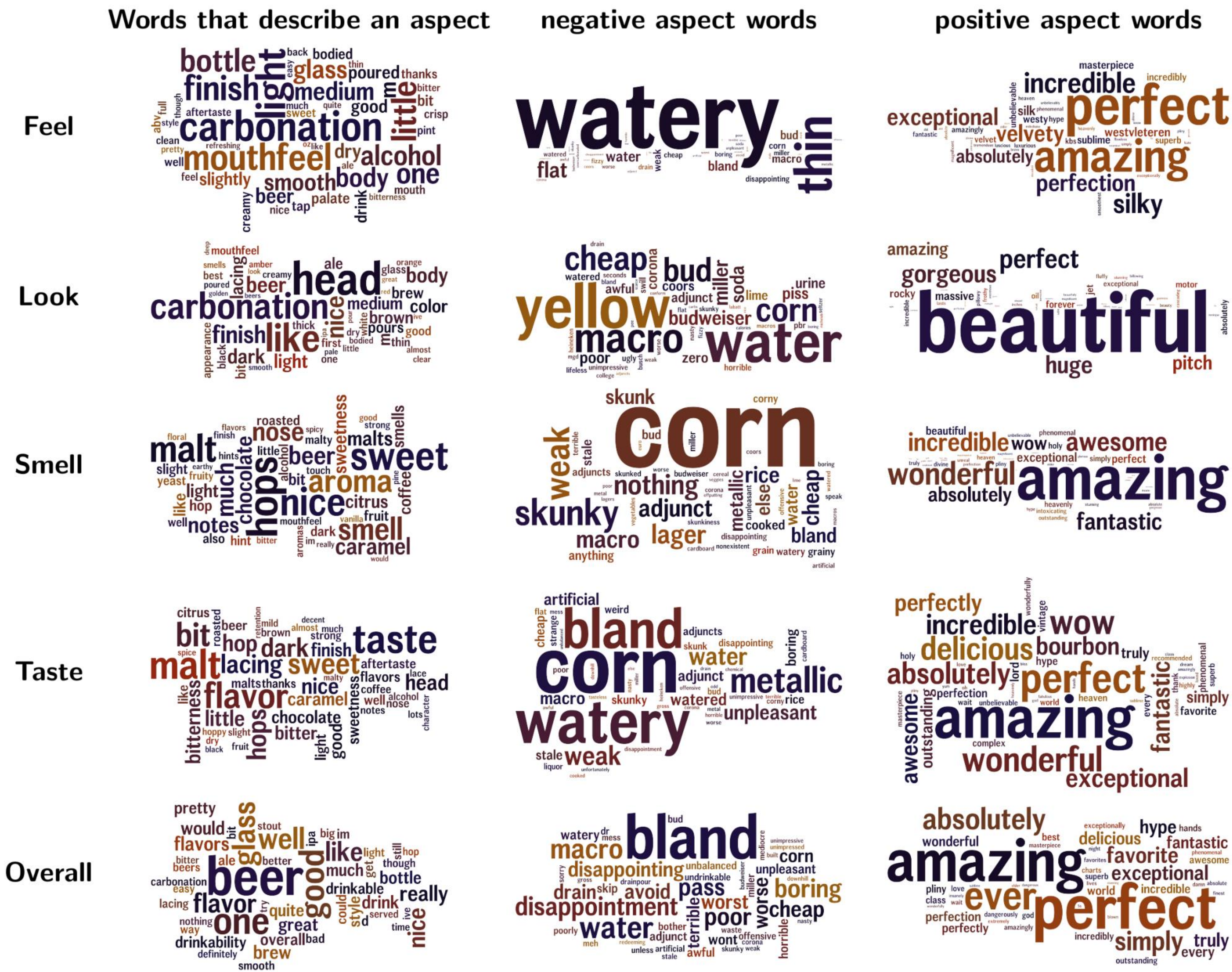
Step 2:
choose the most likely aspect for each sentence

Results

1. Sentence labels predicted by the algorithm have accuracy close to human performance (80% vs. 93% on beer data)

2. Summarization (choosing representative sentences for each aspect) is even more accurate (85% on beer data)

3. Rating completion (inferring aspect ratings from overall ratings+reviews) beats fully-supervised alternatives



Words that describe an aspect
negative aspect words
positive aspect words
Feel
carbonation mouthfeel finish bottle light little medium glass body alcohol smooth one slightly beer palate
watery thin flat water weak bland macro
perfect amazing incredible exceptional velvety perfection silky absolutely
Look
carbonation head like finish dark light body brown nice color
yellow water macro cheap bud corn budweiser miller soda
beautiful gorgeous perfect amazing huge pitch
Smell
malt hops nice sweet aroma smell caramel chocolate notes nose
corn skunky weak nothing adjunct lager macro skunk bland cheap
amazing wonderful incredible awesome fantastic absolutely
Taste
malt taste flavor sweet bit hops dark little nice head chocolate
corn bland watery metallic weak water unpleasant artificial
amazing perfect wonderful delicious incredible absolutely wow bourbon exceptional fantastic
Overall
beer good one flavor glass well like would really nice
bland disappointment macro water disappointing boring pass poor cheap
amazing perfect ever absolutely simply favorite exceptional

Can we discover these aspects automatically?

Hidden factors as hidden topics: understanding rating dimensions with review text

McAuley & Leskovec, RecSys 2013

(link)

Online reviews

We have models for reviews with multiple ratings, but most online reviews aren't like that

What can with only a single rating?

Dataset	#Reviews
citysearch	53K
Yelp	230K
wine	1.57M
movies (amazon)	8.56M
books (amazon)	12.89M
all categories (amazon)	35.28M

Online reviews

We have models for reviews with multiple ratings, but most online reviews aren't like that

What can with only a single rating?

Dataset	#Reviews
citysearch	53K
Yelp	230K
wine	1.57M
movies (amazon)	8.56M
books (amazon)	12.89M
all categories (amazon)	35.28M

Can we discover the aspects (in reviews) that most influence user's ratings?

Model fitting

A 'standard' recommender system decomposes recommendations into user and item latent factors

$$rec(u,i) = \alpha + \beta_u + \beta_i + \gamma_u \cdot \gamma_i$$

which are fit so as to minimize the mean-squared error

$$\arg\min_{\alpha,\beta,\gamma} \frac{1}{|\mathcal{T}|} \sum_{r_{u,i} \in \mathcal{T}} (rec(u,i) - r_{u,i})^2 + \lambda \|\gamma\|_2^2$$

where $r_{u,i} \in \mathcal{T}$ is a training corpus of ratings

Model fitting

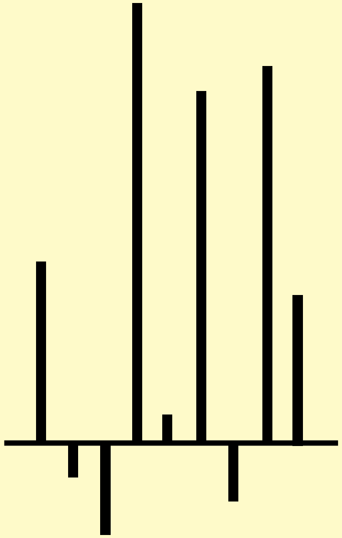
We replace this objective with one that uses the review text as a regularizer:

$$\frac{1}{|\mathcal{T}|}\sum_{r_{u,i}\in\mathcal{T}}\underbrace{(rec(u,i)-r_{u,i})^2}_{\text{rating error}}-\lambda\underbrace{l(\mathcal{T}|\Theta,\phi,z)}_{\text{corpus likelihood}}$$

Here the terms ϕ and z are word distributions and topic assignments, as with LDA

Fusing rating and topic parameters

item parameters



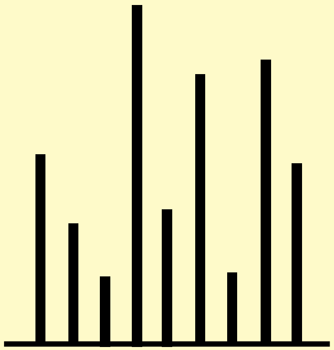
$\gamma_i \in \mathbb{R}^K$

Item parameters ostensibly represent **the extent to which items exhibit certain properties**

Fusing rating and topic parameters

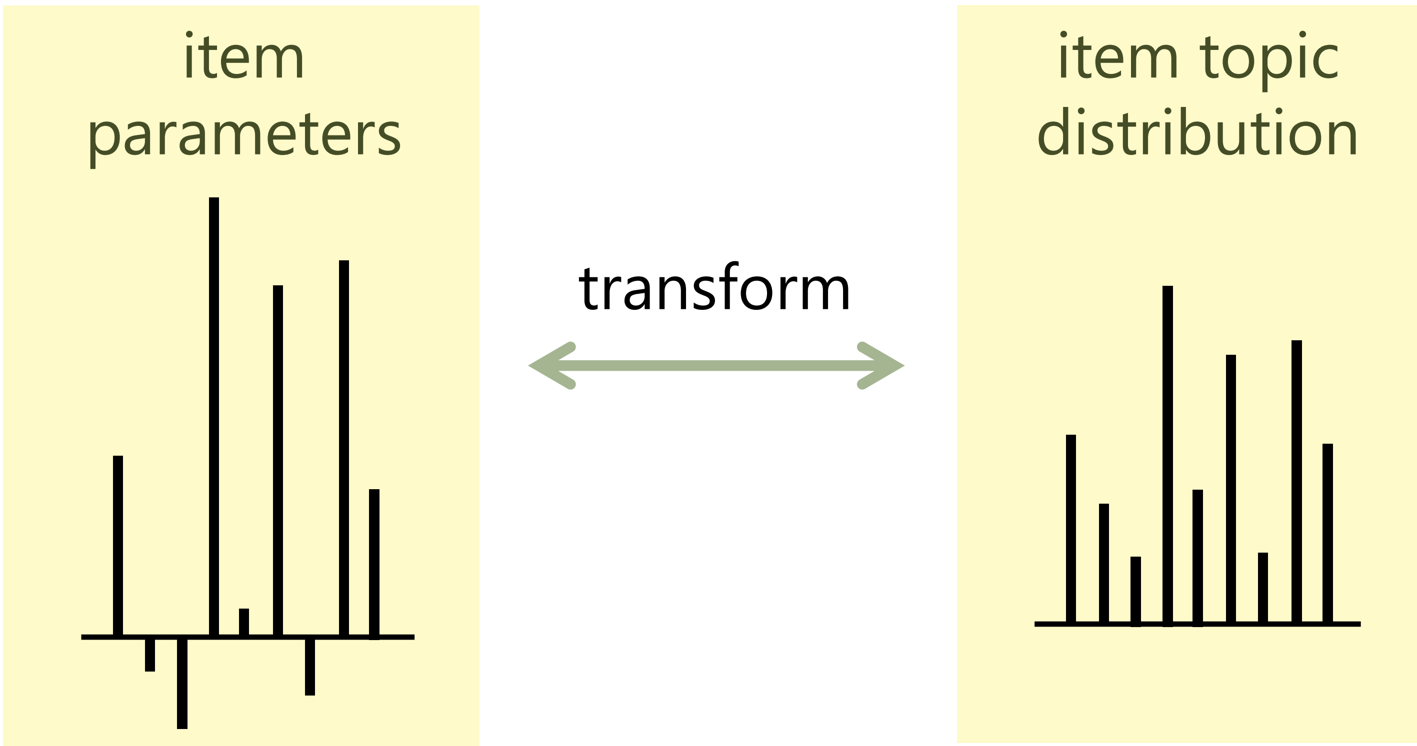
Topic distributions (e.g. in LDA) represent **the extent to which certain sets of words are used in a document**

item topic distribution



$$\theta_i \in \Delta^K \text{ (i.e., } \sum_k \theta_{i,k} = 1)$$

Fusing rating and topic parameters

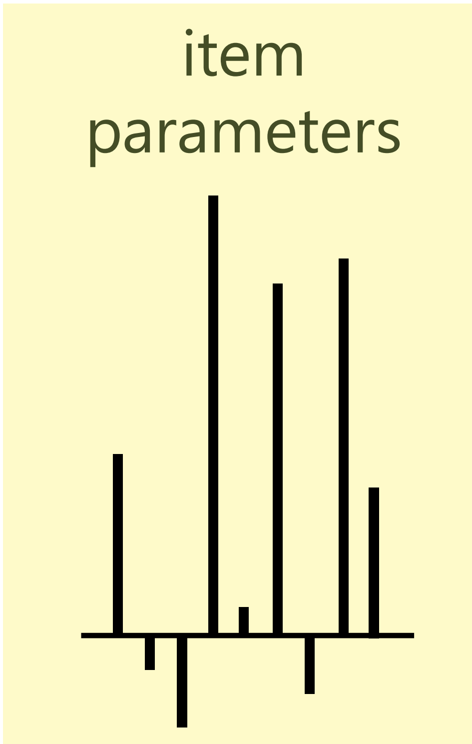


$\gamma_i \in \mathbb{R}^K$

$\theta_i \in \Delta^K \text{ (i.e., } \sum_k \theta_{i,k} = 1)$

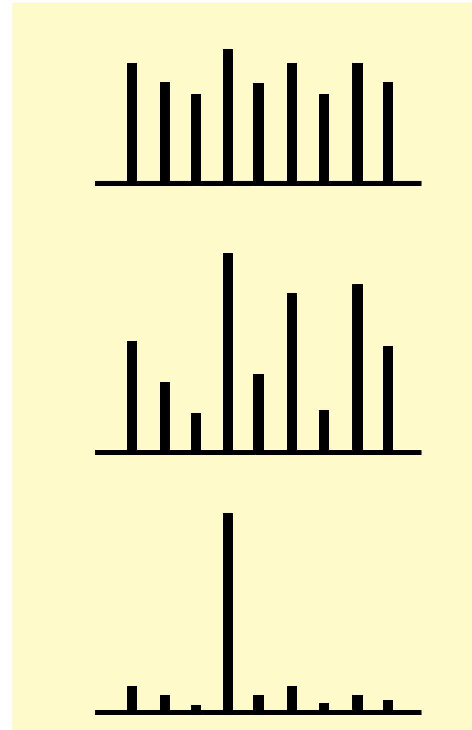
We need to identify a transform between item parameters (real vectors) and topics (stochastic vectors)

Fusing rating and topic parameters



transform

$$\theta_{i,k} = \frac{\exp(\kappa\gamma_{i,k})}{\sum_{k'} \exp(\kappa\gamma_{i,k})}$$



$\kappa \to 0$

$\kappa \to \infty$

$\gamma_i \in \mathbb{R}^K$

$\theta_i \in \Delta^K \text{ (i.e., } \sum_k \theta_{i,k} = 1)$

Model fitting

Repeat steps (1) and (2) until convergence:

$$\arg\min_{\Theta} \frac{1}{|\mathcal{T}|} \sum_{r_{u,i} \in \mathcal{T}} \underbrace{(rec(u,i) - r_{u,i})^2}_{\text{rating error}} - \mu \underbrace{l(\mathcal{T}|\Theta, \phi, z)}_{\text{corpus likelihood}}$$

(solved via gradient ascent using L-BFGS)

Step 1: minimize the MSE using gradient descent

$$\text{sample } z_{d,j} \text{ with probability } p(z_{d,j} = k) = \phi_{k,w_{d,j}}$$

(solved via gibbs sampling)

Step 2: sample topic assignments for each word

Results (selection)

Mean Squared Error on all datasets:

Dataset	offset	Latent factors	HFT (ours)	Improvement
Amazon	1.774	1.423	1.325	6.89%
Beer	0.521	0.371	0.366	1.50%
Wine	0.043	0.029	0.027	4.84%
Citysearch	2.022	1.873	1.731	7.56%
Yelp	1.488	1.272	1.224	3.78%

(link to complete results)

Topics - beer

pale ales	lambics	dark beers	spices	wheat beers
ipa	funk	chocolate	pumpkin	wheat
pine	brett	coffee	nutmeg	yellow
grapefruit	saison	black	corn	straw
citrus	vinegar	dark	cinnamon	pilsner
ipas	raspberry	roasted	pie	summer
piney	lambic	stout	cheap	pale
citrusy	barnyard	bourbon	bud	lager
floral	funky	tan	water	banana
hoppy	tart	porter	macro	coriander
dipa	raspberries	vanilla	adjunct	pils

Topics – musical instruments

drums	**strings**	**wind**	**mics**	**software**
cartridge	guitar	reeds	mic	software
sticks	violin	harmonica	microphone	interface
strings	strap	cream	stand	midi
snare	neck	reed	mics	windows
stylus	capo	harp	wireless	drivers
cymbals	tune	fog	microphones	inputs
mute	guitars	mouthpiece	condenser	usb
heads	picks	bruce	battery	computer
these	bridge	harmonicas	filter	mp3
daddario	tuner	harps	stands	program

Topics – video games

fantasy	nintendo	windows	ea/sports	accessories
fantasy	mario	sims	drm	cable
rpg	ds	flight	ea	controller
battle	nintendo	windows	spore	cables
tomb	psp	xp	creature	ps3
raider	wii	install	nba	batteries
final	gamecube	expansion	football	sonic
battles	memory	program	nhl	headset
starcraft	wrestling	software	basketball	wireless
characters	metroid	mac	madden	controllers
ff	smackdown	sim	hockey	component

Product category discovery

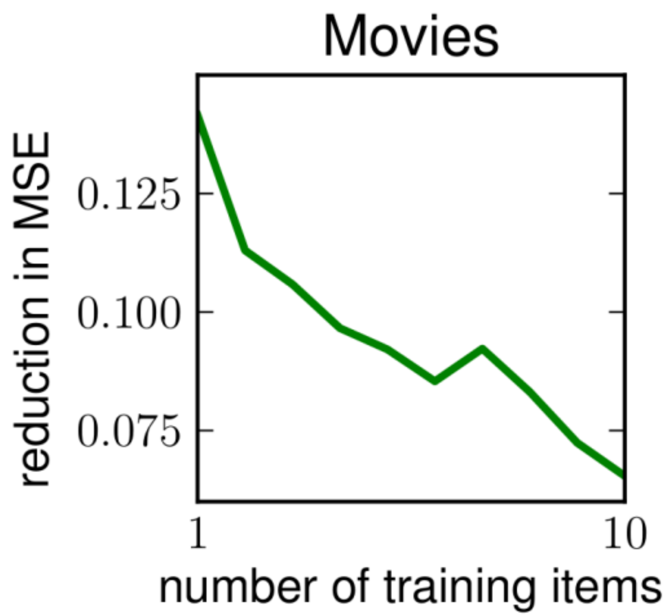
Let each product's 'category' be $c_i = \arg\max \gamma_{i,k}$

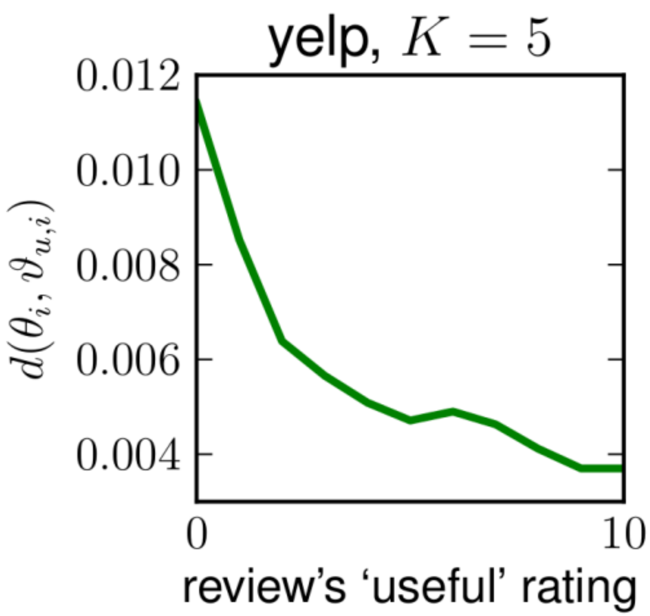
We report the F1 score between the predicted categories and the ground-truth

#topics	lat. factor model	LDA	HFT (ours)	improv. vs lat. factors	improv. vs LDA
5	0.166	0.205	0.412	148%	100%
10	0.097	0.169	0.256	163%	51%
20	0.066	0.091	0.165	151%	81%
50	0.042	0.047	0.199	369%	317%

(yelp businesses)

New reviewers, and good reviewers





We obtain the largest improvements for users/items with few reviews

'Useful' reviews are those that discuss each topic in proportion to its importance

How do users, ratings, and reviews evolve over time?

From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews

McAuley & Leskovec, WWW 2013

(skip section) (link)

Users and products evolve over time

Special effects that were good in 2003 may not be good in 2013

A child who likes Harry Potter in 2003 may have outgrown it by 2013

Even though *today's* children like Harry Potter, the children of 2023 may not

Users and products evolve over time

Special effects that were good in 2003 may not be good in 2013	Age of the product
A child who likes Harry Potter in 2003 may have outgrown it by 2013	Age (development) of the user
Even though *today's* children like Harry Potter, the children of 2023 may not	Age (zeitgest) of the community

Models of user and community evolution

Replace the 'standard' latent factor model

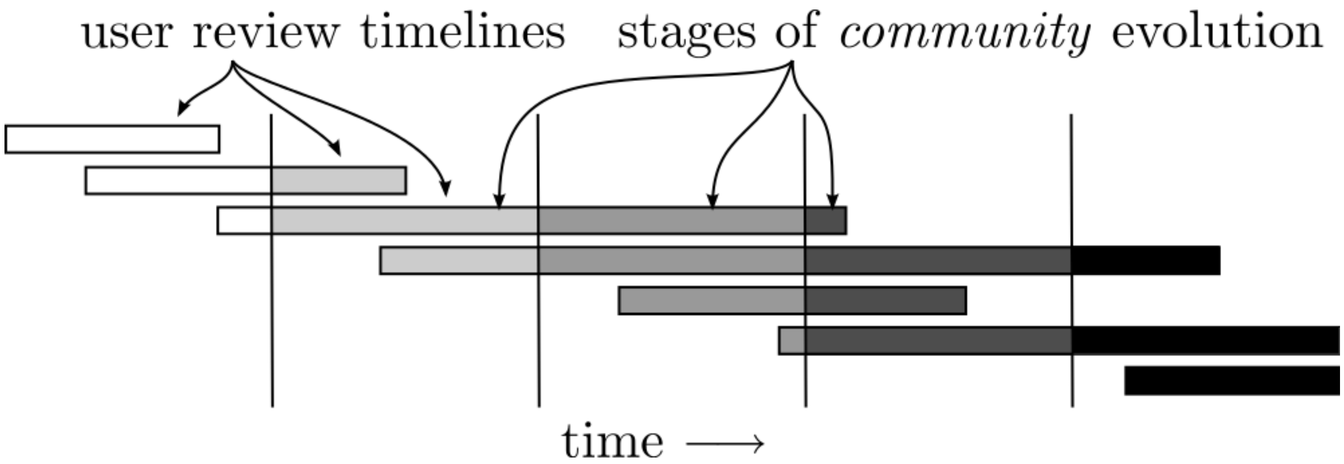
$$rec(u,i) = \alpha + \beta_u + \beta_i + \gamma_u \cdot \gamma_i$$

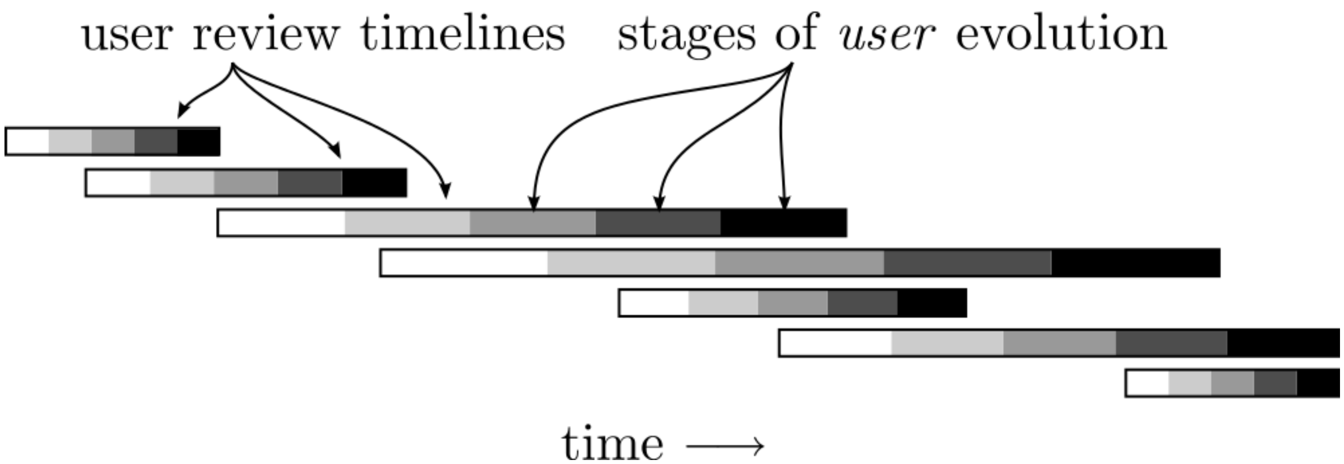
With one whose parameters depend upon the user's experience level (e):

$$rec_e(u,i) = \alpha(e) + \beta_u(e) + \beta_i(e) + \gamma_u(e) \cdot \gamma_i(e)$$

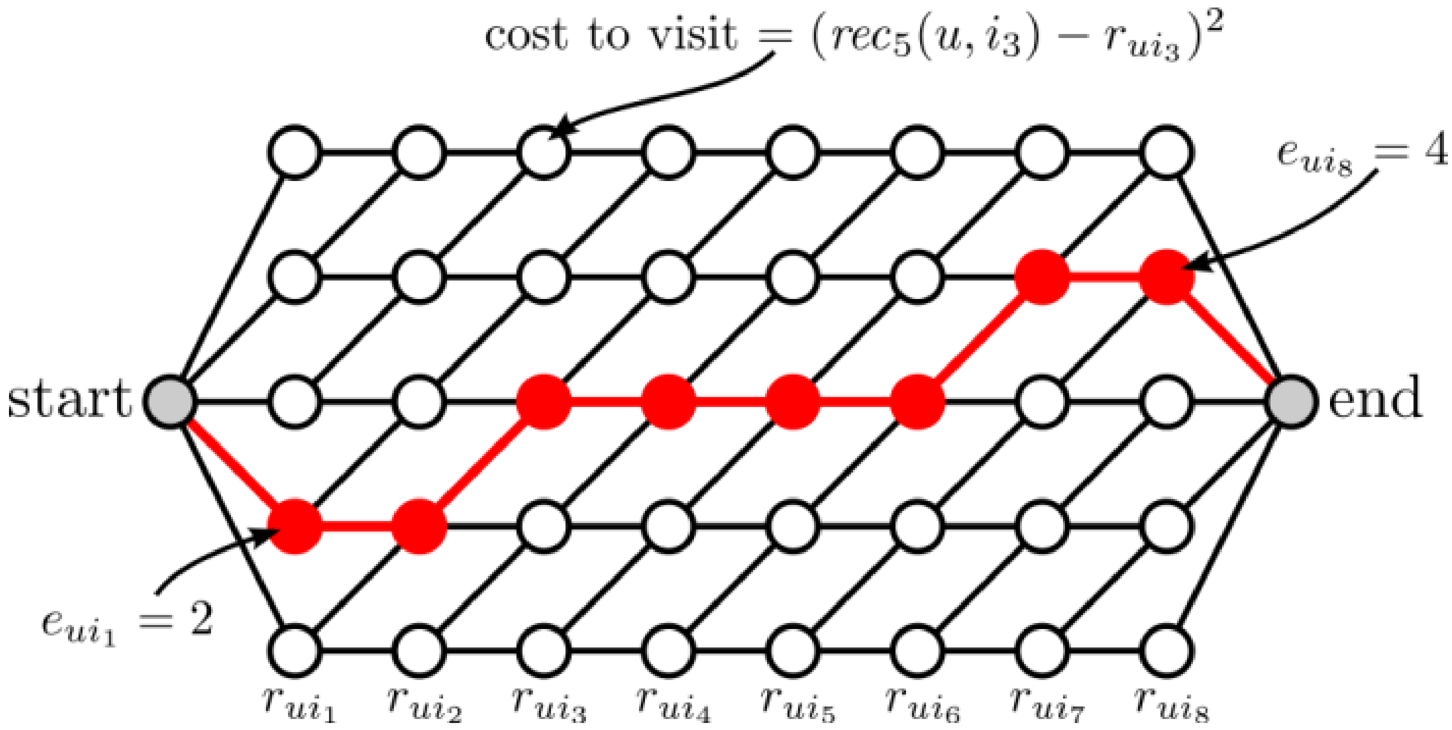
We must now fit users' experience levels, along with model parameters for each level

Models of user and community evolution





Models of user and community evolution



Since users gain experience monotonically, we can fit experience using Dynamic Programming

Model fitting

Repeat steps (1) and (2) until convergence:

$$\arg\min_{\Theta} \frac{1}{|\mathcal{T}|} \sum_{r_{u,i} \in \mathcal{T}} (rec_e(u,i) - r_{u,i})^2 + \Omega(\Theta)$$

(solved via gradient ascent using L-BFGS)

Step 1:
minimize the MSE using gradient descent

$$\arg\min_{\mathcal{E}} \frac{1}{|\mathcal{T}|} \sum_{r_{u,i} \in \mathcal{T}} (rec_e(u,i) - r_{u,i})^2$$

(solved using a DP)

Step 2:
fit experience levels so as to minimize the MSE

Results – rating prediction

We substantially outperform non-temporal models, and alternatives that model temporal information at the level of products or communities

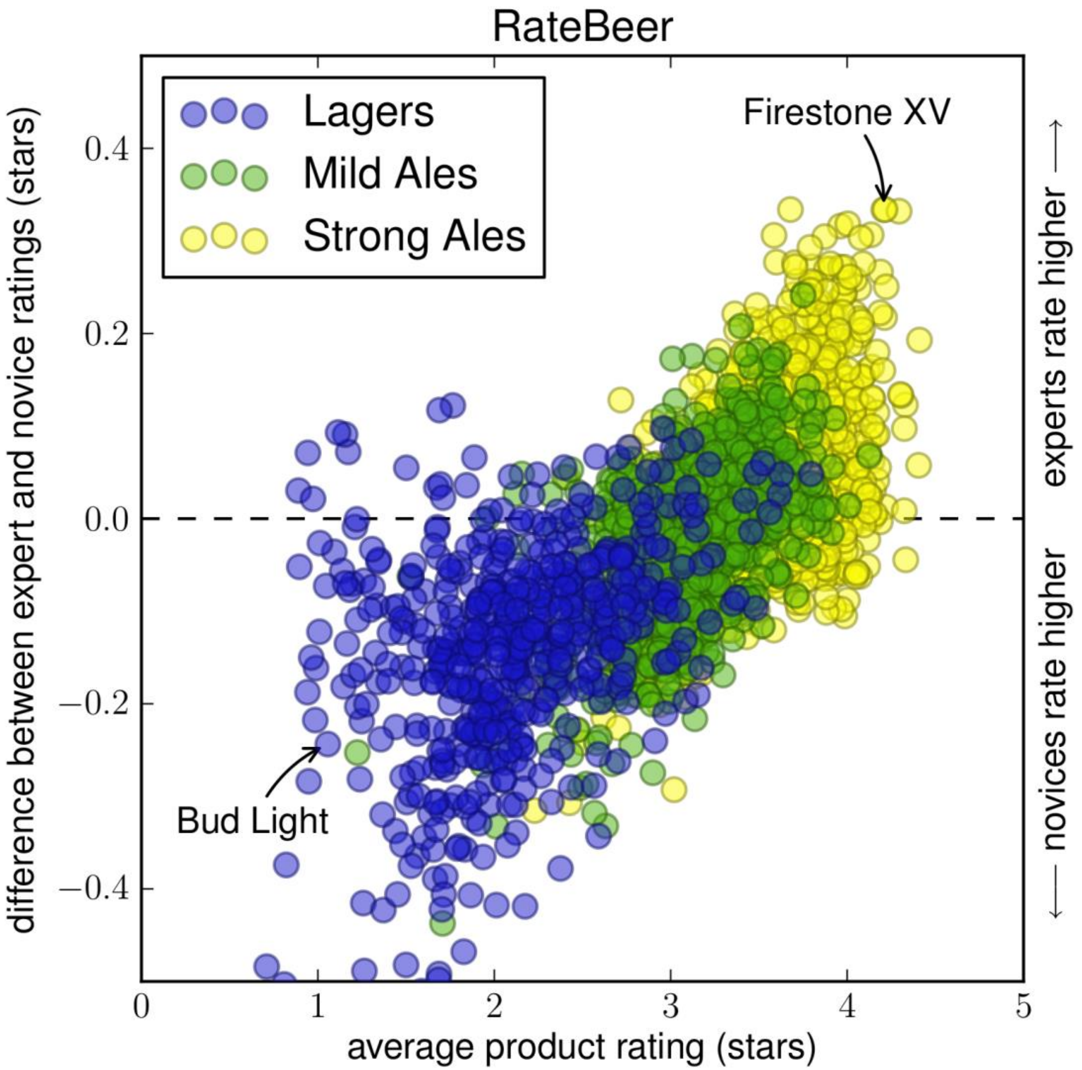
Dataset	Latent factors	community evolution	user evolution	Improvement (over cmty. evolution)
Beer	0.452	0.427	0.400	6.48%
Wine	0.055	0.051	0.045	13.20%
Movies (amazon)	1.379	1.371	1.051	23.34%
Gourmet food	1.582	1.529	1.475	3.53%

(users with 50 or more reviews)

How do beginners differ from experts?

Experts rate the top products more generously, and the bottom products more harshly

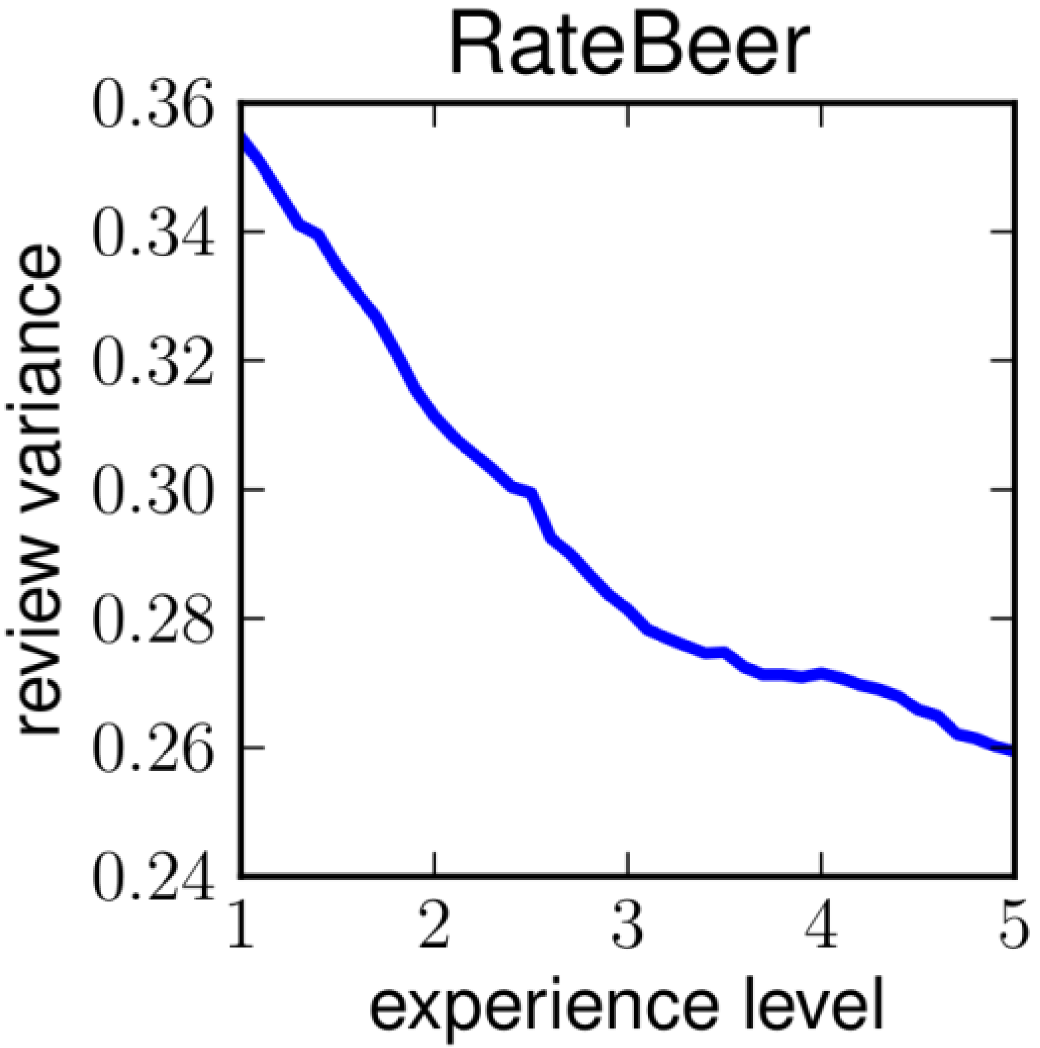
This phenomenon is highly correlated with product categories



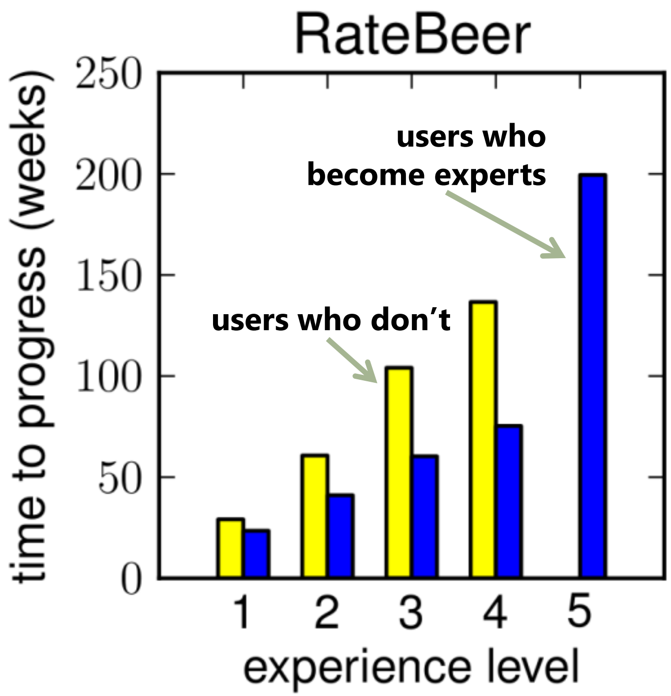
Are we really studying 'expertise'?

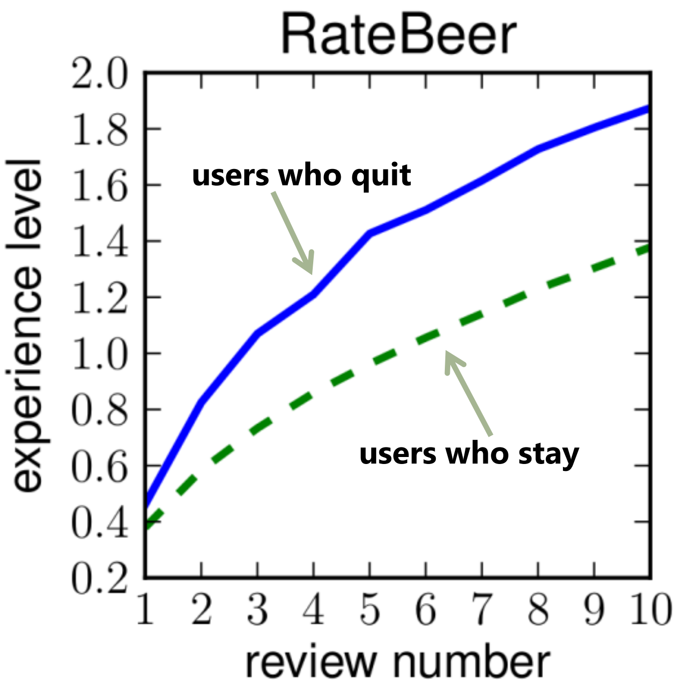
Experts are more predictable than beginners. They are also more inclined to agree with each other (right).

Predictability and agreement are arguably necessary conditions to define them as experts



How do beginners differ from experts?





Users who never eventually become experts progress more quickly

Users who gain expertise slowly are more likely to quit the community

Can we apply these ideas to social media submissions?

Understanding the interplay between titles, content, and communities in social media

Lakkaraju, McAuley & Leskovec, ICWSM 2013

(skip section) (link)

Resubmissions on reddit.com

When social media content is posted,
can we determine

How much of the success was due to the **content itself**

vs.

How much of the success was due to how the content was **marketed**

Resubmissions on reddit.com

When social media content is posted,
can we determine

How much of the success was due to the **content itself**	vs.	How much of the success was due to how the content was **marketed**

Why?
Changing how content is **presented** is easier than changing the content itself!

Resubmissions on reddit.com

[I'm not sure I quite understand this piece]
62 Submitted 2 years ago to pics by xxx
24 comments

[How wars are won]
20 Submitted 18 months ago to WTF by xxx
1 comment

[Murica!]
774 Submitted 1 year ago to funny by xxx
59 comments

[Bring it on England, Bring it on !!]
10 Submitted 10 months ago to pics by xxx
4 comments

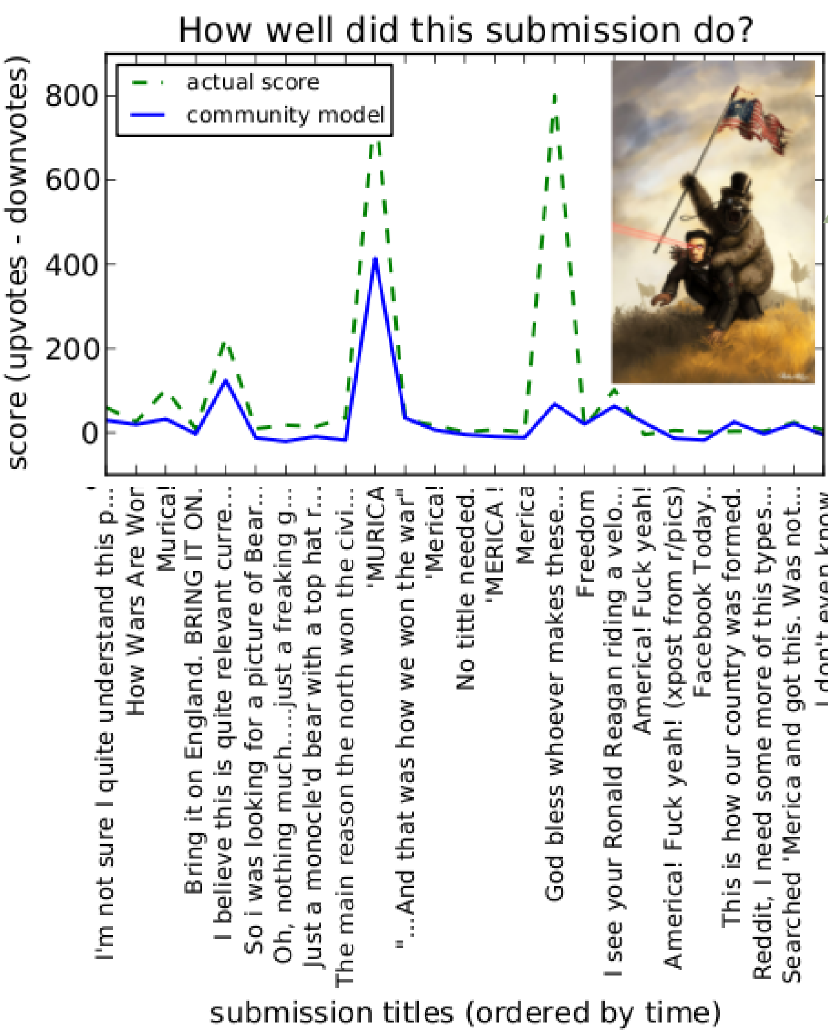
[I believe this is quite relevant currently]
226 Submitted 7 months ago to funny by xxx
15 comments

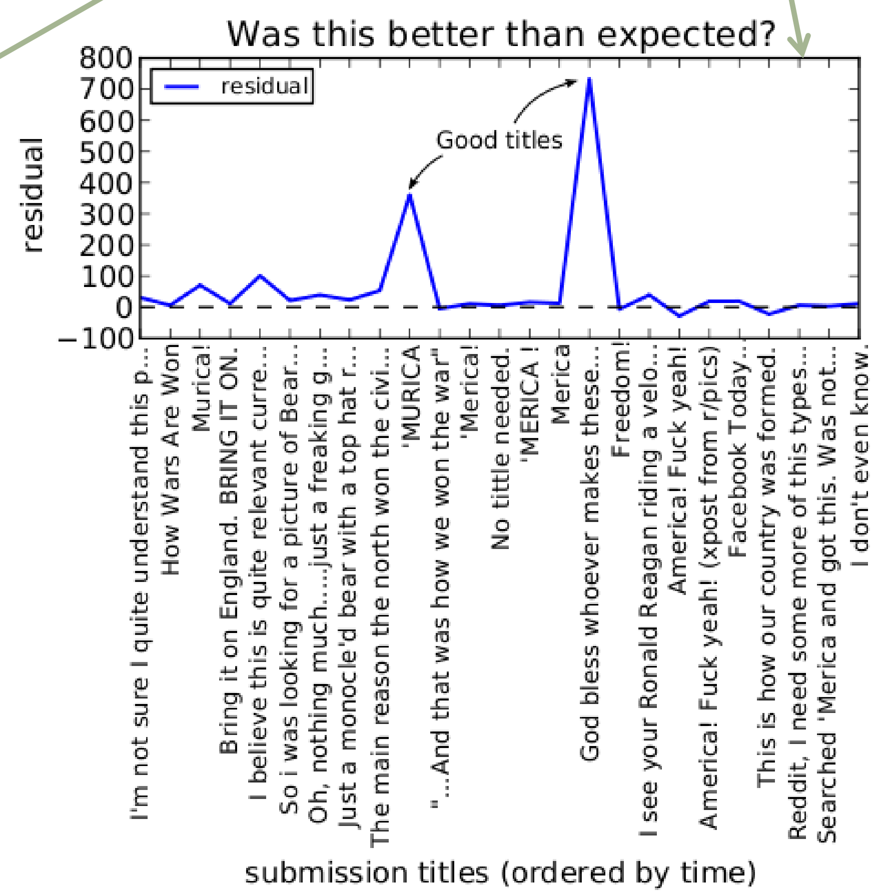
[God bless whoever makes these]
794 Submitted 1 month ago to funny by xxx
34 comments

Resubmissions on reddit.com

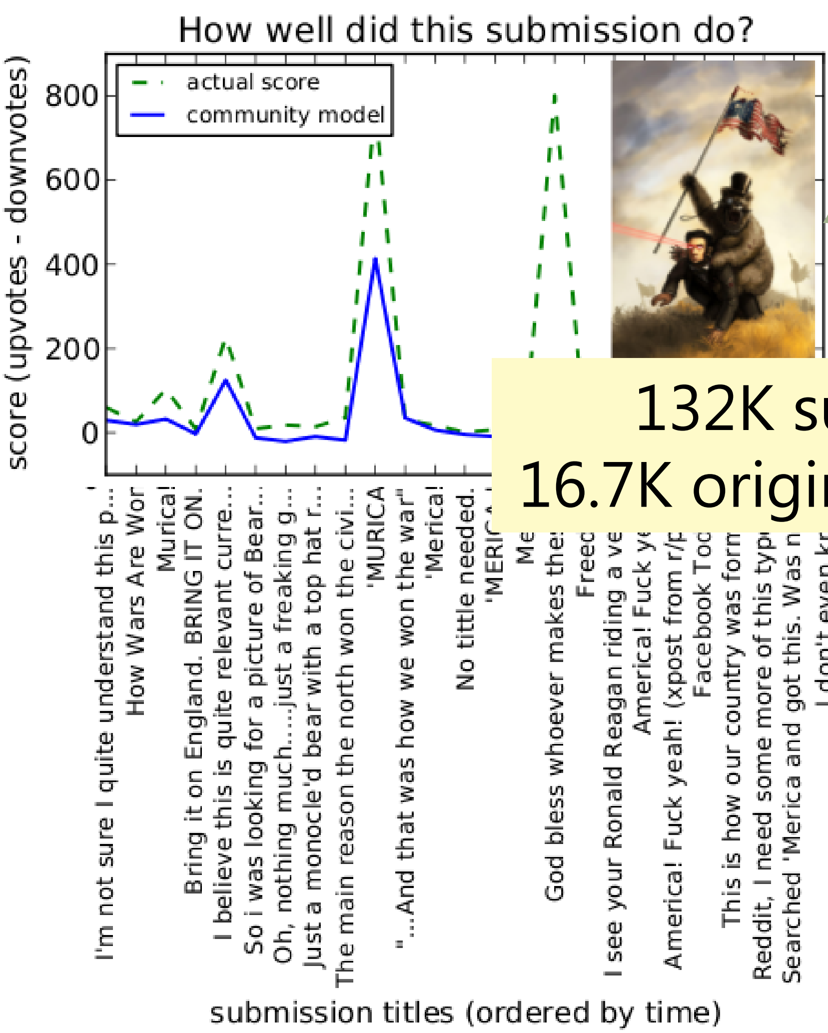
Community effects

Language effects





Resubmissions on reddit.com

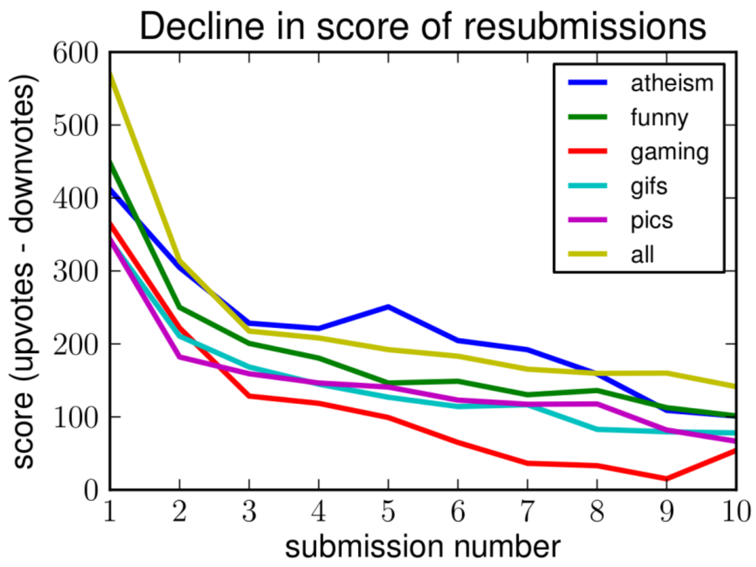


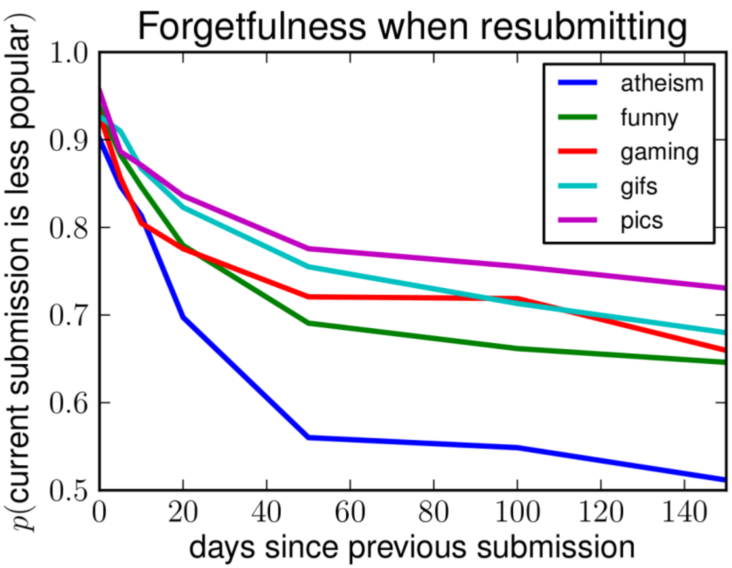
Community effects

Language effects

132K submissions,
16.7K original submissions

Temporal effects on reddit



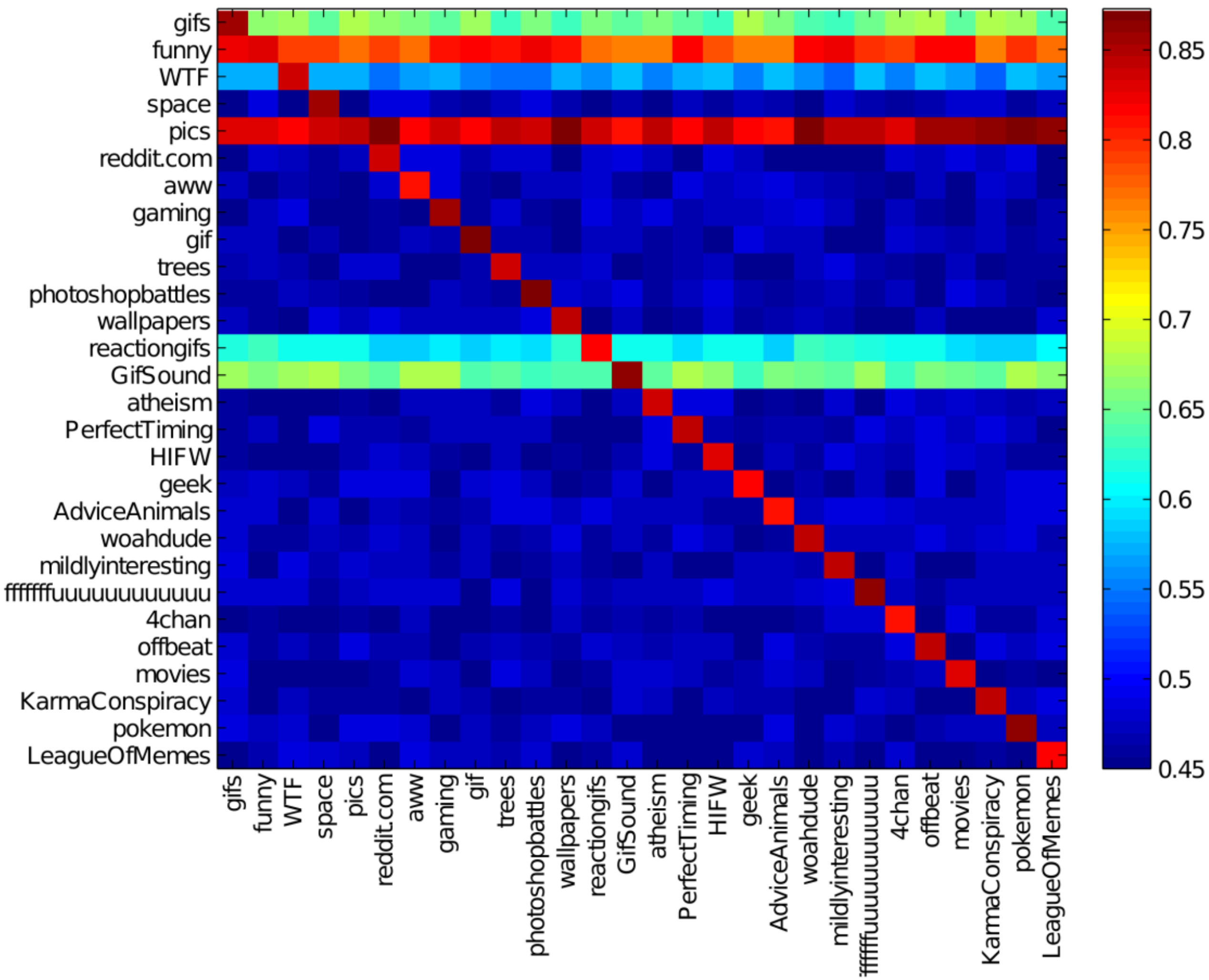


Resubmissions are less popular (left), but can still be popular if we wait long enough (right)

Inter-community temporal effects

Submissions won't be successful in the same community twice (main diagonal)

Submissions won't be successful if they already succeeded in a big community (low-rank structure)



Model (non-title effects)

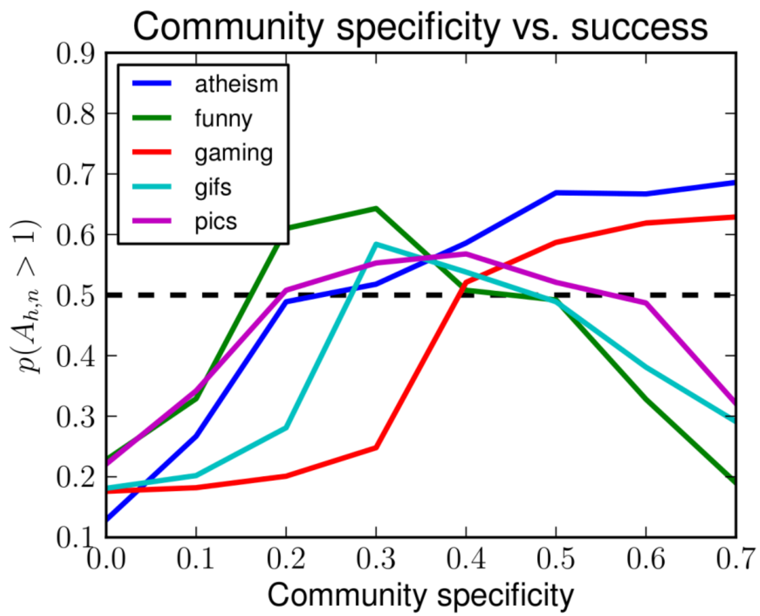
$$\hat{A}_{h,n} = \overbrace{\beta_h + \phi_h}^{\text{inherent popularity}} \underbrace{\exp}_{\text{decay from resubmissions}} \left\{ -\sum_{i=1}^{n-1} \overbrace{\frac{1}{\Delta_{i,n}^{h}}}^{\text{forgetfulness}} \underbrace{(\delta(c_{h,i} \neq c_{h,n})\lambda_{c_{h,i}}}_{\text{other communities}} + \overbrace{\delta(c_{h,i} = c_{h,n})\lambda'_{c_{h,i}})}^{\text{same community twice}} \underbrace{A_{h,i}}_{\text{previous submissions}} \right.$$

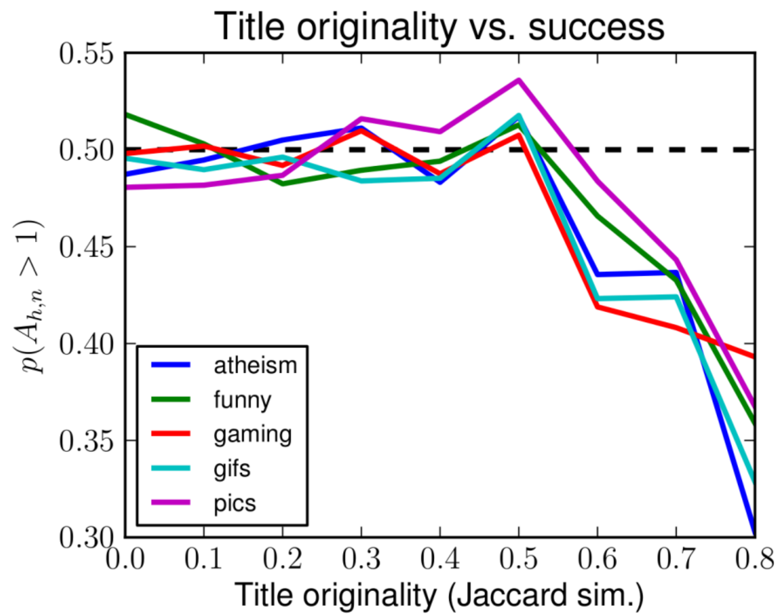
The model is designed to account for five factors:

1. The inherent popularity of the content (i.e., factors other than the title)
2. The decay in popularity due to resubmitting the content
3. This decay should be discounted for old enough submissions
4. A penalty due to resubmitting to another community
5. A penalty due to resubmitting to the same community twice

(we also account for other factors, such as the time of day etc.)

Model (title effects)





Titles should match others in the same community, but should not be too similar

Titles should differ from those previously used for the same content

Regression, and *in situ* evaluation

Performance on held-out test data:

Model	R^2
Community model only	0.528
Language model only	0.081
Community + language	0.618

Regression, and *in situ* evaluation

Performance on held-out test data:

Model	R^2
Community model only	0.528
Language model only	0.081
Community + language	0.618

We generated pairs of titles for 85 submissions, which we submitted simultaneously to two different communities

- The ‘good’ titles garnered three times as many upvotes as the ‘bad’ ones (10,959 vs. 3,438)
- Five good titles reached the front page of their community, and two reached the front page of r/all

Can we recommend communities to users?

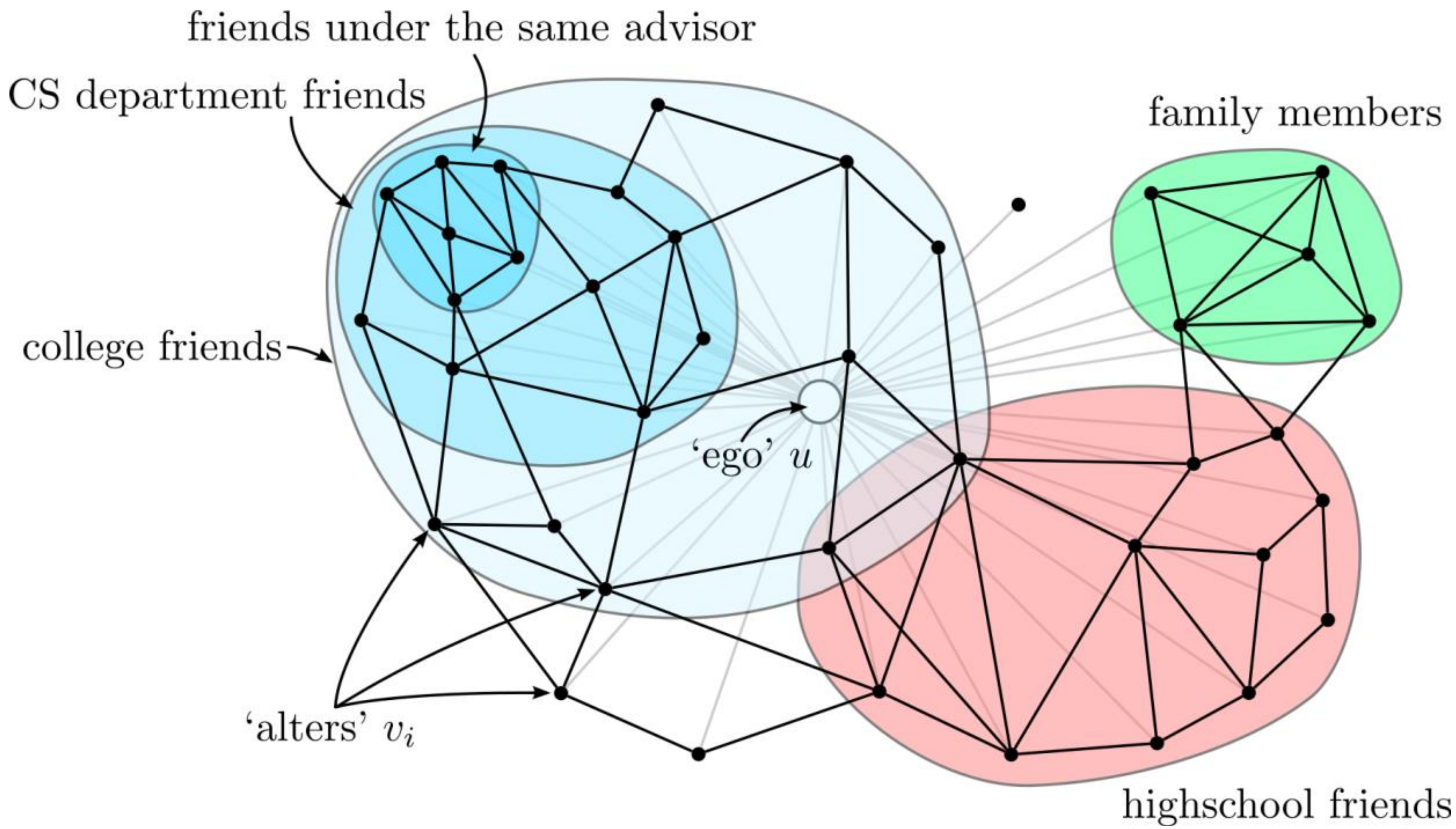
Learning to discover social circles in ego-networks

McAuley & Leskovec, TKDD 2013

McAuley & Leskovec, NIPS 2012

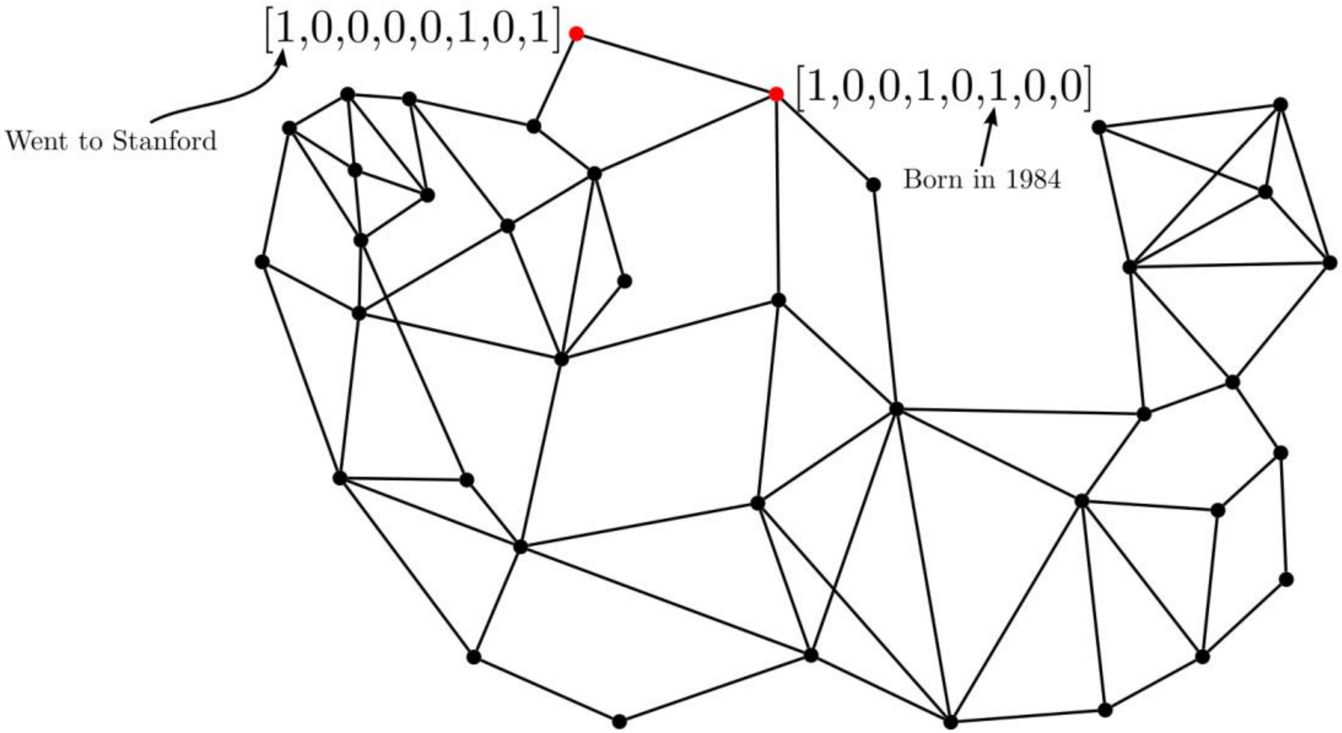
(skip section) (link)

Social circles in ego-networks



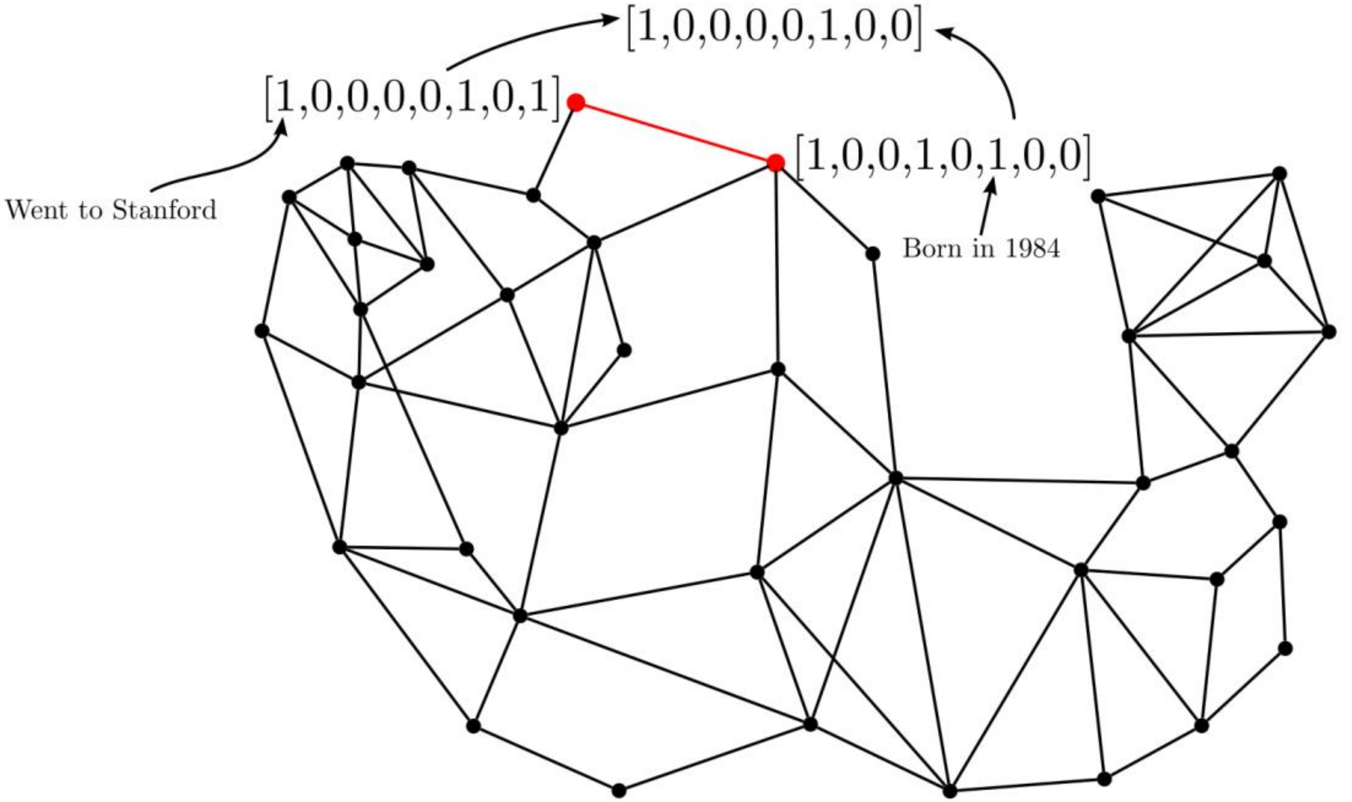
Goal: to recommend circles to users of social networks

Social circles in ego-networks



Goal: to recommend circles to users of social networks

Social circles in ego-networks



Goal: to recommend circles to users of social networks

Social circles in ego-networks

Input:

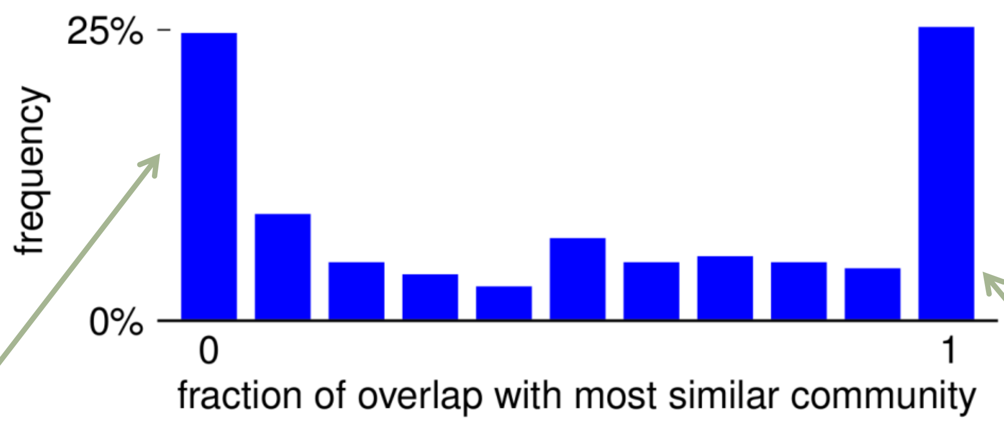
- A graph (directed or undirected)
- Features for each node

Output

- Circles (subsets of the nodes)
- Interpretations of each circle
- Number of circles

dataset	#ego-nets	#circles	#nodes	#edges
Facebook	10	193	4,039	88,234
Google+	133	479	107,614	13,673,453
Twitter	1,000	4,869	81,306	1,768,149

Statistics of social circles



Disjoint communities

Hierarchical communities

Of 193 Facebook circles

- Around 25% share no members with any other circle
- 25% are entirely contained within another circle
- The remaining 50% overlap partially

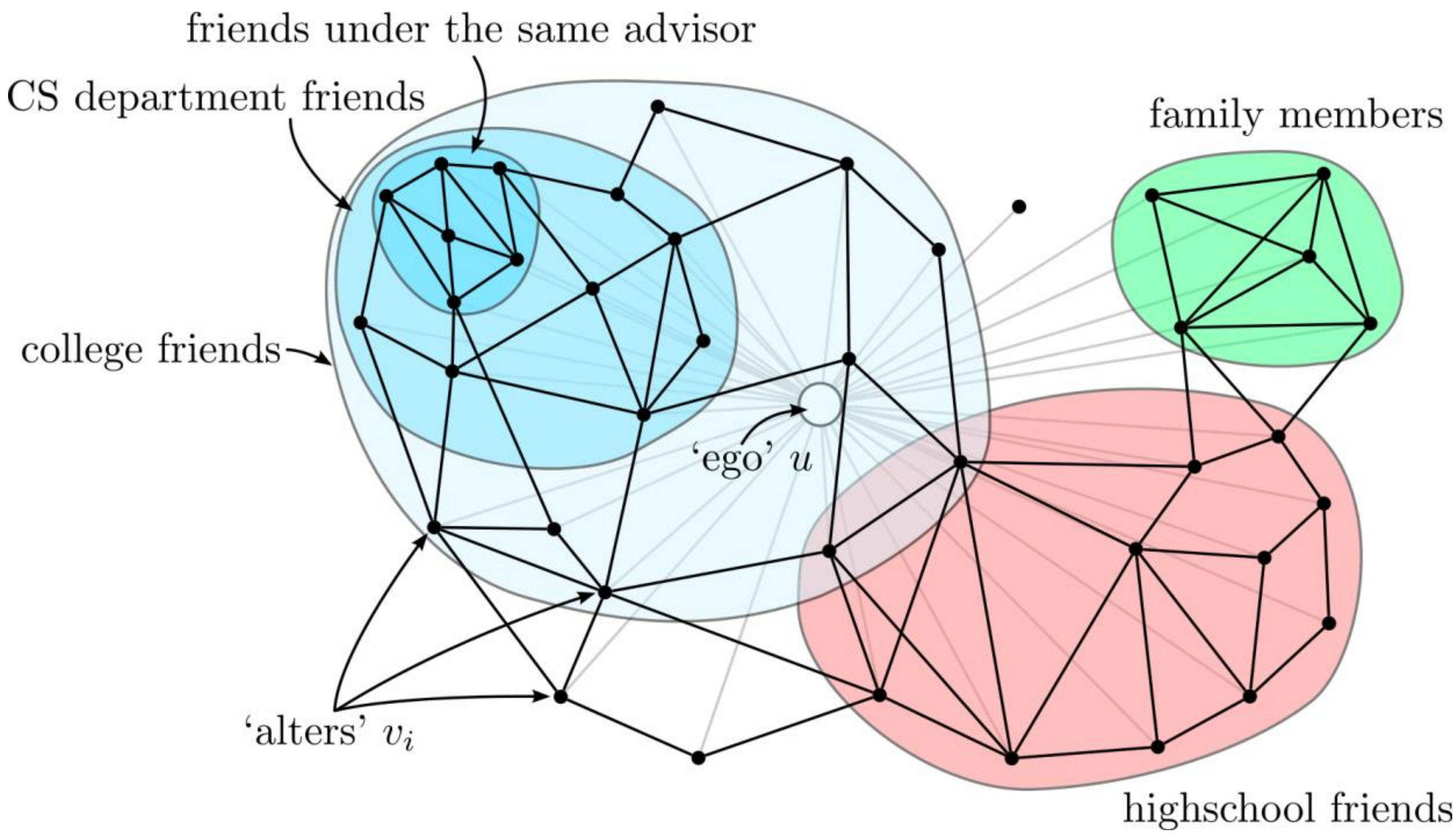
A (too) simple model

$$p((x,y) \in E) \propto \exp\left\{ \underbrace{\sum_{C_k \supseteq \{x,y\}} 1}_{\text{circles containing both nodes}} - \underbrace{\sum_{C_k \nsupseteq \{x,y\}} 1}_{\text{all other circles}} \right\}$$

- Edges are likely between nodes that share many circles
- Edges are unlikely between nodes that share few

A (too) simple model

$$p((x,y) \in E) \propto \exp\Bigg\{ \underbrace{\sum_{C_k \supseteq \{x,y\}} 1}_{\text{circles containing both nodes}} - \underbrace{\sum_{C_k \nsupseteq \{x,y\}} 1}_{\text{all other circles}} \Bigg\}$$



A better model

$$p((x,y) \in E) \propto \exp\left\{ \underbrace{\sum_{C_k \supseteq \{x,y\}} \langle \phi(x,y), \theta_k \rangle}_{\text{circles containing both nodes}} - \underbrace{\sum_{C_k \not\supseteq \{x,y\}} \alpha_k \langle \phi(x,y), \theta_k \rangle}_{\text{all other circles}} \right\}$$

- Reward edges that belong to circles according to $\langle \phi(x,y), \theta_k \rangle$
- Penalize edges that don't belong to circles according to $\langle \phi(x,y), \theta_k \rangle$

Model fitting

Repeat steps (1) and (2) until convergence:

$$\arg\max_{\mathcal{C}} \prod_{x,y \in E} p((x,y) \in E) \prod_{x,y \notin E} (1 - p((x,y) \in E))$$

(solved via gradient ascent using L-BFGS)

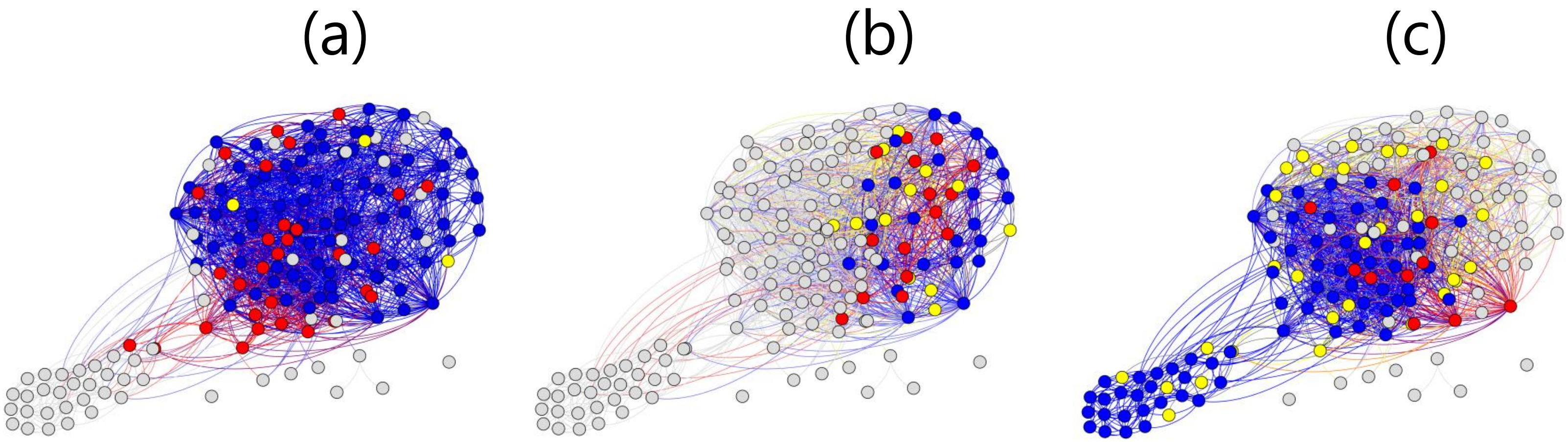
Step 1:
Find circles from circle parameters

$$\arg\max_{\Theta} \prod_{x,y \in E} p((x,y) \in E) \prod_{x,y \notin E} (1 - p((x,y) \in E))$$

(solved via pseudo-boolean optimization)

Step 2:
Find circle parameters from circles

Results



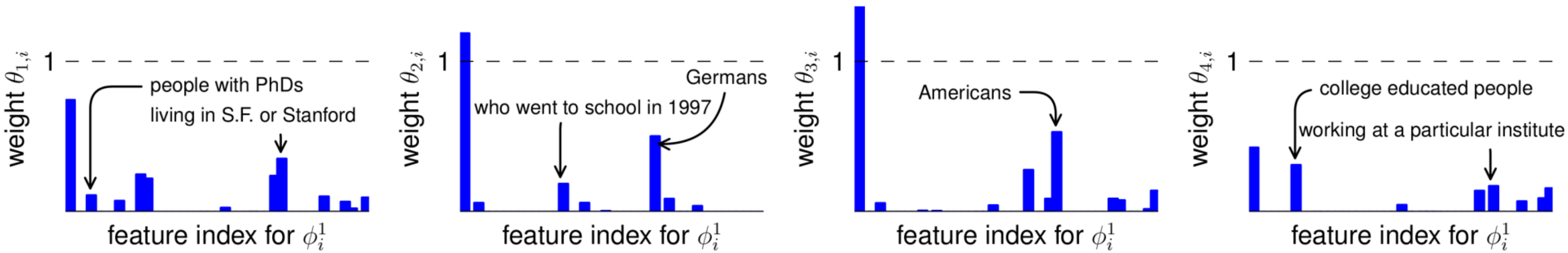
Blue = true positive; grey = true negative;
red = false positive; yellow = false negative

We automatically detect

1. Hierarchically nested circles (e.g. a and b)
2. Disjoint circles (e.g. b and c)
3. Overlapping circles (e.g. a and c)

Results

We also generate automatic 'explanations' for detected circle



We outperform state-of-the-art baselines on all three networks (Facebook, Google+, and Twitter)

- Performance is best on Facebook
- No baseline works well on Twitter or Google+
- Adapting these models to directed networks is the subject of ongoing work

Related work

Aspects: Blair-Goldensohn et al. (2008), Ganu et al. (2009), Titov & McDonald (2008), Lerman et al. (2009), Wang et al. (2010) Automatic aspect discovery: Zhao et al. (2010), Moghaddam & Ester (2011) Latent factor models & LDA: Blei & McAuliffe (2007), Lin & He (2009), Koren and Bell (2011)	aspects, latent factor models
Identifying successful content: Szabo & Huberman (2010), Artzi et al. (2012), Brank & Leskovec (2003) Phrasing vs. success: Danescu-Nicelescu-Mizil et al. (2012), Social dynamics: Hogg & Lerman (2010), Suh et al. (2011), Lerman & Galstyan (2008), Romero et al. (2013)	social media recommen-dation
Clustering social networks: Handcock & Raftery (2007), Overlapping clusters: Yang & Leskovec (2012), Airoldi et al. (2008), Ahn et. al (2010), Palla et al. (2005) Ego-networks: Coscia et al. (2012) Evaluation: Lancichinetti & Fortunato (2009), Yang & Leskovec (2012) Edges & attributes: Chang & Blei (2009), Chang et al. (2009)	community detection

Conclusion

We studied models of users, products, and communities, in order to

- Make better recommendations to users
- Understand the aspects of users' opinions
- Identify useful reviews and expert reviewers

And, we applied these ideas in other applications, to

- Understand the dimensions that determine whether content will be well-received by a given community

Conclusion

We studied models of users, products, and communities, in order to

- Make b
- Underst
- Identify

Thanks!

and thanks to my co-authors, Jure, Dan, and Hima

s

And, we applied these ideas in other applications, to

- Understand the dimensions that determine whether content will be well-received by a given community