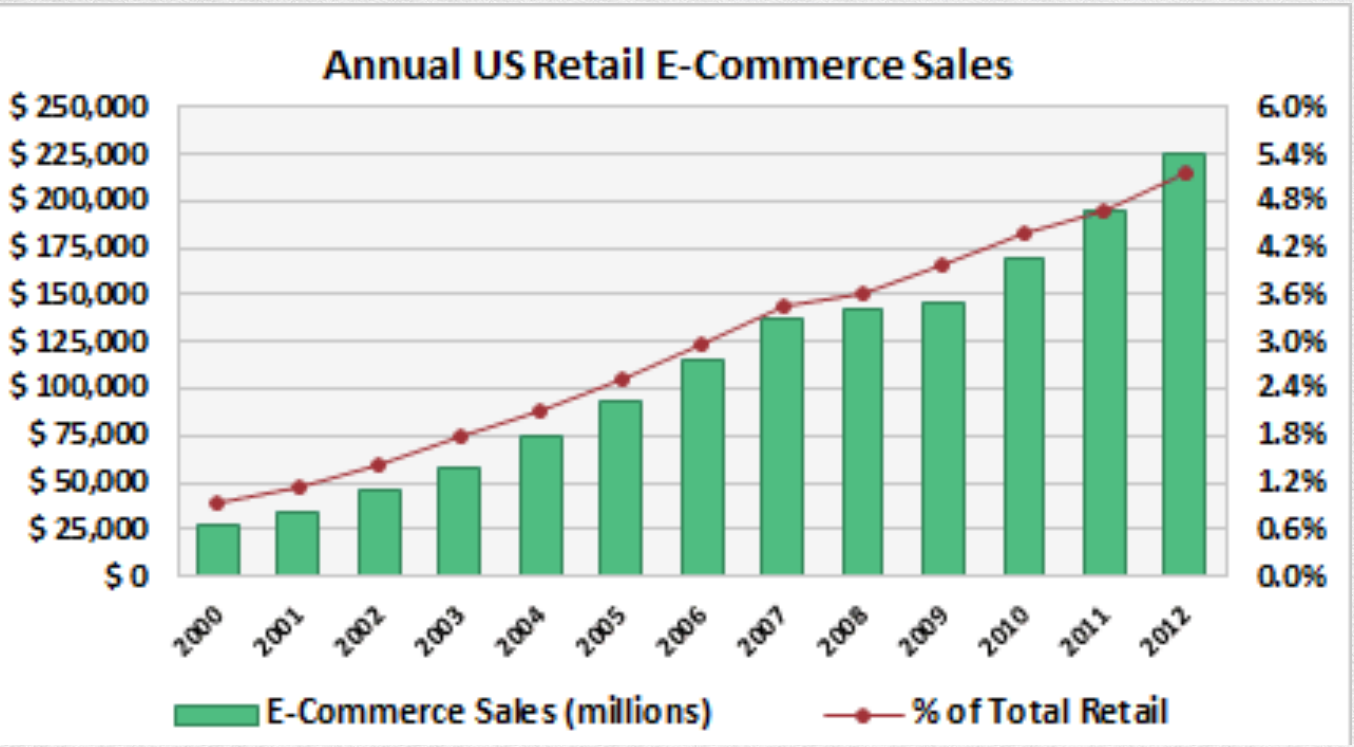
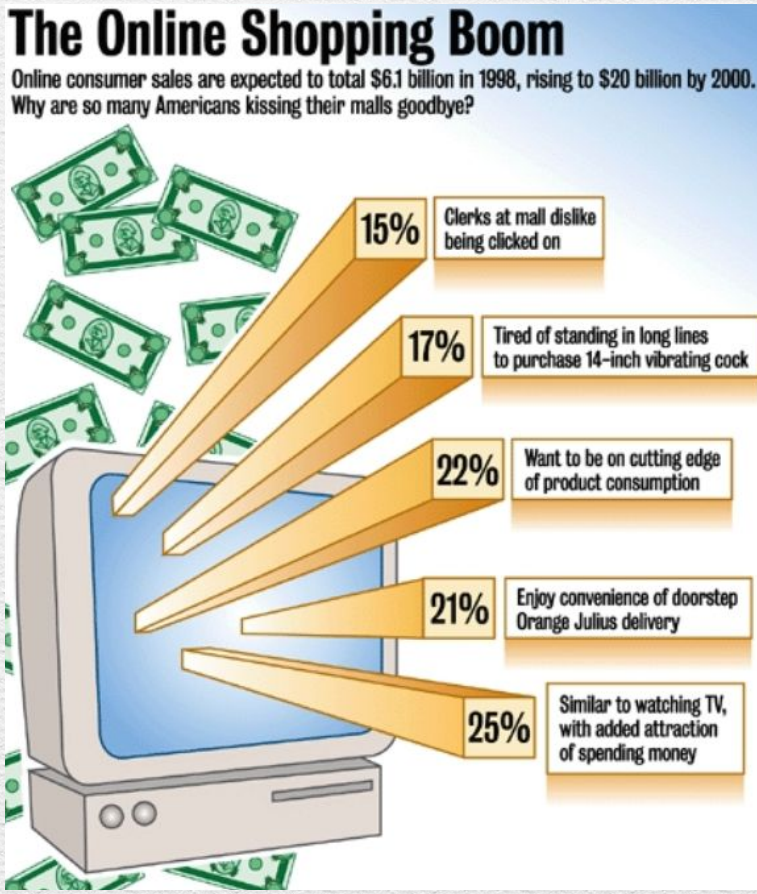
Recommender System

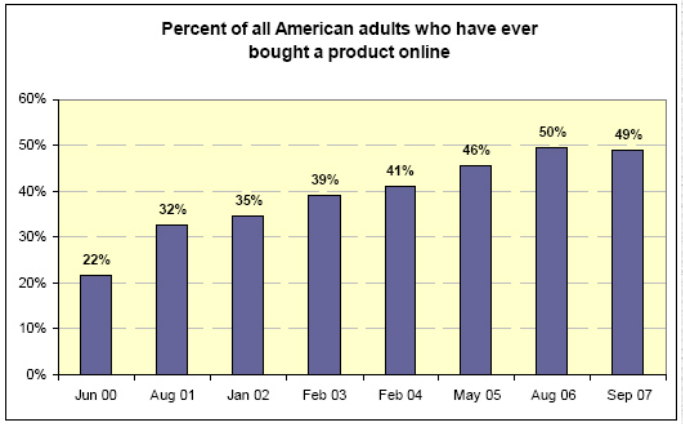
2000's

Data

Online Shopping Data





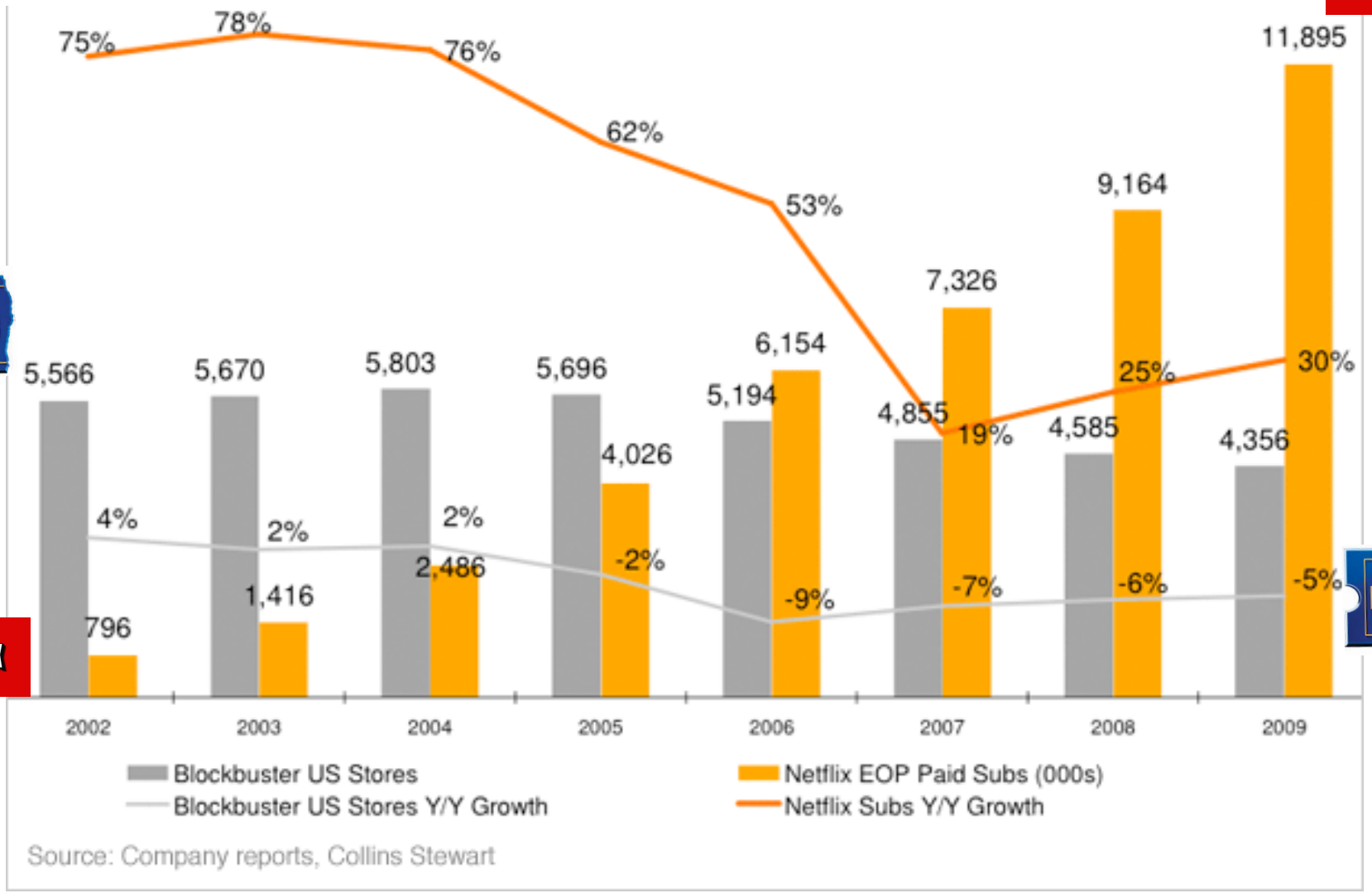


Movie Rental Business

Netflix vs Blockbuster

Moving Bits v.s. Moving Atoms

Movie Rental



https://youtu.be/5sMXR7rK40U?t=6s

in 2013

Book Store

Amazon vs Book Store

amazon

2011


BORDERS
BOOKS • MUSIC • MOVIES • CAFE
STORE CLOSING


Appendix (F)
BORDERS
THIS STORE IS
NOW CLOSED
PLEASE CONTACT:
MCR

THE RISE, FALL
BORDERS®

1997
Borders (Asia-Pacific) opens in Singapore. Later expands to Australia and New Zealand.

2007
Borders opens in Parkway Parade, its second outlet in Singapore.

Feb 2011
Parent company Redgroup Retail placed in voluntary administration.

Aug 2011
Flagship outlet at Wheelock Place suddenly closes.

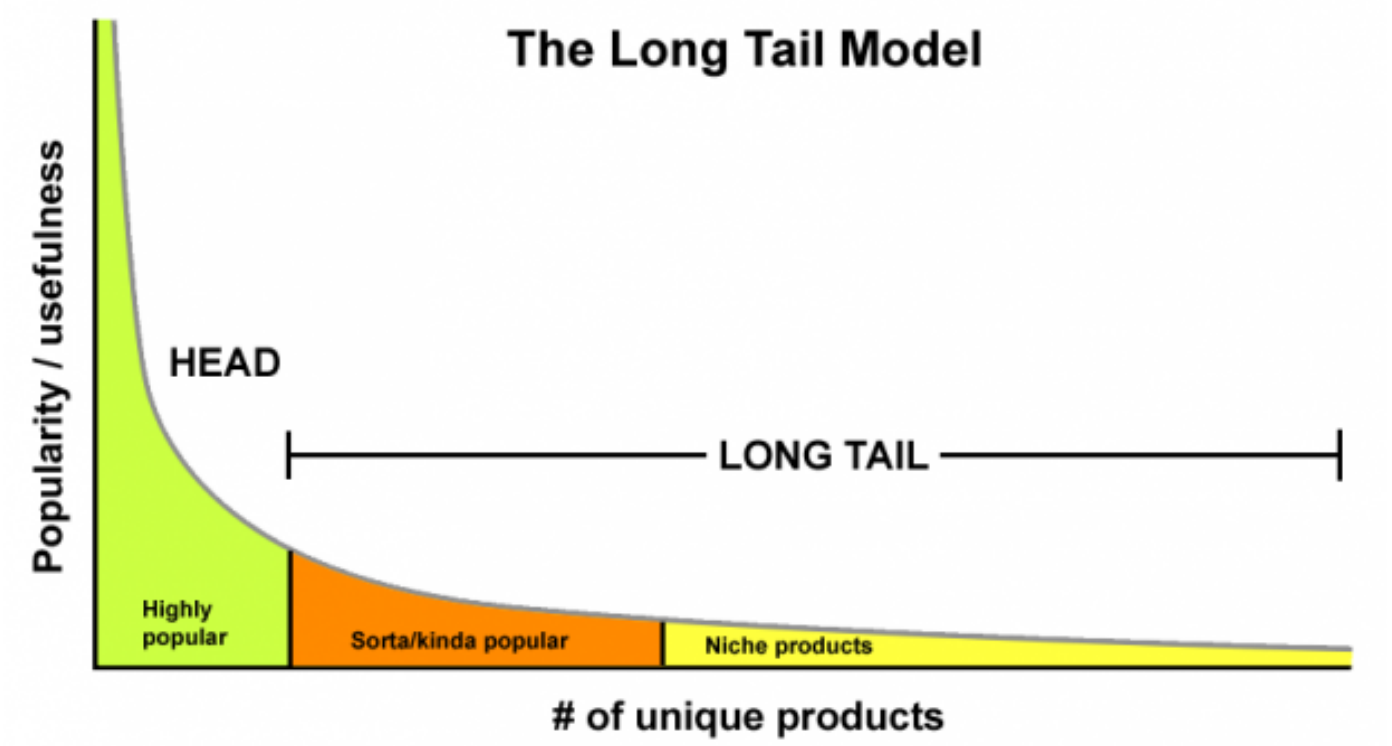
Sept 2011
Remaining store at Parkway Parade also closed down.

Meant to represent the potential for a larger volume of sales in an online bookstore as compared to a physical bookstore. (Renamed from Cadabra.com.)

Book Stores


BESTSELLERS
VALERIE
DOUBLE
GEORGE PELECANOS
CATASTROPHE 1914: EUROPE GOES TO WAR
MAX HASTINGS
THIS TOWN
DARK LIES THE ISLAND
KEVIN BARRY
Cover to Cover
EUROPE WAR

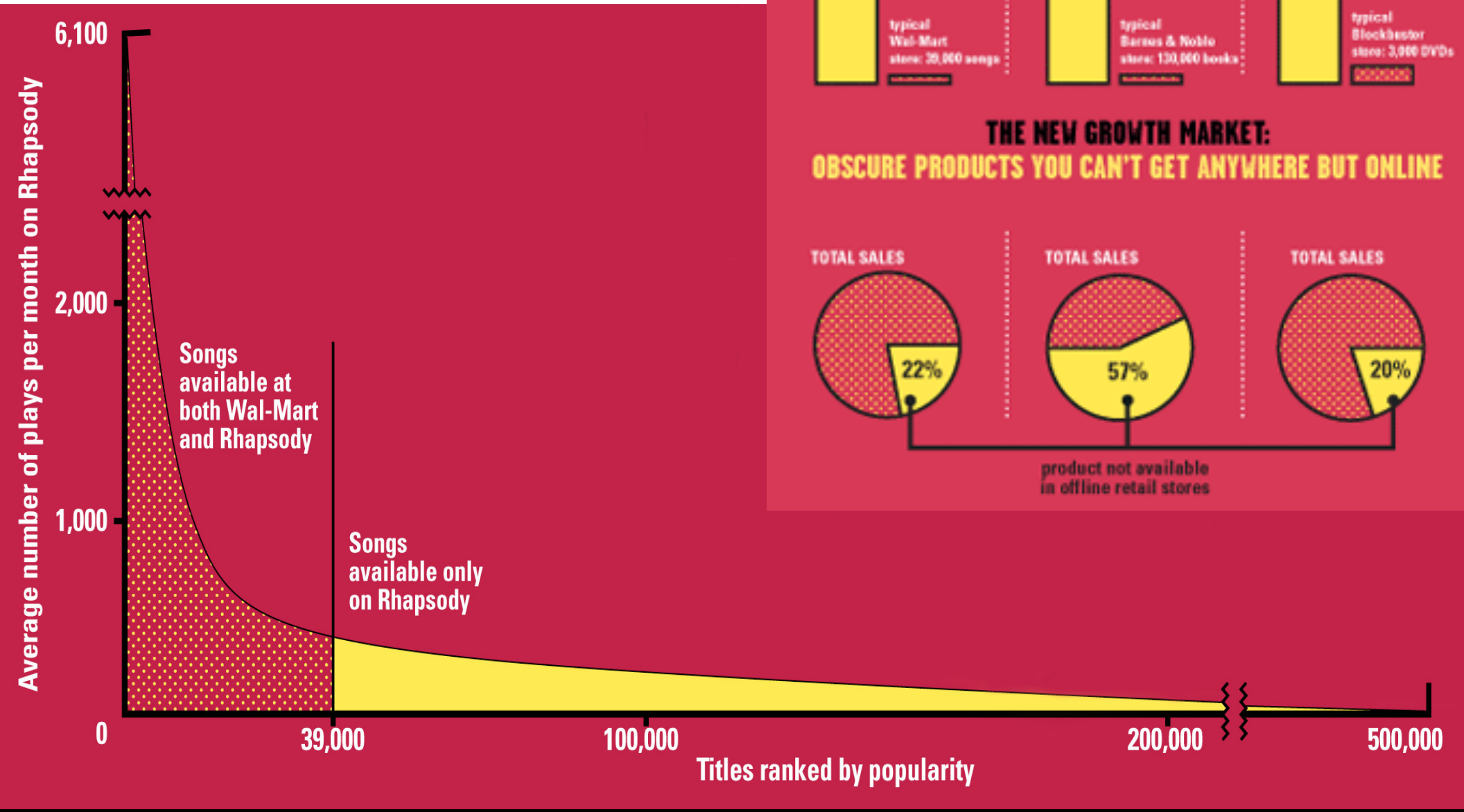
Focus on Popular Products


The Long Tail Model
Popularity / usefulness
HEAD
LONG TAIL
Highly popular
Sorta/kinda popular
Niche products
# of unique products


More Popular Books
Higher Sales Volume
Amazon.com Example in Retail
Sales Volume
Inventory
Less Popular Books
Less Volume Sold

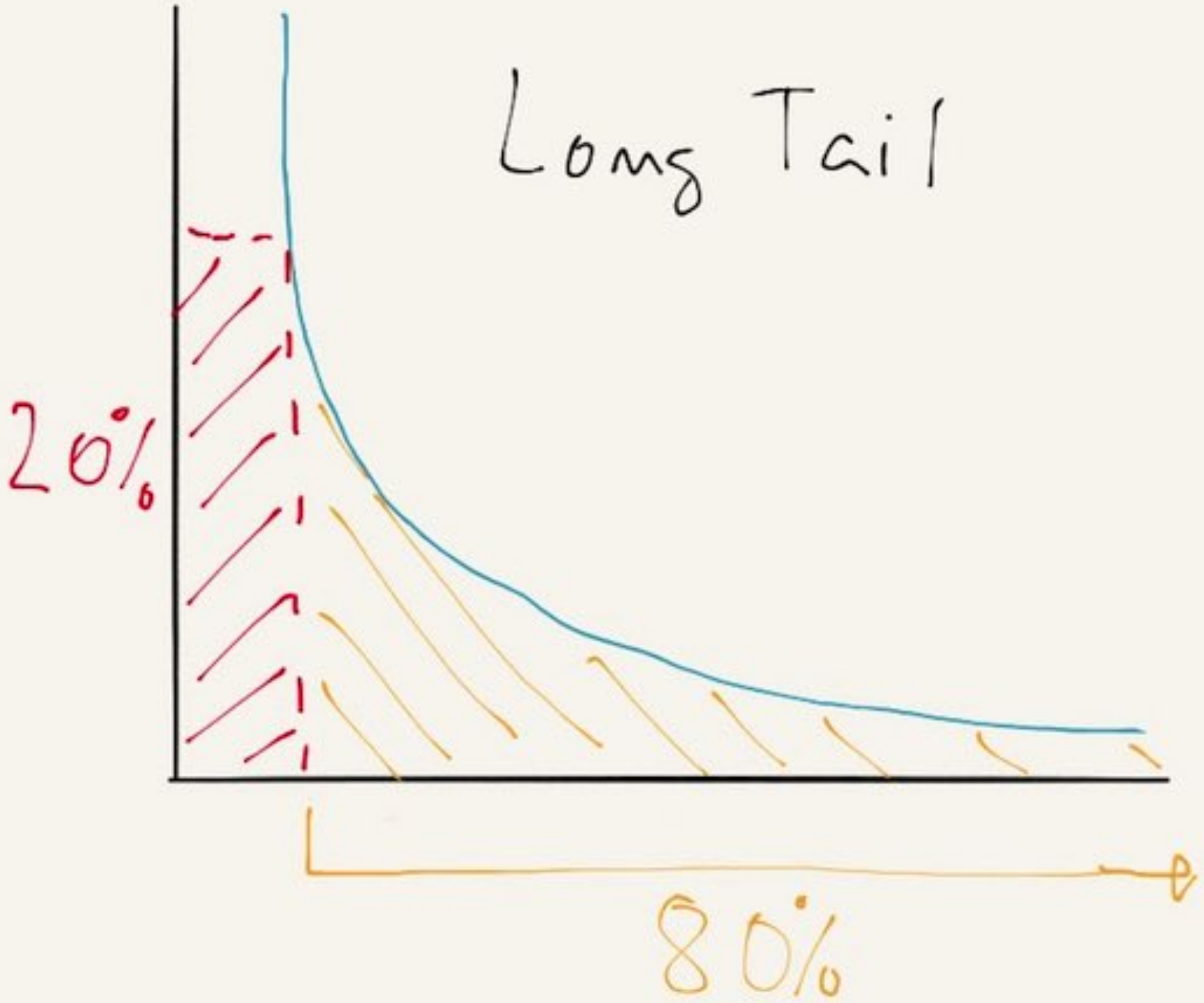
The Long Tail

The Long Tail

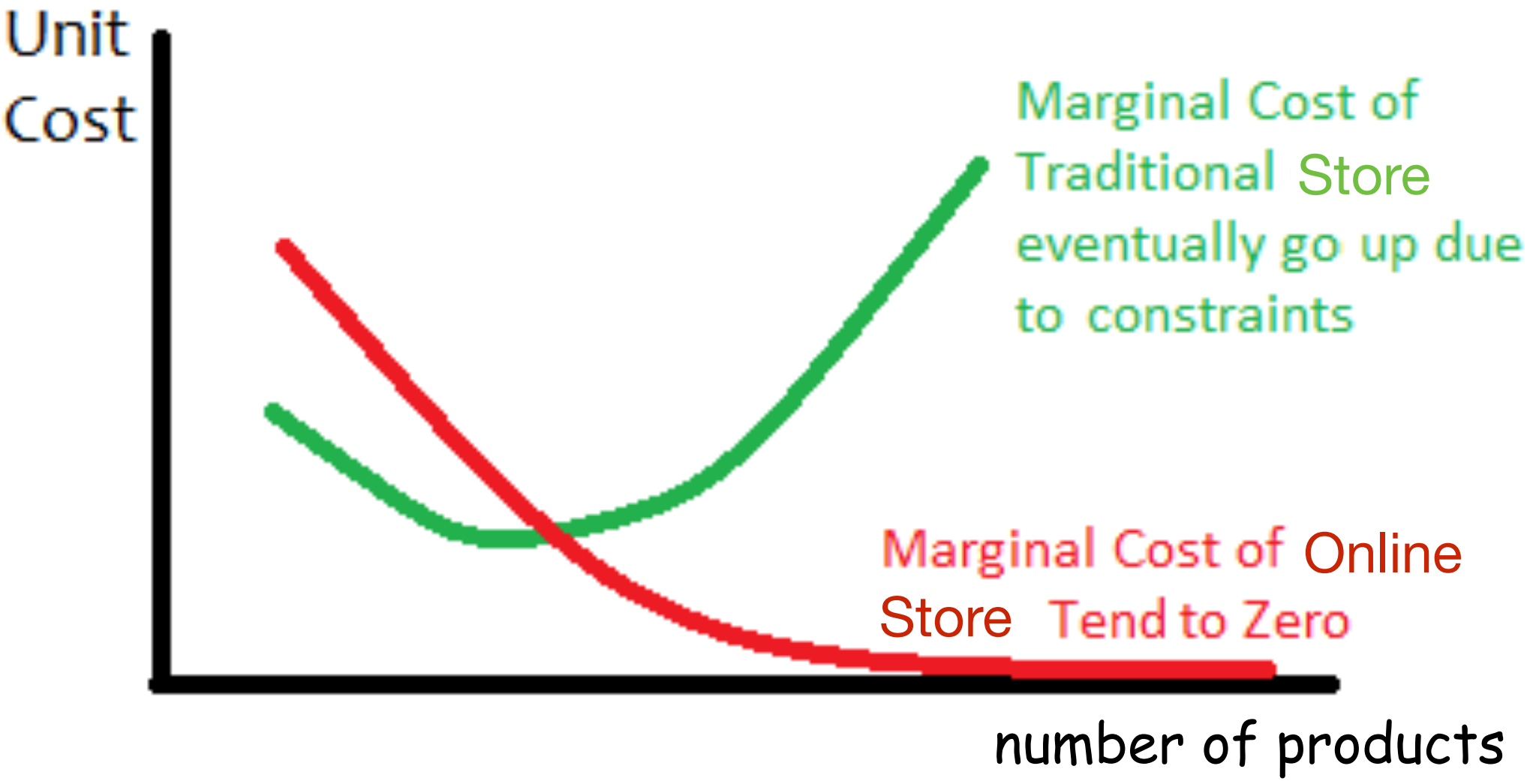


Sources: Erik Brynjolfsson and Jeffrey Hu, MIT, and Michael Smith, Carnegie Mellon; Barnes & Noble; Netflix; RealNetworks

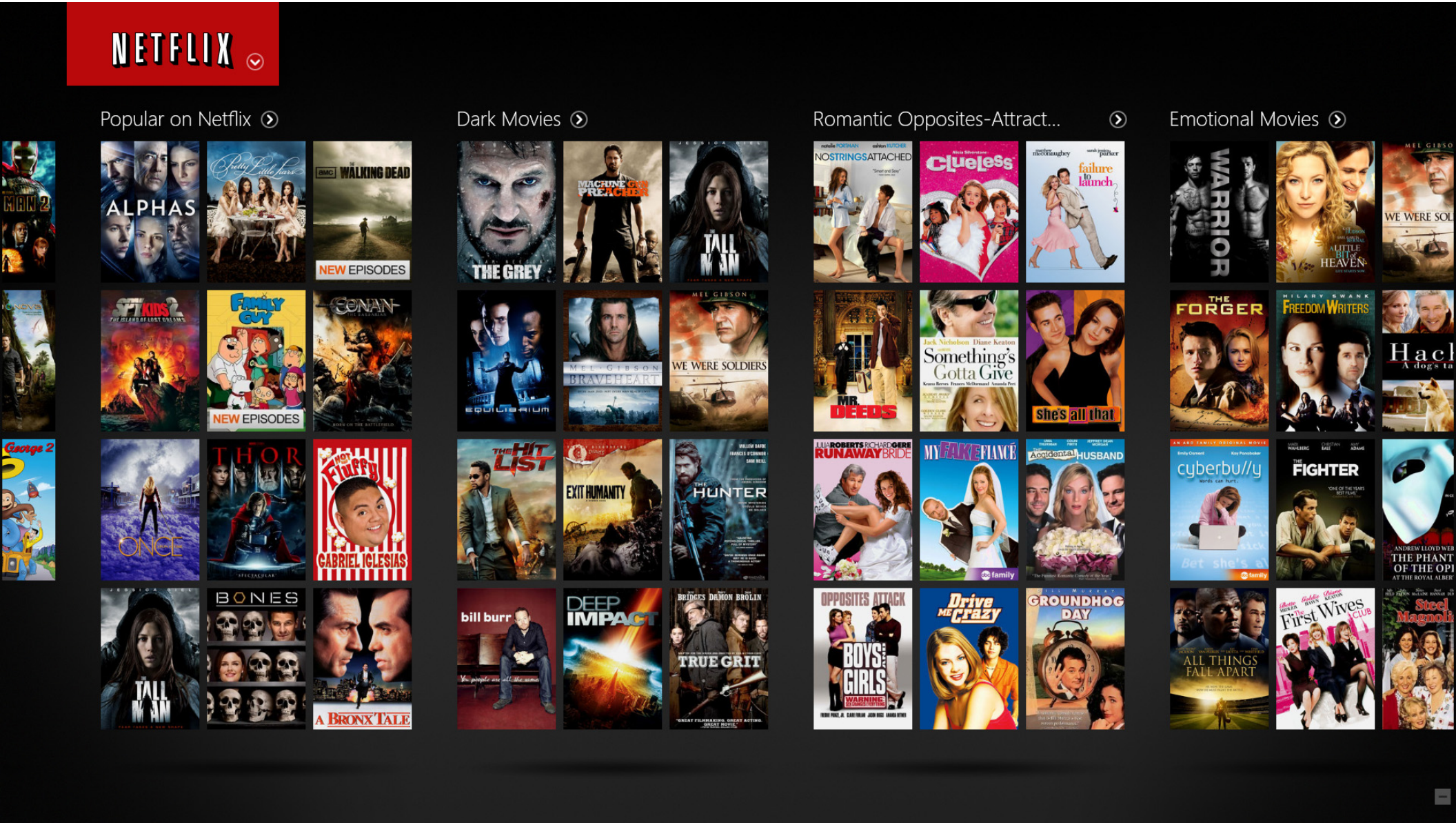
How to make money in long tail?



Marginal Cost of adding a new product

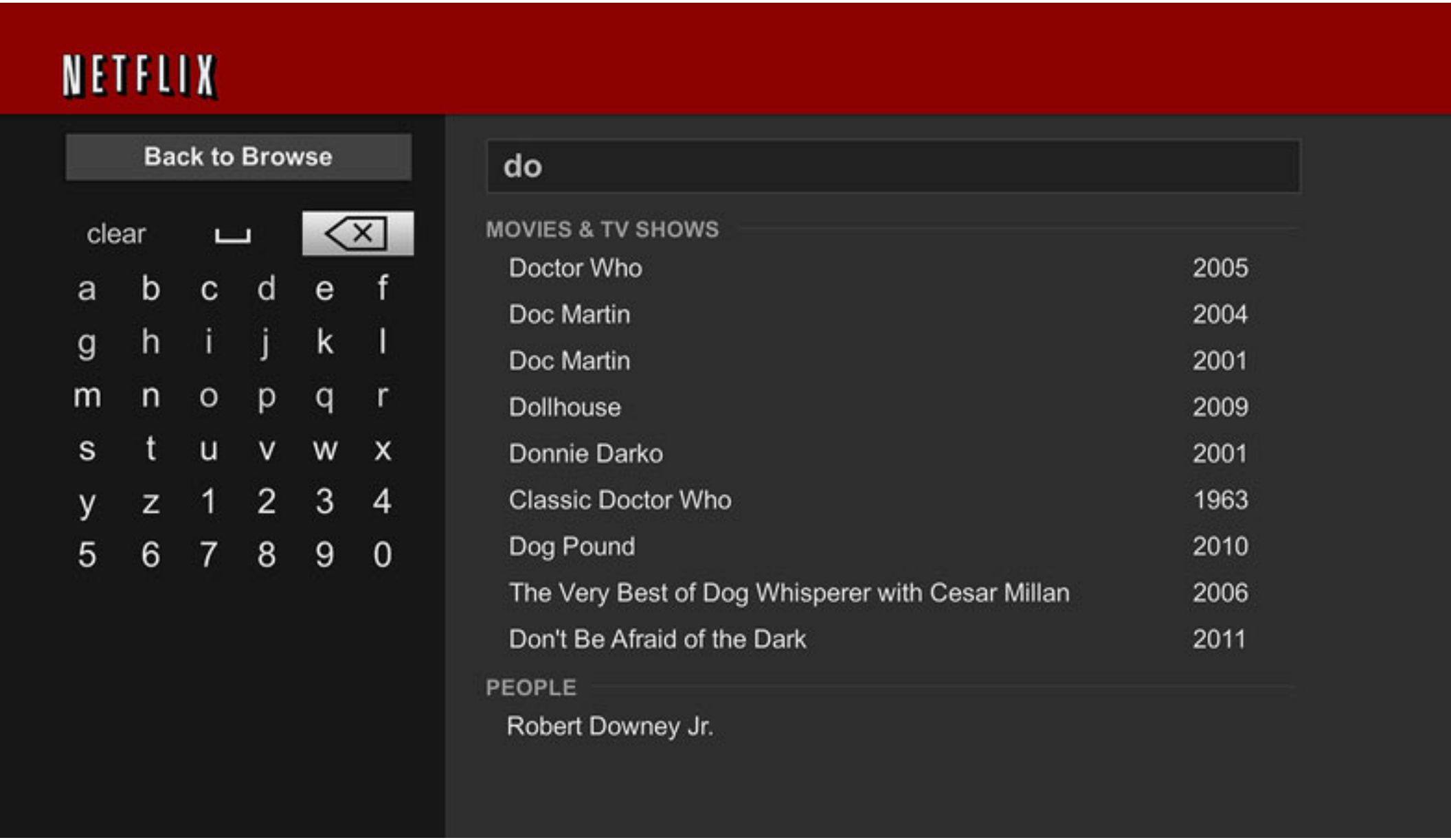


NETFLIX


NETFLIX
Popular on Netflix
Dark Movies
Romantic Opposites-Attract...
Emotional Movies
ALPHAS
Pretty Little Liars
THE WALKING DEAD
NEW EPISODES
THE GREY
MACHINE GUN PREACHER
TALL MAN
NOSTRINGSATTACHED
Clueless
failure to launch
WARRIOR
WE WERE SOL
SPY KIDS 2
FAMILY GUY
NEW EPISODES
CONAN
EQUILIBRIUM
MEL GIBSON
BRAVEHEART
WE WERE SOLDIERS
MR. DEEDS
Jack Nicholson Diane Keaton
Something's Gotta Give
she's all that
THE FORGER
HILARY SWANK
FREEDOM WRITERS
George 2
ONCE
THOR
GABRIEL IGLESIAS
THE HIT LIST
EXIT HUMANITY
THE HUNTER
JULIA ROBERTS RICHARD GERE
RUNAWAY BRIDE
MY FAKE FIANCÉ
Accidental HUSBAND
cyberbu//y
THE FIGHTER
THE PHANT
OF THE OPI
BONES
A BRONX TALE
bill burr
DEEP IMPACT
BRIDGES DAMON BROLIN
TRUE GRIT
OPPOSITES ATTACK
BOYS AND GIRLS
Drive Me Crazy
GROUNDHOG DAY
ALL THINGS FALL APART
First Wives Club
Steel Magnoli

Problem: too many choices

Search


NETFLIX
Back to Browse
clear
a b c d e f
g h i j k l
m n o p q r
s t u v w x
y z 1 2 3 4
5 6 7 8 9 0
do
MOVIES & TV SHOWS
Doctor Who 2005
Doc Martin 2004
Doc Martin 2001
Dollhouse 2009
Donnie Darko 2001
Classic Doctor Who 1963
Dog Pound 2010
The Very Best of Dog Whisperer with Cesar Millan 2006
Don't Be Afraid of the Dark 2011
PEOPLE
Robert Downey Jr.

Problem

Users have to know the name of the book/movie

Users usually know a few popular movies

or the movies they have watched before

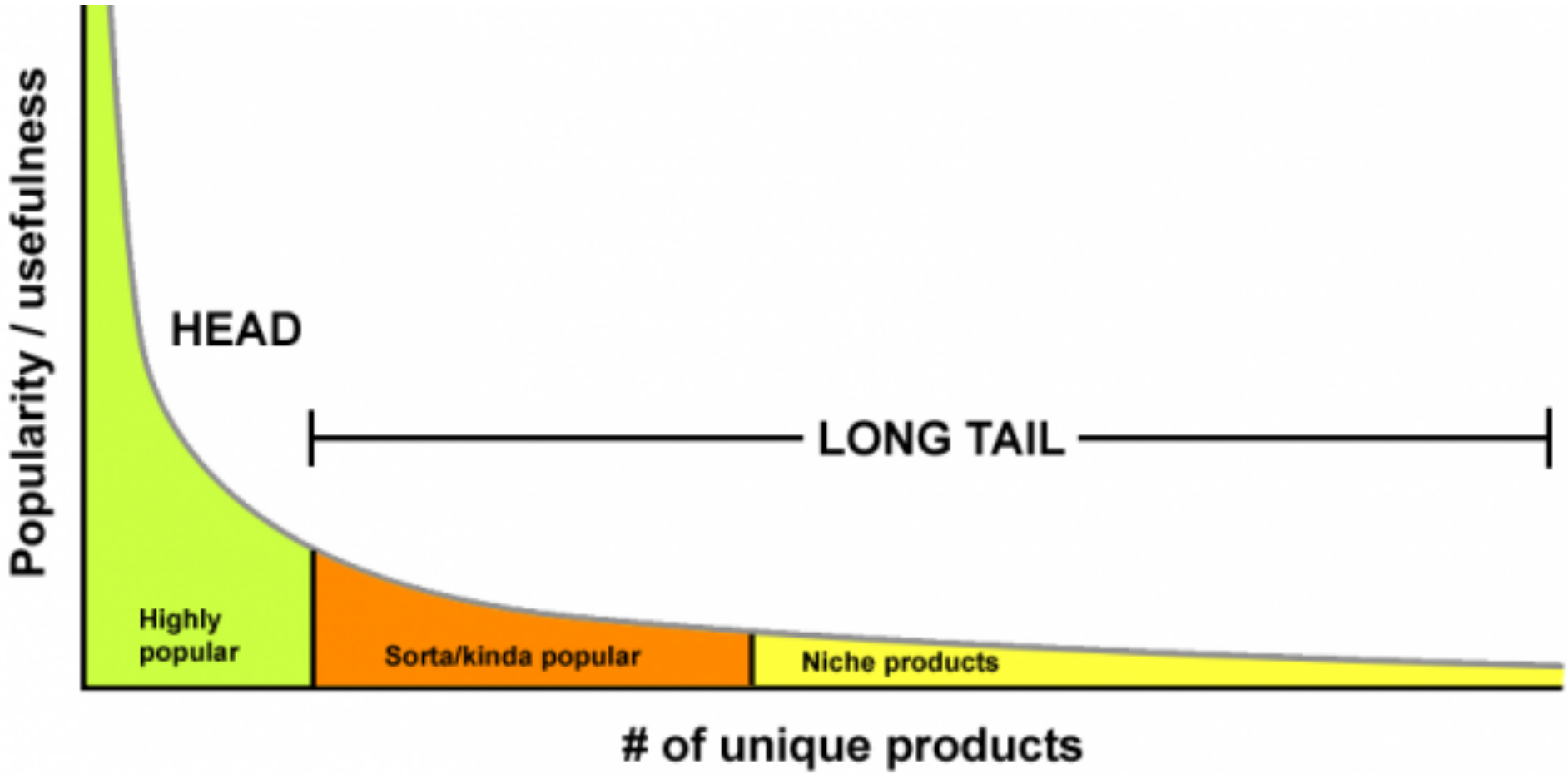
Popular movies are sold in Movie theaters or DVD stores

A user usually won't watch a movie twice

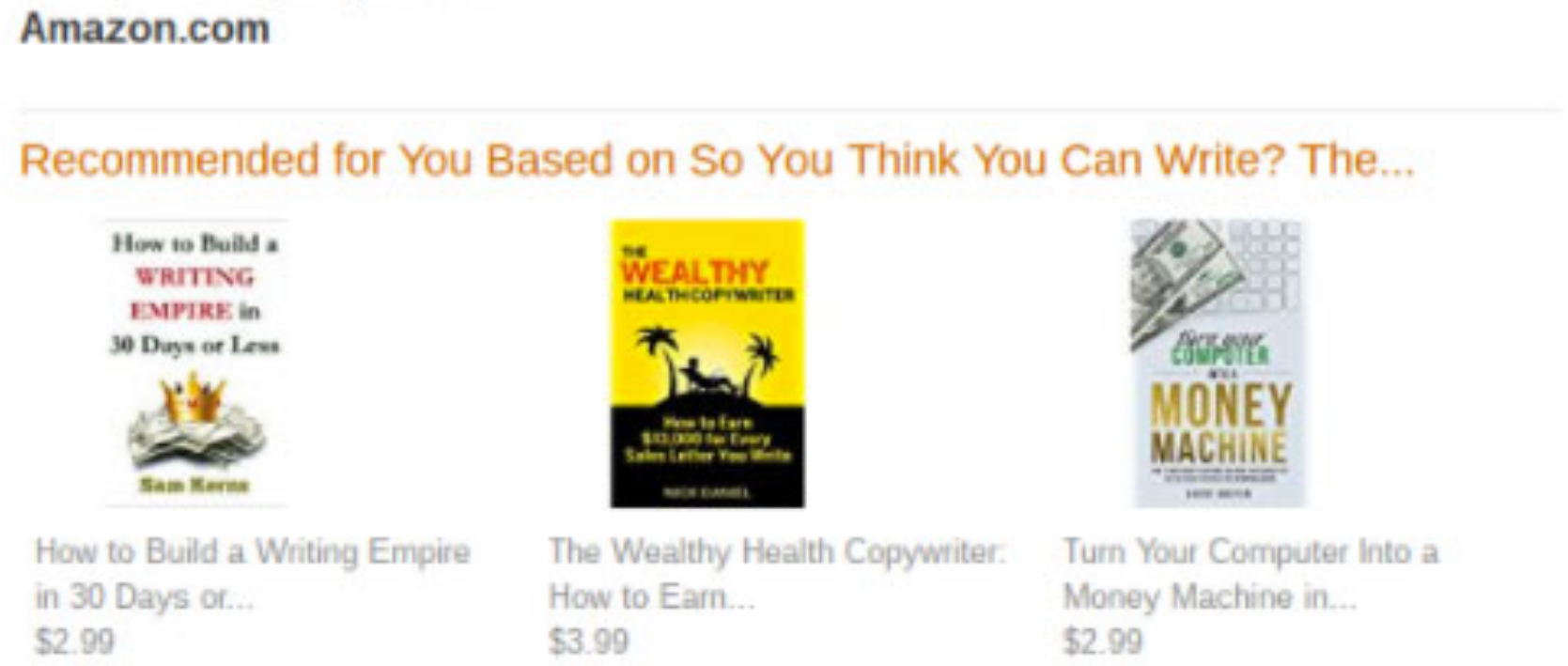
Different users prefer different movies

Problem

How can we predict whether or not a user will like a niche product before the user using the product?



Recommender System



over 75% purchase made through recommendation

Close

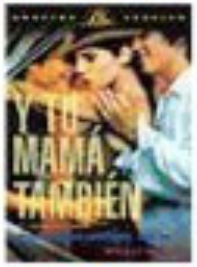
Other Movies You Might Enjoy

Amelie

Add

Not Interested

Y Tu Mama Tambien



Add

Not Interested

Eiken has been added to your Queue at position 2.

This movie is available now.

Move To Top Of My Queue

< Continue Browsing

Visit your Queue >

Guys and Balls

Add

Not Interested

Mostly Martha

Add

Not Interested

Only Human

Add

Not Interested

Russian Dolls

Add

Not Interested

The Netflix Prize

Rating Matrix R

Users

Movies

1	3	4			
	3	5			5
		4	5		5
		3			
		3			
2			2		2
				5	
	2	1			1
	3			3	
1					

Problem to Solve

Users

Movies

1	3	4	?		
	3	5	?		5
		4	5		5
		3	?		
		3	?		
2			2		2
			?	5	
	2	1	?		1
	3		?	3	
1			?		

Item-based Recommendation

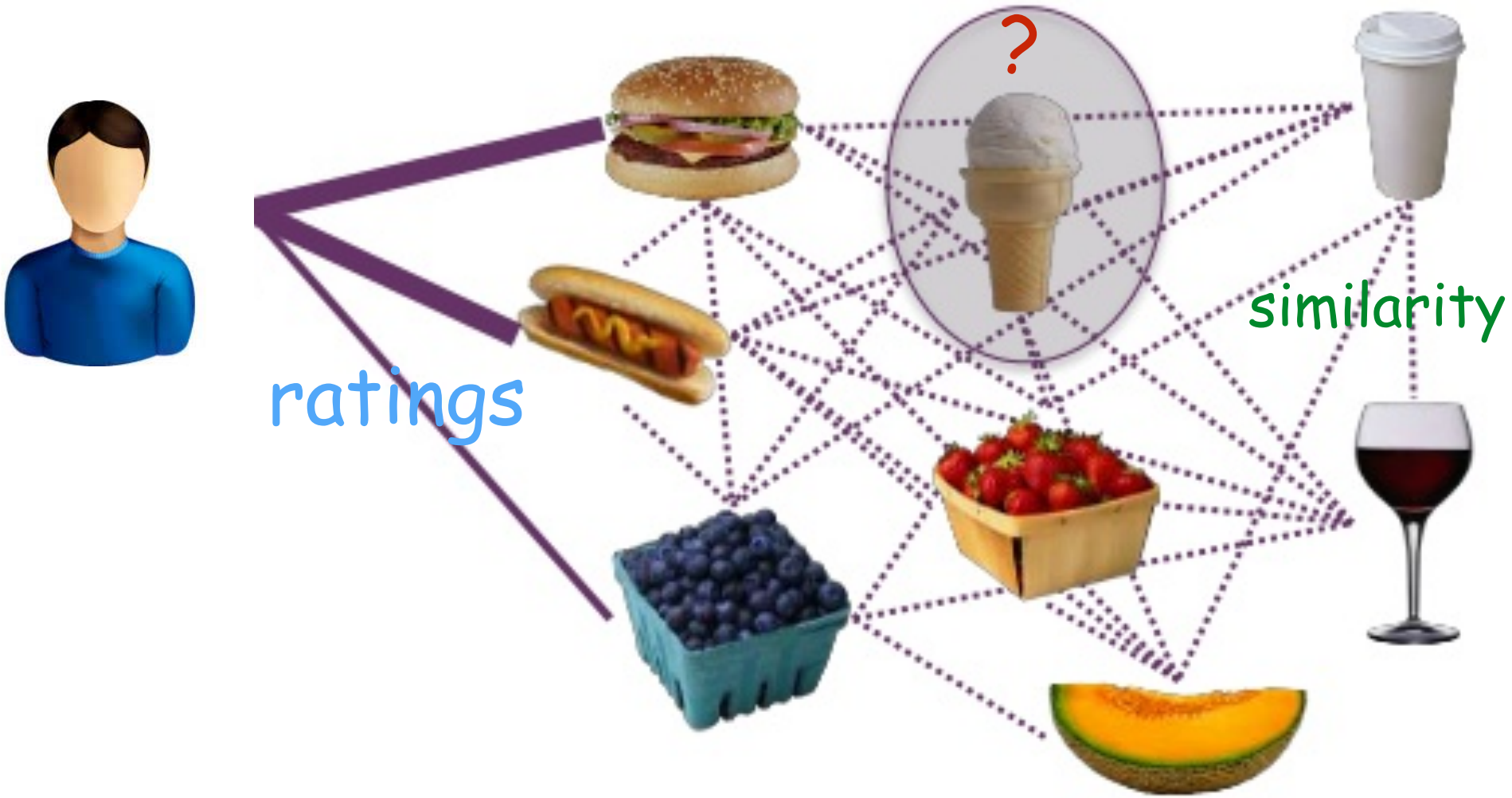
buy

similar

recommend

Item-based CF

1. Define similarity between items
2. select k most similar items rated by the user
3. estimate rating using weighted average



Item-based

Users

Movies

1	3	4			
	3	5			5
		4	5		5
		3			
		3			
2		?	2		2
				5	
	2	1			1
	3			3	
1					

Item-based

Users

Movies

1	3	4			
	3	5			5
		4	5		5
		3			
		3			
2		?	2		2
				5	
	2				1
	3			3	
1					

Similarity to Movie ?

0.4

0.1

0.2

0.5

0.3

Predicted Rating:

$$\frac{0.4X4 + 0.5 X 3}{0.4 + 0.5} = 3.4$$

Item-based

Users

Movies

1	3	4			
	3	5			5
		4	5		5
		3			
		3			
2		3.4	2		2
				5	
	2				1
	3			3	
1					

Similarity to Movie ?

0.4

0.1

0.2

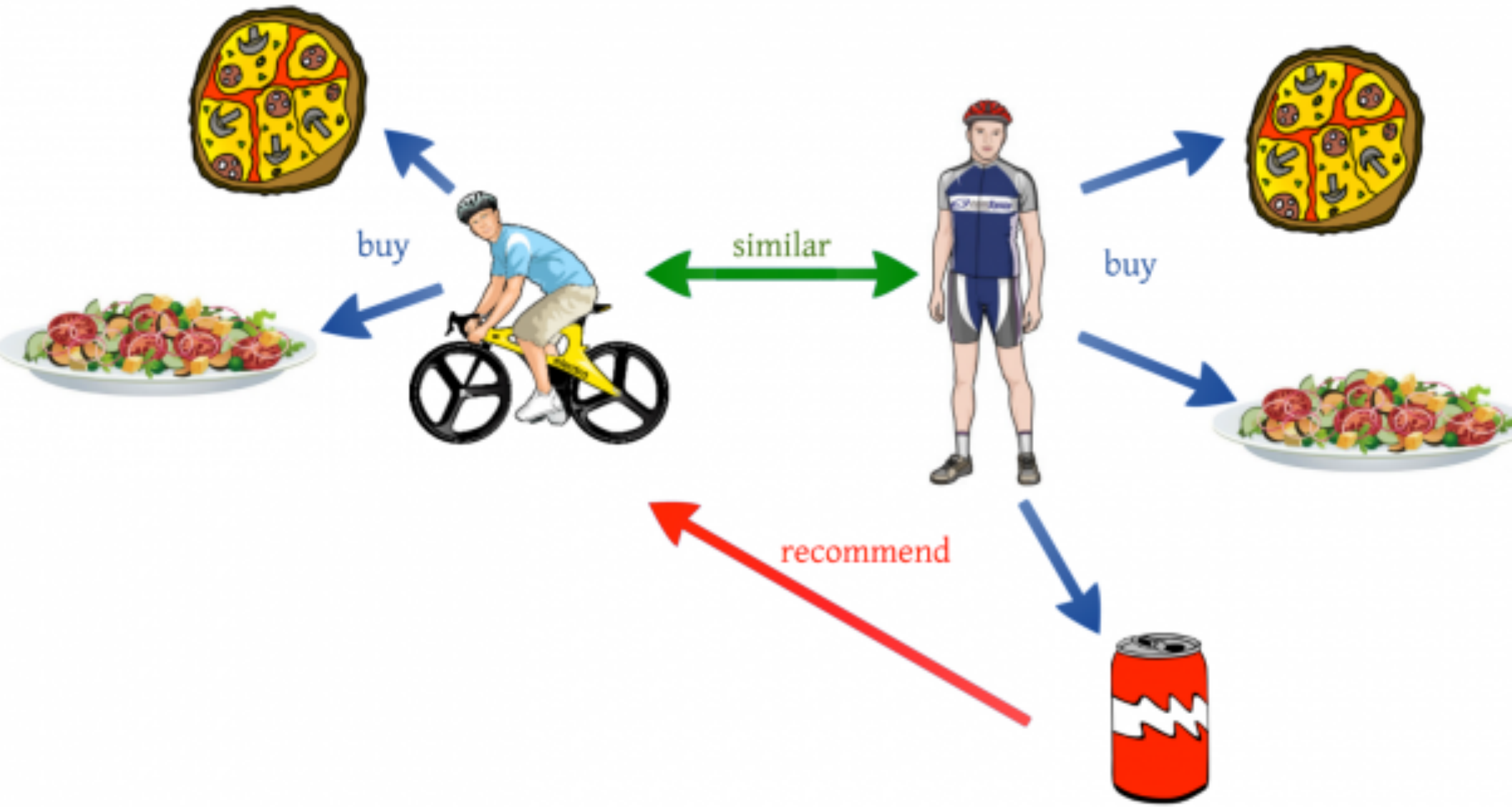
0.5

0.3

Predicted Rating:

$$\frac{0.4X4 + 0.5 X 3}{0.4 + 0.5} = 3.4$$

User-based Recommendation



User-based

Users

Movies

1	3	4			
	3	5			5
		4	5		5
		3			
		3			
2		?	2		2
				5	
	2	1			1
	3			3	
1					

Users

Movies

0.1			0.4		0.5
1	3	4			
	3	5			5
		4	5		5
		3			
		3			
2		?	2		2
				5	
	2	1			1
	3			3	
1					

← Similarity to User ?

Predicted Rating:

$$\frac{0.4X2 + 0.5 X 2}{0.4 + 0.5} = 2$$

Users

Movies

0.1			0.4		0.5
1	3	4			
	3	5			5
		4	5		5
		3			
		3			
2		2	2		2
				5	
	2	1			1
	3			3	
1					

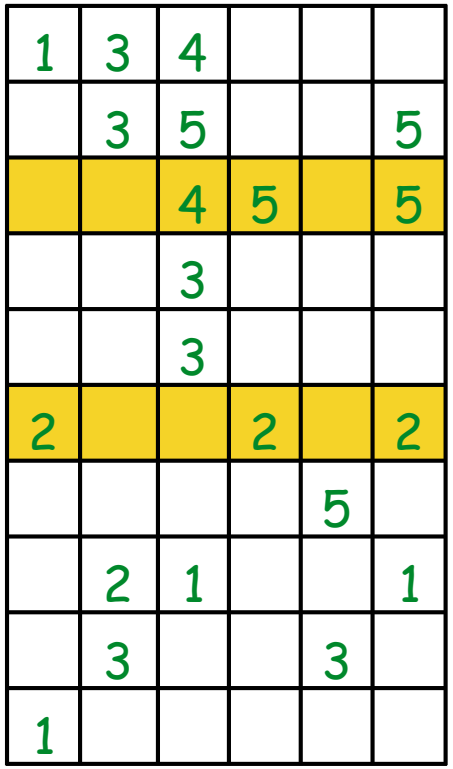
← Similarity to User ?

Predicted Rating:

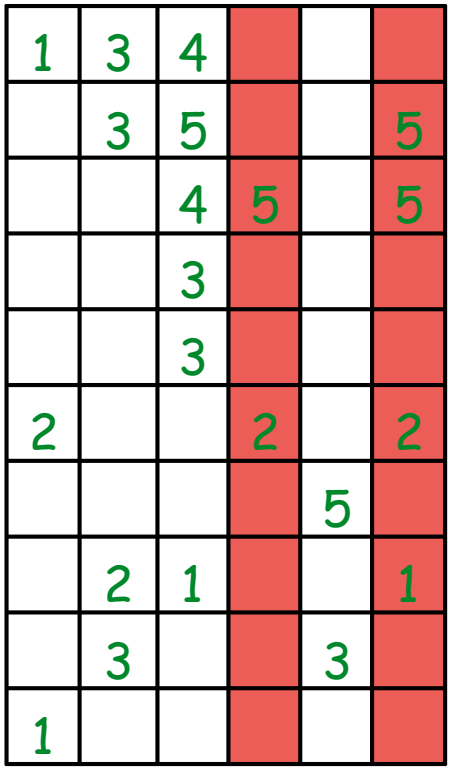
$$\frac{0.4X2 + 0.5 X 2}{0.4 + 0.5} = 2$$

Similarity Metric

Movie Similarity



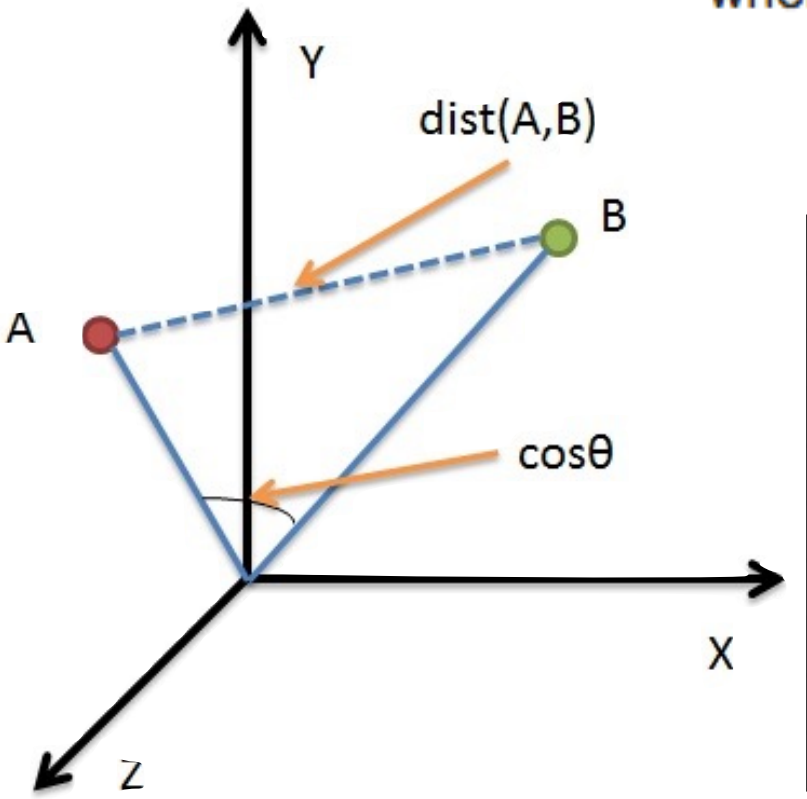
User Similarity



Cosine Similarity

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|_2 \|\mathbf{B}\|_2} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

where A_i and B_i are components of vector A and B



User Similarity

1	3	4			
	3	5			5
		4	5		3
		3			
		3			
2			2		1
				5	
	2	1			1
	3			3	
1					

A, B only include shared movies

A = (5, 2) B = (3, 1)

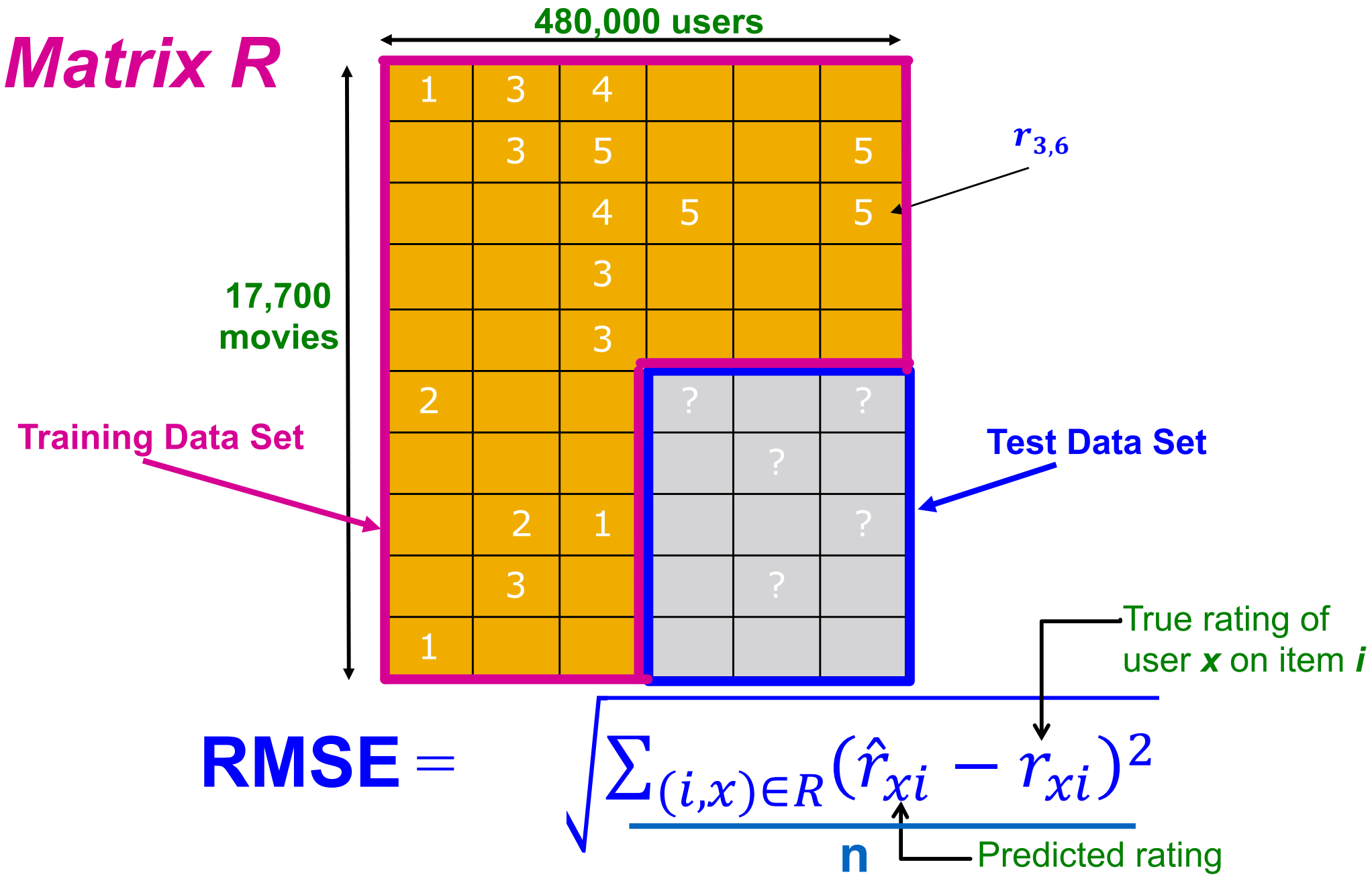
$$\text{sim} = \frac{5X3 + 2X1}{\sqrt{5^2+2^2} X \sqrt{3^2+1^2}}$$

= 0.99

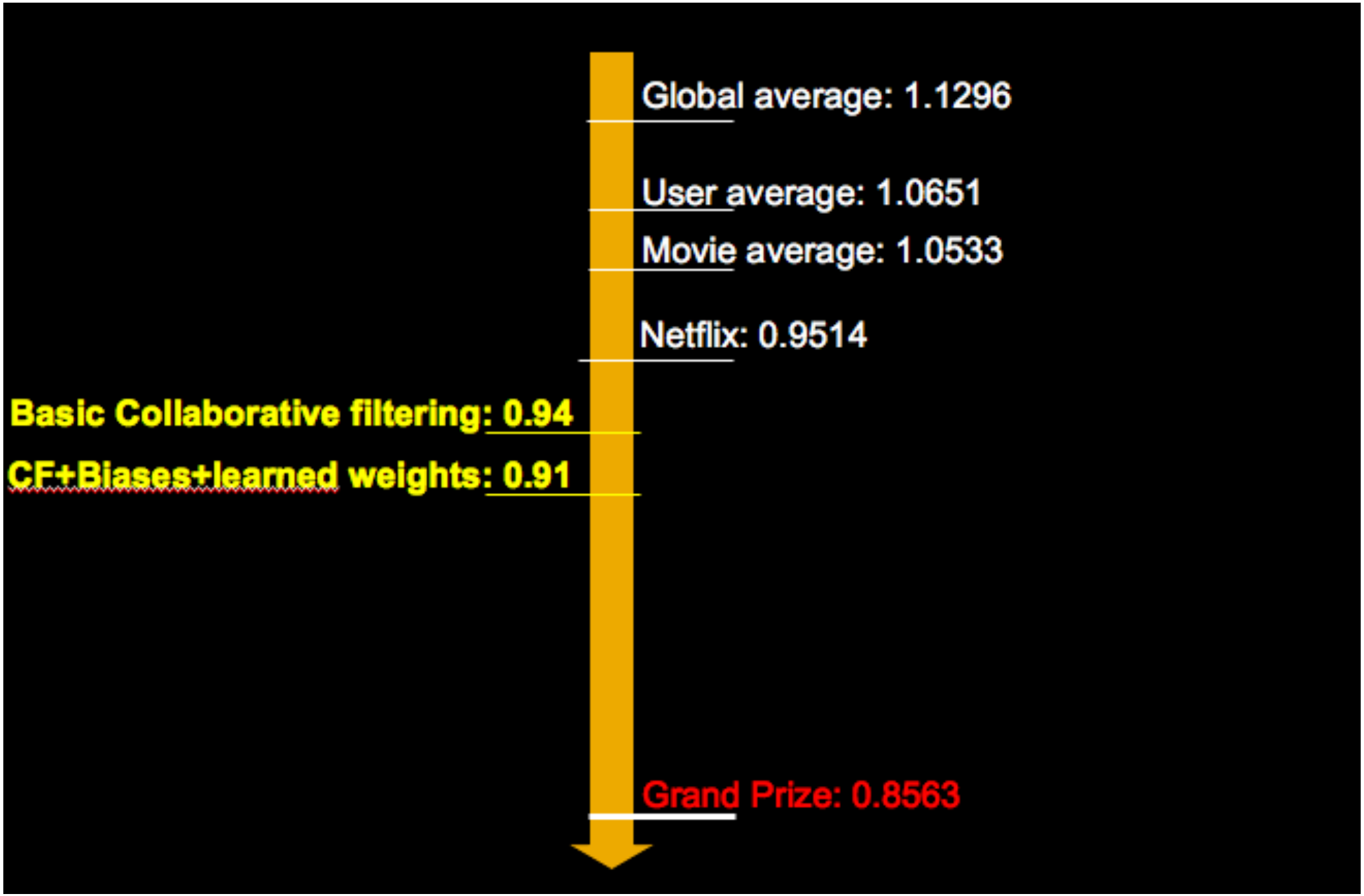
The Netflix Prize

- **Training data**
  - 100 million ratings, 480,000 users, 17,770 movies
  - 6 years of data: 2000-2005
- **Test data**
  - Last few ratings of each user (2.8 million)
  - **Evaluation criterion:** Root Mean Square Error (RMSE) $=$ $\sqrt{\frac{\sum_{(i,x)\in R}(\hat{r}_{xi} - r_{xi})^2}{n}}$
  - **Netflix's system RMSE: 0.9514**
- **Competition**
  - 2,700+ teams
  - **$1 million** prize for 10% improvement on Netflix

Evaluation


Matrix R
480,000 users
17,700 movies
$r_{3,6}$
Training Data Set
Test Data Set
True rating of user x on item i
Predicted rating
n
RMSE =
$\sqrt{\sum_{(i,x)\in R}(\hat{r}_{xi} - r_{xi})^2}$

Performance



Optimization-based CF

$$\min_{R'} \sum_{i,j} \left(R_{i,j} - R'_{i,j} \right)^2$$

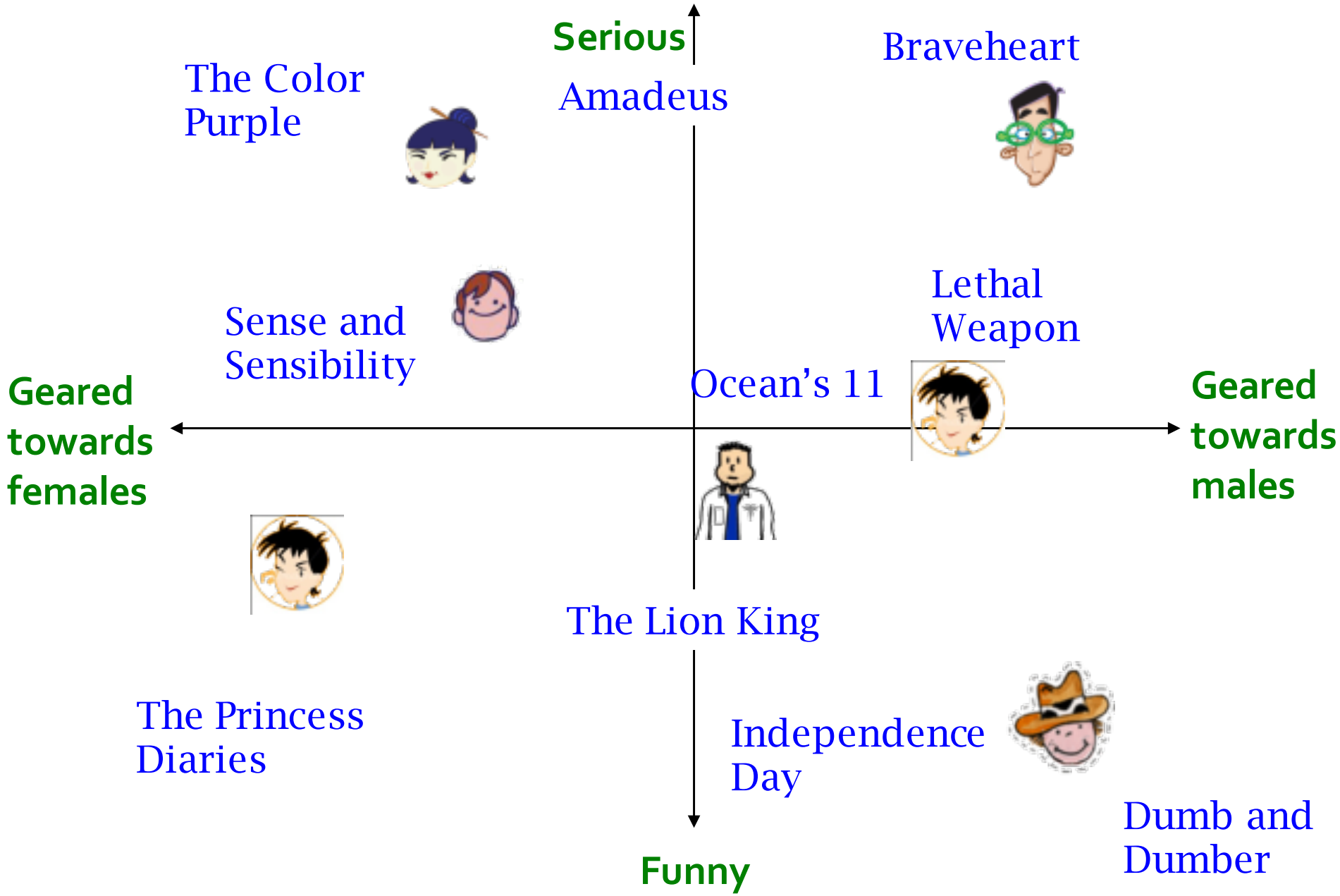
Real Rating

Predicted Rating

i-th movie

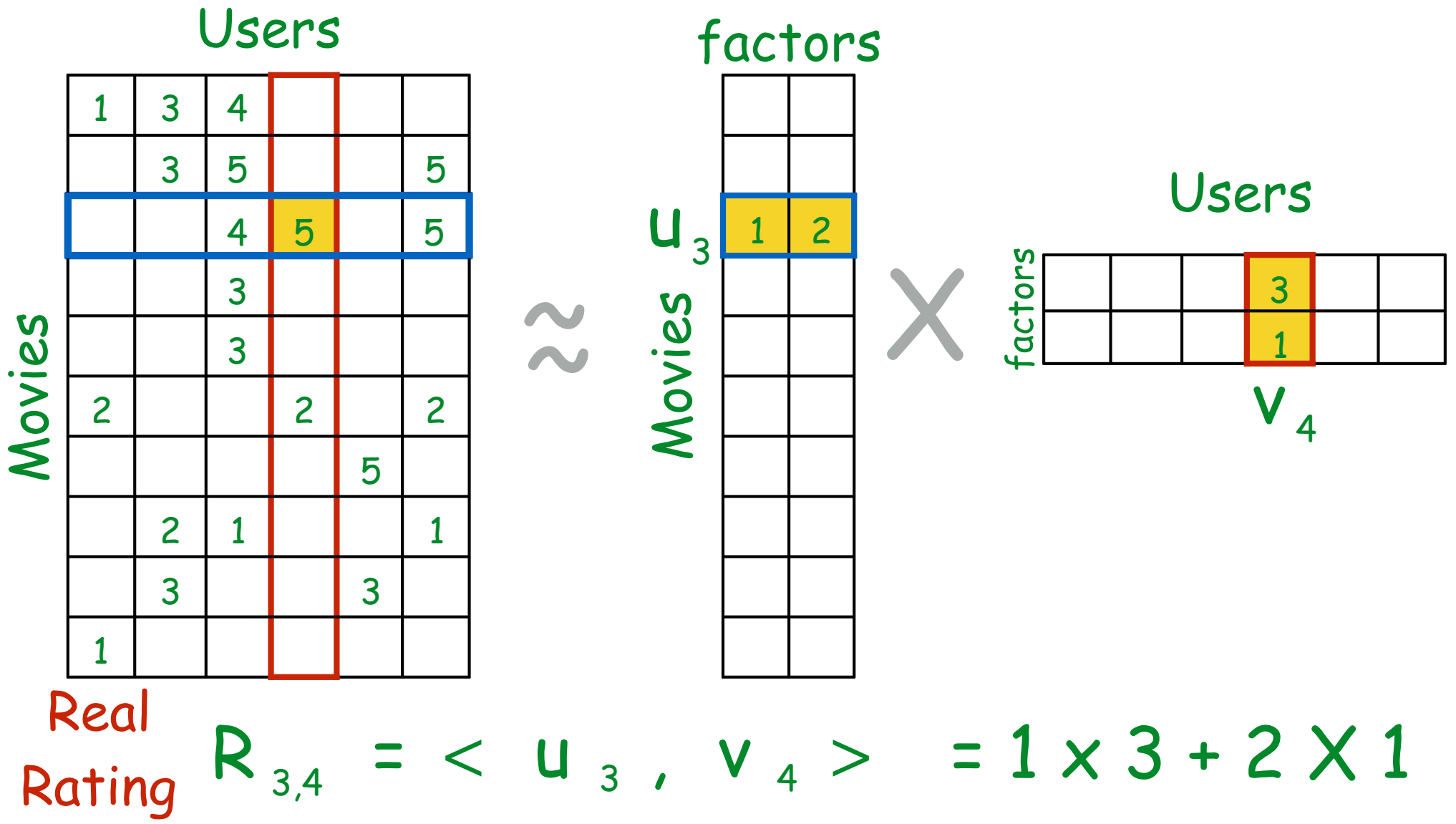
j-th user

Latent Factors



Matrix Decomposition

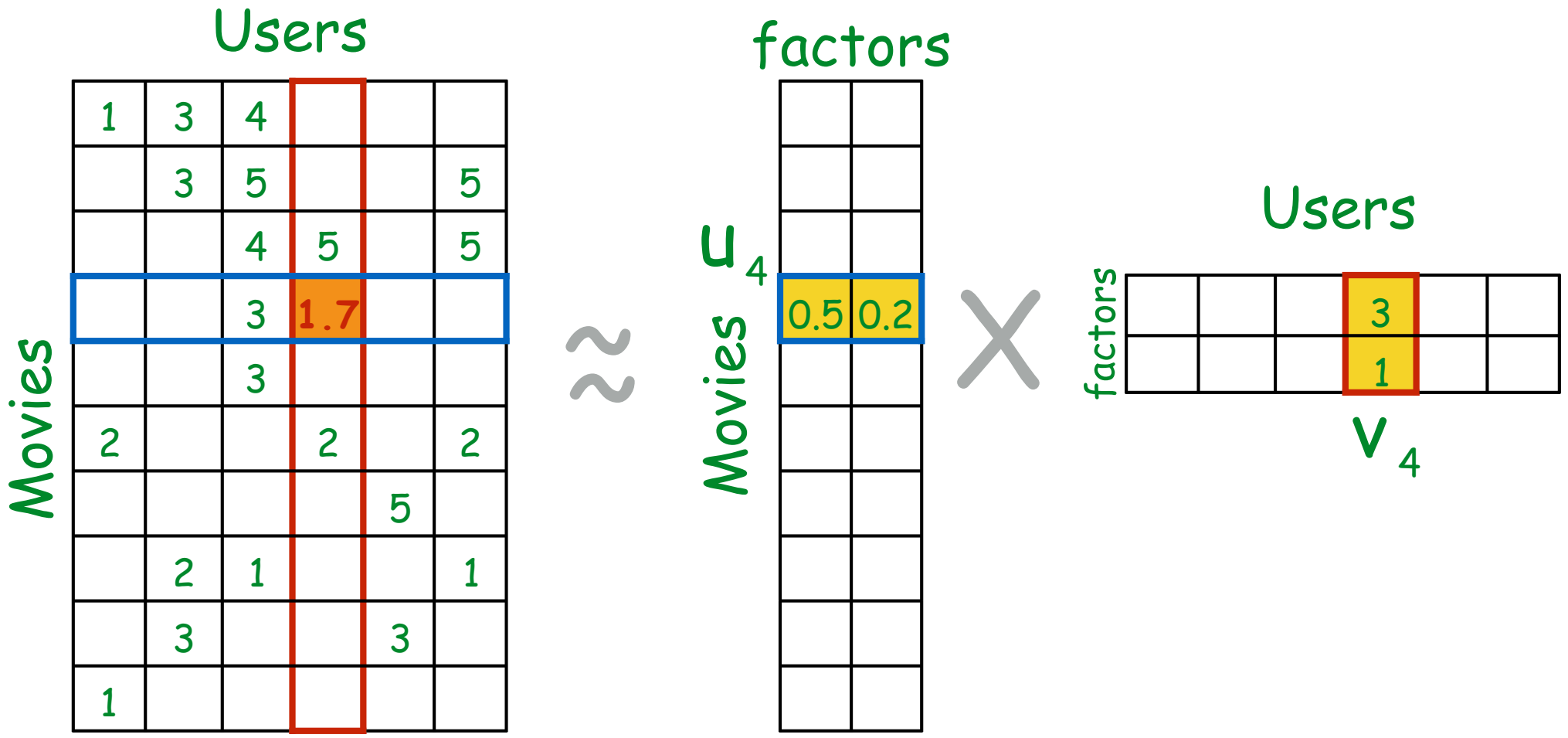
$$R \approx U \times V$$



$$R_{3,4} = < u_3 , v_4 > = 1 \times 3 + 2 \times 1$$

Prediction

$R' \leftarrow U \times V$



Predicted Rating

$R'_{4,4} = < u_4 , v_4 > = .5 \times 3 + .2 \times 1$

Computing Latent Factors

$$\min_{U,V} \sum_{\substack{i,j\text{- has} \\ \text{rating}}} (R_{i,j} - u_i^T v_j)^2$$

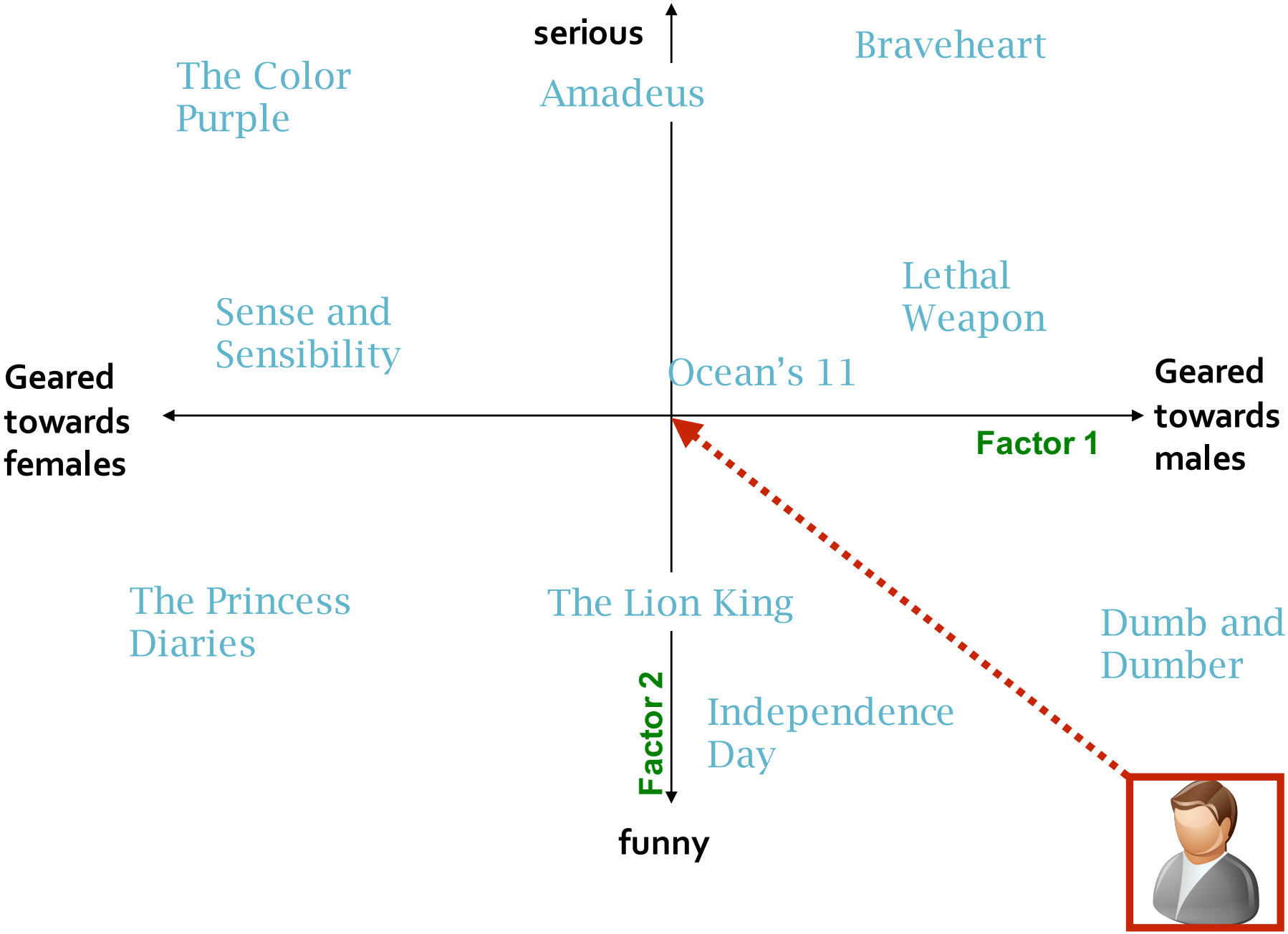
$R_{i,j}$: Real Rating

$u_i^T v_j$: Predicted Rating

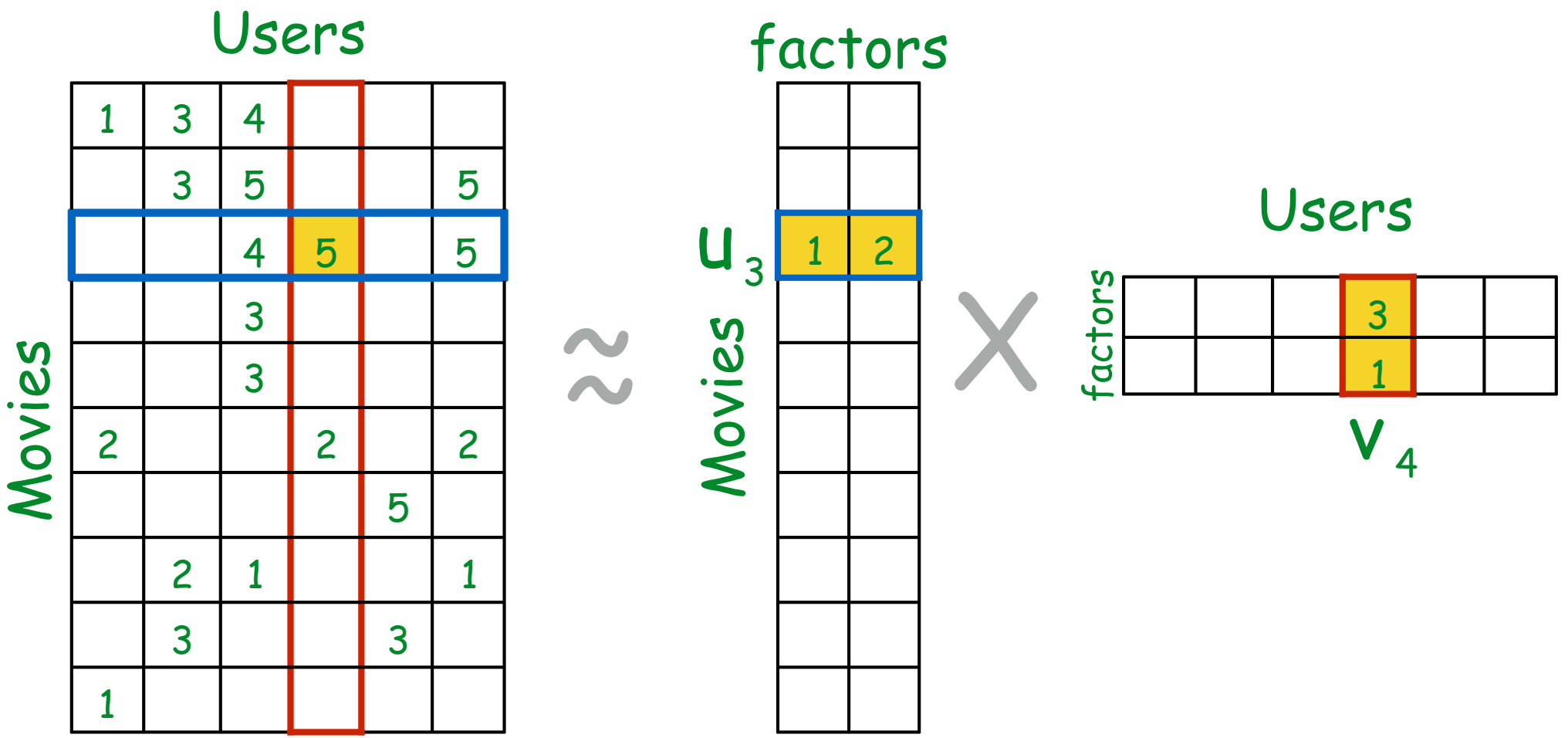
$$+ \mu \left(\sum_i |u_i|^2 + \sum_j |v_j|^2 \right)$$

regularization

Effect of Regularization



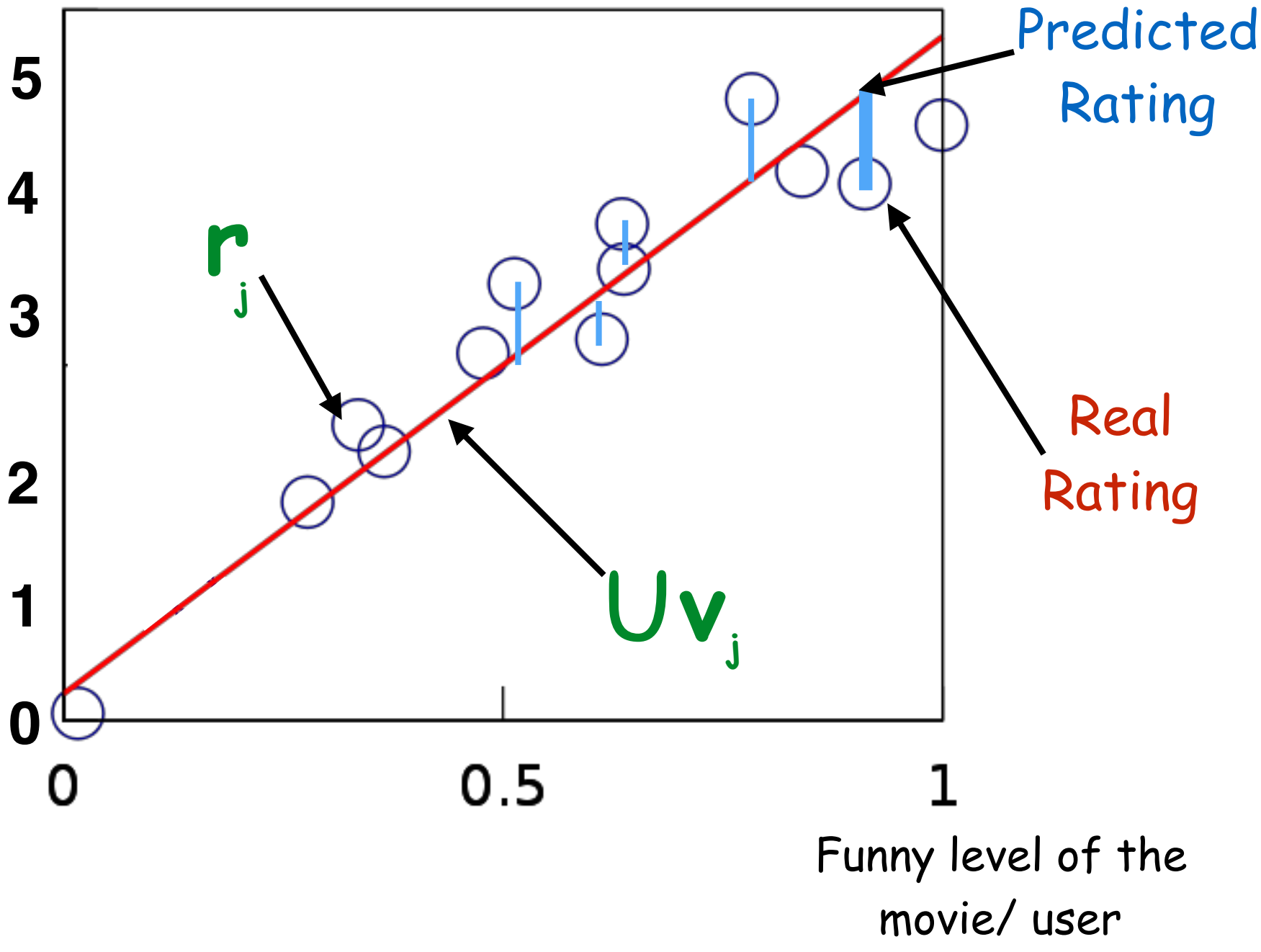
Gradient Descent



Fix u_3 → Update v_4

Update u_3 ← Fix v_4

Update one user/movie



Update one user

$$\min_{v_j} \sum_{\substack{i,j\text{- has} \\ \text{rating}}} (R_{i,j} - u_i^T v_j)^2 + \mu \sum_j |v_j|^2$$

$$\nabla v_j = \sum_{\substack{i,j\text{- has} \\ \text{rating}}} -2\, u_i^T (R_{i,j} - u_i^T v_j) + 2\mu v_j$$

$$v_j' \leftarrow v_j - \beta \nabla v_j$$

Update one movie

$$\min_{u_i} \sum_{\substack{i,j\text{- has} \\ \text{rating}}} (R_{i,j} - u_i^T v_j)^2 + \mu \sum_j |u_i|^2$$

$$\nabla u_i = \sum_{\substack{i,j\text{- has} \\ \text{rating}}} -2(R_{i,j} - u_i^T v_j) v_j^T + 2\mu u_i$$

$$u_i' \leftarrow u_i - \beta \nabla u_i$$

In Matrix Form

update Matrix U while fixing Matrix V

$$\nabla U = -2LV^T + 2\mu U$$

update Matrix V while fixing Matrix U

$$\nabla V = -2U^T L + 2\mu V$$

$$L = (R - UV) \bullet B$$

m - # movies
n - # users
k - # factors

R m by n matrix

U m by k matrix

V k by n matrix

T is the transpose of a matrix

$\bullet$ is element-wise product

B is binary matrix (0/1) of shape m by n

$B_{i,j} = 1$ iff $R_{i,j}$ has rating (>0)

Gradient Descent

Repeat until Convergence (or reach max_step)

update Matrix U for movie latent factors

$$U' \leftarrow U - \beta \nabla U$$

update Matrix V for user latent factors

$$V' \leftarrow V - \beta \nabla V$$

June 2009

NETFLIX

Netflix Prize

Home | Rules | Leaderboard | Register | Update | Submit | Download

Leaderboard

Display top 20 leaders.

Rank	Team Name	Best Score	% Improvement	Last Submit Time
1	BellKor's Pragmatic Chaos	0.8558	10.05	2009-06-26 18:42:37
Grand Prize - RMSE <= 0.8563				
2	PragmaticTheory	0.8582	9.80	2009-06-25 22:15:51
3	BellKor in BigChaos	0.8590	9.71	2009-05-13 08:14:09
4	Grand Prize Team	0.8593	9.68	2009-06-12 08:20:24
5	Dace	0.8604	9.56	2009-04-22 05:57:03
6	BigChaos	0.8613	9.47	2009-06-23 23:06:52
Progress Prize 2008 - RMSE = 0.8616 - Winning Team: BellKor in BigChaos				
7	BellKor	0.8620	9.40	2009-06-24 07:16:02
8	Gravity	0.8634	9.25	2009-04-22 18:31:32
9	Opera Solutions	0.8638	9.21	2009-06-26 23:18:13
10	BruceDengDaoCiYiYou	0.8638	9.21	2009-06-27 00:55:55
11	pengpengzhou	0.8638	9.21	2009-06-27 01:06:43
12	xlvector	0.8639	9.20	2009-06-26 13:49:04
13	xiangliang	0.8639	9.20	2009-06-26 07:47:34
14	Feeds2	0.8641	9.18	2009-06-26 22:51:55
15	Ces	0.8642	9.17	2009-06-24 14:34:14