Principles of Website Functionality & Advertisement

CS57300 - Data Mining
Spring 2016

Instructor: Bruno Ribeiro

Goals

- Introduce a variety of Data Mining applications
- Explain some of the principles
- Give a roadmap for the rest of the course

A/B Testing for Political Campaigns

- Obama's 2008 campaign online effort first of its kind
- Try different strategies to get more donations

Ack: Dan Siroker
optimizely.com

Effectiveness of Distinct Strategies

Probability difference is significant

no. users donated (conversions)

no. users on each bucket

(2)

Download: XML CSV TSV | Print

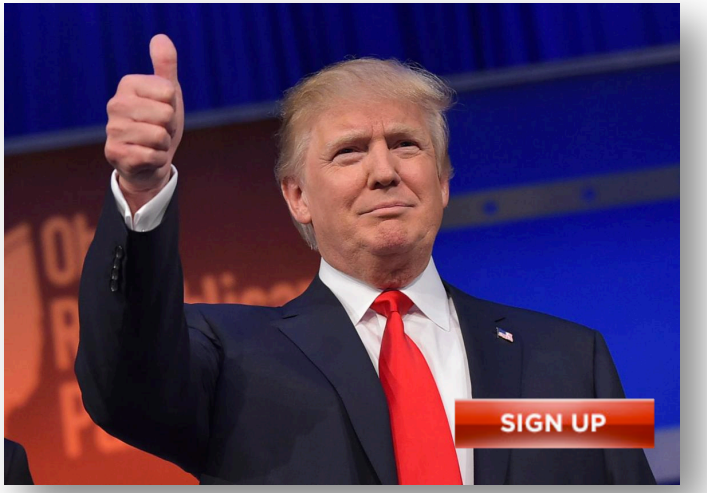
		rate	Chance to Beat Orig.	Observed Improvement	Conv./Visitors
		0.2%	—	—	5851 / 77858
		0.2%	100%	18.6%	6927 / 77729
		0.2%	73.5%	1.37%	5915 / 77644
		0.2%	13.7%	-2.38%	5660 / 77151
		0.2%	—	—	4425 / 51794
		0.2%	100%	13.1%	4996 / 51696
	Change Image	8.87% ± 0.2%	92.2%	3.85%	4595 / 51790
	Barack's Video	7.76% ± 0.2%	0.04%	-9.14%	3992 / 51427
	Sam's Video	6.29% ± 0.2%	0.00%	-26.4%	3261 / 51864
	Springfield Video	5.95% ± 0.2%	0.00%	-30.3%	3084 / 51811

Ack: Dan Siroker
optimizely.com

Political Donations

- This election cycle the candidates will be doing similar experiments

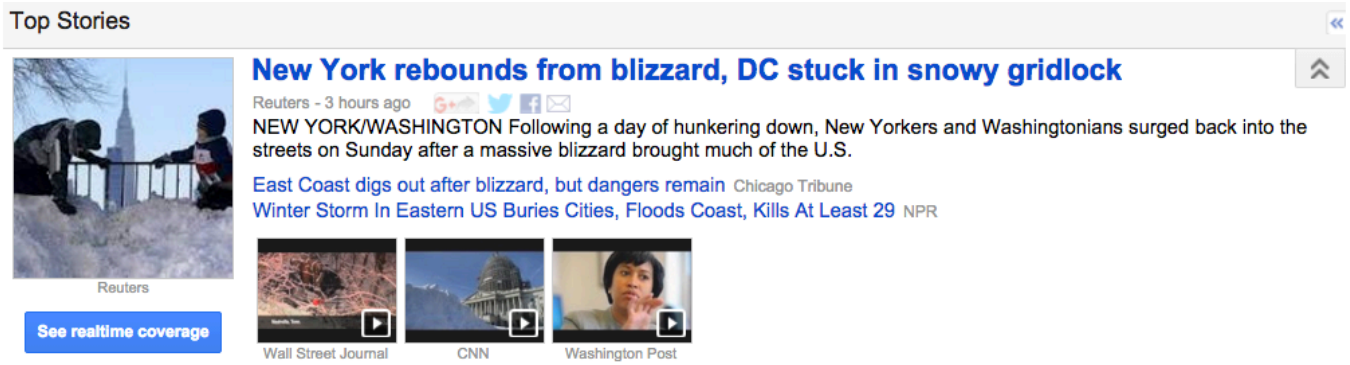
v.s.



- Politician statements are similarly tested
  - No surprises on the impact on likely voters

Showing News

- Yahoo! News
- Google News



- What is relevant for a user?
  - How do we it find out?

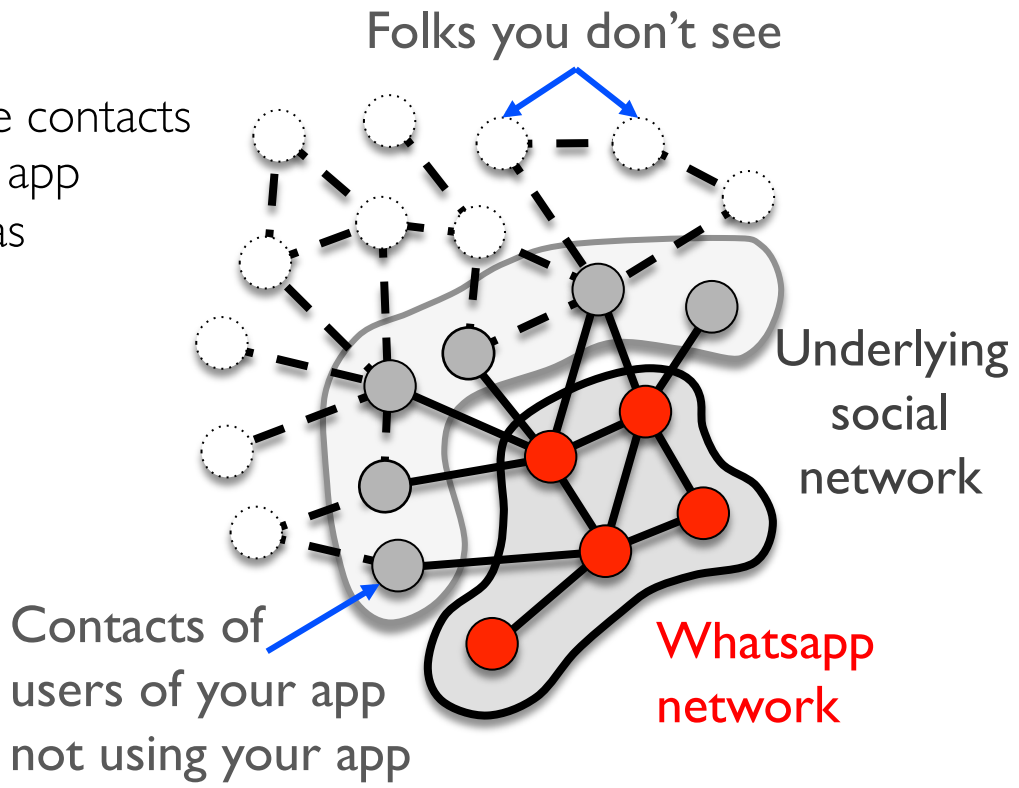
Hypothesis Testing (Sequential Analysis)

- **Hypotheses** are tentative statements of the expected relationships between two or more variables
- Formulate null and alternative hypothesis
  - H_0: Angry Trump donations = Calm Trump donations
  - H_1: Angry Trump donations ≠ Calm Trump donations
- Gather a sample statistic (e.g., μ =estimate of Angry Trump donations)
- Determine the sampling distribution for the statistic under the null hypothesis
- Use the sampling distribution to calculate the probability of obtaining the observed value of μ , given H_0
  - If the probability is low, reject H_0 in favor of H_1

Problems in Network Analysis

Why Social Networks Grow

- Facebook / Twitter growth
  - Suggests new users (link prediction)
- Whatsapp growth:
  - Populate user list with existing phone contacts
  - Suggests user can send messages via app
  - Message arrives to non-subscribers as invitation to subscribe
- Recruit users via social networks
  - Online & Offline social networks
- Network effect:
  - Value of service monotonically increases with no. users
  - Positive feedback loop
  - How useful is having an email address if nobody has one?

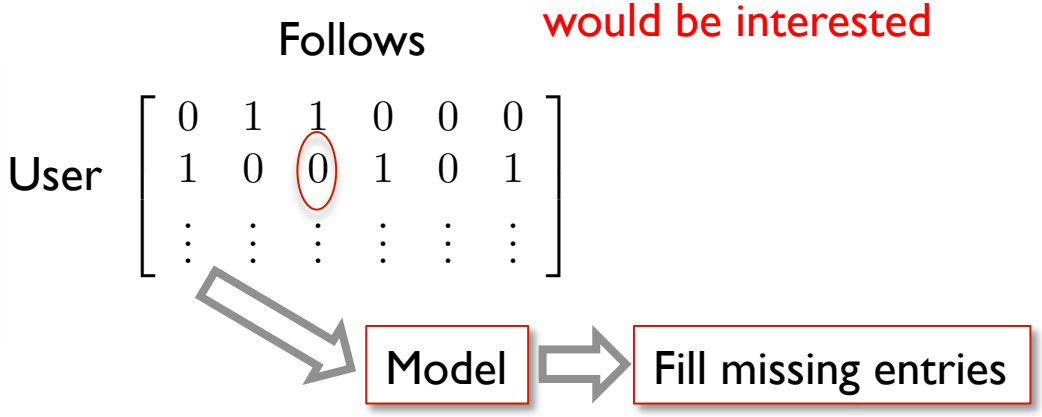


Asks current users (red) to invite friends (gray)

Facebook / Twitter

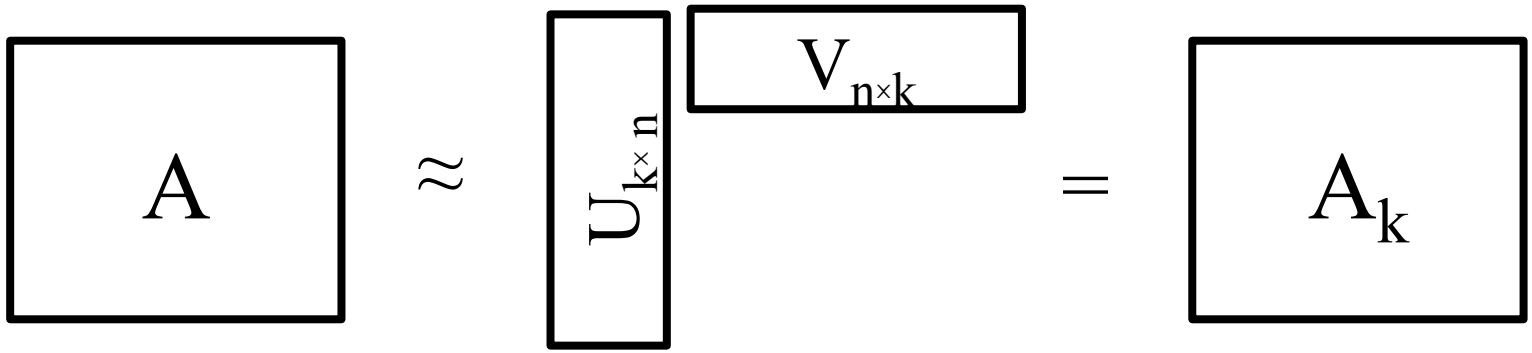
- Link prediction

Zero represents 2 things:
1) not interested
2) not aware but would be interested



Low Rank Reconstruction

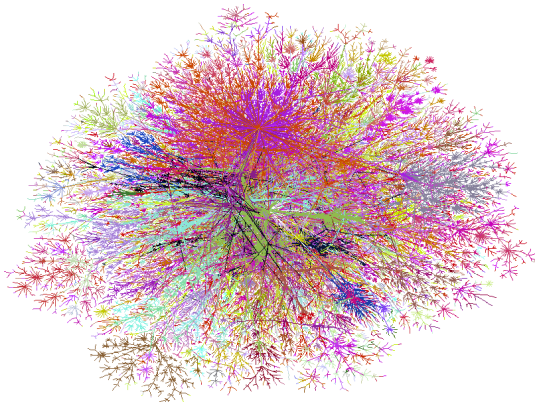
- Represent the adjacency matrix A with a lower rank matrix A_k.



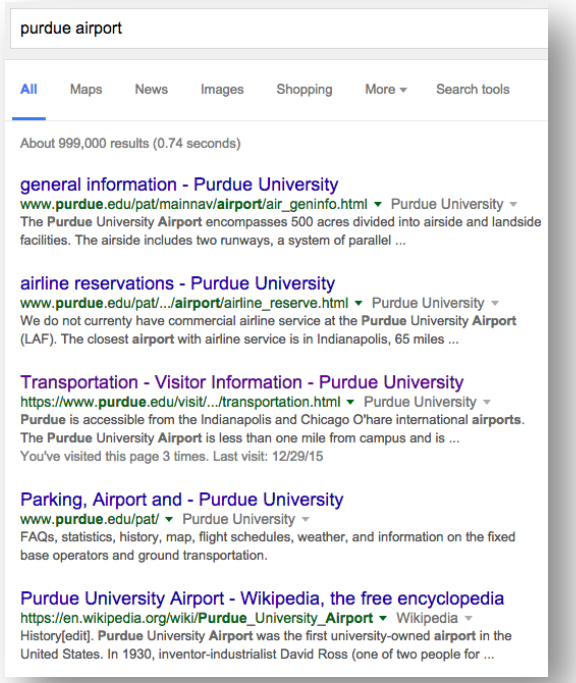
- If $A_k(u,v)$ has large value for a missing $A(u,v)=0$, then recommend link (u,v)

Graph Applications

- Node Importance



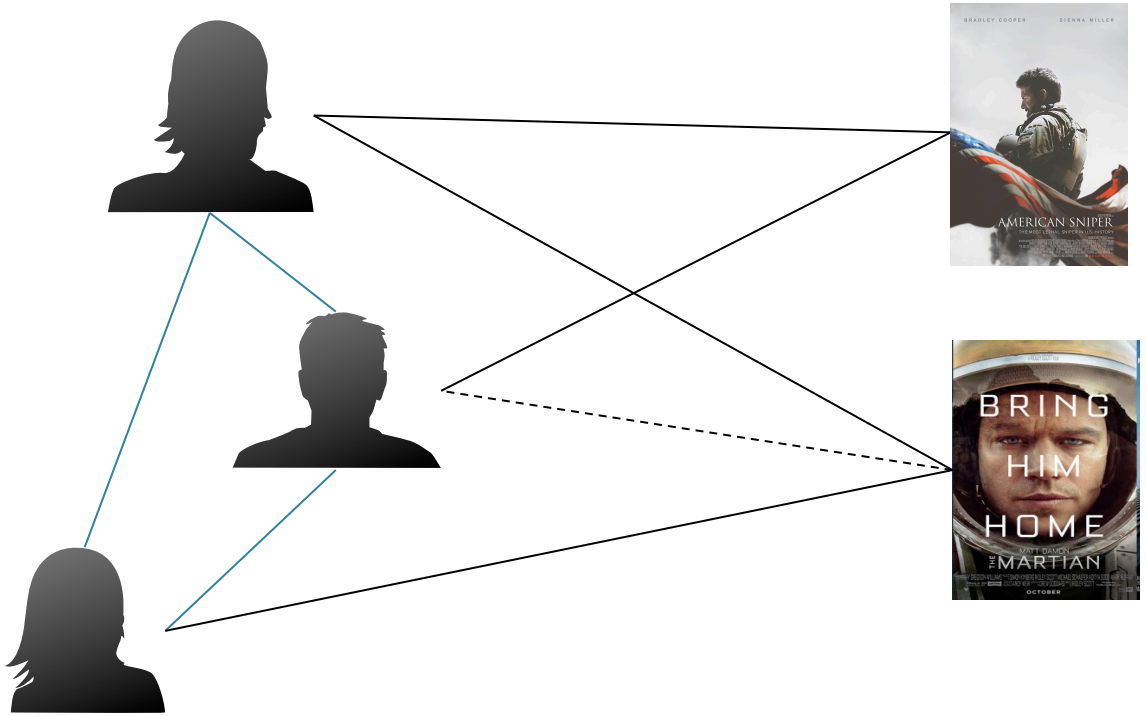
Web Graph



- Spammer Detection
  - Twitter spammers tend to form strongly connected components

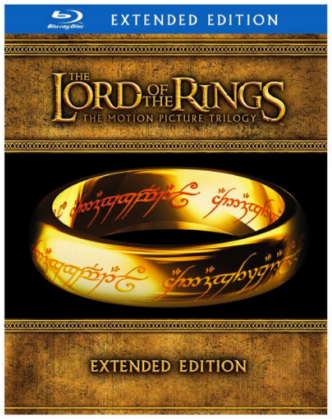
Collaborative filtering (Graph-based Approaches)

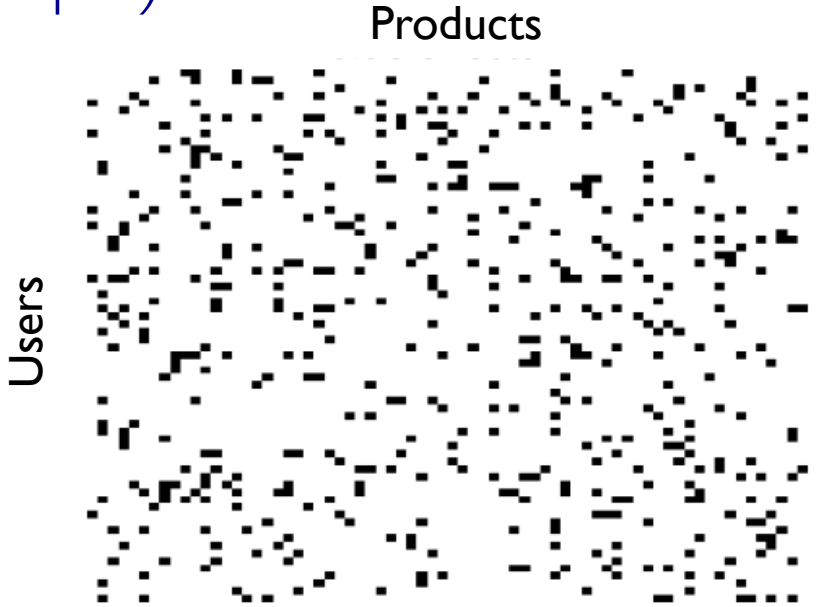
- (Social) Collaborative Filtering
  - - Based on homophily: items or users
  - - Rely on user serendipity
  - + Bias recommendations towards truly relevant items



Collaborative filtering (example)

amazon





- When searching for action movies, should we show LORT Blue Ray edition?

Amazon (Fraud)

- Detecting Fraud

Review Fraud?

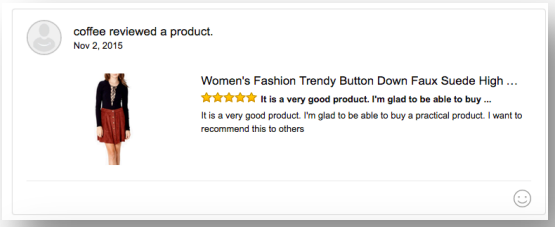
5 star 50%
4 star 0%
3 star 0%
2 star 0%
1 star 50%

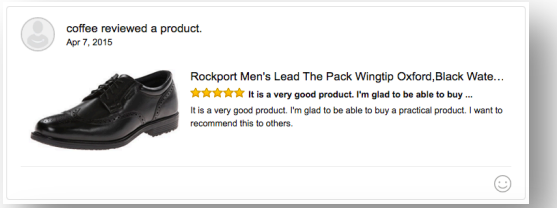
See both customer reviews ›

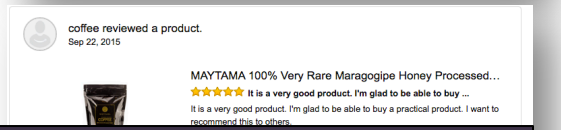
Most Helpful Customer Reviews

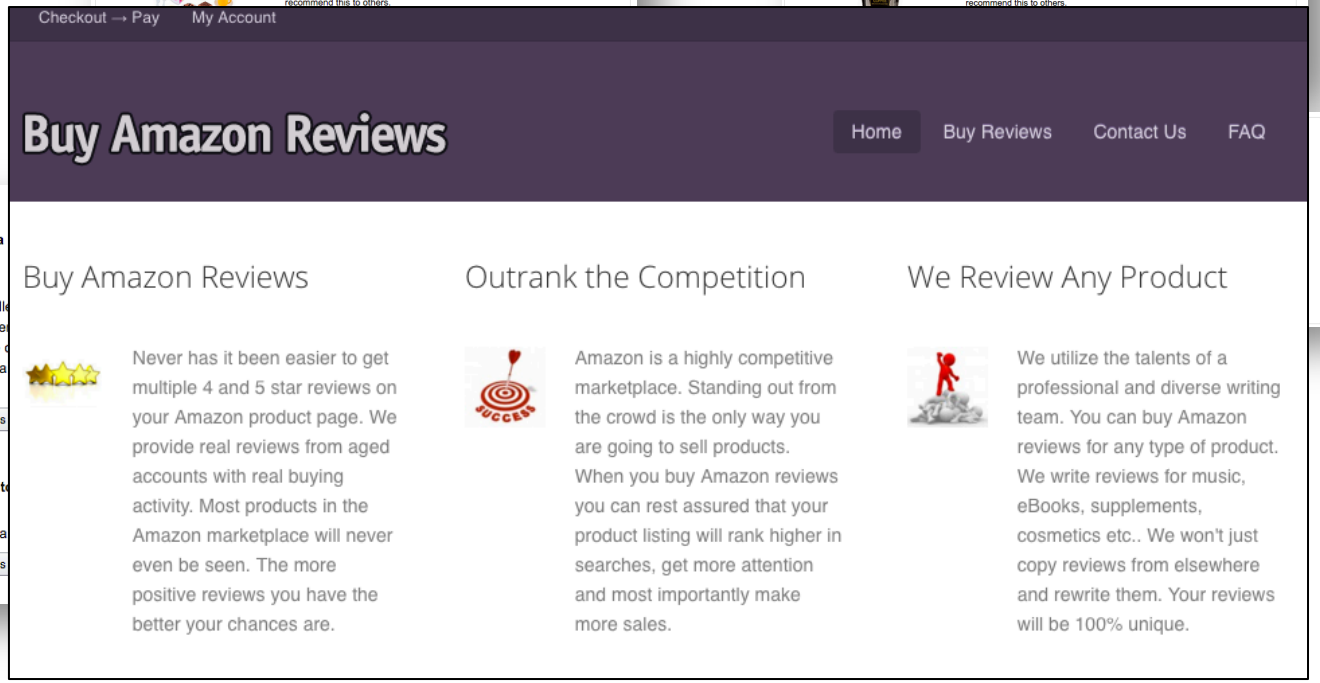
Realistic Looking Security Camera
By Kim S on September 5, 2015
Verified Purchase
This is the second on of these from two different sell
replaced. It is not convenient to keep replacing batte
one that had a light, but this is no good, just like the
flashes and I only have to replace the battery once a
length of time.
Comment | Was this review helpful to you? Yes

0 of 1 people found the following review helpful
It is a very good product. I'm glad t
By coffee on April 1, 2015
It is a very good product. I'm glad to be able to buy a
Comment | Was this review helpful to you? Yes



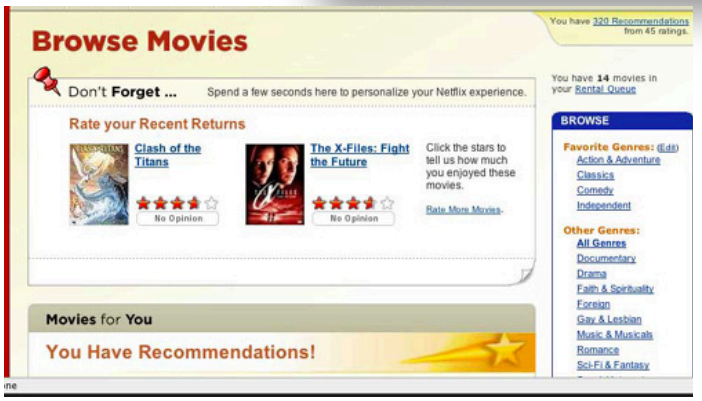


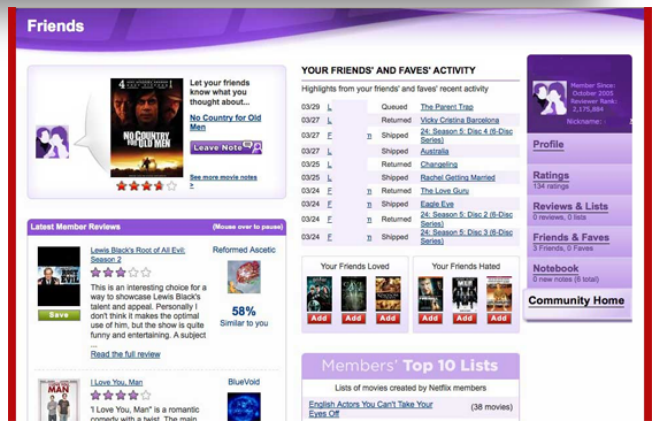


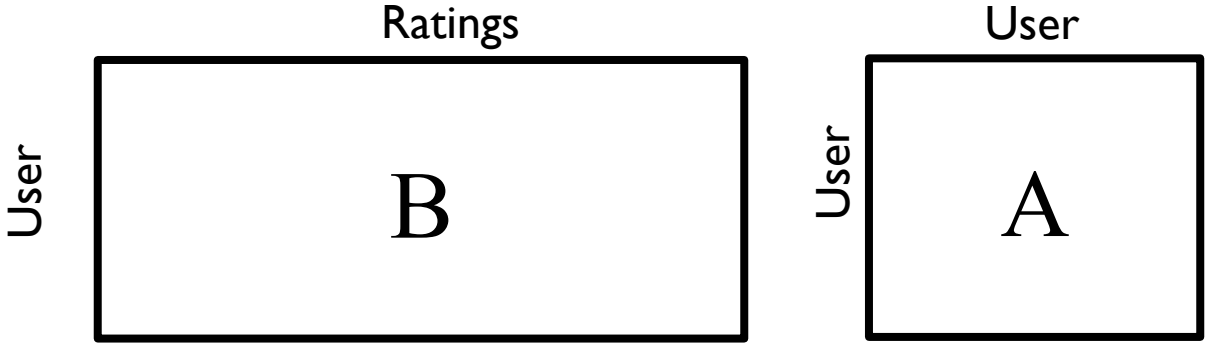


Netflix

- Joining multiple sources of information: Co-factorization



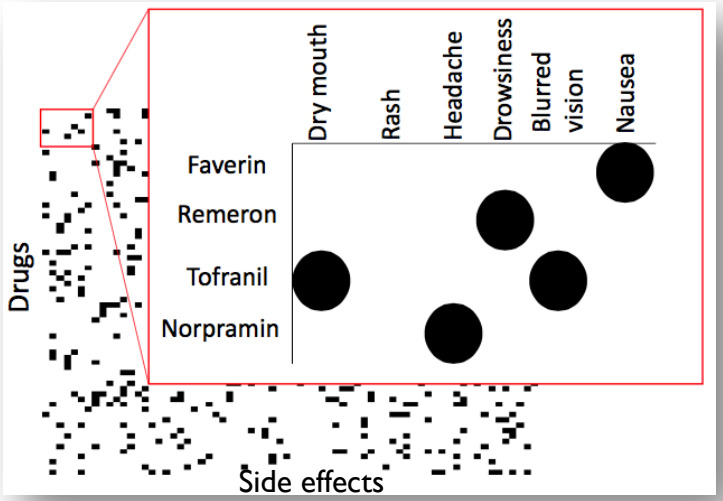




Other forms of data

Retail & Healthcare

- Classification (drug safe or not safe, user buys or does not buy)



Bryan Hooi, Hyun Ah Song, Evangelos Papalexakis, Rakesh Agrawal, Christos Faloutsos, PAKDD'16

DrugBank dataset: http://www.drugbank.ca/downloads

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

Ack: Neville

- Descriptive vs Predictive
  - Tylenol and Advil are good for pain (descriptive)
  - Drug will reduce fever (predictive)

Modeling approaches

Descriptive vs. predictive modeling

- Descriptive models summarize the data
  - Provide insights into the domain
  - Focus on modeling joint distribution $P(X)$
  - May be used for classification, but prediction is not the primary goal

- Predictive models predict the value of one variable of interest given known values of other variables
  - Focus on modeling the conditional distribution $P(Y \mid X)$ or on modeling the decision boundary for Y

Example: SPAM

- I was reading a little more about Tsalling entropy and trying to figure out whether it would be appropriate for relational learning problems. One possibility is to use it for exponential random graph models, which have features like the number of triangles in the graph. Since these grow with graph size, it seems to be an "extensive" property that the Tsalling entropy is trying to model…

- Don't Be Silly To Pay Hundred$ Or Thousand$ You Can Have The Exactly Same Licensed Software At 5%-10% Of The Retail Price : All popular softwares for PC & MAC : Language available: English, Deutsch, French, Italian, Spanish : Buy & start downloading right after you paid : You will be given your dedicated PERSONAL LICENSE right after you paid….

Data representation

- Class label: isSpam {+, -}

- Attributes?
  - Convert email text into a set of attributes

isSpam	$word_1$	$word_2$	$word_3$	...	$word_n$
+	1	0	1	...	1
-	0	0	0	...	1

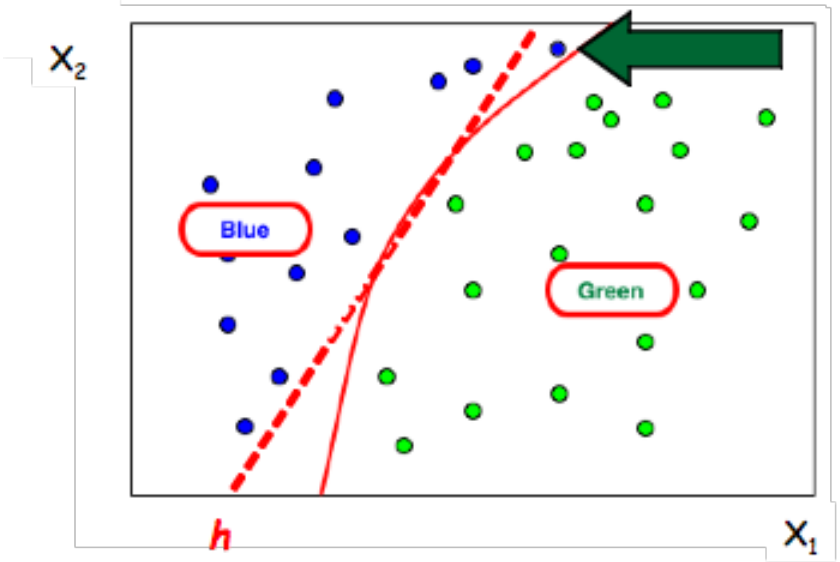
Predictive modeling

- Data representation:
  - Training set: Paired attribute vectors and class labels $<y(i), \mathbf{x}(i)>$
    or n×p tabular data with class label (y) and p-1 attributes ($\mathbf{x}$)

- Task: estimate a predictive function $f(\mathbf{x}; \theta)=y$
  - Assume that there is a function $y=f(x)$ that **maps** data instances ($\mathbf{x}$) to class labels (y)
  - Construct a model that approximates the mapping
    - Classification: if y is categorical
    - Regression: if y is real-valued

Classification

- In its simplest form, a classification model defines a decision boundary (h) and labels for each side of the boundary
- Input: $\mathbf{x}=\{x_1,x_2,...,x_n\}$ is a set of attributes, function f assigns a label y to input $\mathbf{x}$, where y is a discrete variable with a finite number of values



Classification output

- Different classification tasks can require different kinds of output
  - Each requires progressively more accurate models (e.g., a poor probability estimator can still produce an accurate ranking)
- Class labels — Each instance is assigned a single label
  - *Model only need to decide on crisp class boundaries*
- Ranking — Instances are ranked according to their likelihood of belonging to a particular class
  - *Model implicitly explores many potential class boundaries*
- Probabilities — Instances are assigned class probabilities $p(y|\mathbf{x})$
  - *Allows for more refined reasoning about sets of instances*

Discriminative classification

- Model the decision boundary directly
- Direct mapping from inputs **x** to class label y
- No attempt to model probability distributions
- May seek a discriminant function $f(\mathbf{x}; \theta)$ that maximizes measure of separation between classes
- Examples:
  - Perceptrons, nearest neighbor classifiers, support vector machines, decision trees

Probabilistic classification

- Model the underlying probability distributions
  - Posterior class probabilities: $p(y|\mathbf{x})$
  - Class-conditional and class prior: $p(\mathbf{x}|y)$ and $p(y)$
- Maps from inputs $\mathbf{x}$ to class label y indirectly through posterior class distribution $p(y|\mathbf{x})$
- Examples:
  - Naive Bayes classifier, logistic regression, linear regresssion, probability estimation trees