JOÃO VINAGRE

DSPT MEETUP #60

"I KNOW WHAT YOU WANT!": THE MAGIC, THE SCIENCE AND THE FUTURE OF RECOMMENDER SYSTEMS

RECOMMENDER SYSTEMS

Google

NETFLIX

YouTube

amazon

Spotify

twitter

PANDORA
INTERNET RADIO

facebook

Instagram

LinkedIn

Quora

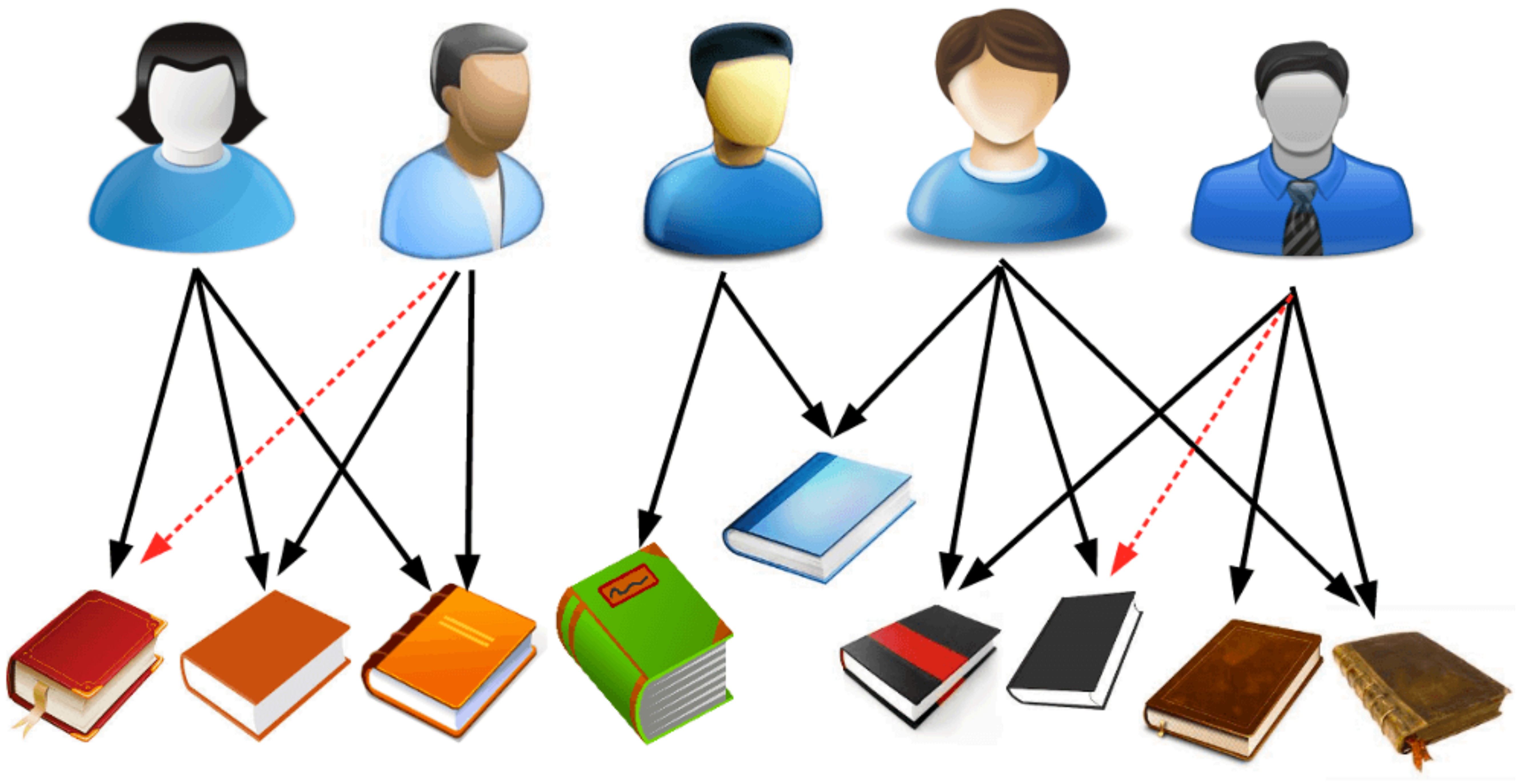
iTunes

last.fm
the social music revolution

ebay

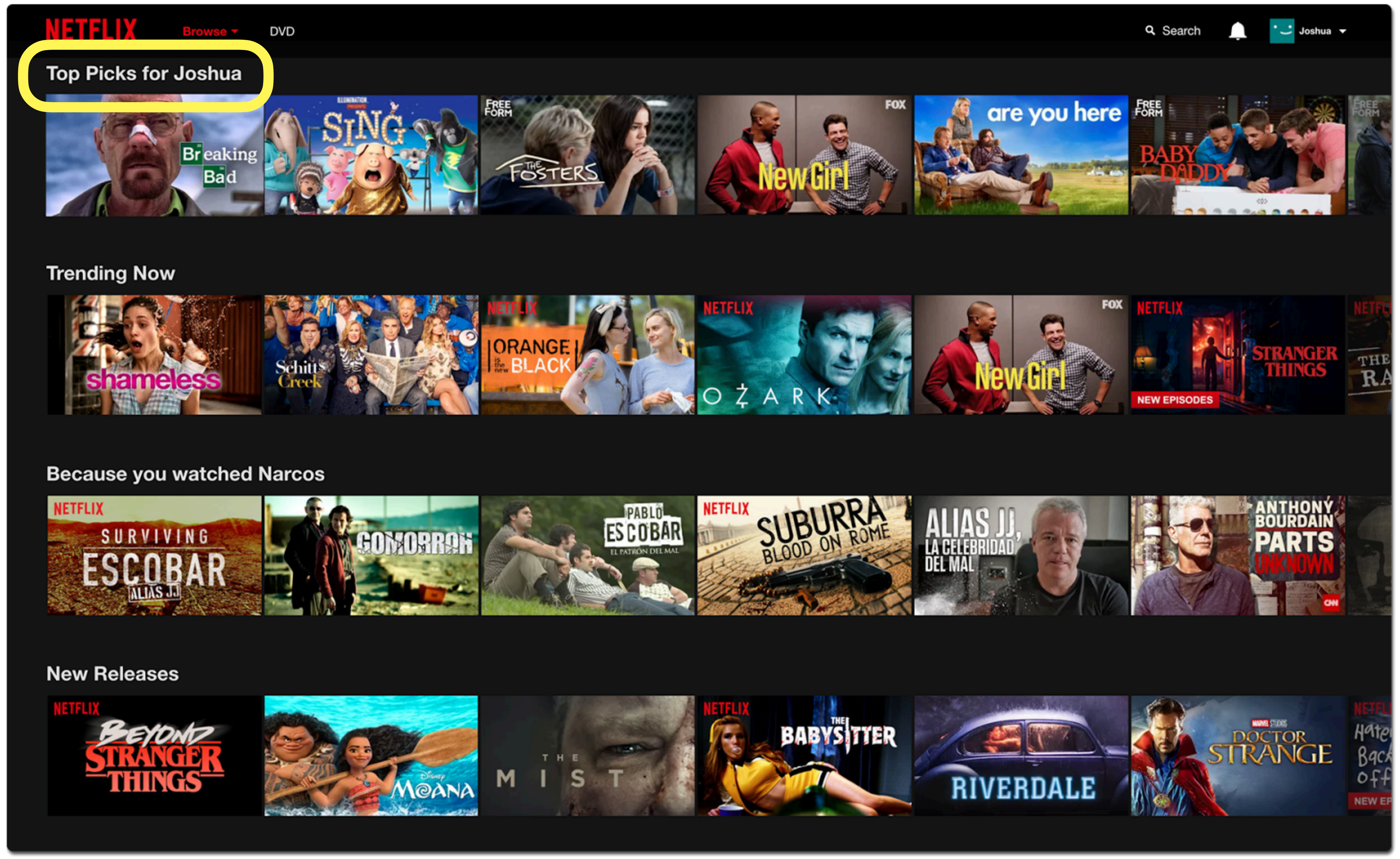
RECOMMENDER SYSTEMS

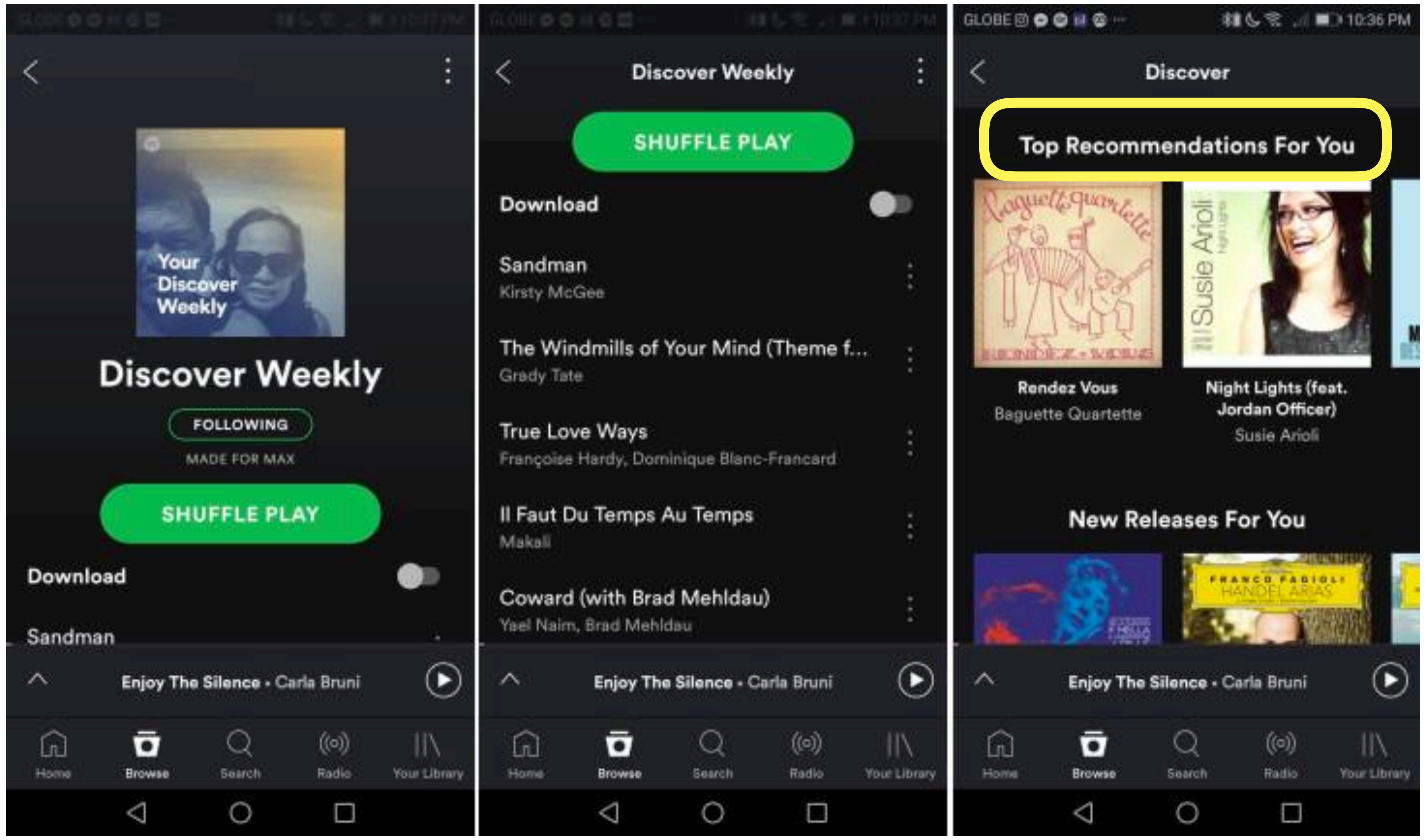
Algorithms that guess what we like



TASK

To provide **personalised** recommendations

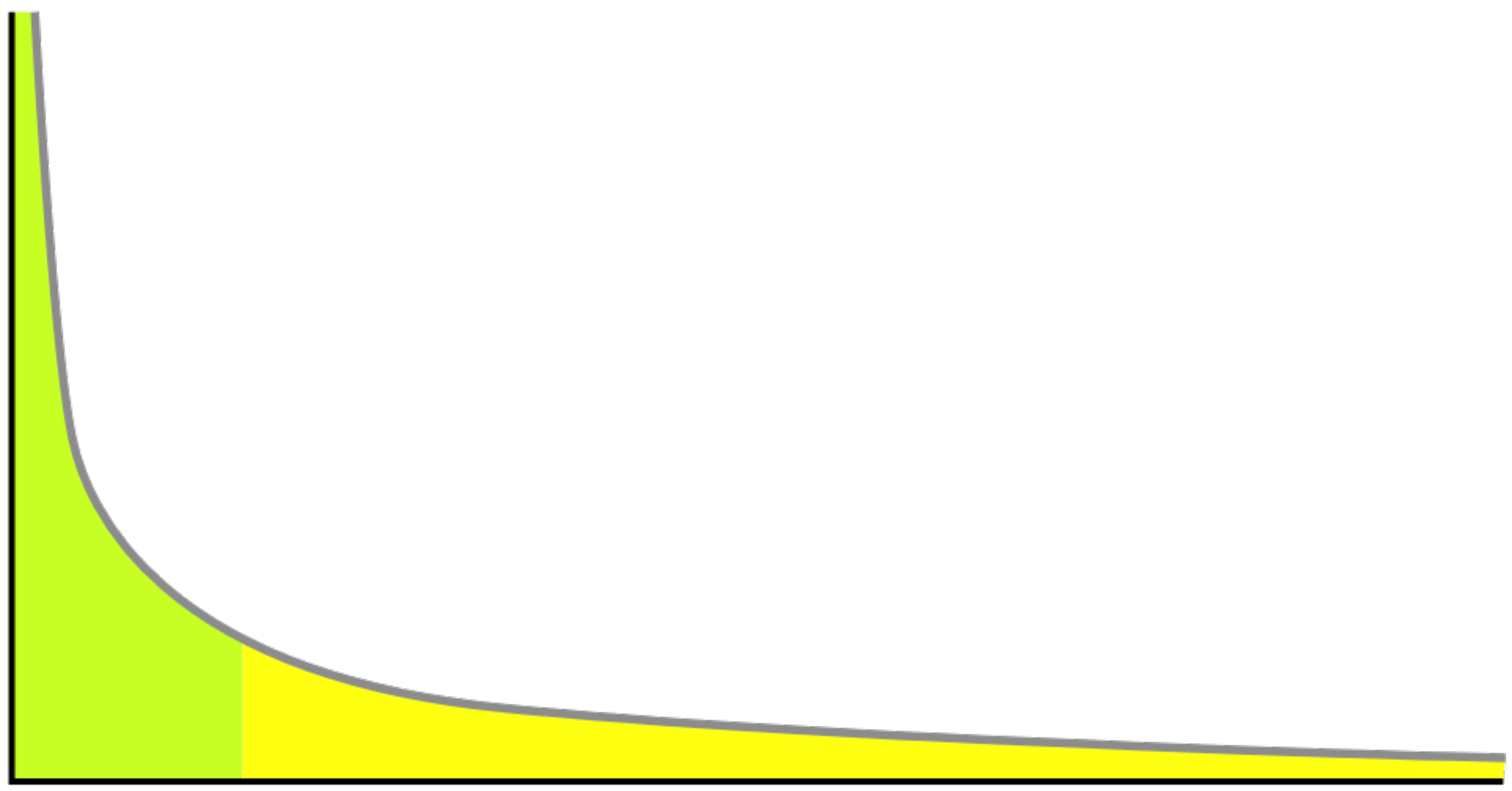




VALUE

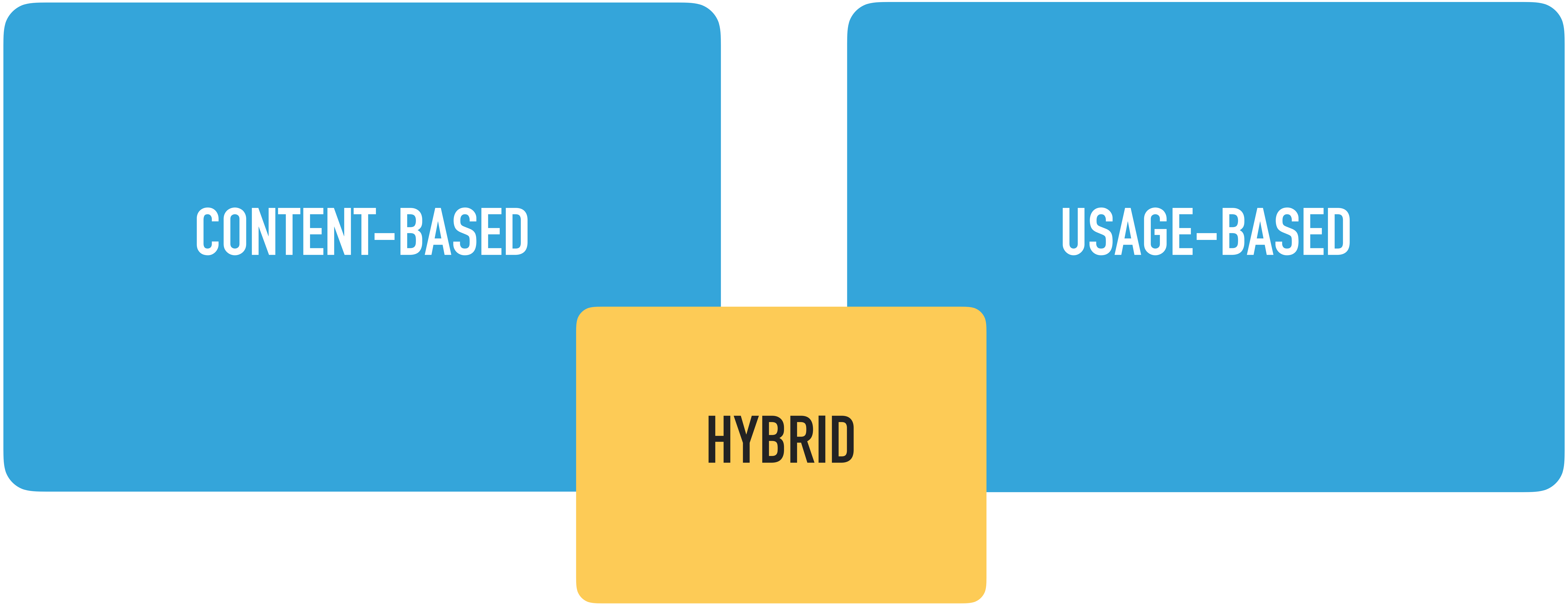
- Netflix
  - 75% Traffic (Mckinsey & Co., 2013),
  - ~1B USD (Gomez-Uribe et al, 2016)
- Youtube
  - 60% Traffic (Davidson et al, 2010)
- Amazon
  - 35% Purchases (McKinsey & Co., 2013)

WHY ARE THEY SO EFFECTIVE?



THE LONG TAIL

TYPES OF RECOMMENDER SYSTEMS



CONTENT-BASED ALGORITHMS

- Use content analysis techniques
  - Computer vision
  - Audio analysis
  - NLP
- Domain-dependent
- Research is mostly industry-driven

USAGE-BASED ALGORITHMS

- Analyse user-item interactions
  - Neighborhood-based techniques
  - Latent factor models (MF, TF, NN)
- Domain-independent
- Research is driven by both industry and academia

CLASSIC COLLABORATIVE FILTERING

Given a set of known ratings, predict the missing ones.

	i_1	i_2	i_3	i_4	…	i_n
u_1	1		5			
u_2			4			2
u_3				3		
u_4		1		5		
…						
u_m			2			

DO WE REALLY NEED RATINGS?

- Ratings (explicit) data:

★★★★☆

- Implicit data:
  - like/share buttons
  - web access logs / spam flag
  - music listening / playlisting
  - shopping history
  - news reading
  - event participation
  - …
- More widely available and less intrusive (for both users and system)

CLASSIC COLLABORATIVE FILTERING

Implicit: given a set of interactions, predict the likely ones.

	i_1	i_2	i_3	i_4	…	i_n
u_1	√		√			
u_2			√			√
u_3				√		
u_4		√		√		
…						
u_m			√			

NEIGHBORHOOD-BASED

Take users (rows) as vectors

	i_1	i_2	i_3	i_4	…	i_n
u_1	√		√			
u_2			√			√
u_3				√		
u_4		√		√		
…						
u_m			√			

$$\mathrm{sim}(u_1, u_2) = \cos(\vec{u}_1, \vec{u}_2) = \frac{|I_{u_1} \cap I_{u_2}|}{|I_{u_1}||I_{u_2}|}$$

NEIGHBORHOOD-BASED

Or Items (columns)!

	i_1	i_2	i_3	i_4	…	i_n
u_1	√		√			
u_2			√			√
u_3				√		
u_4		√		√		
…						
u_m			√			

$$\mathrm{sim}(i_1, i_2) = \cos(\vec{i}_1, \vec{i}_2) = \frac{|U_{i_1} \cap U_{i_2}|}{|U_{i_1}||U_{i_2}|}$$

- Good alternatives
  - Pearson correlation
  - Jaccard index
- Bad alternatives
  - Euclidean metrics

MATRIX FACTORIZATION

R

	i_1	i_2	i_3	i_4	…	i_n
u_1	1		5			
u_2			4			2
u_3				3		
u_4		1		5		
…						
u_m			2			

=

P

	f_1	f_2	…	f_k
u_1	·	·	·	·
u_2	·	·	·	·
u_3	·	·	·	·
u_4	·	·	·	·
…	·	·	·	·
u_m	·	·	·	·

x

Q^T

	i_1	i_2	i_3	i_4	…	i_n
f_1	·	·	·	·	·	·
f_2	·	·	·	·	·	·
…	·	·	·	·	·	·
f_k	·	·	·	·	·	·

$$\hat{r}_{ui} = p_u q_i^T = \vec{p}_u \cdot \vec{q}_i$$

- Model consists of matrices P and Q
- k is the number of latent features (the magic)

STOCHASTIC GRADIENT DESCENT

$$\min_{P\cdot,Q\cdot} \sum_{(u,i)\in D} (r_{ui} - p_u q_i^T)^2 + \lambda(||p_u||^2 + ||q_i||^2)$$

initialize P and Q **with values close to** 0

repeat *iter* **times**:

shuffle D

for each *(u,i,r)* **in** D:

$$\epsilon \leftarrow r - p_u q_i^T$$

$$p_u \leftarrow p_u + \eta(\epsilon q_i - \lambda p_u)$$

$$q_i \leftarrow q_i + \eta(\epsilon p_u - \lambda q_i)$$

HYPER-PARAMETERS:

k - no. latent features

iter - no. iterations (epochs)

η - learn rate (aka step size)

λ - regularization factor

STOCHASTIC GRADIENT DESCENT WITH IMPLICIT DATA

	i_1	i_2	i_3	i_4	…	i_n
u_1	√		√			
u_2			√			√
u_3				√		
u_4		√		√		
…						
u_m			√			

	i_1	i_2	i_3	i_4	…	i_n
u_1	1		1			
u_2			1			1
u_3				1		
u_4		1		1		
…						
u_m			1			

Apply SGD

STOCHASTIC GRADIENT DESCENT WITH IMPLICIT DATA

$$\min_{P_{\cdot}, Q_{\cdot}} \sum_{(u,i)\in D} (1 - p_u q_i^T)^2 + \lambda(||p_u||^2 + ||q_i||^2)$$

initialize P and Q **with values close to** 0

repeat *iter* **times**:
  shuffle D
  for each (u,i) **in** D:

$$\epsilon \leftarrow 1 - p_u q_i^T$$
$$p_u \leftarrow p_u + \eta(\epsilon q_i - \lambda p_u)$$
$$q_i \leftarrow q_i + \eta(\epsilon p_u - \lambda q_i)$$

HYPER-PARAMETERS:

feat - no. latent features

iter - no. iterations (epochs)

η - learn rate (aka step size)

λ - regularization factor

INCREMENTAL MF

- SGD ***is*** incremental
- Multiple passes and shuffling are only useful when data is finite [LeCunn et al., 1996; Bottou, 2003]

initialize P and Q **with values close to** 0

for each *(u,i)* **in** D**:**
repeat *iter* **times:**

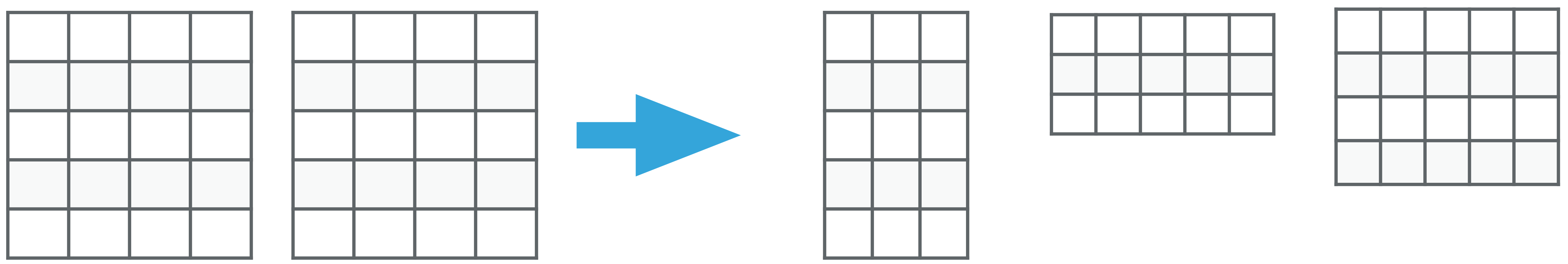
$$\epsilon \leftarrow 1 - p_u q_i^T$$
$$p_u \leftarrow p_u + \eta(\epsilon q_i - \lambda p_u)$$
$$q_i \leftarrow q_i + \eta(\epsilon p_u - \lambda q_i)$$

HYPERPARAMETERS:

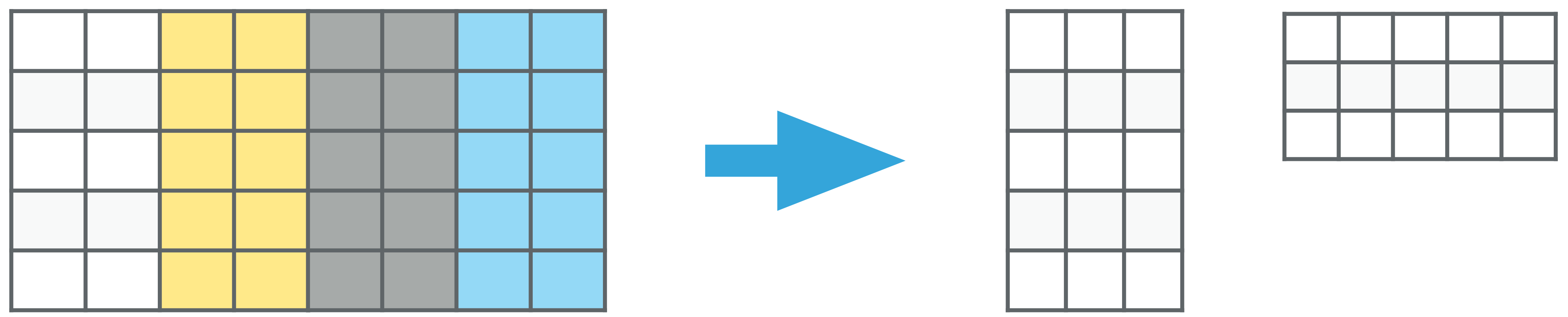
feat - no. latent features
iter - no. iterations (epochs)
η - learn rate (aka step size)
λ - regularization factor

CONTEXT-AWARENESS & SIDE INFORMATION

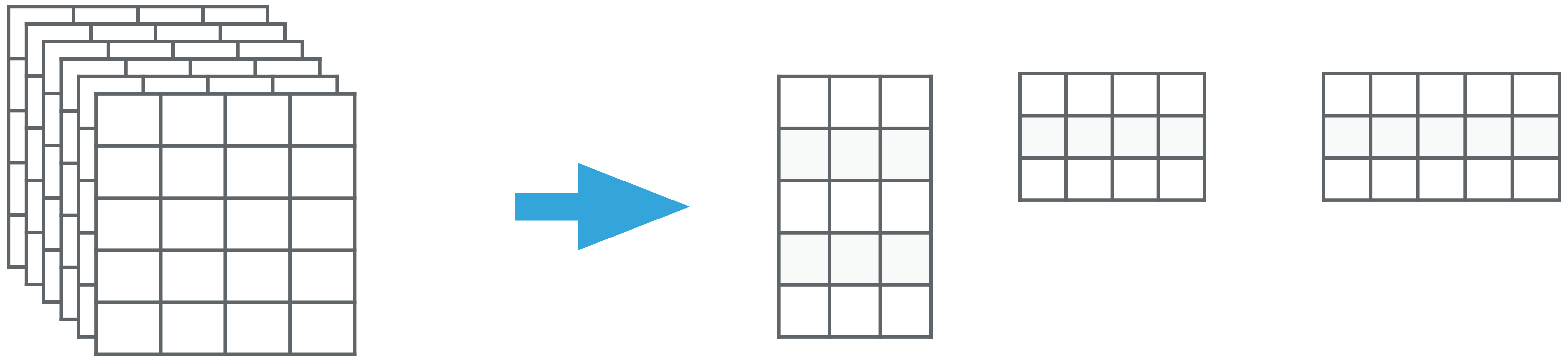
- Matrix co-factorization



- Factorization machines



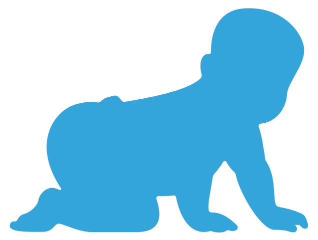
- Tensor factorization



- Data pre-processing (DaVI, etc)

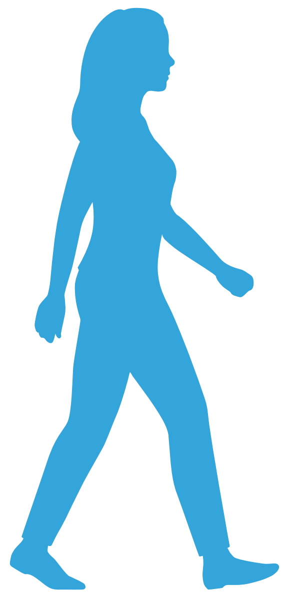
EVALUATION

- Offline batch
  - ML method: train-test, cross-validation



- Offline dynamic (stream-based)
  - Prequential
  - Holdout

- Online
  - A/B testing
  - Multivariate tests
  - User surveys



EVALUATION - METRICS

- Accuracy metrics:
  - RMSE (!!)
  - Relevance metrics (Precision, Recall, F1)
  - Top-N metrics (MAP, NDCG)
- Other dimensions
  - Novelty / Serendipity
  - Diversity
  - Coherence
  - Coverage
  - Trust / Confidence

PROBLEMS I: POPULARITY BIAS

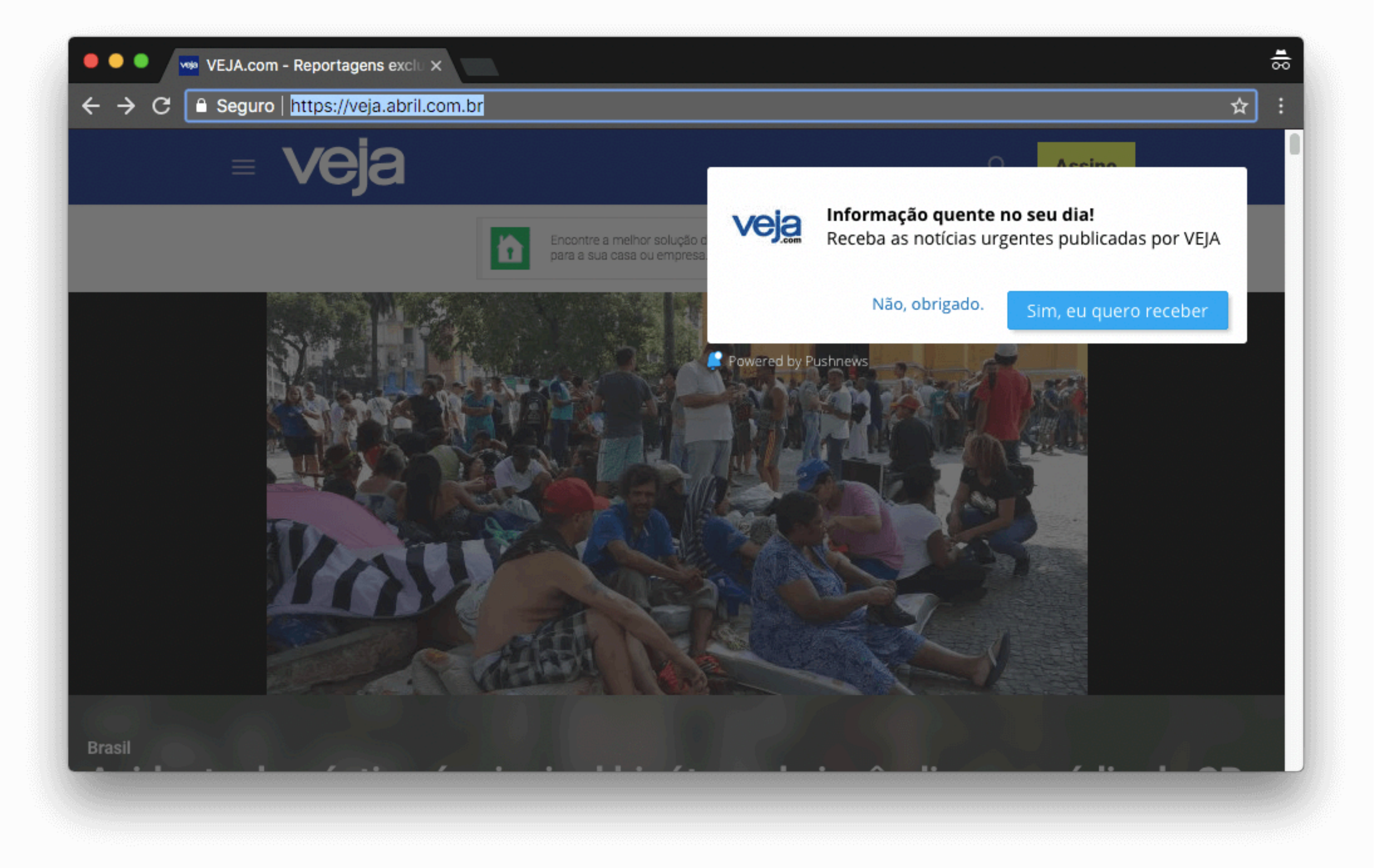
THE LONG TAIL (AGAIN)

PROBLEMS II: COLD-START

- User cold-start
  - Questionnaires
  - Session-based recommendation
    - RNNs, LSTMs
- Item cold-start
  - Content-based approaches

PROJECTS I: AD PERSONALISATION

- Problem
  - Short message dissemination
  - Increase ENGAGEMENT
    - Right MESSAGE
    - Right MOMENT
    - Right USER
    - Right CHANNEL
- Approach
    - Content-based
    - NLP
    - Big Data (Kafka + NoSQL)



- Impact
  - Massive reduction of no. of messages (for same return)
  - Increased subscriber loyalty (lower opt-out rates)

PROJECTS II: PALCO 3.0

- Problem
  - Online music streaming and playlists
  - Complete music recommendation solution
  - Global and user-based blacklisting
  - Recommendation as a Service (RaaS)
- Approach
  - Collaborative filtering (neighborhood-based)
  - Content-based filtering (audio analysis)
  - Hybrid
  - Server-client approach
- Impact
  - Increased traffic
  - Automatic online radios

FUTURE

CONTENT-BASED

USAGE-BASED

HYBRID

FUTURE

CONTE

ASED

HYBRID

+

CONTEXT

+

SIDE INFORMATION

+

DYNAMIC

FUTURE

- Interactive RS
    - User-customizable
- Meta-learning
    - Self-parameter tuning
    - Algorithm selection
- Better offline evaluation proxies

FUTURE

- Privacy and security
  - Cryptographic protocols
  - Data ownership
  - Profile storage and portability
- Ethics and trustworthiness
  - Fair, explainable algorithms
  - International and national regulation compliance

THANKS