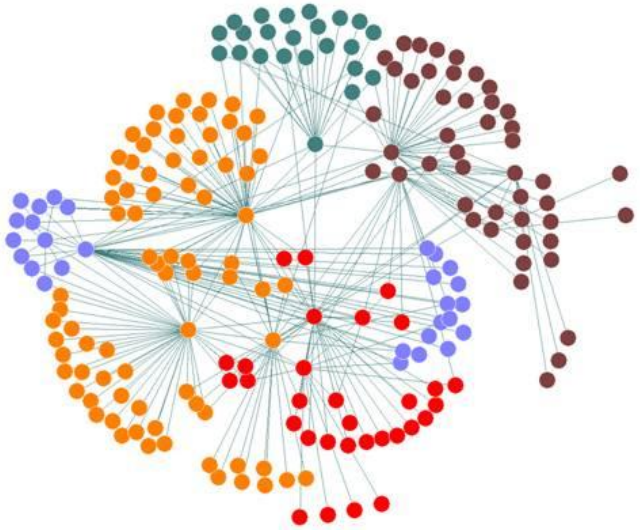
TEL AVIV UNIVERSITY

# Algorithms and Applications in Social Networks



2019/2020, Semester B

Slava Novgorodov

# Lesson #10

- Feed Generation Algorithms
- Advertisement in Social Networks
- Data Privacy and Leakage
- Riddles, riddles, riddles…

# Feed Generation Algorithms

# News Feeds in Social Networks

- The news feed is the primary system through which the users of the social network are exposed to the content and the updates

- We focus on Facebook, as it has 2B active users around the world and Facebook's newsfeed is the most viewed and the most influential aspect of the news industry

# Disclaimer

- All the information here is based on publically available information

- The author has no information about the internals of the algorithms

# Facebook before 2006

- Facebook introduced the newsfeed in September 2006. Before that it mainly focused on the users profiles and the connections…

[ thefacebook ]
home search global social net invite faq logout
Puget S
Mark Zuckerberg's Profile
quick search go
My Profile [ edit ]
My Friends
My Parties
My Messages
My Account
My Privacy
Picture
Send Mark a Message
Poke Him!
Connection
This is you.
Mutual Friends
You have 19 friends in comm
Information
Account Info:
Name: Mark Zuckerberg
Member Since: The beginning (February 4, 2004)
Last Update: July 19, 2005
Basic Info:
School Harvard
Status: Student
Sex: Male
Residence: Todd 311
Birthday: May 14, 1984
Home Town: Dobbs Ferry, NY
High School: Phillips Exeter Academy
Contact Info:
Email:
Screenname:
Mobile:

# Facebook after 2006

- In September 2006 Facebook announced the new home page, called the “News Feed”

- Received critics because it was too detailed (people felt it violates their privacy

facebook
home search browse invite help logout
Welcome Mark!
Search
My Profile edit
My Friends
My Photos
My Notes
My Shares (1)
My Groups
My Events
My Messages (14)
My Account
My Privacy
News Feed
David uploaded a mobile photo. 6:25pm
Share
Mobile Uploads
2 photos
by David
James created a group. 8:05pm
Share
Brandee joined the group Apple Students. 5:33pm
Share
Jim is thinking the Giants have a prayer! 8:19am
April joined the group Chase +1. 7:39am
Share
Kasey commented on Brandee's photo. 7:19am
I "heart" Hootie and totally think she needs a group fan club!
See Wall-to-Wall.
Alison is charging it. 6:54am
Chris joined the group Feingold for President 2008. 5:21pm
Share
Katie uploaded a mobile photo. 12:44pm
Share
Launch!
1 photo
by Katie
Location: HQ
Linh Doan added new photos. 11:16pm
Share
Requests
∞ friend requests
1 group invitation
Messages
13 new messages
My Status edit
Mark is at work.
Updated on Friday
Share Inbox see all
1 new shared item
Group: Chase +1
From Carolyn Abram
Photo Album: DS Weekend
From Chris Hughes
Birthdays see all
No upcoming birthdays.
The Next Step see all
How well do you know your friends? Play the Friend Game to find out.

# The algorithm

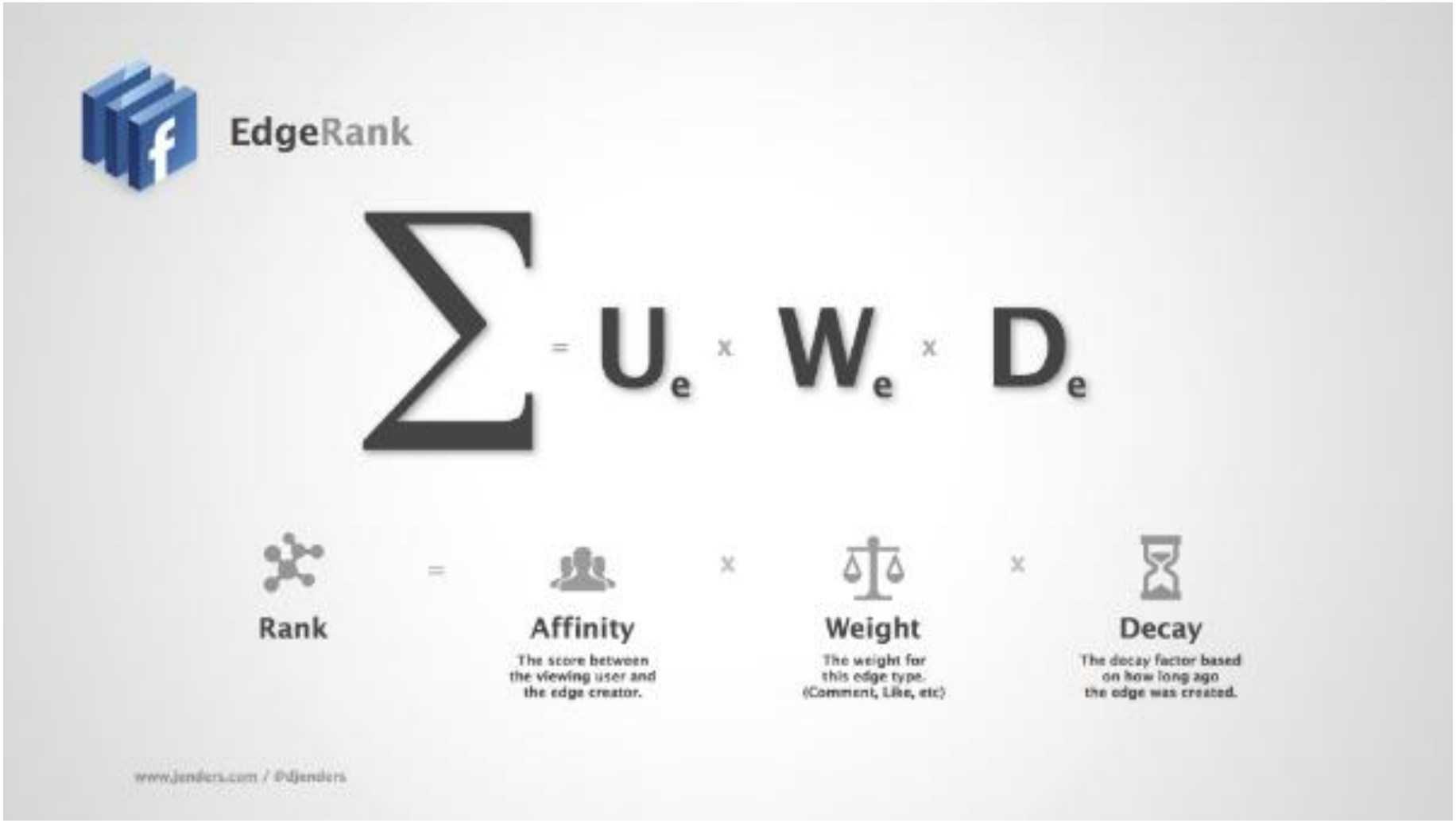
- Once the news feed became the main place where the users are spending most of the time, it’s neccesarry to decide how to sort and filter the news feed

- The problem became even more critical with the growth of the numbers of users (the content should be more relevant)

# EdgeRank

- Facebook's proprietary algorithm called EdgeRank is used to sort and filter the news feed. Simplified version is build from 3 elements:
  - Affinity – the proximity of the user to "like" the content
  - Weight – the weight of the content
  - Decay – the time-based decay factor

# EdgeRank


EdgeRank
$\sum = U_e \times W_e \times D_e$
Rank
=
Affinity
The score between the viewing user and the edge creator.
x
Weight
The weight for this edge type. (Comment, Like, etc)
x
Decay
The decay factor based on how long ago the edge was created.
www.jenders.com / @djenders

# EdgeRank

Each time you open Facebook it needs to decide between ~2000 potential content units (posts/photos…)

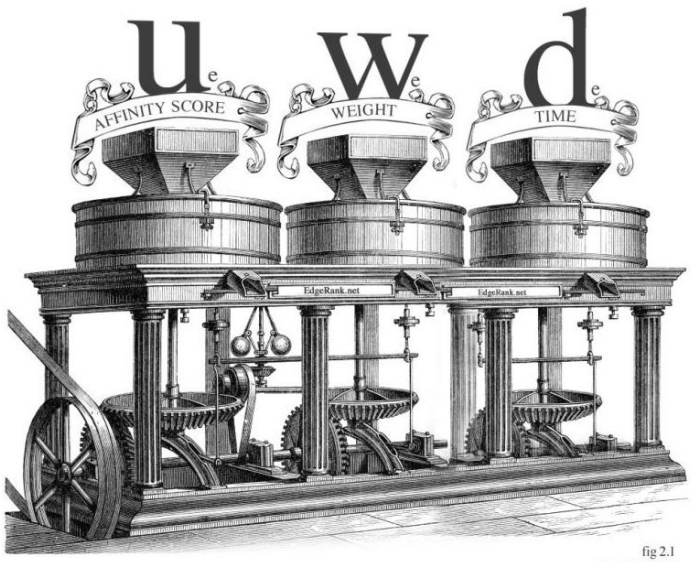



fig 2.1
Source: EdgeRank.net

$$\sum_{edges\ e} u_e w_e d_e$$

$u_e$ ~ affinity score between viewing user and edge creator
$w_e$ ~ weight for this edge type (create, connect, like, tag, ect.)
$d_e$ ~ time decay factor based on how long ago the edge was created

# EdgeRank

- In addition to the 3 main factors, there are 4 additional factors that are increasing the chance of the content to show-up:
  - Previous interactions with the author
  - Previous interactions with this type of posts
  - Reactions of users who already saw the post
  - Amount of negative feedback

## 1 YOUR PREVIOUS INTERACTIONS WITH THE AUTHOR

THE MORE YOU ENGAGE WITH A FRIEND OR PAGE, THE MORE LIKELY YOU ARE TO SEE THEIR POSTS.

## 2 YOUR PREVIOUS INTERACTIONS WITH THE POST TYPE

IF YOU OFTEN ENGAGE WITH A CERTAIN TYPE OF POST, YOU'RE MORE LIKELY TO SEE POSTS OF THAT TYPE

## 3 REACTIONS FROM USERS WHO ALREADY SAW THE POST

THE MORE FACEBOOK USERS ENGAGE WITH A POST, THE MORE LIKELY YOU ARE TO SEE THAT POST.

## 4 AMOUNT OF COMPLAINTS OR NEGATIVE FEEDBACK*

AS MORE USERS GIVE NEGATIVE FEEDBACK, THE LESS LIKELY YOU ARE TO SEE THAT POST.

# Another side of the moon

- 5 tips for content makers to increase the reach
  - Post frequently
  - Use pictures
  - Be active
  - Stay clean (avoid reports)
  - Focus on quality

# Click-bait control

- In 2014 Facebook added the click-bait control to reduce the number of clickbaits in users newsfeeds

# The timeline of algorithm updates

- 2006 – The launch of the News Feed
- 2009 – Change from chronological sort to popularity sort
- 2013 – New design
- 2014 – Click-bait control
- 2014 – Video content prioritization
- 2015 – Friends posts prioritization
- 2017 – Reactions valued more then “likes”
- …

# Advertisement in Social Networks

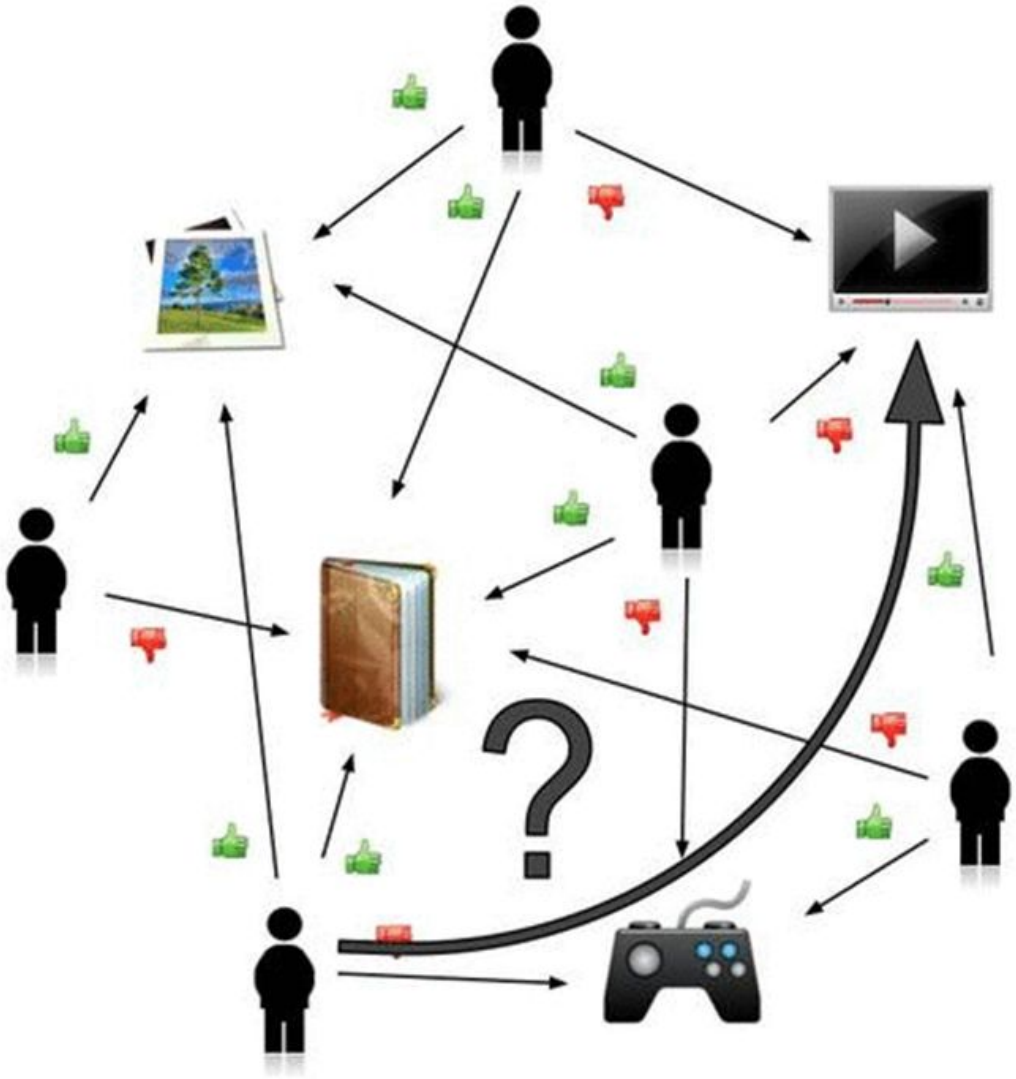
# Ads in Web

- Many different techniques for deciding which ad to show to whom
- The most popular technique – collaborative filtering

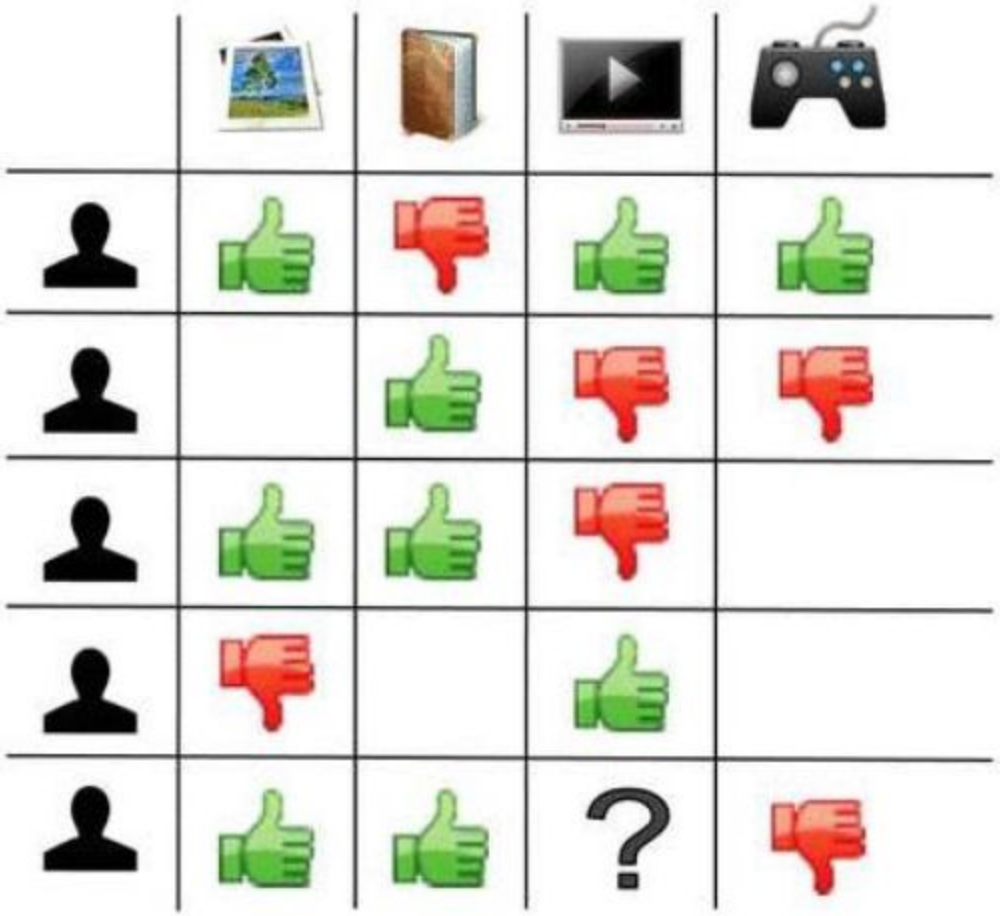
# CF in Brief

- The data is modeled as a matrix of Users x Ads
- Each cell is filled with the interaction value between the user and the ad
- The missing cells can be “filled” based on the similarity between users of between the ads

# CF in Brief



# CF in Brief

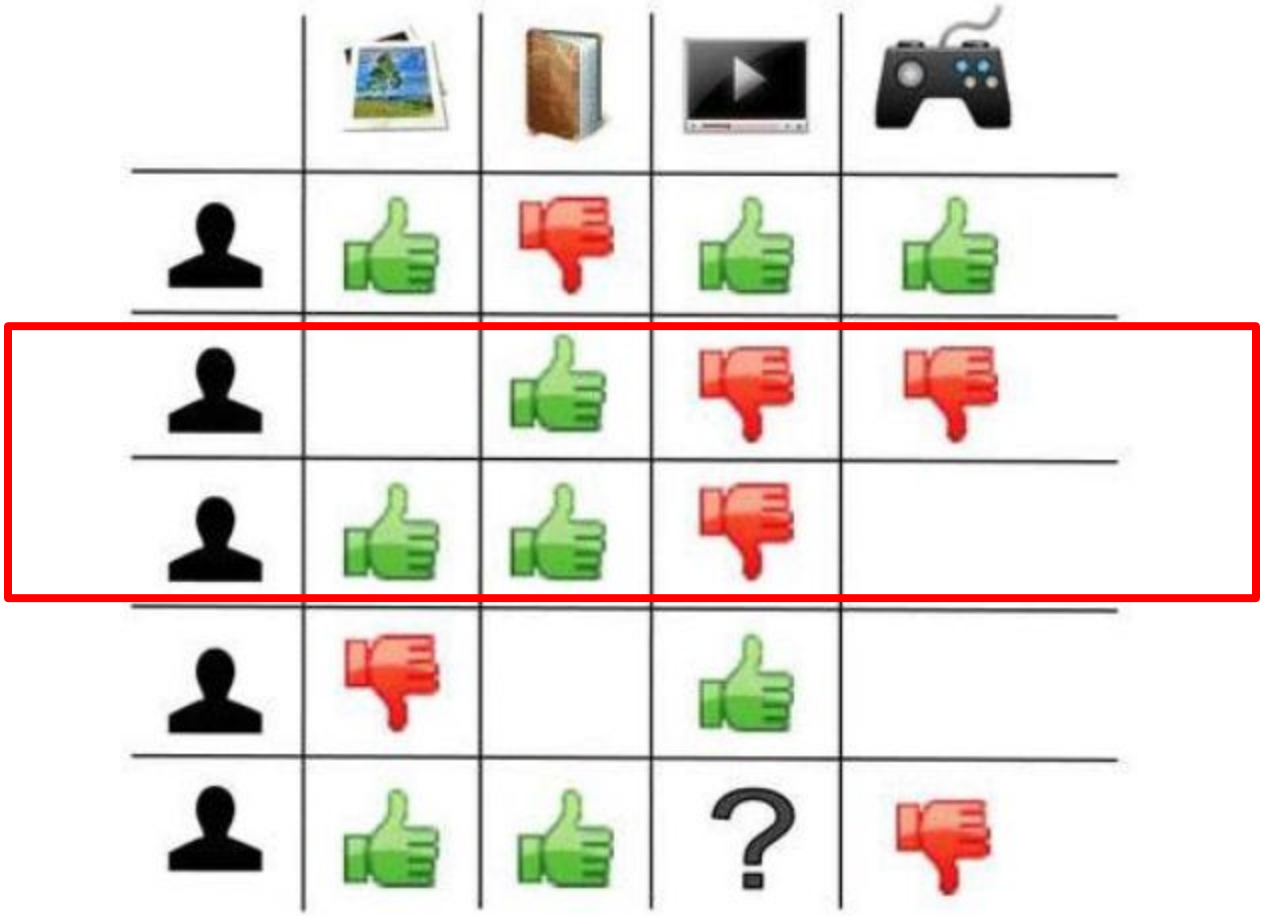


# CF in Brief

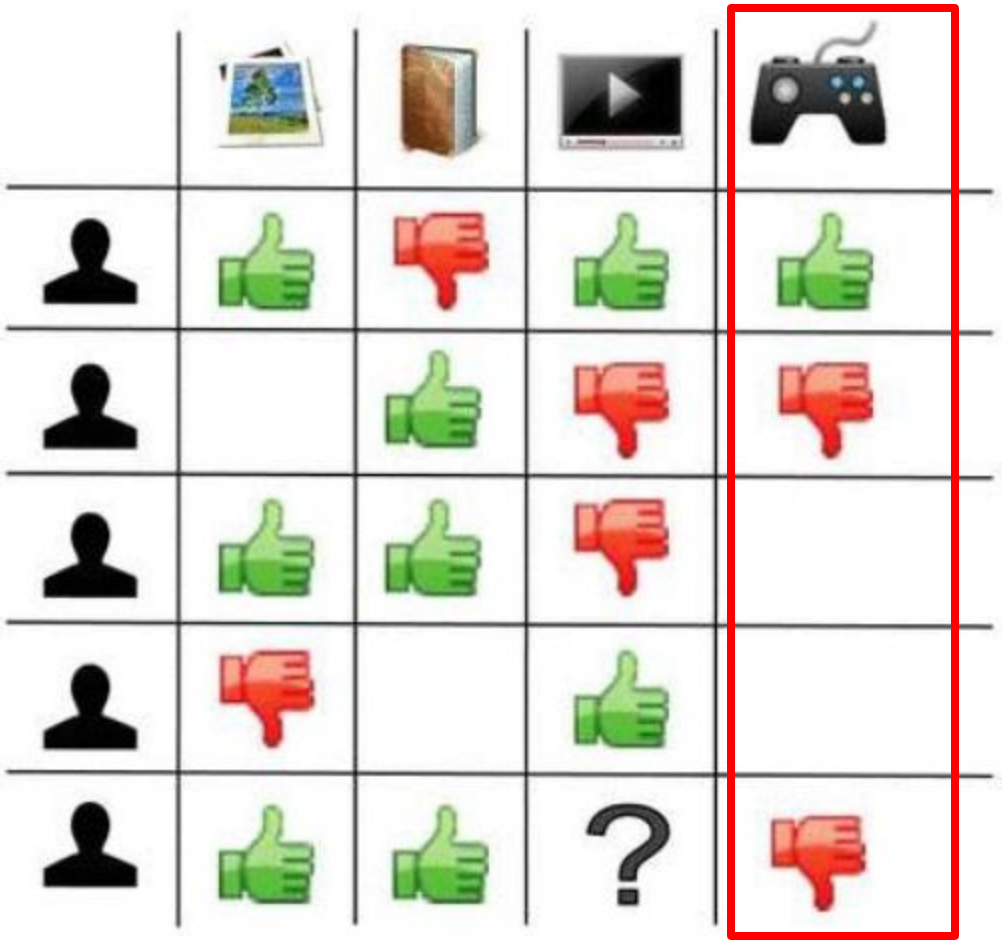
- User based – find similarity between users and “fill” with the action that the similar users did on this ad

- Item based – find similar items and “fill” the action that this user did on the similar ads

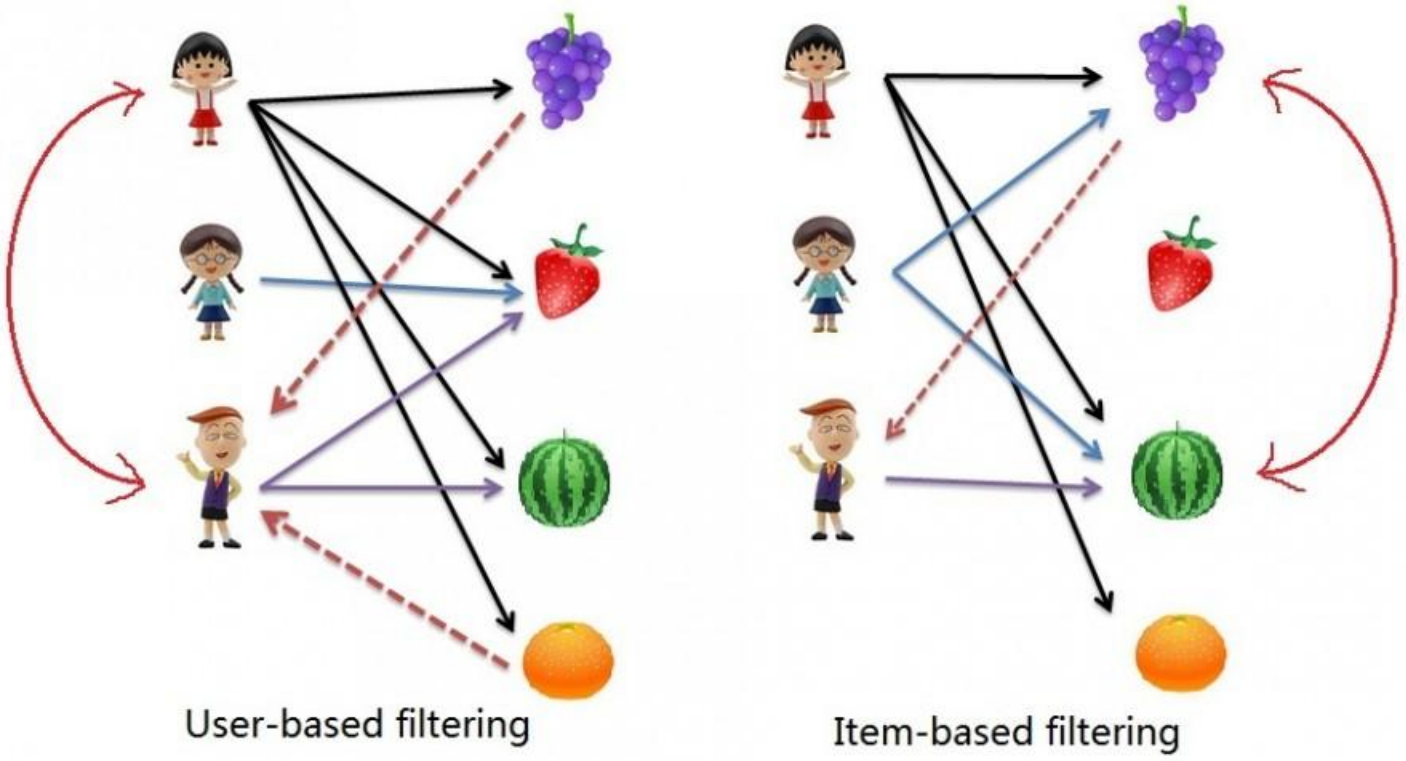
# User Based



# Item Based



# User Based vs Item Based CF

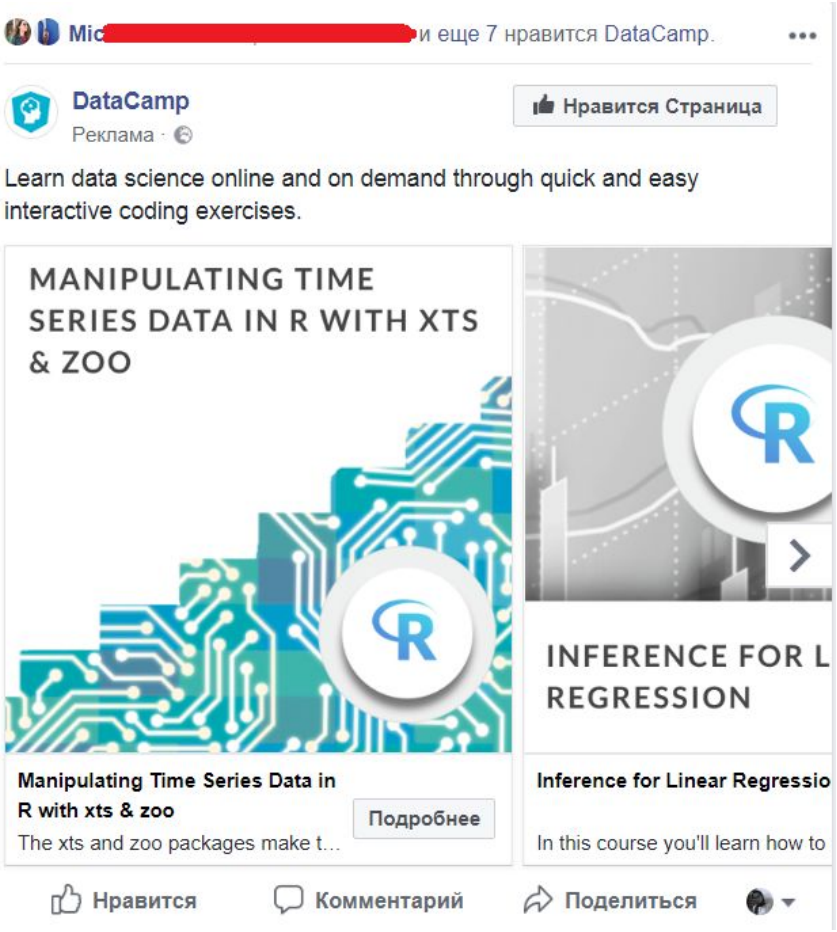


# Similarity Functions

- Similar to the functions we saw in the previous lectures:
  - Cosine similarity between the vectors
  - Jaccard similarity
  - Euclidian distance
  - etc

# Ads on Facebook

- “Organic”, part of the news feed



# Data Privacy and Leakage

# Data Privacy and Leakage

- After we understood how the news feed is generated and how the ads work in general on the Web, let's talk about data

# Information Sharing

- For internal content is easy, if you like/search/visit some FB page, Facebook has information about it

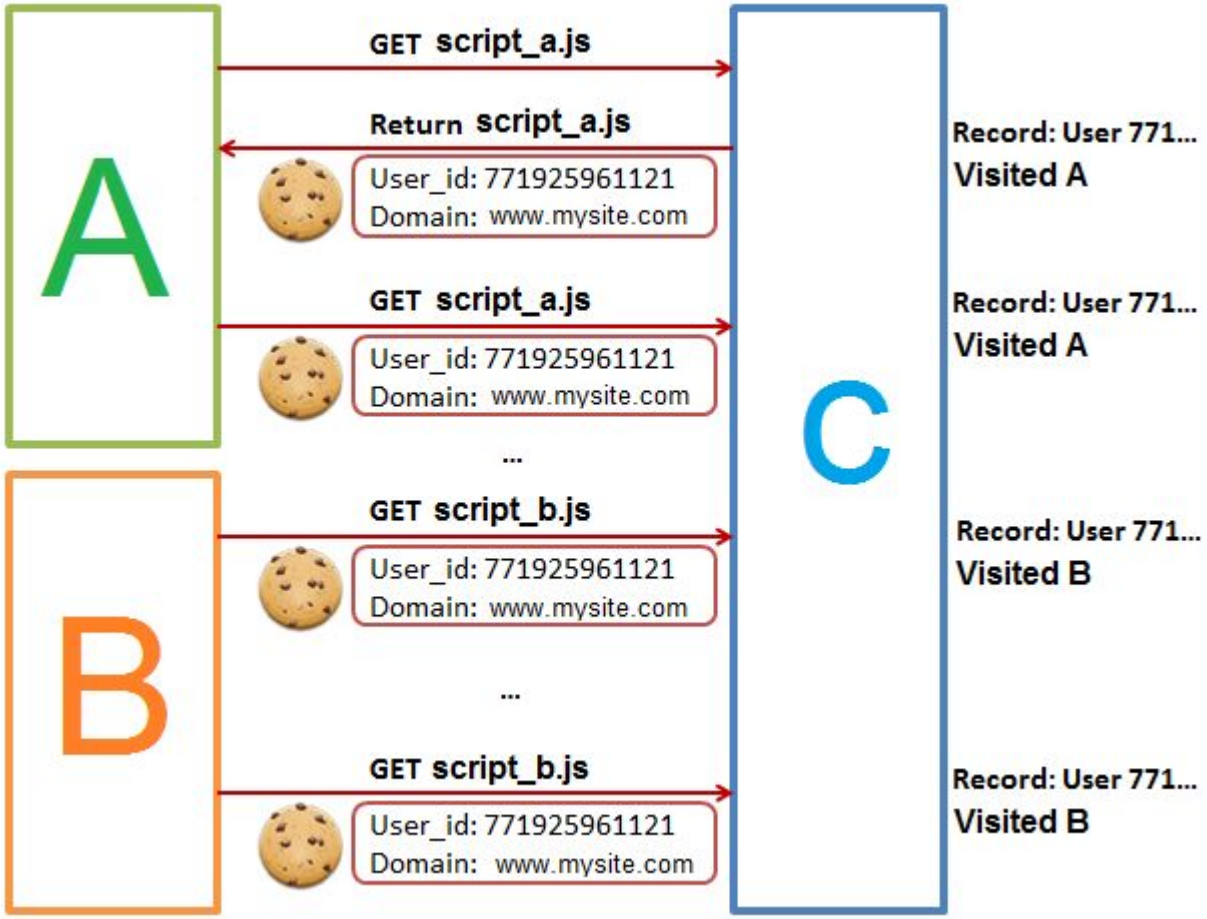
- But how do they know your “external” behavior

# Cookies

- A **HTTP cookie** (also called **web cookie**, **Internet cookie**, **browser cookie** or simply **cookie**), is a small piece of data sent from a website and stored in a user's web browser while the user is browsing that website.
- Every time the user loads the website, the browser sends the cookie back to the server to notify the website of the user's previous activity

# Tracking Cookies

Mainly used for users tracking over the web (ads, statistics, etc…)



# Ads on Facebook

- So, if user visits some external website that cooperates with Facebook (put's FB's script/pixel) – Facebook knows about it!

- This data is enough for the deciding which ad to choose

# Data leakage

- Facebook contains a lot of personal information and allow advertisers to create "targeted campaigns"

- Targeting options include (directly and indirectly) many properties, including the "collected externally" properties

Example: Person visits "Dr. Baby", "Babies", "Toys R'Us"  Has kids

# Data leakage

- Targeting and analyzing the people interacting with the ad afterwards

- People who interacted – targeting group (i.e. has the property) – not private anymore

# Data leakage

# Data leakage

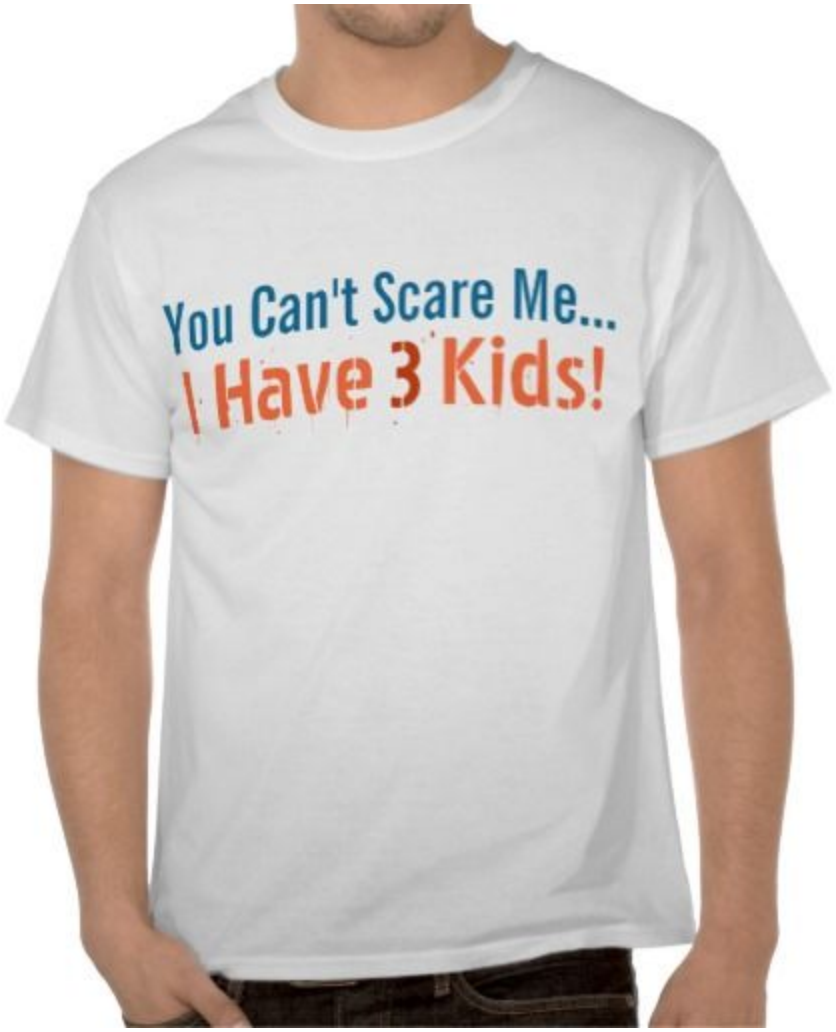
Example, based on personal experience:

- I started to get cool designed shirts with "personalized" writings
- The most surprising was "Cool people are born in November **1989**"
- When I came to the comments of the ad I saw 2K+ likes and hundrents of comments – most of them was of this age!
- People also tagged their friends and so on…

# Data leakage

# Data leakage

# Data leakage



# Data leakage

# Data leakage

- Other types of leakage: groups created based on a property:
  - “Ford Mustang drivers” □ has Ford Mustang
  - “Calculus 1 15/16 TAU” □ studies in TAU, Math/CS, just finished his B.Sc
  - …

# Data leakage

- To summarize – the problem of leakage is not dependent on Facebook, but on our interaction with these ads.

- Facebook works strictly by law, but we are doing the extra step and providing this information to the advertisers

# Riddles, riddles, riddles…

# Thank you!

## Questions?