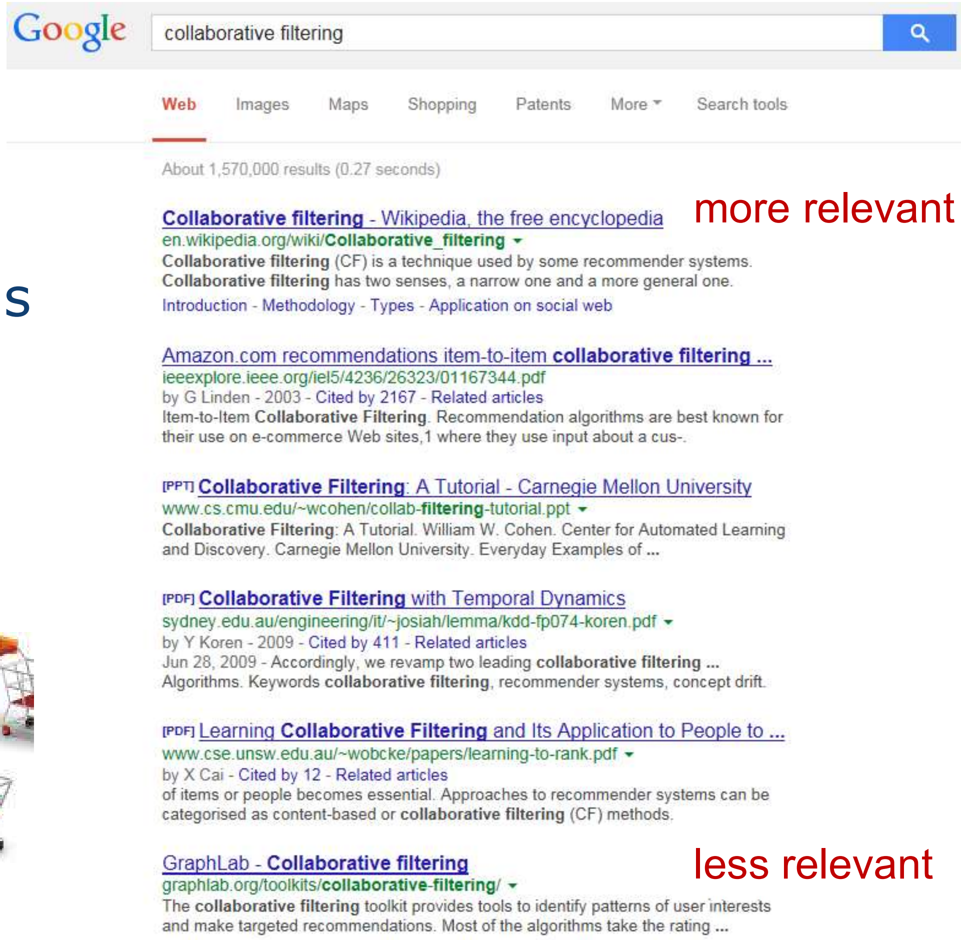
Recommendation Algorithms

Lecturer: Dr. Bo Yuan

E-mail: yuanb@sz.tsinghua.edu.cn

Overview

- TF-IDF
- Vector Space Model
- Latent Semantic Analysis
- PageRank
- Collaborative Filtering

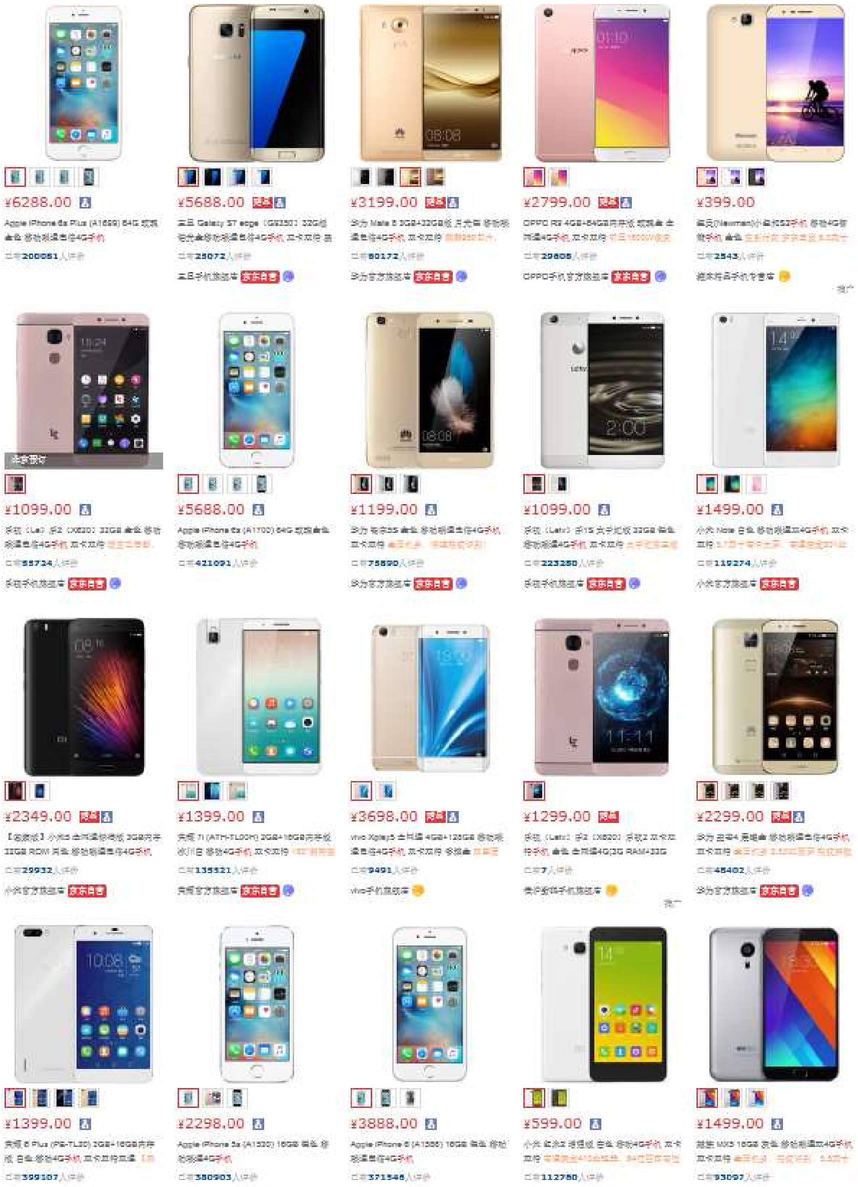


more relevant

less relevant

IN 60 SECONDS...
1 NEW DEFINITION IS ADDED ON urban
1,600+ READS ON Scribd.
13,000+ HOURS MUSIC STREAMING ON PANDORA
12,000+ NEW ADS POSTED ON craigslist
370,000+ MINUTES VOICE CALLS ON skype
98,000+ TWEETS
320+ NEW twitter ACCOUNTS
100+ NEW LinkedIn ACCOUNTS
1 associatedcontent NEW ARTICLE IS PUBLISHED
THE WORLD'S LARGEST COMMUNITY CREATED CONTENT!!
6,600+ NEW PICTURES ARE UPLOADED ON flickr
50+ WORDPRESS DOWNLOADS
125+ PLUGIN DOWNLOADS
695,000+ facebook STATUS UPDATES
79,364 WALL POSTS
510,040 COMMENTS
1,700+ Firefox DOWNLOADS
694,445 SEARCH QUERIES
Google
Google Search
168 MILLION EMAILS ARE SENT
60+ NEW BLOGS
1,500+ BLOG POSTS
70+ DOMAINS REGISTERED
25+ HOURS TOTAL DURATION
600+ NEW VIDEOS
You Tube
QUESTIONS ASKED ON THE INTERNET...
100+ Answers.com
40+ YAHOO! ANSWERS
13,000+ iPhone APPLICATIONS DOWNLOADED
20,000+ NEW POSTS ON tumblr.
THE LARGEST SOCIAL READING PUBLISHING COMPANY!!
New Craigslist Ads
GO-Globe.com

Information Overload



Today a person is subjected to more new information in a day than a person in the middle ages in his entire life!

Recommendation Systems

- A system that predicts a user's rating or preference to an item.
  - Help people discover interesting or informative stuff that they wouldn't have thought to search for.
  - One of the most influential applications of data mining.

- Content-Based Filtering
  - Focuses on the characteristics of items.
  - Recommends items similar to those that a user liked in the past.

- Collaborative Filtering
  - Predicts what users will like based on their similarity to other users.
  - Similar to asking the opinions of friends.
  - Does not rely on machine analysable contents.

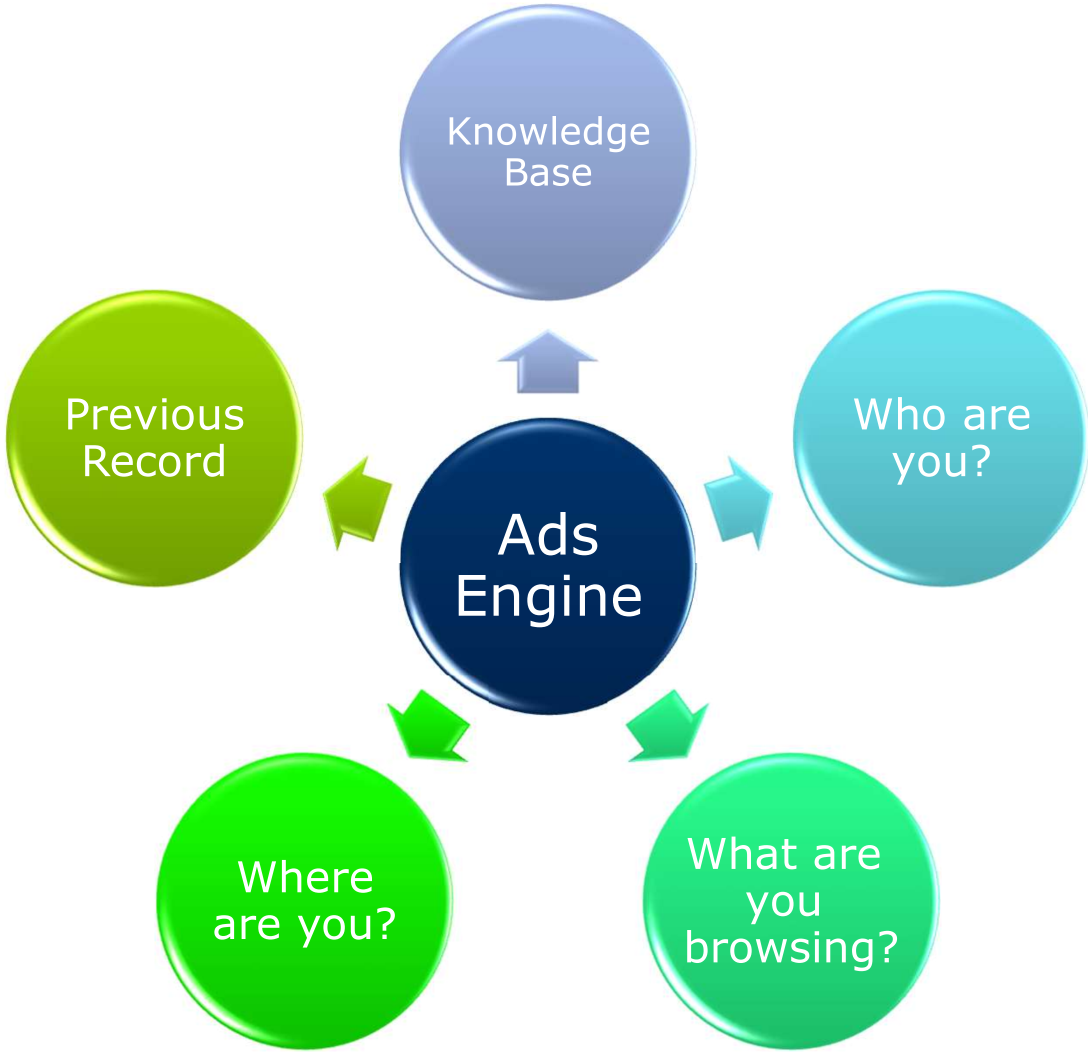
PANDORA internet radio   amazon.com   movielens helping you find the *right* movies

Junk Advertisement

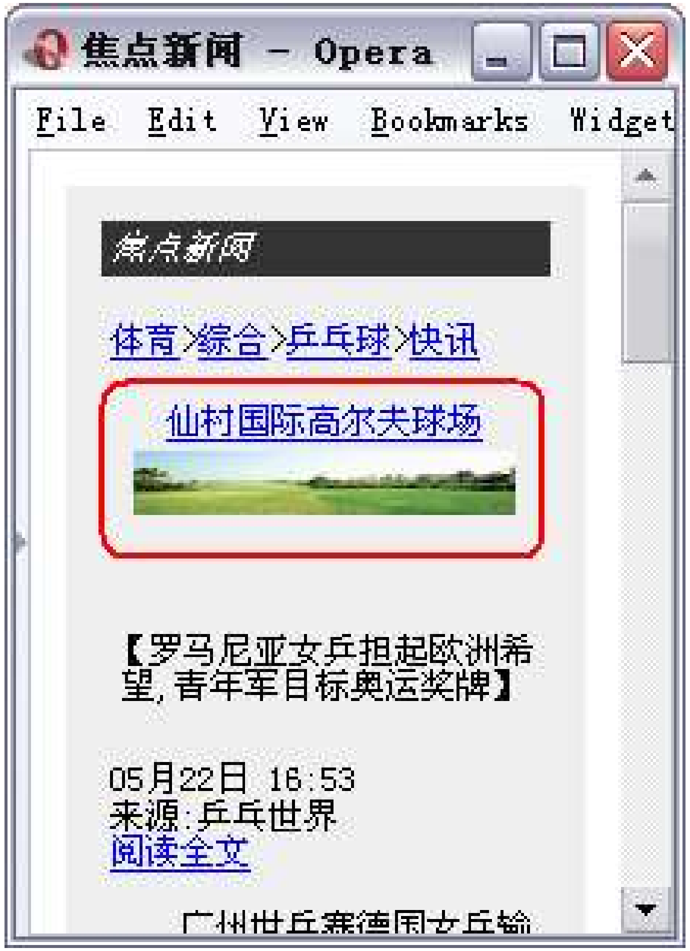
CITROËN

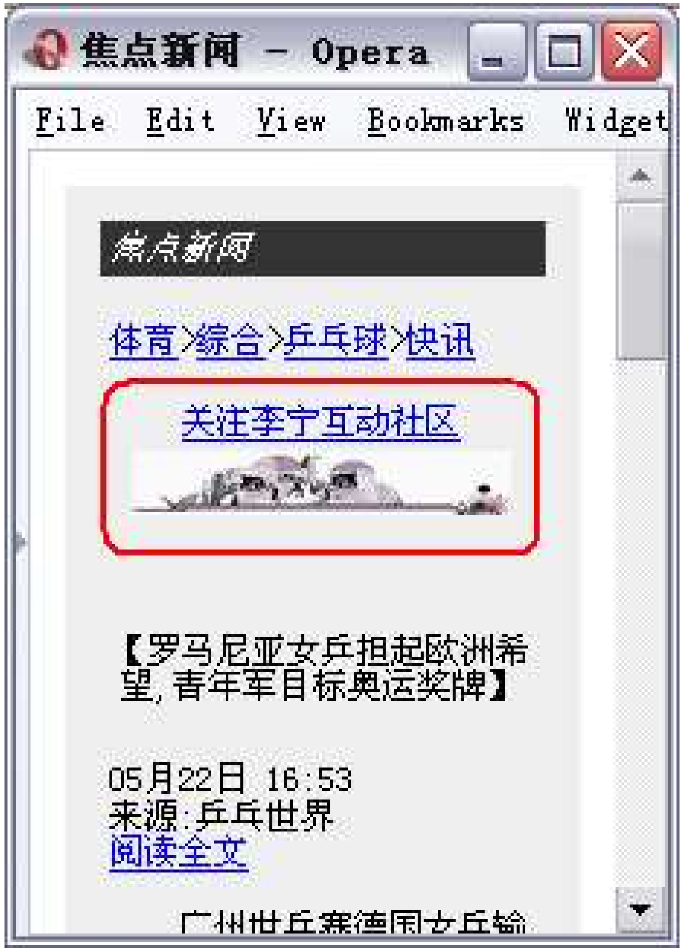
Your Trash Can Be Someone's Treasure!

Targeted Advertisement



Mobile Advertisement Platform





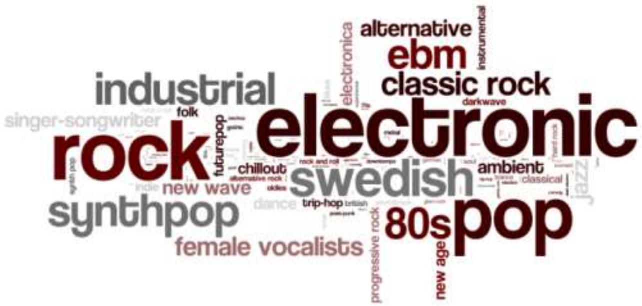
Music Recommendation

Keywords

Preference

Popularity

Ranking



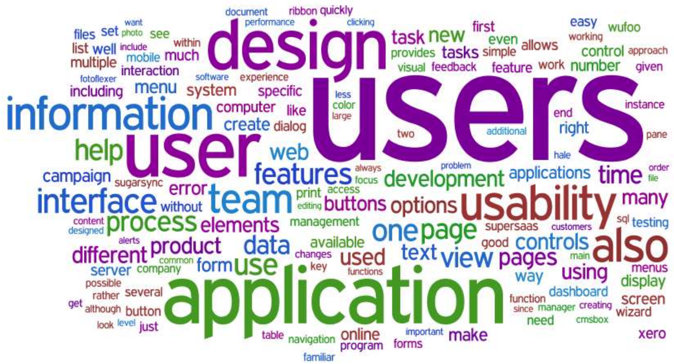
Feedback (rating, like vs. dislike)

Tf-idf

- Given a collection of documents and a query word, how relevant is a document to the word?

- Some words appear more frequently than others.

- Term Frequency (TF)
  - Raw frequency
  - $tf(t,d) = \frac{n_{t,d}}{\sum_k n_{k,d}}$



- Inverse Document Frequency (IDF)
  - $idf(t,D) = \log \frac{|D|}{|\{d \in D : t \in d\}|}$

- Tf-idf
  - tf-idf (t, d, D) = tf(t, d)×idf(t, D)

Tf-idf

- Multiple query words

$$Score(q,d) = \sum_{t \in q} tf-idf(t,d,D)$$

	Doc 1	Doc 2	Doc 3	Doc 4
the	20	10	15	8
best	0	1	0	2
car	3	5	0	0

Term-Document Matrix

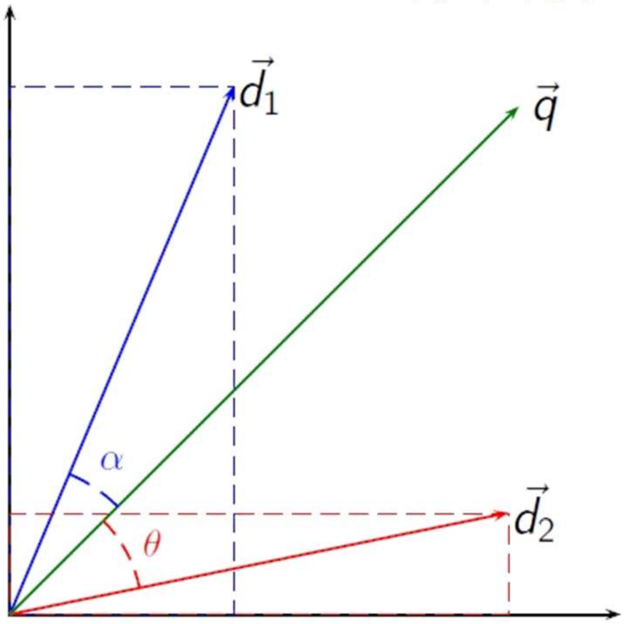
Vector Space Model

- An algebraic model for representing text documents as vectors.

$$p = \left(w_{1,p}, w_{2,p}, \ldots, w_{t,p}\right)$$

- Cosine Similarity

$$sim(p,q) = cos(\theta) = \frac{p \cdot q}{|p| \cdot |q|}$$



Vector Space Model

- Synonymy
  - Different words, same meaning
  - Car, Vehicle, Automobile
  - Small cosine values → unrelated
  - Poor recall
- Polysemy
  - One word, different meanings
  - Apple Computer vs. Apple Juice
  - Large cosine values → related
  - Poor precision

- Let’s work in a more informative space.
  - Merge dimensions with similar meanings.
  - Singular Value Decomposition

Latent Semantic Analysis

$$X = TSD^T$$

$$X : m \times n;\ T : m \times r;\ S : r \times r;\ D : n \times r;\ r = rank(X)$$

$XX^T = (TSD^T)(TSD^T)^T = T(SS^T)T^T$,

T is the eigenvectors of XX^T (dot products of terms)

Rows of TS: Coordinates of terms

$X^T X = (TSD^T)^T (TSD^T) = D(S^T S)D^T$,

D is the eigenvectors of $X^T X$ (dot products of documents)

Rows of DS: Coordinates of documents

Latent Semantic Analysis

Technical Memo Example

Titles:

c1: *Human* machine *interface* for Lab ABC *computer* applications
c2: A *survey* of *user* opinion of *computer system response time*
c3: The *EPS user interface* management *system*
c4: *System* and *human system* engineering testing of *EPS*
c5: Relation of *user*-perceived *response time* to error measurement

m1: The generation of random, binary, unordered *trees*
m2: The intersection *graph* of paths in *trees*
m3: *Graph minors* IV: Widths of *trees* and well-quasi-ordering
m4: *Graph minors*: A *survey*

Original Matrix

Terms	Documents								
	c1	c2	c3	c4	c5	m1	m2	m3	m4
human	1	0	0	1	0	0	0	0	0
interface	1	0	1	0	0	0	0	0	0
computer	1	1	0	0	0	0	0	0	0
user	0	1	1	0	1	0	0	0	0
system	0	1	1	2	0	0	0	0	0
response	0	1	0	0	1	0	0	0	0
time	0	1	0	0	1	0	0	0	0
EPS	0	0	1	1	0	0	0	0	0
survey	0	1	0	0	0	0	0	0	1
trees	0	0	0	0	0	1	1	1	0
graph	0	0	0	0	0	0	1	1	1
minors	0	0	0	0	0	0	0	1	1

Decomposition

T =

-0.2214	-0.1132	0.2890	-0.4148	-0.1063	-0.3410	-0.5227	0.0605	0.4067
-0.1976	-0.0721	0.1350	-0.5522	0.2818	0.4959	0.0704	0.0099	0.1089
-0.2405	0.0432	-0.1644	-0.5950	-0.1068	-0.2550	0.3022	-0.0623	-0.4924
-0.4036	0.0571	-0.3378	0.0991	0.3317	0.3848	-0.0029	0.0004	-0.0123
-0.6445	-0.1673	0.3611	0.3335	-0.1590	-0.2065	0.1658	-0.0343	-0.2707
-0.2650	0.1072	-0.4260	0.0738	0.0803	-0.1697	-0.2829	0.0161	0.0539
-0.2650	0.1072	-0.4260	0.0738	0.0803	-0.1697	-0.2829	0.0161	0.0539
-0.3008	-0.1413	0.3303	0.1881	0.1148	0.2722	-0.0330	0.0190	0.1653
-0.2059	0.2736	-0.1776	-0.0324	-0.5372	0.0809	0.4669	0.0363	0.5794
-0.0127	0.4902	0.2311	0.0248	0.5942	-0.3921	0.2883	-0.2546	0.2254
-0.0361	0.6228	0.2231	0.0007	-0.0683	0.1149	-0.1596	0.6811	-0.2320
-0.0318	0.4505	0.1411	-0.0087	-0.3005	0.2773	-0.3395	-0.6784	-0.1825

S =

3.3409	0	0	0	0	0	0	0	0
0	2.5417	0	0	0	0	0	0	0
0	0	2.3539	0	0	0	0	0	0
0	0	0	1.6445	0	0	0	0	0
0	0	0	0	1.5048	0	0	0	0
0	0	0	0	0	1.3064	0	0	0
0	0	0	0	0	0	0.8459	0	0
0	0	0	0	0	0	0	0.5601	0
0	0	0	0	0	0	0	0	0.3637

Decomposition

D =

-0.1974	-0.0559	0.1103	-0.9498	0.0457	-0.0766	-0.1773	0.0144	0.0637
-0.6060	0.1656	-0.4973	-0.0286	-0.2063	-0.2565	0.4330	-0.0493	-0.2428
-0.4629	-0.1273	0.2076	0.0416	0.3783	0.7244	0.2369	-0.0088	-0.0241
-0.5421	-0.2318	0.5699	0.2677	-0.2056	-0.3689	-0.2648	0.0195	0.0842
-0.2795	0.1068	-0.5054	0.1500	0.3272	0.0348	-0.6723	0.0583	0.2624
-0.0038	0.1928	0.0982	0.0151	0.3948	-0.3002	0.3408	-0.4545	0.6198
-0.0146	0.4379	0.1930	0.0155	0.3495	-0.2122	0.1522	0.7615	-0.0180
-0.0241	0.6151	0.2529	0.0102	0.1498	0.0001	-0.2491	-0.4496	-0.5199
-0.0820	0.5299	0.0793	-0.0246	-0.6020	0.3622	-0.0380	0.0696	0.4535

T*S*D'

1.0000	0.0000	0.0000	1.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
1.0000	0.0000	1.0000	0.0000	0.0000	0.0000	-0.0000	-0.0000	-0.0000
1.0000	1.0000	0.0000	-0.0000	-0.0000	0.0000	0.0000	0.0000	0.0000
-0.0000	1.0000	1.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
-0.0000	1.0000	1.0000	2.0000	0.0000	-0.0000	-0.0000	-0.0000	-0.0000
-0.0000	1.0000	0.0000	0.0000	1.0000	-0.0000	-0.0000	0.0000	0.0000
-0.0000	1.0000	0.0000	0.0000	1.0000	-0.0000	-0.0000	0.0000	0.0000
-0.0000	0.0000	1.0000	1.0000	0.0000	0.0000	0.0000	-0.0000	-0.0000
-0.0000	1.0000	0.0000	-0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
-0.0000	0.0000	-0.0000	-0.0000	0.0000	1.0000	1.0000	1.0000	0.0000
-0.0000	-0.0000	-0.0000	-0.0000	0.0000	0.0000	1.0000	1.0000	1.0000
-0.0000	0.0000	-0.0000	-0.0000	0.0000	0.0000	0.0000	1.0000	1.0000

Rank K Approximation

K=2

T(:, 1:2)

-0.2214	-0.1132
-0.1976	-0.0721
-0.2405	0.0432
-0.4036	0.0571
-0.6445	-0.1673
-0.2650	0.1072
-0.2650	0.1072
-0.3008	-0.1413
-0.2059	0.2736
-0.0127	0.4902
-0.0361	0.6228
-0.0318	0.4505

S(1:2, 1:2)

3.3409	0
0	2.5417

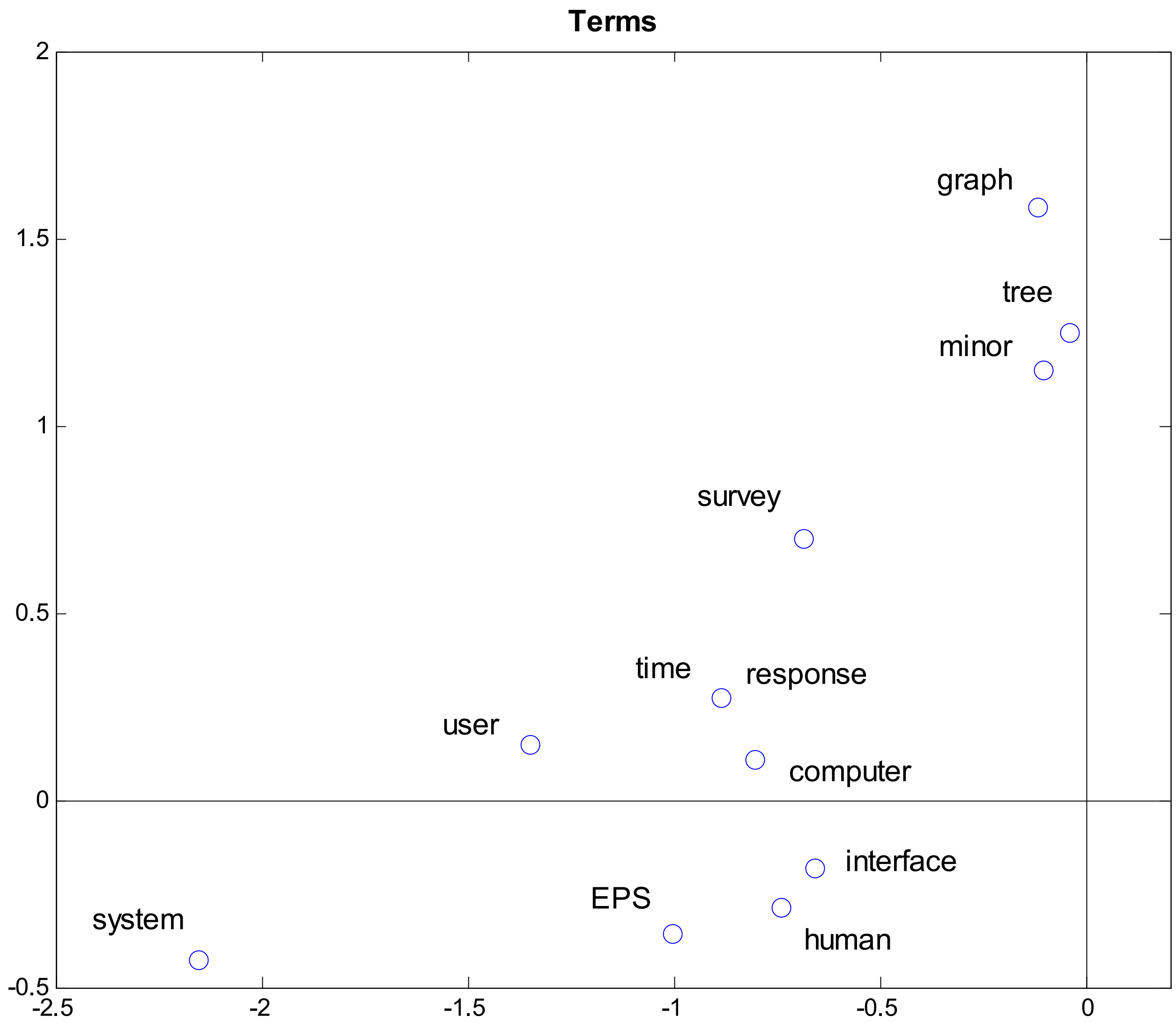
D(:, 1:2)'

-0.1974	-0.6060	-0.4629	-0.5421	-0.2795	-0.0038	-0.0146	-0.0241	-0.0820
-0.0559	0.1656	-0.1273	-0.2318	0.1068	0.1928	0.4379	0.6151	0.5299

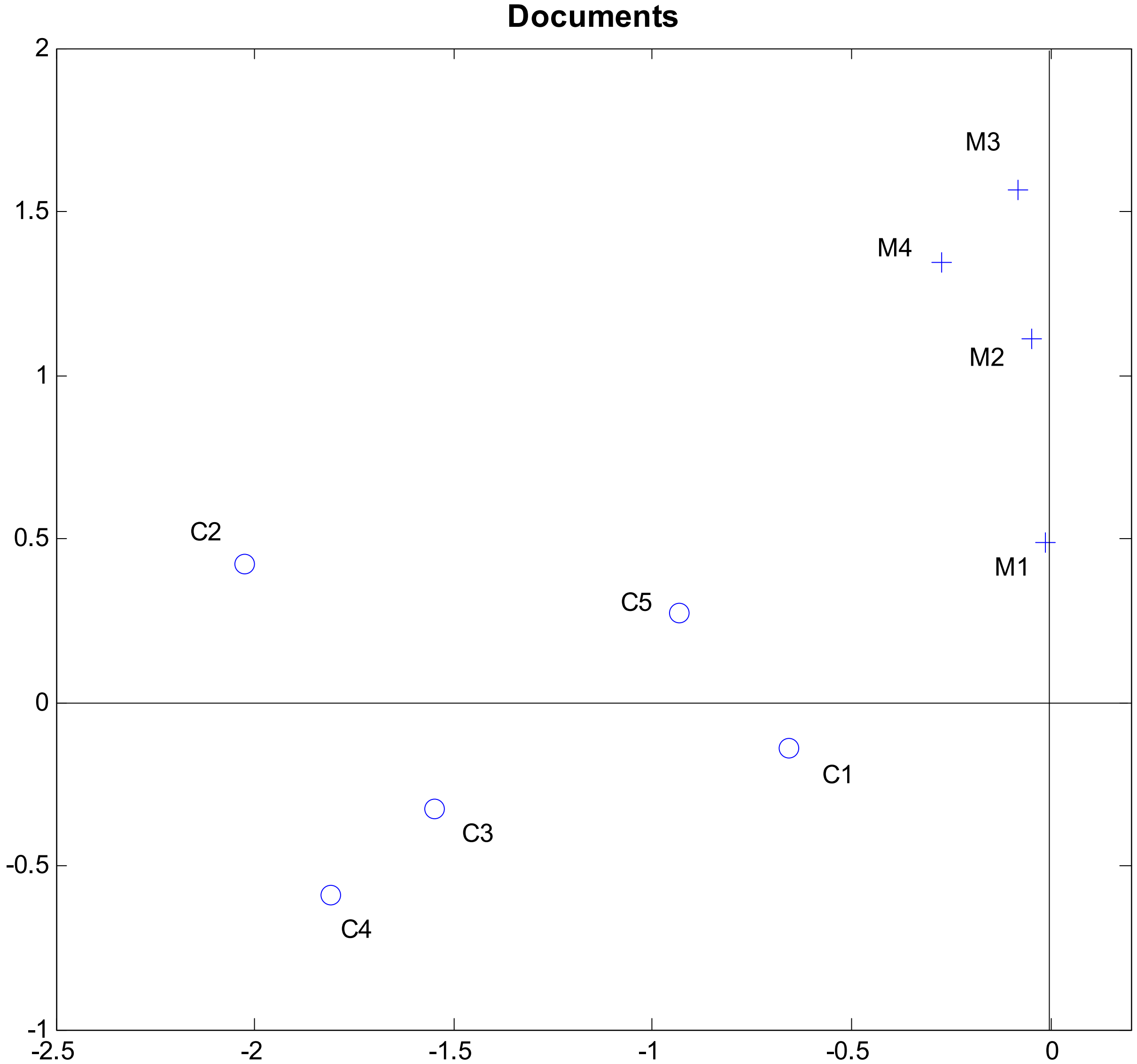
$\hat{X}$

0.1621	0.4005	0.3790	0.4676	0.1760	-0.0527	-0.1151	-0.1591	-0.0918
0.1406	0.3698	0.3290	0.4004	0.1650	-0.0328	-0.0706	-0.0968	-0.0430
0.1524	0.5050	0.3579	0.4101	0.2362	0.0242	0.0598	0.0869	0.1240
0.2580	0.8411	0.6057	0.6974	0.3923	0.0331	0.0832	0.1218	0.1874
0.4488	1.2344	1.0509	1.2658	0.5563	-0.0738	-0.1547	-0.2096	-0.0489
0.1596	0.5817	0.3752	0.4169	0.2765	0.0559	0.1322	0.1889	0.2169
0.1596	0.5817	0.3752	0.4169	0.2765	0.0559	0.1322	0.1889	0.2169
0.2185	0.5496	0.5110	0.6281	0.2425	-0.0654	-0.1425	-0.1966	-0.1079
0.0969	0.5321	0.2299	0.2118	0.2665	0.1368	0.3146	0.4444	0.4250
-0.0613	0.2321	-0.1389	-0.2656	0.1449	0.2404	0.5461	0.7674	0.6637
-0.0647	0.3353	-0.1456	-0.3014	0.2028	0.3057	0.6949	0.9766	0.8487
-0.0431	0.2539	-0.0967	-0.2079	0.1519	0.2212	0.5029	0.7069	0.6155

Items in 2D Space



Documents in 2D Space



Document Cosine Similarity

0	0.9142	1.0000	0.9948	0.8799	-0.1852	-0.1676	-0.1600	-0.0117
0.9142	0	0.9166	0.8681	0.9970	0.2289	0.2463	0.2537	0.3945
1.0000	0.9166	0	0.9942	0.8827	-0.1793	-0.1617	-0.1541	-0.0057
0.9948	0.8681	0.9942	0	0.8268	-0.2845	-0.2673	-0.2599	-0.1137
0.8799	0.9970	0.8827	0.8268	0	0.3040	0.3210	0.3282	0.4648
-0.1852	0.2289	-0.1793	-0.2845	0.3040	0	0.9998	0.9997	0.9848
-0.1676	0.2463	-0.1617	-0.2673	0.3210	0.9998	0	1.0000	0.9878
-0.1600	0.2537	-0.1541	-0.2599	0.3282	0.9997	1.0000	0	0.9889
-0.0117	0.3945	-0.0057	-0.1137	0.4648	0.9848	0.9878	0.9889	0

Transformed

0	0.2357	0.2887	0.2357	0	0	0	0	0
0.2357	0	0.4082	0.3333	0.7071	0	0	0	0.2357
0.2887	0.4082	0	0.6124	0.2887	0	0	0	0
0.2357	0.3333	0.6124	0	0	0	0	0	0
0	0.7071	0.2887	0	0	0	0	0	0
0	0	0	0	0	0	0.7071	0.5774	0
0	0	0	0	0	0.7071	0	0.8165	0.4082
0	0	0	0	0	0.5774	0.8165	0	0.6667
0	0.2357	0	0	0	0	0.4082	0.6667	0

Original

Query

$$Query = "human\ response"$$

$$q = [1,0,0,0,0,1,0,0,0,0,0,0]^T$$

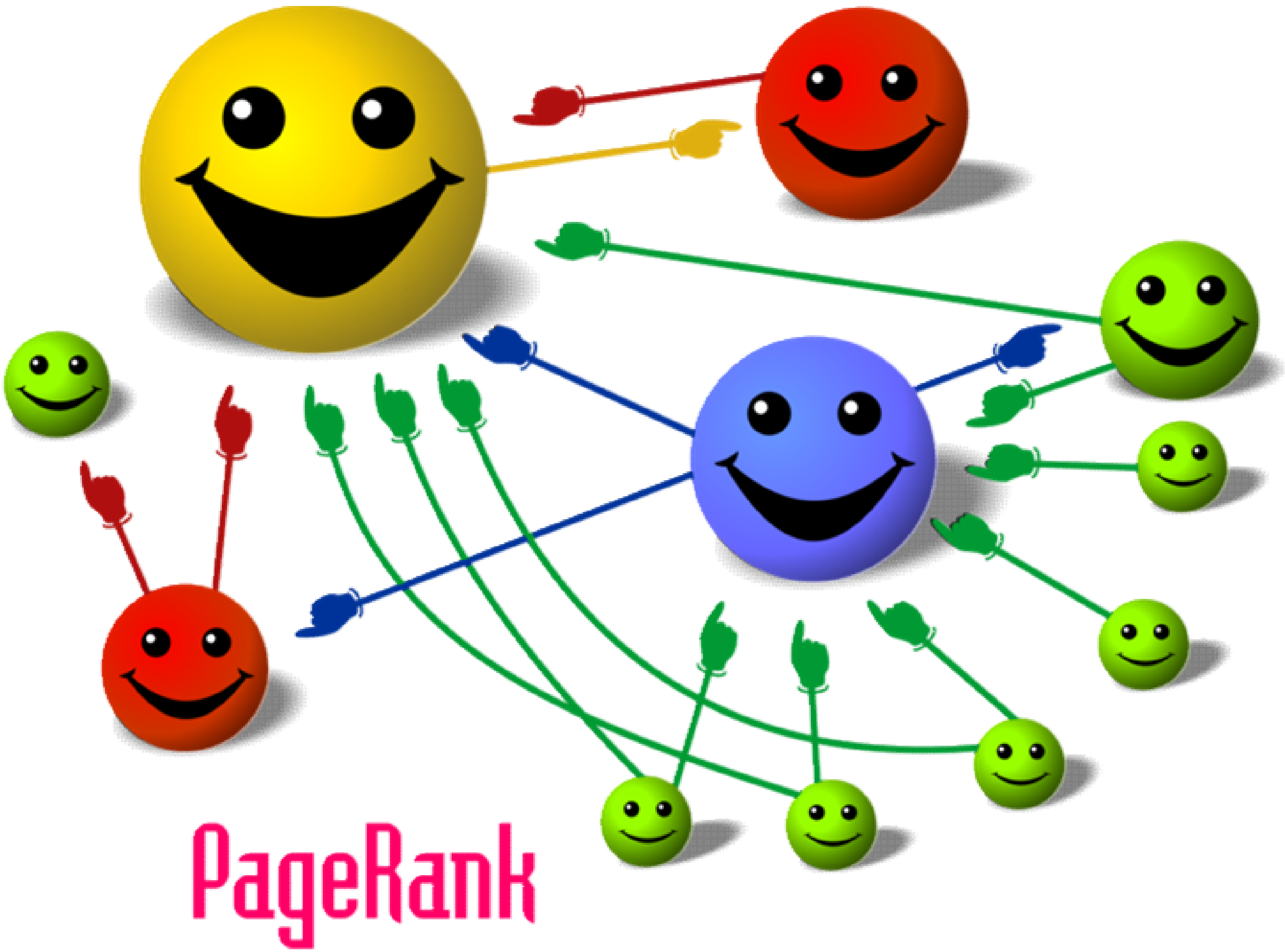
$$\hat{q} = S_k^{-1} T_k^T q = [-0.1456, -0.0024]^T$$

$$\hat{q}^T S = [-0.4864, -0.0060]$$

$$\overbrace{0.9809 \quad 0.9756 \quad 0.9820 \quad 0.9559 \quad 0.9556}^{C} \quad \overbrace{0.0097 \quad 0.0276 \quad 0.0353 \quad 0.1832}^{M}$$

Cosine Similarity to Current Documents

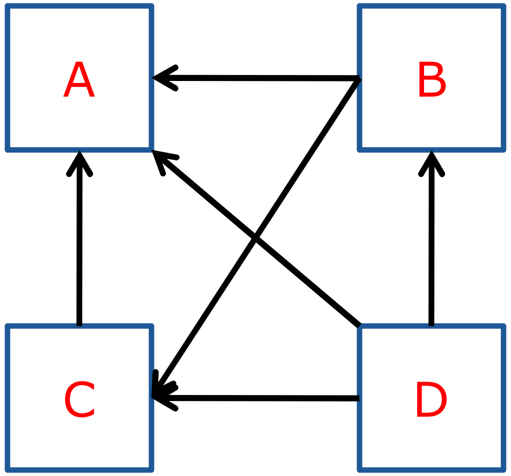
Linked Documents



PageRank

- Given a set of hyperlinked documents, how to evaluate the relative importance of each document?
- A hyperlink to a page counts as a vote of support.
  - The importance of vote from a page depends on its own PageRank and the number of outbound links.
- The PageRank of a page is determined by the number and PageRank metric of all pages that link to it.
- The outbound links of a page do not affect its PageRank value.
  - Difficult to manipulate inbound links.
- A key factor determining a page’s ranking in the search results of Google.

PageRank



$$PR(A) = \frac{PR(B)}{2} + \frac{PR(C)}{1} + \frac{PR(D)}{3}$$

$$PR(P_i) = \sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$$

$$\boxed{PR(P_i; t+1) = \frac{1-d}{N} + d \sum_{p_j \in M(p_i)} \frac{PR(p_j; t)}{L(p_j)}}$$

d: damping factor (0.85)

PageRank

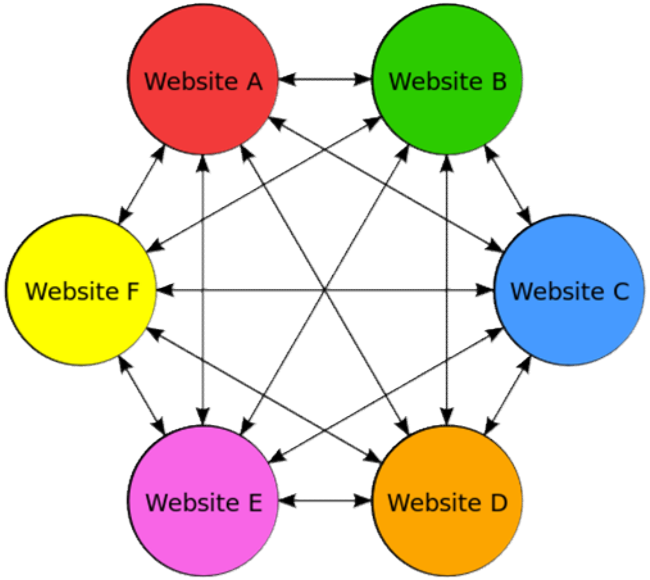
$$R(t+1) = dMR(t) + \frac{1-d}{N} l$$

$$R_i(t) = PR(p_i;t) \quad PR(p_i;0) = \frac{1}{N} \qquad d = 0.85$$

$$M_{ij} = \begin{cases} 1/L(p_j), & \text{if } j \text{ links to } i \\ 0, & \text{otherwise} \end{cases} \qquad l = \text{ones}(N,1)$$

$$R = dMR + \frac{1-d}{N} l, \qquad \text{for } t \to \infty$$

$$\boxed{R = (I - dM)^{-1} \frac{1-d}{N} l}$$



Monetary Success

- Stanford University received 1.8 million shares for allowing Google Inc. to use this technique.
  - Sergey Brin: US$ 37 billion (2016)
  - Larry Page: US$ 38 billion (2016)

- Made totally US$ 336 million in return by 2005.
  - Within two years after Google's IPO
  - Around US$ 187 per share
  - How about if the shares are sold today?

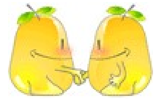
Google Inc
NASDAQ: GOOG - 27/11 4:00 pm ET
1,063.11 ↑ 4.70 (0.44%)

- Current Endowment: US$ 22 billion

- One of the largest single academic licensing transactions
  - Cloning Technology: US$ 225 million in royalties

Collaborative Filtering

- Core Idea:
  - People get the best recommendation from others with similar tastes.



- Workflow:
  - Creates a rating or purchase matrix.
  - Finds similar people by matching their ratings.
  - Recommends items that similar people rate highly.

- Memory-Based CF
  - User-Based vs. Item-Based

- Model-Based CF

- Things to know:
  - Gray Sheep
  - Shilling Attack
  - Cold Start

User-Based CF

	I_1	I_2	I_3	I_4
U_1	4	?	5	5
U_2	4	2	1	
U_3	3		2	4
U_4	4	4		
U_5	2	1	3	5

$$w_{u,v} = \frac{\sum_{i \in I} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I} (r_{v,i} - \bar{r}_v)^2}}$$

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_u) \cdot w_{a,u}}{\sum_{u \in U} |w_{a,u}|}$$

User-Based CF

	I_1	I_2	I_3	I_4
U_1	4	?	5	5
U_2	4	2	1	
U_3	3		2	4
U_4	4	4		
U_5	2	1	3	5

$$P_{1,2} = \overline{r}_1 + \frac{\sum_u (r_{u,2} - \overline{r}_u) \cdot w_{1,u}}{\sum_u |w_{1,u}|}$$

$$= \overline{r}_1 + \frac{(r_{2,2} - \overline{r}_2) w_{1,2} + (r_{4,2} - \overline{r}_4) w_{1,4} + (r_{5,2} - \overline{r}_5) w_{1,5}}{|w_{1,2}| + |w_{1,4}| + |w_{1,5}|}$$

$$= 4.67 + \frac{(2 - 2.5)(-1) + (4 - 4)0 + (1 - 3.33)0.756}{1 + 0 + 0.756}$$

$$= 3.95.$$

Item-Based CF

$$w_{i,j} = \frac{\sum_{u \in U} (r_{u,i} - \overline{r_i})(r_{u,j} - \overline{r_j})}{\sqrt{\sum_{u \in U} (r_{u,i} - \overline{r_i})^2} \sqrt{\sum_{u \in U} (r_{u,j} - \overline{r_j})^2}}$$

U: Users that have rated both *i* and *j*.

$$P_{a,i} = \frac{\sum_{j \in I} w_{i,j} \cdot r_{a,j}}{\sum_{j \in I} \left| w_{i,j} \right|}$$

I: All items that have been rated by User *a*.

Item-Based CF

$$P_{a,i} = \overline{r_a} + \frac{1}{|U|}\sum_{u \in U}(r_{u,i} - \overline{r_u})$$

U: Users that have rated *i*.

$$dev_{i,j} = \frac{1}{|U|}\sum_{u \in U}(r_{u,i} - r_{u,j})$$

U: Users that have rated both *i* and *j*.

$$P_{a,i} = \frac{1}{|I|}\sum_{j \in I}(dev_{i,j} + r_{a,j})$$

I: Items that the user has rated and have *dev* values.

Item-Based CF

Customer	Item 1	Item 2	Item 3
John	5	3	2
Mark	3	4	Didn't rate it
Lucy	Didn't rate it	2	5

$$P_{Lucy,1} = \frac{2+5}{2} + \frac{5-2.5+3-4}{2} = 4.25$$

$$dev_{1,2} = \frac{2-1}{2} = 0.5 \qquad dev_{1,3} = \frac{3}{1} = 3$$

$$P_{Lucy,1} = \frac{1}{2}(0.5+2+3+5) = 5.25$$

$$P_{Lucy,1} = \frac{2\times 2.5+1\times 8}{2+1} = 4.33$$

Slope One

Model-Based CF

Class Label

Attributes

	I_1	I_2	I_3	I_4
U_1	4	?	5	5
U_2	4	2	1	
U_3	3		2	4
U_4	4	4		
U_5	2	1	3	5

Training Samples

$$class = \underset{j \in classSet}{\arg\max}\, P(class_j) \prod_{o} P(X_o = x_o | class_j)$$

Model-Based CF

	I_1	I_2	I_3	I_4
U_1	4	?	5	5
U_2	4	2	1	
U_3	3		2	4
U_4	4	4		
U_5	2	1	3	5

$$class = \underset{c_j \in \{1,2,3,4,5\}}{\arg\max} P(c_j) P(U_2 = 2 | c_j) P(U_4 = 4 | c_j) P(U_5 = 1 | c_j)$$

$$= \underset{c_j \in \{1,2,3,4,5\}}{\arg\max} \{0, 0, 0, 0.0031, 0.0019\} = 4$$

$$P(5)P(U_2 = 2|5)P(U_4 = 4|5)P(U_5 = 1|5)$$

$$= \frac{2}{3} \times \frac{0+1}{2+5} \times \frac{0+1}{2+5} \times \frac{0+1}{2+5} = 0.0019$$

$$P(4)P(U_2 = 2|4)P(U_4 = 4|4)P(U_5 = 1|4)$$

$$= \frac{1}{3} \times \frac{0+1}{1+5} \times \frac{1+1}{1+5} \times \frac{0+1}{1+5} = 0.0031$$

Model-Based CF

	I_1	I_2	I_3	I_4	I_5
U_1	Like	Dislike	Dislike		Like
U_2	Dislike			Dislike	Dislike
U_3		Like	Like		Like
Class Label	Like	Dislike	Like	Like	?

	I_1	I_2	I_3	I_4	I_5
U_1 like	1	0	0	0	1
U_1 dislike	0	1	1	0	0
U_2 like	0	0	0	0	0
U_2 dislike	1	0	0	1	1
U_3 like	0	1	1	0	1
U_3 dislike	0	0	0	0	0
Class Label	Like	Dislike	Like	Like	?

Netflix Prize

- A public company providing DVD-rental service
- Target:
  - To predict whether someone will enjoy a movie based on how much they liked or disliked other movies.
  - To improve the score of its own Cinematch by 10%
  - RMSE (Root Mean Squared Error)

- Training Set:
  - <user, movie, date of grade, grade>
  - 480,189 users, 17,770 movies,100,480,507 ratings

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22

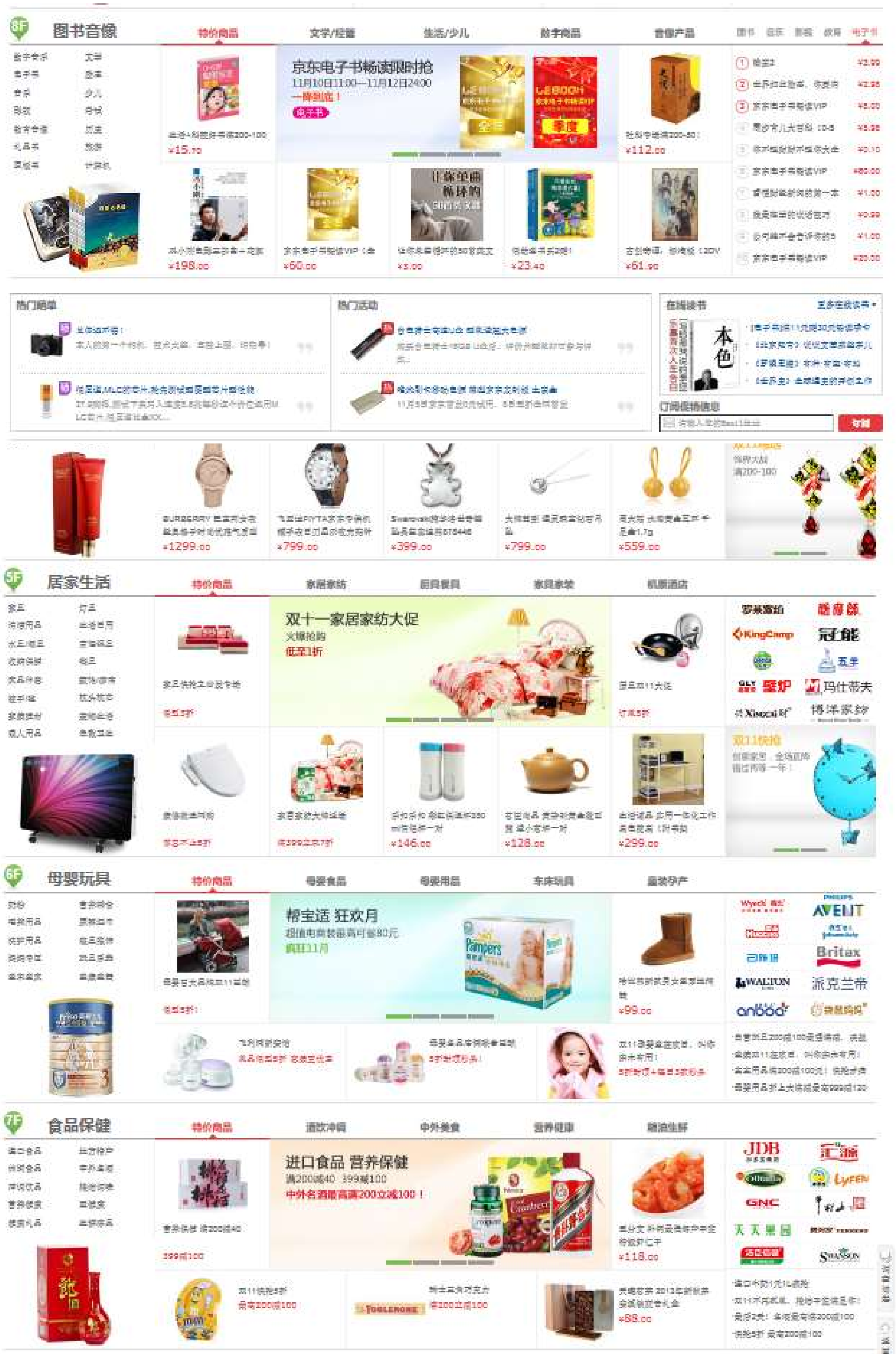
KDD Cup


Tencent
2012
KDD Cup
Sponsored by Tencent Inc.


Jobs You May Be Interested In beta
Principal Development...
Microsoft - Beijing, CN
Business Development Director
NorCap China Internships - Beijing
Director - Shenzhen
Michael Page China - Shenzhen
Feedback | See more »
Get hired faster with Job Seeker Premium


Groups You May Like
Memetic Algorithms
Join - Professional Group
Artificial Intelligence in Healthcare and Life Sciences (AIHLS)
Join - Professional Group
HAIS Series of Conferences
Join - Conference Group
Feedback | See more »

双11快抢
快抢京光 最后1天
JD.COM 京东
全部商品分类
首页
服饰城
京东超市
团购
夺宝岛
在线商城
服饰鞋帽优惠返场
双11快抢
11.12 男欢专场
最后1天
GO >
手机疯抢
5折起
家电
低至5折
电脑最高
直降5000
京东快报
家电通讯
特价商品
大家电
小家电
手机通讯
汽车五金
电脑数码
笔记本
数码影音
DIY装机
办公打印
汪疯了
再High一把
狂欢最后一天一夜！
外设满购全城
服饰鞋包
品牌男装
品牌女装
鞋靴箱包
运动户外
海外购



图书音像
特价商品
文学/经管
生活/少儿
数字商品
音像产品
京东电子书畅读限时抢
热门榜单
热门活动
在线读书
订阅促销信息
居家生活
家居家纺
厨房餐具
家具家装
机票酒店
双十一家居家纺大促
火爆抢购
低至1折
母婴玩具
母婴食品
母婴用品
车床玩具
备孕孕产
帮宝适 狂欢月
食品保健
酒饮冲调
中外美食
营养健康
粮油生鲜
进口食品 营养保健
满200减40 399减100
中外名酒都满200立减100！


新蛋11周年
因为你
百货优惠专场
买满即送 限时限量 抢到就是赚到！
活动时间：11月5日-11月8日


13周年
店庆月
惹火行动
第3波
2012.11.5-2012.11.11
家居爆品低至2折
最高直减400元
满减飓风
满149减50
满300减100
满600减220
满1000立减400
家纺
厨具
生活日用
家具装饰
宠物


13周年
店庆月
惹火行动
第3波
童装童鞋 百家名店
当当价基础上5折
仅3天
满500送100现金券


1元抢大牌
不畏光棍节 狠狠爱自己
LV,Gucci,Chanel，Coach血拼
一年等一回!


光棍节 让你不孤独
百货全场满200减50
最高满减
9900


13周年
店庆月
惹火行动
第3波
11.5-11
家电品牌旗舰店
秒杀
低至1元
大牌暴降
1000元
周年庆钜惠

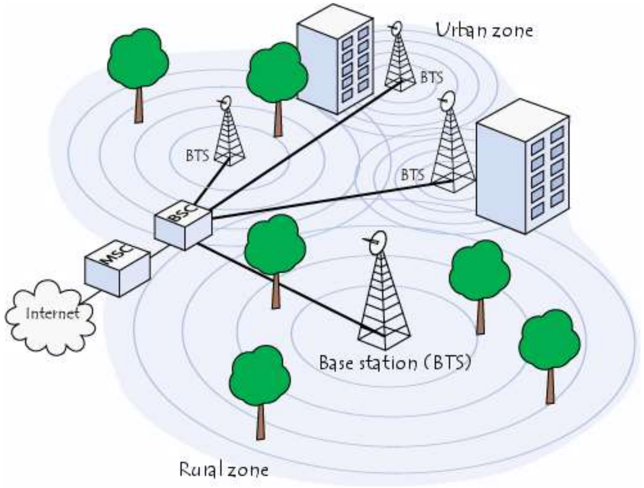

2012.11.11
支付宝总交易额 191亿元
其中 天猫 132亿元
淘宝 59 亿元
11.11 购物狂欢节
上天猫 就购了

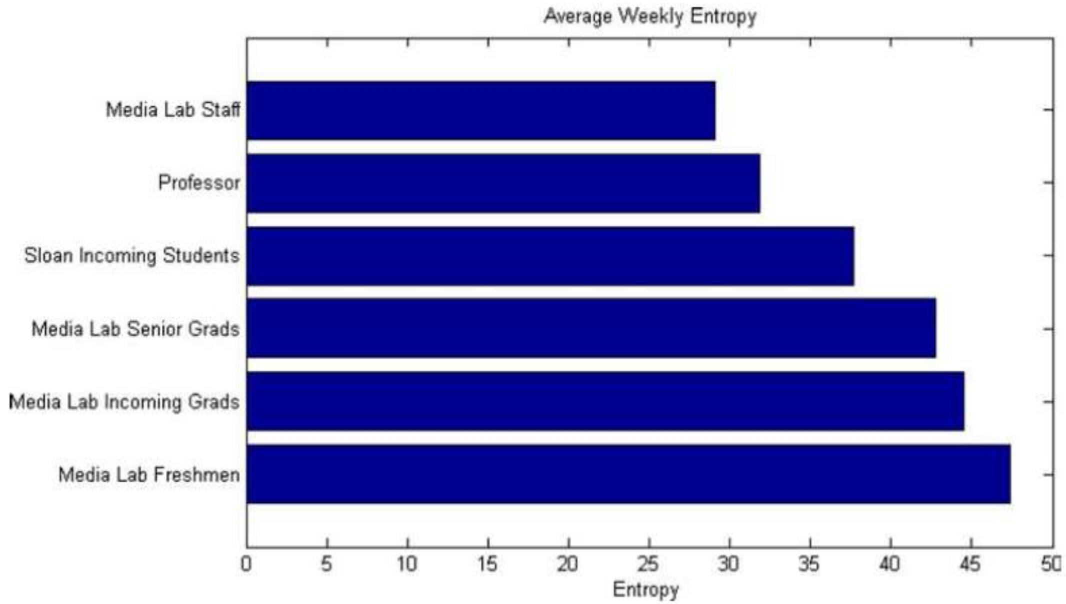

13时38分
支付宝总交易额
100 亿元
11.11 购物狂欢节
上天猫 就购了
淘宝论坛

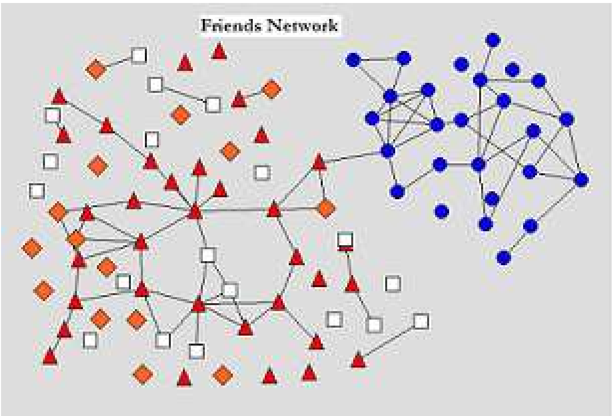

11时18分
支付宝总交易额
79 亿元
11.11 购物狂欢节
上天猫 就购了
淘宝论坛

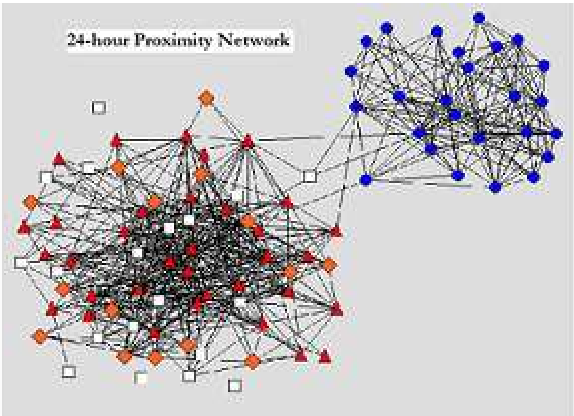

双11狂欢 奇迹大揭密
6分07秒
支付宝交易额超10亿
38分05秒
支付宝交易额超50亿
50亿
1小时
文胸成交1600000件
叠放=3个珠穆朗玛峰高度
1小时零10秒
手机淘宝的支付宝交易额
突破10亿
凌晨3点
天猫双11
03:00
广东人民还在奋力拼搏
成为全国省份购买力榜首
交易金额超9亿
13点04分
支付宝交易额
突破191亿
超越2012年
2013
2012
20xx
2
1
3
13点17分
同时在线购物人数超1700万
=香港人口的2.5倍
香港
截止24:00
支付宝交易额
突破350.19亿

Reality Mining

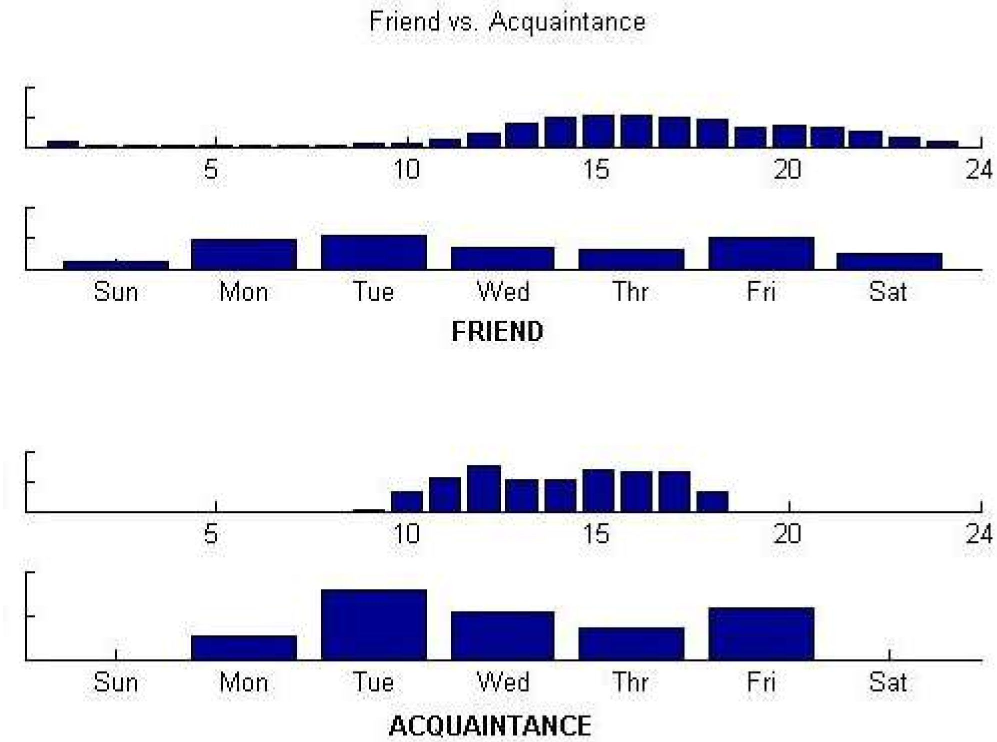








Reality Mining

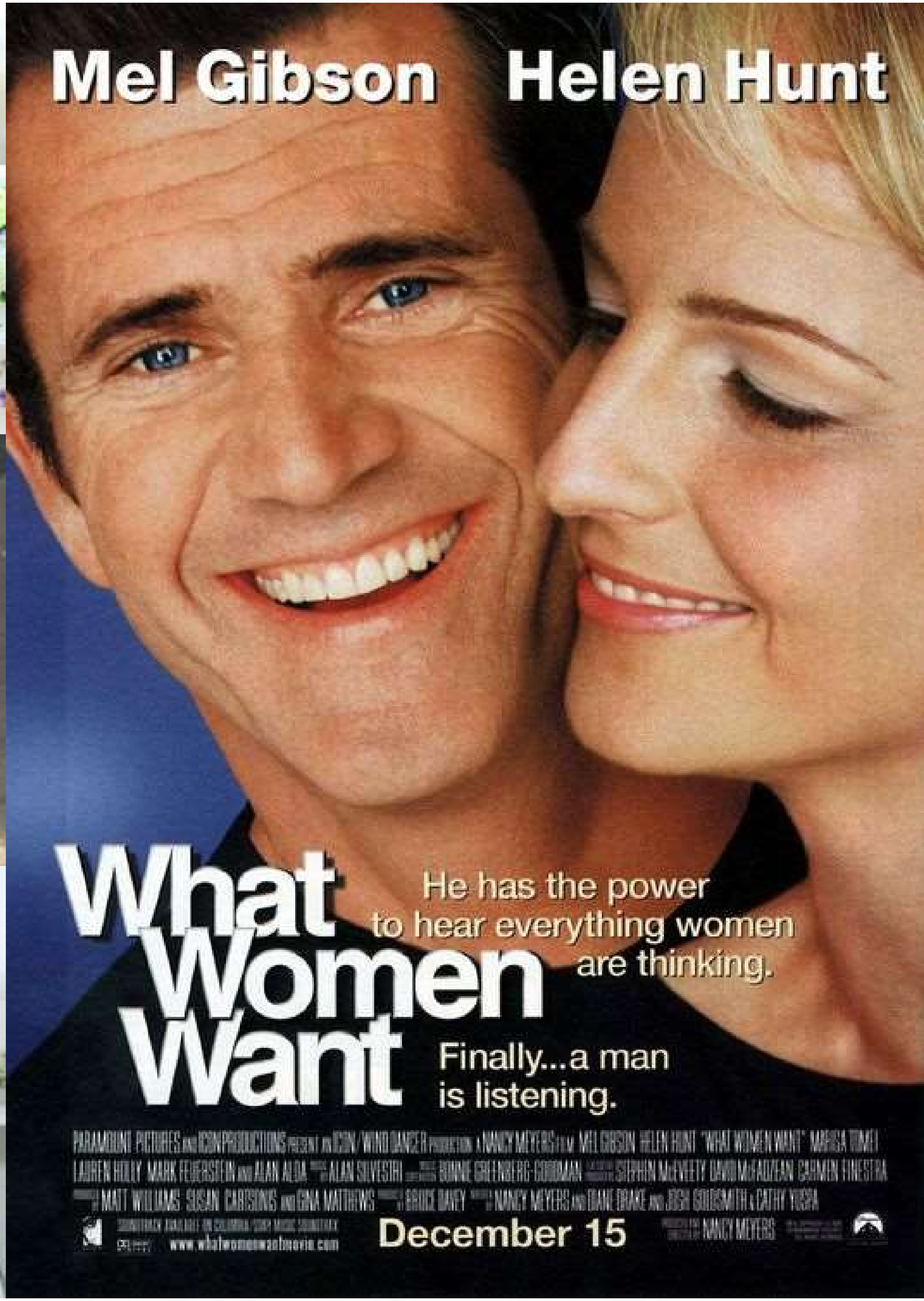



缤点


HanYid


雪贝蕾
xuebeilie


水晶玛俪


Mel Gibson
Helen Hunt
What Women Want
He has the power to hear everything women are thinking.
Finally...a man is listening.
December 15

Open Questions

- Customers are analyzed based solely on purchasing records.
  - More dimensions are to be added.
  - Just image how to recommend something to your friend …
- People have different personalities.
  - Different strategies may be required.
  - Selling insurance: Emotional vs. Mature

- People buy things for different reasons.
  - Impulse Buying vs. Planned Buying
  - Time Variant

Reading Materials

- P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "Grouplens: an Open Architecture for Collaborative Filtering of Netnews", in Proceedings of the ACM Conference on Computer Supported Cooperative Work, pp. 175–186, 1994.
- D. Billsus and M. Pazzani, "Learning Collaborative Information Filters", in Proceedings of the 15th International Conference on Machine Learning, pp. 46-54, 1998.
- B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms", in Proceedings of the 10th international Conference on World Wide Web, pp. 285-295, 2001.
- X. Su and T. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques", Advances in Artificial Intelligence, Article ID: 421425, 2009.
- L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank Citation Ranking: Bringing Order to the Web", Technical Report, Stanford InfoLab, 1999.
- S. Deerwester, S. Dumais, G. Furnas, T. Landauer, and R. Harshman, "Indexing by Latent Semantic Analysis", JASIS, vol. 41(6), pp. 391-407, 1990.
- E. Nathan and A. Pentland, "Reality Mining: Sensing Complex Social Systems", Personal and Ubiquitous Computing, vol. 10(4), pp. 255-268, 2006.