Understanding Opinions and Preferences in Product Networks

Julian McAuley, UCSD

Preferences and Opinions

27 of 28 people found the following review helpful

Essential guide to the academic job search, August 30, 2008

By **Laura Malisheski "Career Counselor for PhDs"** (Cambridge, MA USA) - See all my reviews

REAL NAME

This review is from: The Academic Job Search Handbook (Paperback)

The Academic Job Search Handbook is the essential guide for anyone pursuing an academic career. Now in its 4th edition, this book is recognized amongst graduate career professionals not only as a classic in our field, but also as an up-to-date guide book to preparing for and applying to faculty positions. Through straightforward advice coupled with sensitivity toward individual and field-specific differences, Julie Vick and Jennifer Furlong provide extensive coverage of the academic job search fundamentals, including many examples of successful CVs, cover letters, and other application materials. The recently expanded chapters on a variety of special circumstances (e.g. those who are pregnant or new parents on the job market, dual career couples, or older candidates) provide real-life success stories that encourage all applicants to capitalize on their own personal strengths while offering specific strategies to help ameliorate potential concerns of search committees.

As a career counselor for graduate students and PhDs, I experience, through my clients, the enormous anxiety inherent in such a competitive job market. The Academic Job Search Handbook provides an indispensable touchstone to help turn unproductive angst into thoughtful, confident action, through concrete and specific advice. I recommend it to all those aspiring to and navigating an academic career.

Help other customers find the most helpful reviews

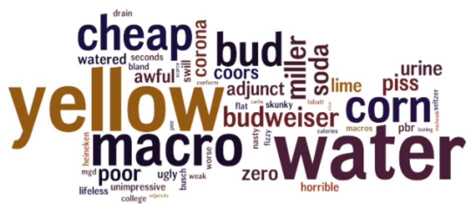
Was this review helpful to you? Yes No

Report abuse | Permalink

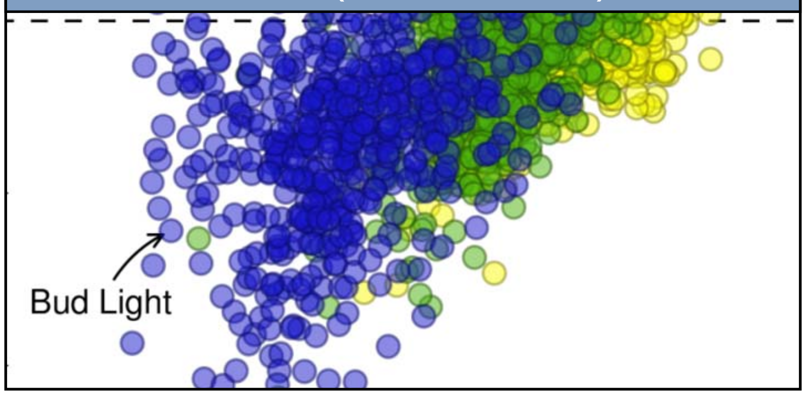
Comment

Preferences and Opinions – Tasks

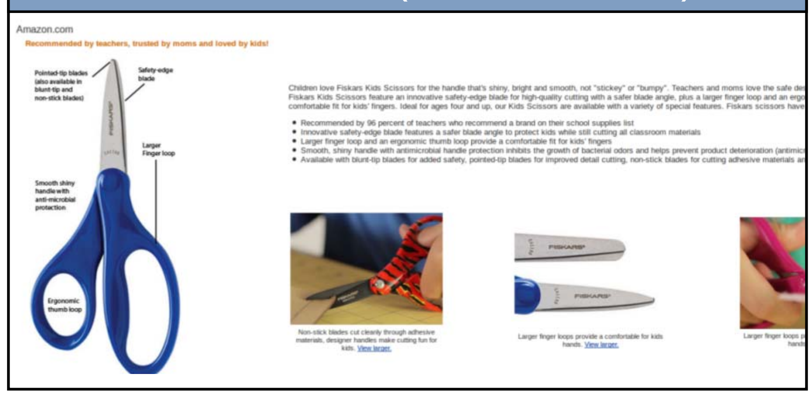
Understanding opinions through reviews (ICDM 2012, RecSys 2013)



Understanding how people acquire tastes (WWW 2013)



Predicting opinions from text and social media (under review)



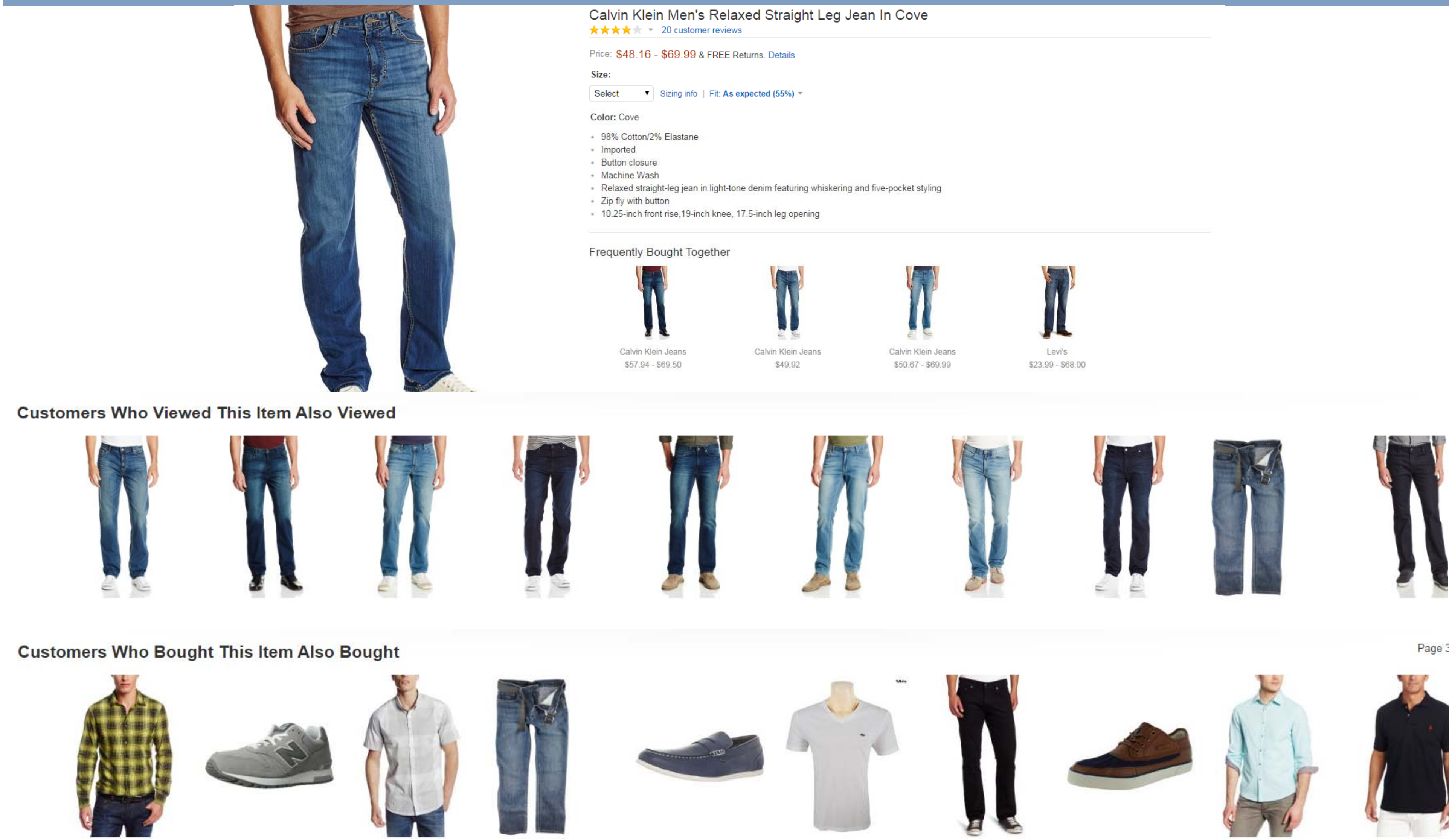
Preferences and Opinions – Applications

Recommend a product/story that I'll like

Estimate what I'll purchase

Identify 'useful' opinions

Opinions in Networks



Opinions in Networks

browsed together

bought together

Opinions in Networks – why?

1. To understand the notions of **substitute** and **complement goods**

Opinions in Networks – why?

2. To generate **explanations**
of why certain products are preferred

People prefer this because:

"Good quality, soft, **light weight, the colors** are beautiful and exactly like the picture!"

Opinions in Networks – why?

3. To **recommend**
baskets of related items

Query:

Suggested outfit:

Query:

Suggested outfit:

Data

Amazon product network:

- thousands of **categories**
- 9 million **products**
- 21 million **users**
- 140 million **reviews**
- 300 million **relationships**

Data

Four types of relationship:

1) People who **viewed** X also **viewed** Y
2) People who **viewed** X eventually **bought** Y

3) People who **bought** X also **bought** Y
4) People **bought** X and Y **together**

Substitutes (1 and 2), and **Complements** (3 and 4)

Networks of **text** and **images**

Part 1: Understanding product networks with **text**

Modeling: Can we use the text of product reviews to model relationships between products?
Understanding: Can we explain **why** people tend to prefer certain products over other

Part 2: Understanding product networks with **images**

Modeling: Can we understand which products have compatible visual "styles", and use this to recommend baskets of products to people?
Understanding: Can we discover competing styles of products, and understand the visual features common to each?

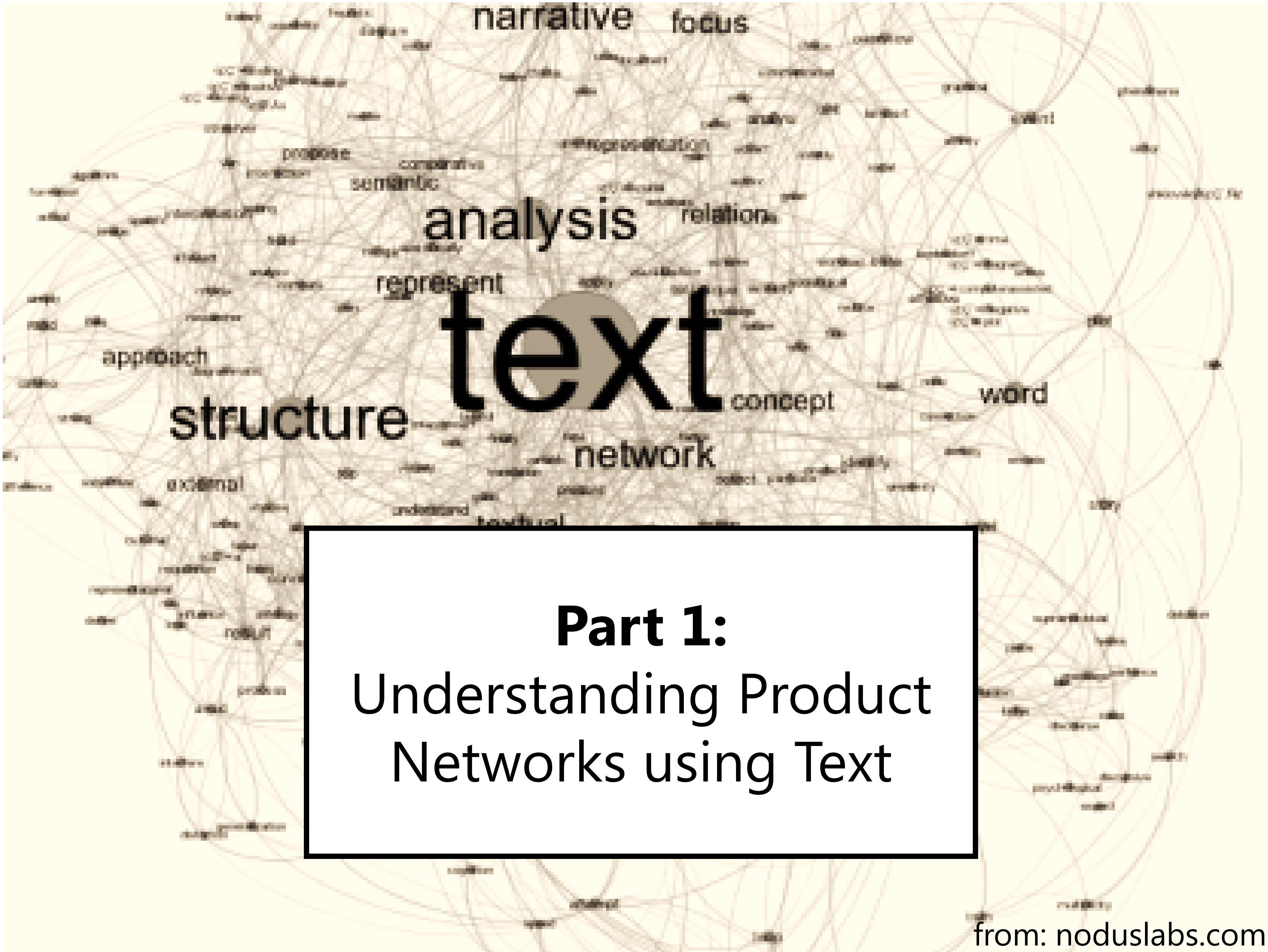
Inspiration

"I personally find Amazon's recommendation system for **books and music** to be very, very good... With domains like **shirts or shoes**, it's murkier semantically, and they have less data, and so it's much poorer"

Mike Jordan (IEEE Spectrum)

"You need to dress better now that you're a professor"

Laura



Part 1:
Understanding Product Networks using Text

Problem setting

Binary prediction task:
Given a pair of products, **x and y**, predict whether they were purchased together, or whether they were chosen randomly

$$p(x \text{ and } y \text{ are related}) \sim -d(x, y)$$

Problem setting

But we are not **given** a distance function:
We need to **learn** the concept of similarity from data:

$$p_\theta(x \text{ and } y \text{ are related}) \sim -d_\theta(x, y)$$

Train θ by maximum likelihood:

$$\theta = \arg\max_{\theta'} \prod_{\text{edges } (x,y)} p_\theta(x \text{ and } y \text{ are related}) \prod_{\text{non-edges } (x,y)} (1 - p_\theta(x \text{ and } y \text{ are related}))$$

What are we actually learning?

How did Amazon generate their ground-truth data?

Given a product:  Let U_i be the set of users who viewed it

for every product in the corpus...

U_1 U_2 U_3 $\cdots$

What are we actually learning?

How did Amazon generate their ground-truth data?

Given a product:  Let U_i be the set of users who viewed it

Rank products according to: $\frac{|U_i \cap U_j|}{|U_i \cup U_j|}$ ('Jaccard index')

.86 .84 .82 .79 ...

Linden, Smith, & York (2003)

Attempt 1: features derived from words

Reviews of product *i*:

$\mathbf{x_i} =$ [0,0,0,0,0,0,0,1,0,5,0,0,0, … ,0,1,0,0,0,0,0,1,2]

Reviews of product *j*:

$\mathbf{x_j} =$ [0,0,0,1,0,0,0,0,0,0,0,1,0, … ,0,0,0,0,0,0,0,1,0]

aardvark

zoetrope

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sum_w (\mathbf{x}_{i,w} - \mathbf{x}_{j,w})^2$$

Attempt 1: features derived from words

Reviews of product *i*:

$$\mathbf{x}_i = \quad [0,0,0,0,0,0,0,1,0,5,0,0,0,\ \ldots\ ,0,1,0,0,0,0,0,1,2]$$

Reviews of product *j*:

$$\mathbf{x}_j = \quad [0,0,0,1,0,0,0,0,0,0,0,1,0,\ \ldots\ ,0,0,0,0,0,0,0,1,0]$$

aardvark

zoetrope

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sum_w \theta_w (\mathbf{x}_{i,w} - \mathbf{x}_{j,w})^2$$

Attempt 1: features derived from words

Reviews of product *i*:

$$\mathbf{x}_i = [0,0,0,0,0,0,0,1,0,5,0,0,0, \ldots ,0,1,0,0,0,0,0,1,2]$$

Reviews of product *j*:

$$\mathbf{x}_j = [0,0,0,1,0,0,0,0,0,0,0,1,0, \ldots ,0,0,0,0,0,0,0,1,0]$$

aardvark

zoetrope

- High-dimensional
- Prone to overfitting
- Too fine-grained

Attempt 2: features derived from topics

Topic models:

87 of 102 people found the following review helpful

☆☆☆☆☆ **You keep what you kill**, December 27, 2004

By **Schtinky "Schtinky"** (Washington State) - See all my reviews

VINE™ VOICE

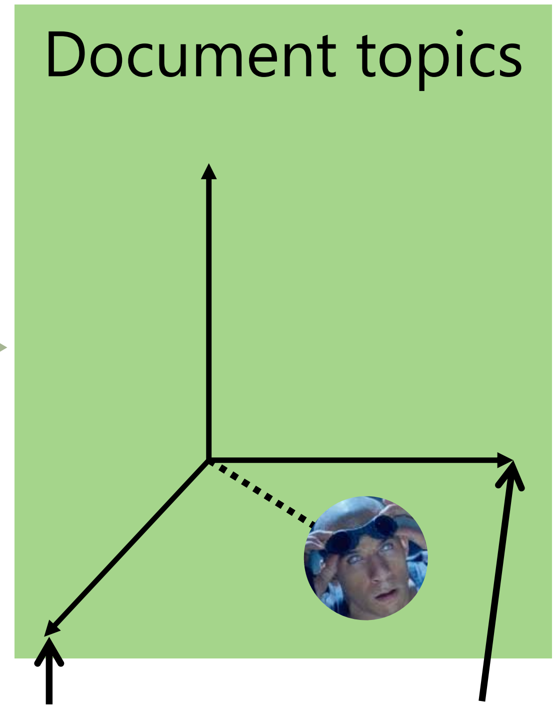
This review is from: **The Chronicles of Riddick (Widescreen Unrated Director's Cut) (DVD)**

Even if I have to apologize to my Friends and Favorites, and my family, I have to admit that I really liked this movie. It's a Sci-Fi movie with a "Mad Maxx" appeal that, while changing many things, left Riddick from `Pitch Black' to be just Riddick. They did not change his attitude or soften him up or bring him out of his original character, which was very pleasing to `Pitch Black' fans like myself.

First off, let me say that when playing the DVD, the first selection to come up is Convert or Fight, and no explanation of the choices. This confused me at first, so I will mention off the bat that they are simply different menu formats, that each menu has the very same options, simply different background visuals. Select either one and continue with the movie.

(review of “The Chronicles of Riddick”)

LDA →



Action:
action, loud, fast, explosion,...

Sci-fi
space, future, planet,...

Blei & McAuliffe (2007)

Attempt 2: features derived from topics

Reviews of product *i*:

$\mathbf{x}_i =$ [0.1, 0.4, 0.2, 0.1, 0.2]

Reviews of product *j*:

$\mathbf{x}_j =$ [0.3, 0.1, 0.3, 0.2, 0.1]

action romance

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sum_k \theta_k (\mathbf{x}_{i,k} - \mathbf{x}_{j,k})^2$$

Attempt 2: features derived from topics

Reviews of product *i*:

$$\mathbf{x}_i = [0.1, 0.4, 0.2, 0.1, 0.2]$$

Reviews of product *j*:

$$\mathbf{x}_j = [0.3, 0.1, 0.3, 0.2, 0.1]$$

action

romance

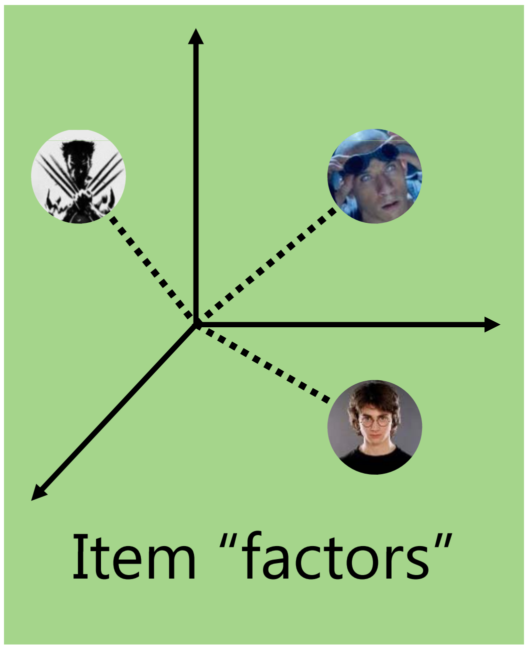
On the right track, but are the topics we're discovering **relevant** to link prediction?

Attempt 3: directly learn 'good' topics

Learn to discover topics that
explain the graph structure

Attempt 3: directly learn 'good' topics

Previously…



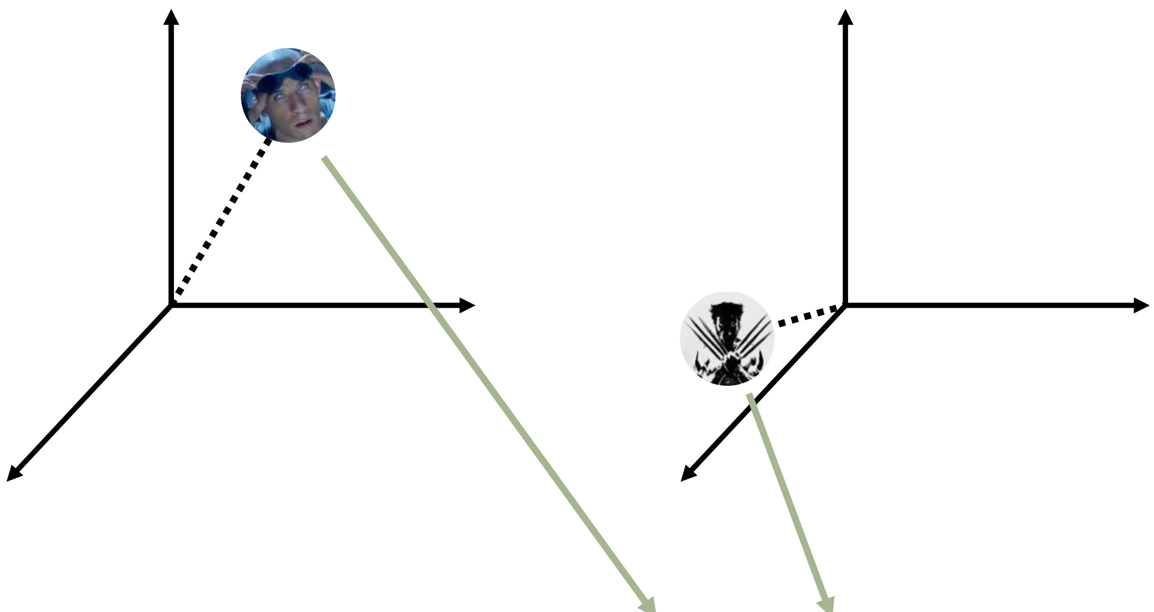
transform

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

We learned to discover topics that "explained" people's ratings

McAuley & Leskovec (2013)

Attempt 3: directly learn 'good' topics



$$d(\mathbf{x}_i, \mathbf{x}_j) = \|\theta_i - \theta_j\|_2^2$$

Learn to project documents (reviews) into topic space such that related products are nearby

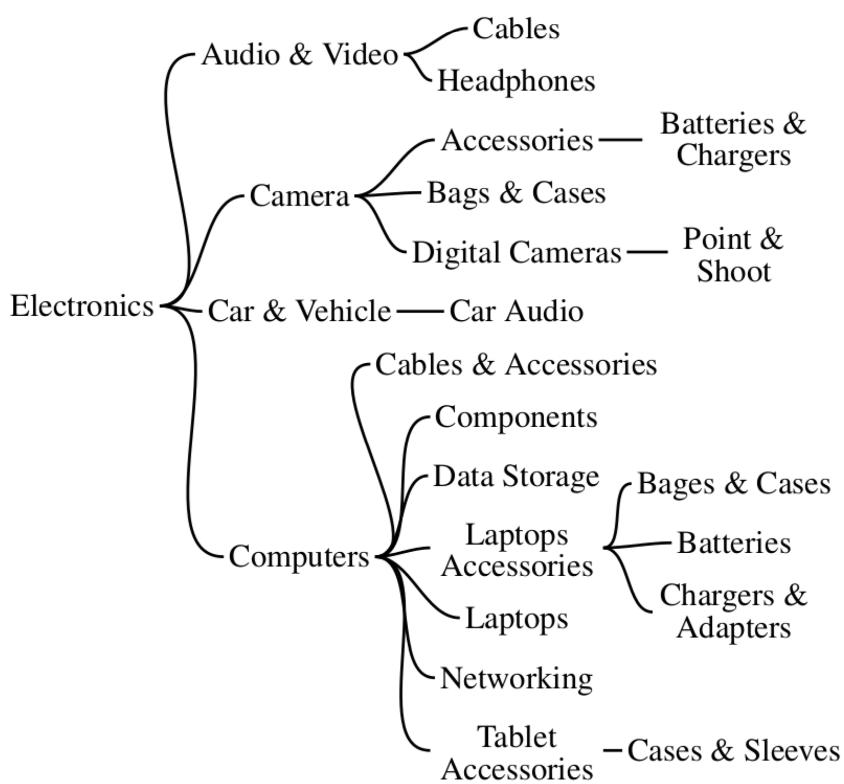
Training

$$\theta = \arg\max_{\theta'} \overbrace{\prod_{\text{edges } (x,y)} p_\theta(x \text{ and } y \text{ are related}) \prod_{\text{non-edges } (x,y)} (1 - p_\theta(x \text{ and } y \text{ are related}))}^{\text{logistic regression}} \times \underbrace{p(\text{review corpus}|\theta)}_{\text{topic model}}$$

Not tractable

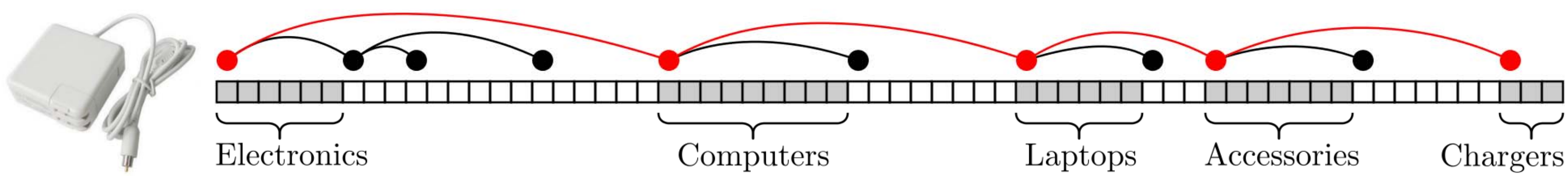
This works well, but has **a lot** of parameters,
e.g. **hundreds** of topics multiplied by **millions** of products

Idea: use the category hierarchy to sparsify the model



Not tractable

Associate each node in the category tree with a small number of topics:



Now we can fit models with **hundreds** of topics but only 10-20 are **active** per product

Results

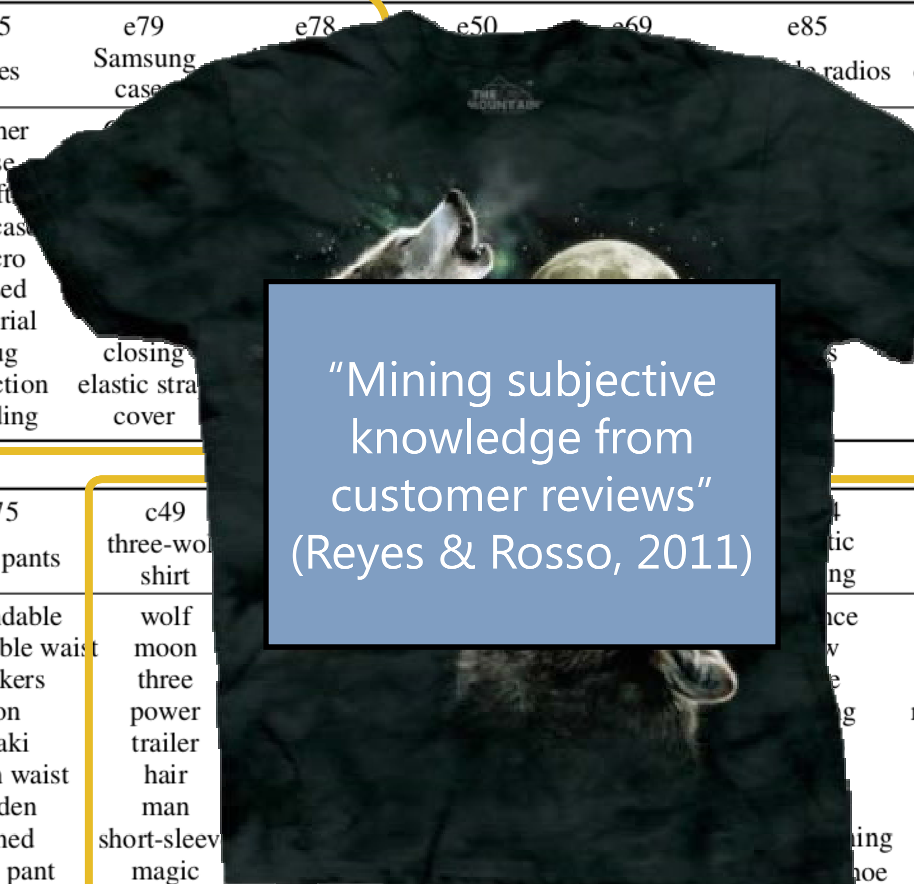
	Word counts	Topic model	Topic model, trained to identify predictive topics
Books	86.5%	86.5%	95.1%
Movies	86.7%	86.5%	93.7%
Music	86.7%	87.2%	97.4%
Electronics	86.2%	86.3%	93.8%
Men's Clothing	83.6%	83.5%	95.2%
Women's Clothing	82.9%	83.1%	95.2%
	85.4%	**85.5%**	**95.1%**

Results

Electronics

e111	e92	e75	e79	e78	e50	e69	e85	e96	e89	e99
cameras	portable spakers	cases	Samsung case				le radios	car radios	high-end headphones	budget headphones
camera	little speaker	leather						radio	bass	bass
zoom	bose	case						eer	Sennheiser	Skullcandy
pictures	portable speaker	soft						y	Bose	sound
Kodak	small speaker	Roocas						ra	Shure	bud
Canon	sound	velcro						ord	Beats	outside noise
flash	iHome	closed						lash	Koss	another pair
digital	bass	material						Honda	Akg	comfortable
optical	wireless speaker	snug	closing				s	Jeep	music	gym
taken	great speaker	protection	elastic stra					wiring	classical	Beats
picture	mini speaker	standing	cover					deck	Klipsch	head

c44	c107	c75	c49				4	c133	c24	c9
dress shirts	dress shoes	dress pants	three-wo shirt				tic ng	sports shoes	generic clothing	generic clothing
sleeves	leather	expandable	wolf				ce	court	dry	same
arms	sole	expandable waist	moon				w	play	cold	durable
neck	dress	Dockers	three				e	Nike	working	store
shoulders	brown	iron	power				g	running shoe	short	different
dress shirt	dress shoe	khaki	trailer					running	hot	two
dress	polish	stretch waist	hair					games	weather	brand
jacket	brown pair	hidden	man					light shoe	tight	comfort
long sleeve	toe	ironed	short-sleev				ing	great shoe	cool	fine
iron	looking shoe	dress pant	magic				oe	support	down	tight
tucked	formal	elastic waist	powerful	fitted shirt	Rangers	run	cross	miles	regular	another pair



"Mining subjective knowledge from customer reviews" (Reyes & Rosso, 2011)

Explaining user preferences

$$p(x \text{ and } y \text{ are related}) \sim -d(x,y)$$

why do people why view
X eventually buy **Y**?

{baby, pants, pyjamas, colorful}

Explaining user preferences

$$p(\text{edge flows from } x \text{ to } y \mid x \text{ and } y \text{ are related})$$

why do people why view
X eventually buy **Y**?

"**true size,** fits well, items are the same color and type as on the picture"

Explaining user preferences

Current product:

Straight Edge (non-skid) Socks

Products to buy instead:

Luvable Friends 8 Pack Newborn Socks
view recommendations for this product
They **look a little small,** but they are newborn socks, which is what I wanted.
Belongs to micro-category: 4

Products to buy together:

i play. Baby-Girls Infant Classic Birm Sun Protection Hat
view recommendations for this product
It's a **little big** but I would rather it be a **little big** then a little small.
Belongs to micro-category: 110

i play. Babywear Sunhat
view recommendations for this product
Our little one should be able to wear this for a **long time.**
Belongs to micro-category: 110

jmcauley.ucsd.edu/amazon/

Extensions

Training on cold-start items:

Editorial Reviews

Review

'The listener is clutched by the throat' Daily Express 'Stephen Fry is "the Potter perfect narrator" ' The Times --*This text refers to the Audio CD edition.*

About the Author

J. K. Rowling is the author of the beloved, bestselling, record-breaking Harry Potter series. She started writing the series during a delayed Manchester to London King's Cross train journey, and during the next five years, outlined the plots for each book and began writing the first novel. Harry Potter and the Sorcerer's Stone was published in the United States by Arthur A. Levine Books in 1998, and the series concluded nearly ten years later with Harry Potter and the Deathly Hallows, published in 2007. J. K. Rowling is the recipient of numerous awards and honorary degrees including an OBE for services to children's literature, France's Légion d'Honneur, and the Hans Christian Andersen Literature Award. She supports a wide number of causes through her charitable trust Volant, and is the founder of Lumos, a charity working to transform the lives of disadvantaged children. J. K. Rowling lives in Edinburgh with her husband and three children.

Kazu Kibuishi is the creator of the *New York Times* bestselling *Amulet* series and *Copper*, a collection of his popular webcomic. He is also the founder and editor of the acclaimed Flight anthologies. *Daisy Kutter: The Last Train*, his first graphic novel, was listed as one of the Best Books for Young Adults by YALSA, and *Amulet, Book One: The Stonekeeper* was an ALA Best Book for Young Adults and a Children's Choice Book Award finalist. Kazu lives and works in Alhambra, California, with his wife and fellow comics artist, Amy Kim Kibuishi, and their two children. Visit Kazu online at www.boltcity.com.

Mary GrandPré has illustrated more than twenty beautiful books for children, including the American editions of the Harry Potter novels. Her work has also appeared in the *New Yorker*, the *Atlantic Monthly*, and the *Wall Street Journal*, and her paintings and pastels have been shown in galleries across the United States. Ms. GrandPré lives in Sarasota, Florida, with her family.

--This text refers to an alternate Paperback edition.

More About the Author

› Visit Amazon's J. K. Rowling Page

Biography

J K (Joanne Kathleen) Rowling was born in the summer of 1965 at Yate General Hospital in England and grew up in Chepstow, Gwent where she went to Wyedean Comprehensive. Jo left Chepstow for Exeter University, where she earned a French and Classics degree, and where her course included one year in Paris. As a postgraduate she moved to London to work at Amnesty International, doing research into human rights abuses in Francophone Africa. She started writing the Harry Potter series during a Manchester to London King's Cross train journey, and during the next five years, outlined the plots for each book and began writing the first novel. Jo then moved to northern Portugal, where she taught English as a foreign language. She married in October

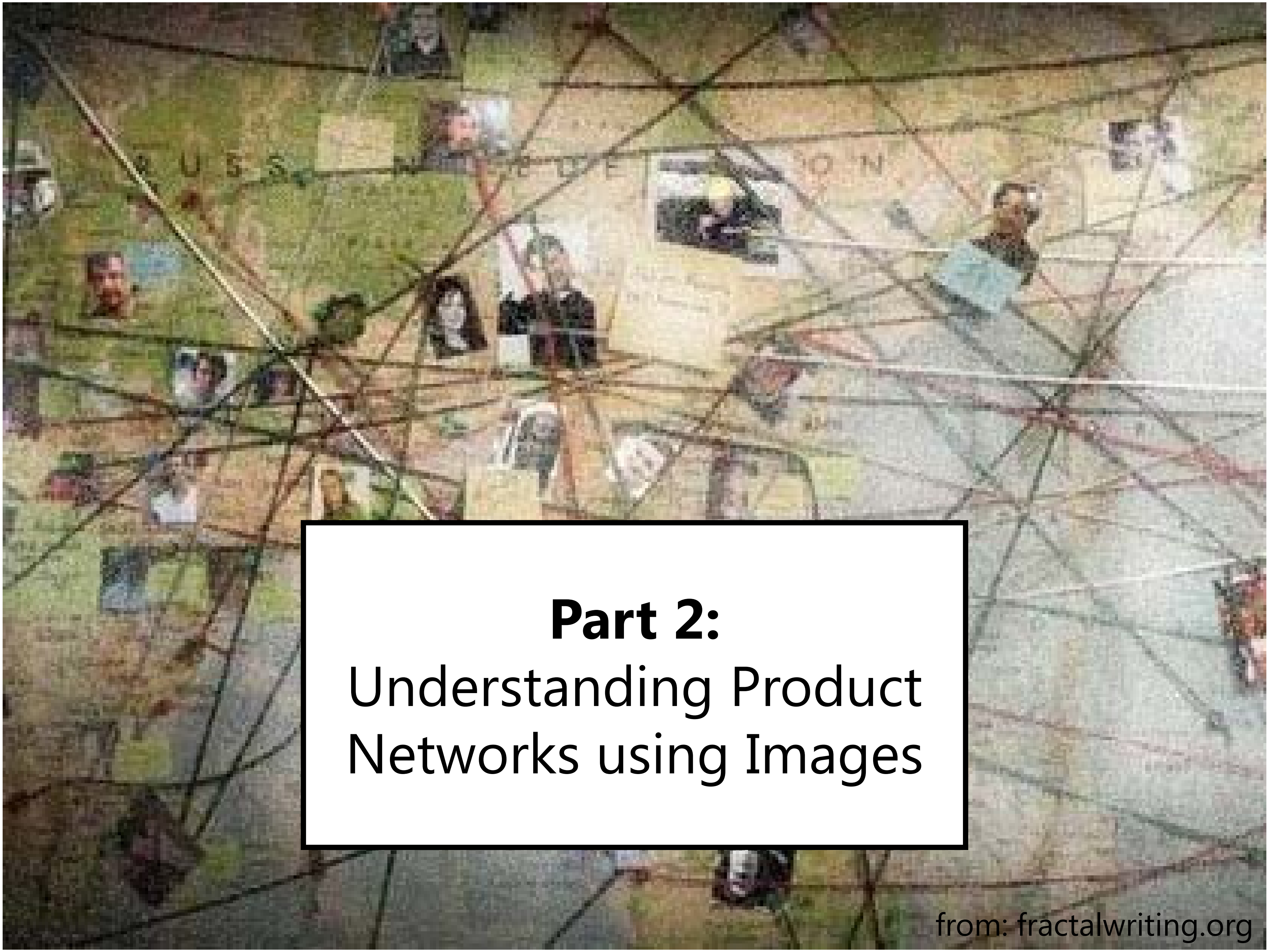
Extensions

Friend recommendation on Yelp

Extensions

Identifies friends/not-friends with 93.3% accuracy

y21 Scotland	y28 Vegas hotels	y27 dim sum	y5 great!	y18 Thai
Royal Mile	hotel	Chinese broccoli	great pizza	Thai restaurant
wee	Aria	dim sum	great staff	flavorful
mum	Venetian	Taiwanese	great service	yellow curry
Scottish	Cosmopolitan	dim	great atmosphere	outstanding
homely	MGM	topped	good service	tom
wee bit	forum	tender	excellent service	prompt
quirky	Carlo	onions	great	requested
keen	Hollywood	hot pot	friendly service	tender
Victoria	Wicked Spoon	fried	friendly staff	ordered
Royal	XS	Korean	great breakfast	contained



Part 2:
Understanding Product Networks using Images

from: fractalwriting.org

Why might images be useful

- Cold-start problems
- Visual explanations might be more intuitive
- The image is the most important feature for many categories

Problem setting (again)

$$p(x \text{ and } y \text{ are related}) \sim -d(x, y)$$

[0.723845, 0.153926, 0.757238, 0.983643, ...]

[0.456353, 0.898354, 0.123342, 0.234253, ...]

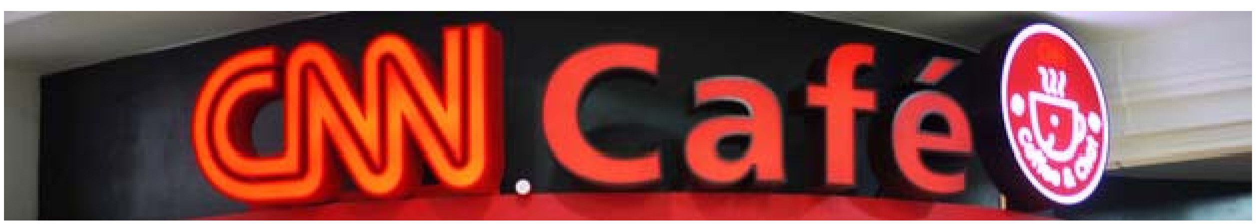
image features

Problem setting (again)

[0.723845, 0.153926, 0.757238, 0.983643, …]

4096-dimensional image features

We used **Caffe**, a convolutional neural net trained on **ImageNet**



http://caffe.berkeleyvision.org/

Attempt 1: distance between features

Features of (image of) product *i*:

$$\mathbf{x}_i = \quad [0.723845, 0.153926, 0.757238, 0.983643, \ldots]$$

Features of product *j*:

$$\mathbf{x}_j = \quad [0.456353, 0.898354, 0.123342, 0.234253, \ldots]$$

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sum_k \theta_k (\mathbf{x}_{i,k} - \mathbf{x}_{j,k})^2$$

Attempt 1: distance between features

Features of (image of) product *i*:

$$\mathbf{x}_i = \quad [0.723845, 0.153926, 0.757238, 0.983643, \ldots]$$

Features of product *j*:

$$\mathbf{x}_j = \quad [0.456353, 0.898354, 0.123342, 0.234253, \ldots]$$

At best we'll discover visual **similarity,**
but visual relationships are more subtle

Attempt 2: Mahalanobis distance

$$d(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j) M (\mathbf{x}_i - \mathbf{x}_j)^T$$

$$M = \begin{array}{cl} \begin{array}{cccc} & \text{texture} & & \end{array} & \\ \left(\begin{array}{cccc} 0.1 & 0.2 & \cdots & 0.1 \\ 0.3 & 0.0 & & 0.2 \\ \vdots & & \ddots & \vdots \\ 0.3 & \boxed{0.6} & \cdots & 0.1 \end{array}\right) & \text{color} \end{array}$$

Attempt 2: Mahalanobis distance

$$d(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j) M (\mathbf{x}_i - \mathbf{x}_j)^T$$

- High-dimensional
- Prone to overfitting
- Too slow!

Attempt 3: Low-rank Mahalanobis

$$d(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j) M (\mathbf{x}_i - \mathbf{x}_j)^T$$

Replace *M* by an approximation of low rank

$$d(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j) U U^T (\mathbf{x}_i - \mathbf{x}_j)^T$$

Attempt 3: Low-rank Mahalanobis

let $\mathbf{s}_i = \mathbf{x}_i U$

$(1 \times K)$ $(1 \times F)$ $(F \times K)$

then $d(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{s}_i - \mathbf{s}_j\|_2^2$

We call this the 'style space' embedding of **x**

Training

$$U = \arg\max_{U'} \prod_{\text{edges } (x,y)} p_U(x \text{ and } y \text{ are related}) \prod_{\text{non-edges } (x,y)} (1 - p_U(x \text{ and } y \text{ are related}))$$

Results

Books

K	buy after viewing	also viewed	also bought	bought together	average
1	66.3%	66.1%	66.7%	60.7%	65.0%
10	72.4%	71.6%	72.1%	68.8%	71.2%
100	73.5%	72.4%	73.6%	69.0%	72.1%

Electronics

K	buy after viewing	also viewed	also bought	bought together	average
1	68.4%	74.7%	64.5%	72.3%	67.5%
10	83.4%	80.4%	77.6%	78.0%	79.9%
100	85.7%	84.0%	82.3%	82.4%	83.6%

Results

Clothing

K	also viewed	also bought	bought together	average
1	78.7%	75.4%	78.9%	77.7%
10	88.2%	86.8%	90.7%	88.6%
100	90.0%	90.8%	93.8%	91.5%

Shoes

K	also viewed	also bought	bought together	average
1	78.4%	78.9%	89.5%	82.3%
10	94.1%	95.3%	96.1%	95.2%
100	96.6%	97.6%	97.9%	97.4%

Visualizing 'style space'

We've projected images into a low dimensional space encoding their style, what are the "extreme" points?

Visualizing 'style space'

Visualizing 'style space'

Visualizing 'style space'

Which styles are at **opposite** ends of the spectrum?

Generating recommendations

How can we use the system to generate recommendations?

Query:

Suggested outfit:

Generating recommendations

How can we use the system to generate recommendations?

Query:

Suggested outfit:

Generating recommendations

Conclusion
and future work

Relationships via text and images

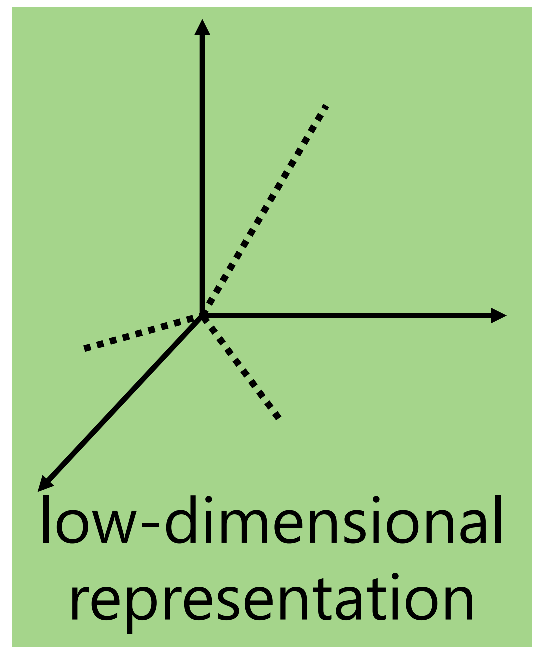
11 of 11 people found the following review helpful

★★★★★ **Great pants for the money**

By kdogg74 on August 5, 2011

Color Name: Navy | Size Name: 38W x 32L | Verified Purchase

Having to dress for work every day I was pleasantly surprised with the quality and look of these pants. They are comfortable enough to wear every day yet have a look refined enough for the office environment. They wash up well and exhibit minimal shrinkage in the length. They do tend to shrink a bit in the waist. I can live with this as I always buy trousers a waist size bigger. The need for ironing after a wash is minimal as well, but I still give them a quick once over with the iron just for good measure. A word of fashion advice to those of us men... please don't buy your trousers too long long. I have seen many men with pant legs pooled and gathered around their shoes. The pant leg should have a slight break at the top of the instep. It just looks cleaner and presents a more professional appearance. I have seen these same pants at department stores for $70+, I got them for half that on Amazon. Now all Amazon has to do is get a better selection of dress shirts at a reasonable price.



Future work

- How have visual styles evolved over time?
- Can text and images be **combined** to model relationships?

- How can we **personalize** these models to predict opinions?

Data & code

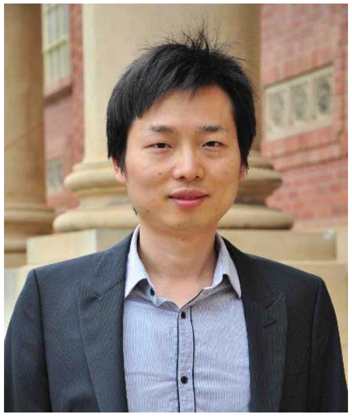
Data and code ~~is~~ **will be made** available on cseweb.ucsd.edu/~julian/

Thanks!

Anton van
den Hengel

Gary
Cottrell

Jure
Leskovec



Qinfeng
"Javen" Shi

e-mail: jmcauley@cs.ucsd.edu