

# Recommender Systems

Marie Al Ghossein

June 10, 2024

 Crossing Minds

 Dauphine | PSL   
UNIVERSITÉ PARIS

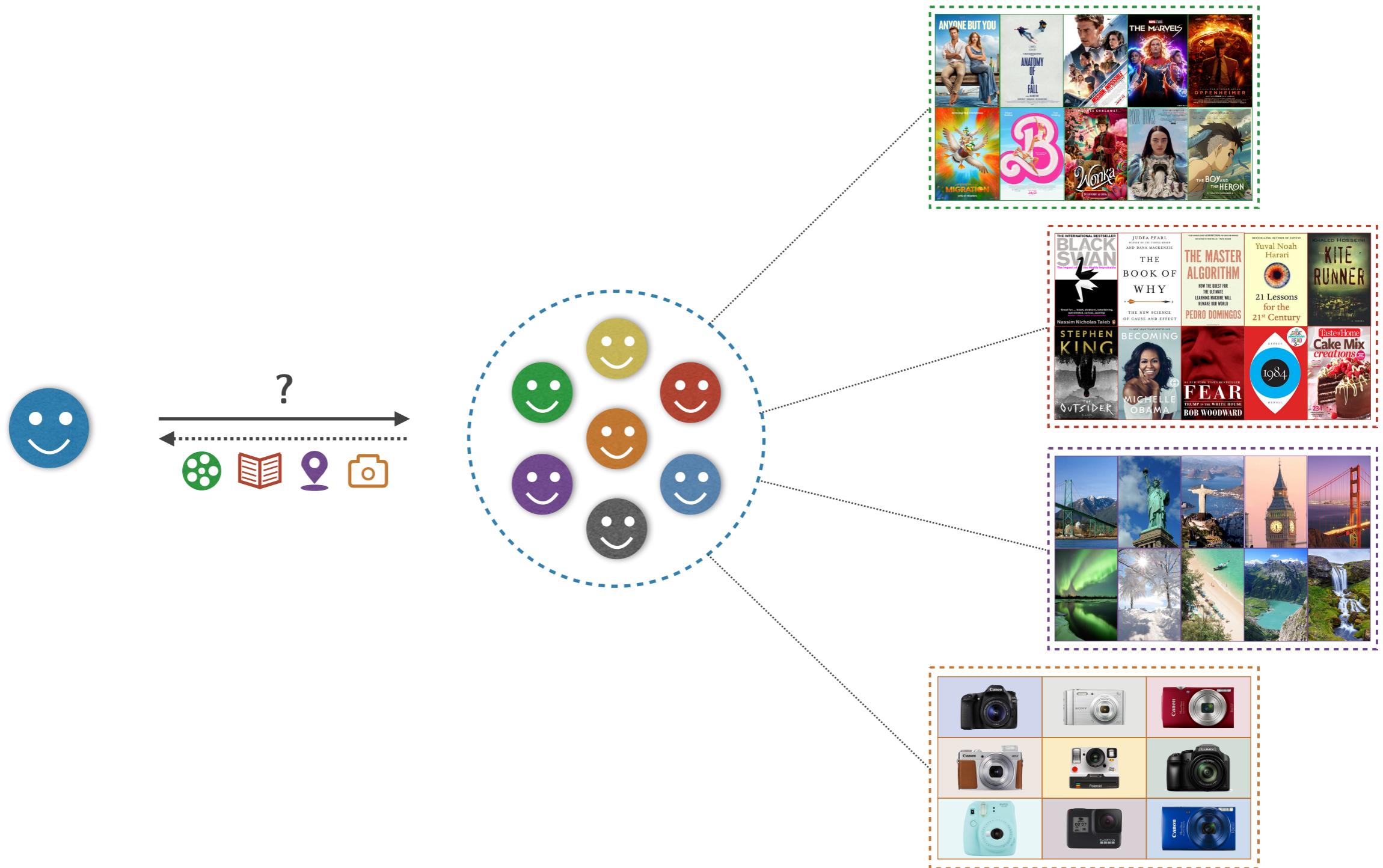
# Recommender systems

## Outline

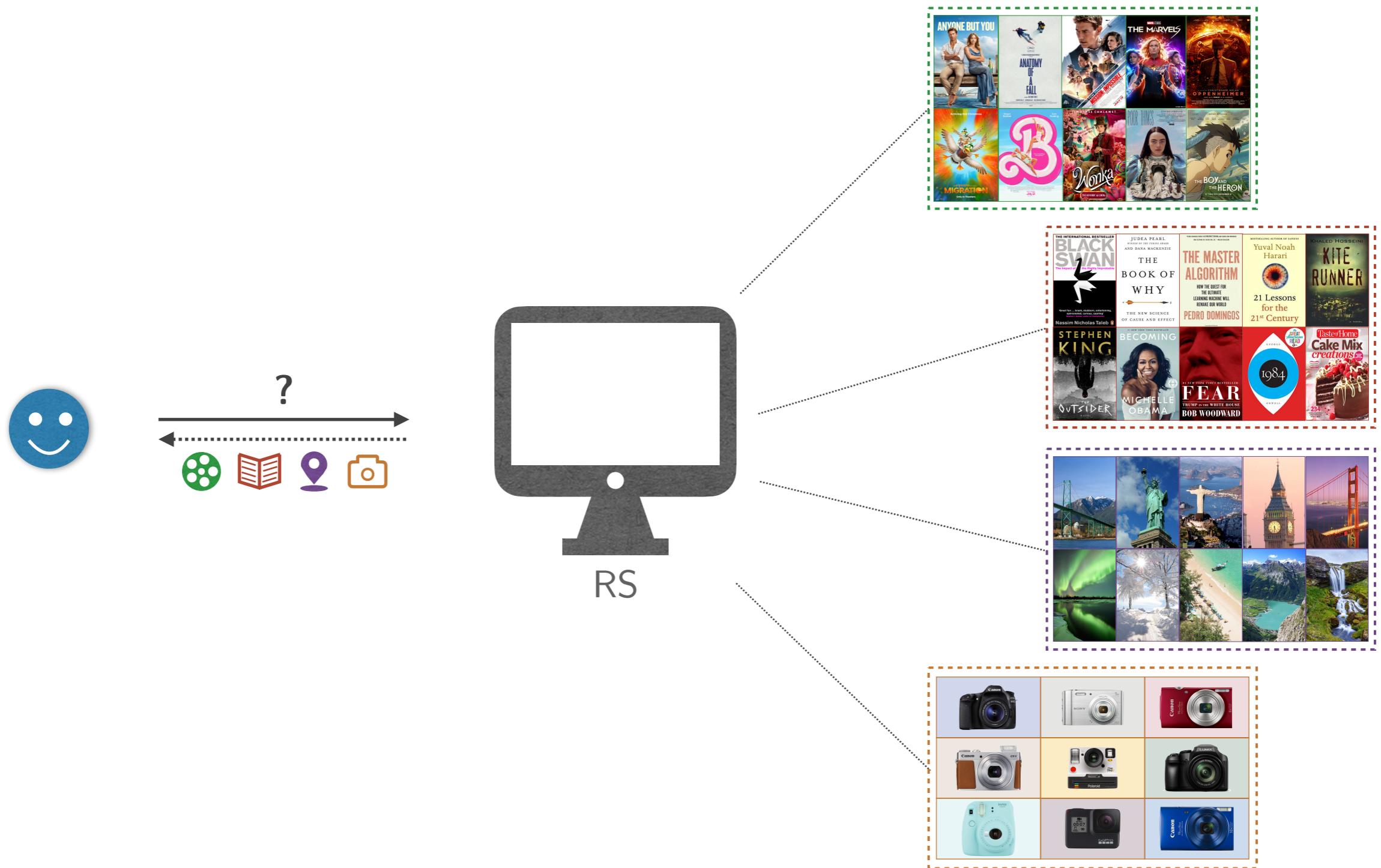
1. Introduction
2. The Recommendation Problem
3. User Feedback
4. Challenges
5. Evaluation of RS
6. Collaborative Filtering
7. Content-Based Filtering
8. Hybrid Approaches
9. Context-Aware Recommendation
10. Reinforcement Learning for Recommender Systems
11. Deep Learning for Recommender Systems
12. Short Intro to Large Language Models for Recommender Systems
13. A Practical Example: Recommendation in Location-Based Social Networks

# Introduction

# The (black) art of recommendation



# The (black) art of recommendation



# The value of recommender systems

## For consumers

- Find good and relevant products
- Improve decision making
- Help them overcome the *information overload*

# The value of recommender systems

## For consumers

- Find good and relevant products
- Improve decision making
- Help them overcome the *information overload*

## For service providers

- Increase the number of products sold
- Sell more diverse products
- Increase the user satisfaction and fidelity
- Learn more about consumers

# Personalization at Spotify

The screenshot shows the Spotify Home page with a dark theme. On the left, there's a sidebar with navigation icons for Home, Browse, and Radio, and sections for Your Library (Made For You, Recently Played, Liked Songs, Albums, Artists, Podcasts) and Playlists (On Repeat, Your Top Songs 2018, Your Top Songs 2017, Discover Weekly). The main content area features a "Your Decade Wrapped" section with cards for 2019, 2018, 2017, 2016, and "le meilleur de la décennie". Below this is a "Made for you" section with cards for Daily Mix 5, Daily Mix 6, Discover Weekly, Release Radar, Your Summer Rewind, and Your Top Songs 2018. At the bottom is a "À ne pas manquer aujourd'hui!" section with cards for PVNCHLNRS, La vie est belle, Hits du Moment, Fresh Variété, Maximum, and Pop Urbaine.

**Your Decade Wrapped**

- Your Top Songs 2019**  
Your Top Songs 2019  
The songs you loved most this year, all wrapped up.
- Your Top Songs 2018**  
Your Top Songs 2018  
The songs you loved most this year, all wrapped up.  
1 FOLLOWER
- Your Top Songs 2017**  
Your Top Songs 2017  
The songs you loved most this year, all wrapped up.  
PLAYLIST • BY SPOTIFY
- Your Top Songs 2016**  
Your Top Songs 2016  
We collected all the songs you loved the most this year, wrapped them up, and are...
- le meilleur de la décennie**  
le meilleur de la décennie  
Quelques-uns des meilleurs titres des dix dernières années réunis dans une...

**Made for you**

- Your Daily Mix 5**  
Daily Mix 5  
Jascha Heifetz, Daria van den Bercken, Hilary Hahn and more
- Your Daily Mix 6**  
Daily Mix 6  
Carter Burwell, Hans Zimmer, Angelo Badalamenti and more
- Your Discover Weekly**  
Your weekly mixtape of fresh music. Enjoy new discoveries and deep cuts...  
PLAYLIST • BY SPOTIFY
- Your Release Radar**  
Release Radar  
Never miss a new release! Catch all the latest music from artists you follow, plu...
- Your Summer Rewind**  
Your Summer Rewind  
Time for Your Summer Rewind! We've made you a new playlist featuring your...
- Your Top Songs 2018**  
Your Top Songs 2018  
The songs you loved most this year, all wrapped up.  
1 FOLLOWER

**À ne pas manquer aujourd'hui !**

- PVNCHLNRS**  
PVNCHLNRS  
La première playlist de rap français entre en scène : réserve ta place dès...  
935,857 FOLLOWERS
- La vie est belle**  
La vie est belle  
Le meilleur de la musique d'hier et d'aujourd'hui pour une journée parfaite.  
490,050 FOLLOWERS
- Hits du Moment**  
Hits du Moment  
Angèle au sommet de la première playlist de France.  
1,545,775 FOLLOWERS
- Fresh Variété**  
Fresh Variété  
Les meilleures nouveautés de la pop française. Photo : Gims  
125,168 FOLLOWERS
- Maximum**  
Maximum  
Le mix parfait ! Photo : David Guetta x MORTEN  
292,679 FOLLOWERS
- Pop Urbaine**  
Pop Urbaine  
Tous les hits pop urbaine et afropop. Photo : Jul  
478,409 FOLLOWERS

# The value of recommender systems



“Our recommender system influences choice for about 80% of hours streamed at Netflix. [...] We think the combined effect of personalization and recommendations save us more than \$1B per year.” (2015)

C. A. Gomez-Uribe and N. Hunt. *The Netflix Recommender System: Algorithms, Business Value, and Innovation*. ACM Trans. Management Inf. Syst. 6(4), 2015.

# The value of recommender systems



“Our recommender system influences choice for about 80% of hours streamed at Netflix. [...] We think the combined effect of personalization and recommendations save us more than \$1B per year.” (2015)

C. A. Gomez-Uribe and N. Hunt. *The Netflix Recommender System: Algorithms, Business Value, and Innovation*. ACM Trans. Management Inf. Syst. 6(4), 2015.



“35% of what consumers purchase on Amazon comes from product recommendations.” (2013)

<https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers>

# The value of recommender systems



“Our recommender system influences choice for about 80% of hours streamed at Netflix. [...] We think the combined effect of personalization and recommendations save us more than \$1B per year.” (2015)

C. A. Gomez-Uribe and N. Hunt. *The Netflix Recommender System: Algorithms, Business Value, and Innovation*. ACM Trans. Management Inf. Syst. 6(4), 2015.



“35% of what consumers purchase on Amazon comes from product recommendations.” (2013)

<https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers>



“Recommendations account for about 60% of all video clicks from the homepage.” (2010)

J. Davidson et al. *The YouTube Video Recommendation System*. In RecSys'10, 2010.

# The value of recommender systems



“Our recommender system influences choice for about 80% of hours streamed at Netflix. [...] We think the combined effect of personalization and recommendations save us more than \$1B per year.” (2015)

C. A. Gomez-Uribe and N. Hunt. *The Netflix Recommender System: Algorithms, Business Value, and Innovation*. ACM Trans. Management Inf. Syst. 6(4), 2015.



“35% of what consumers purchase on Amazon comes from product recommendations.” (2013)

<https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers>



“Recommendations account for about 60% of all video clicks from the homepage.” (2010)

J. Davidson et al. *The YouTube Video Recommendation System*. In RecSys'10, 2010.



“Recommendations generate 38% more clicks compared to popularity-based recommendations.” (2009)

A. S. Das et al. *Google News Personalization: Scalable Online Collaborative Filtering*. In WWW '07, 2007.

# Measuring the business value of recommender systems

- Click-Through Rate (CTR)

# Measuring the business value of recommender systems

- Click-Through Rate (CTR)
- Conversion rates

# Measuring the business value of recommender systems

- Click-Through Rate (CTR)
- Conversion rates
- Sales and revenue

# Measuring the business value of recommender systems

- Click-Through Rate (CTR)
- Conversion rates
- Sales and revenue
- Effects on sales distributions

# Measuring the business value of recommender systems

- Click-Through Rate (CTR)
- Conversion rates
- Sales and revenue
- Effects on sales distributions
- User engagement and behavior

# What is a recommender system?

- A Recommender System (RS) provides personalized suggestions of **items** for **users**, the items being drawn from a large catalog, based on knowledge of the user, the items, and **interactions** between users and items.

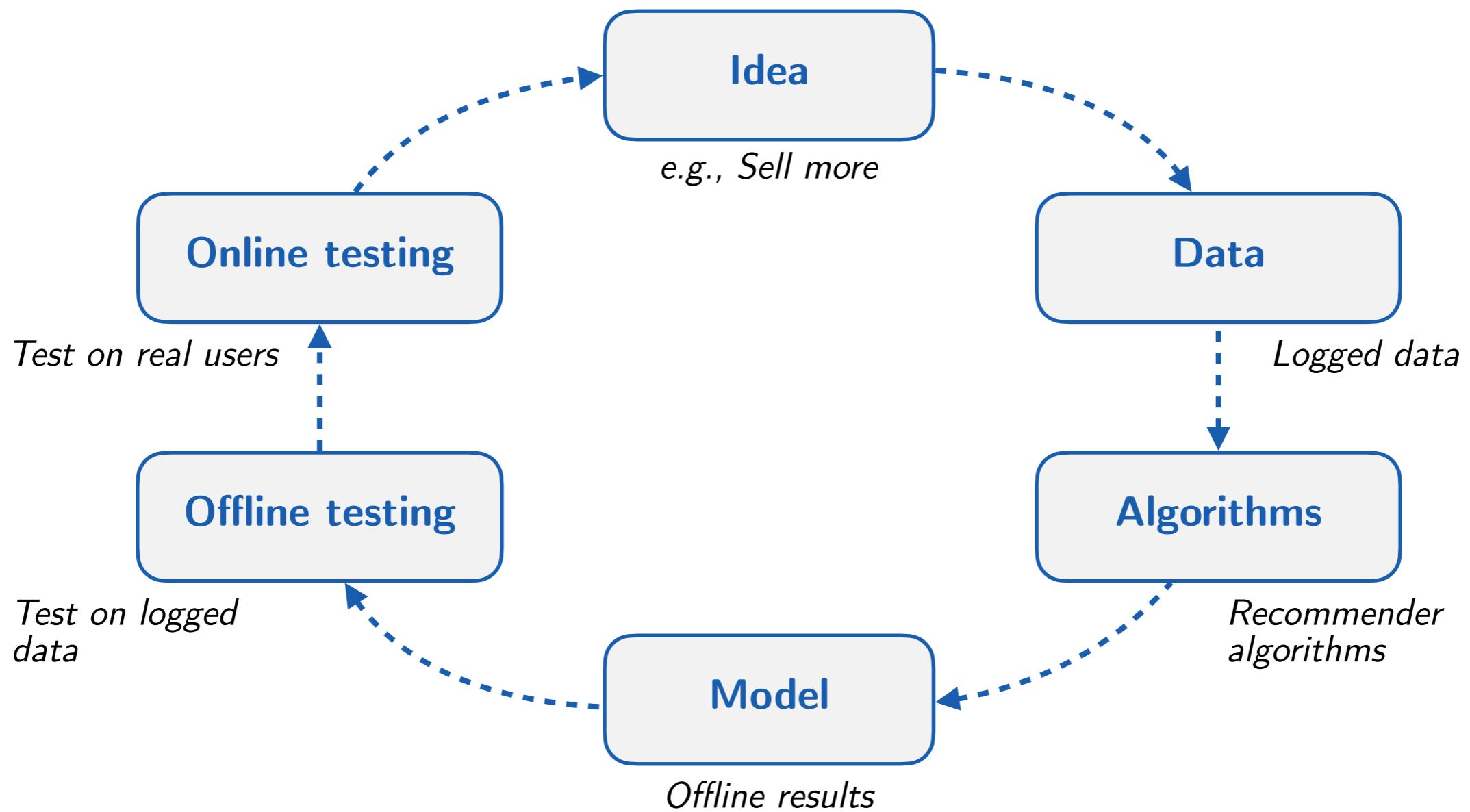
# What is a recommender system?

- Knowledge of the user
  - ▶ Historical data - interactions with items
    - Purchases, clicks, views, ratings,...
  - ▶ Demographics
    - Age, gender, nationality, political views, income,...
  - ▶ Context
    - Device, location, time of day/week/year, alone or in a group,...

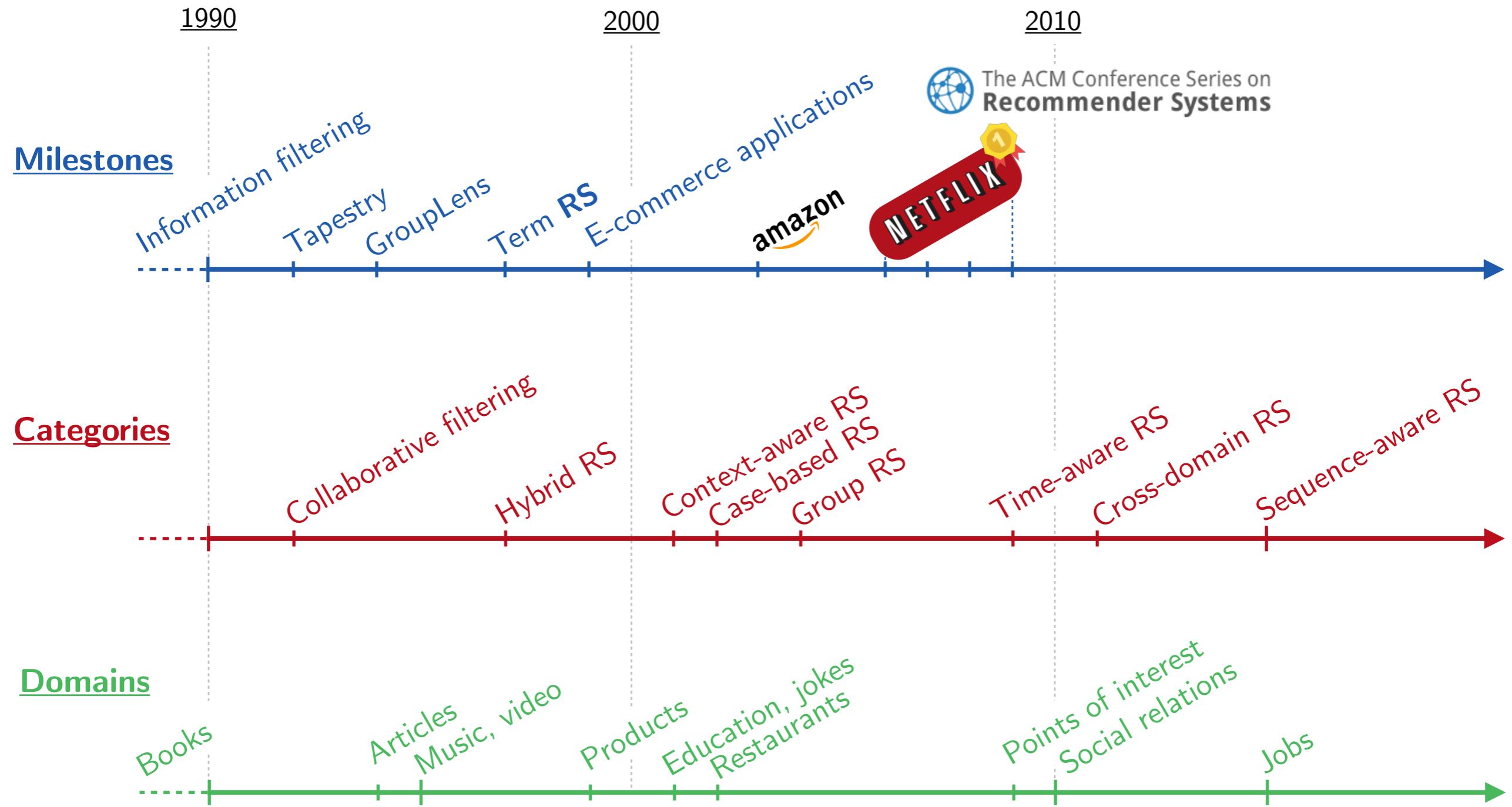
# What is a recommender system?

- Knowledge of the **user**
  - ▶ Historical data - interactions with items
    - Purchases, clicks, views, ratings,...
  - ▶ Demographics
    - Age, gender, nationality, political views, income,...
  - ▶ Context
    - Device, location, time of day/week/year, alone or in a group,...
- Knowledge of the **item**
  - ▶ Content information
    - Features, tags, text, images, multimedia signals,...

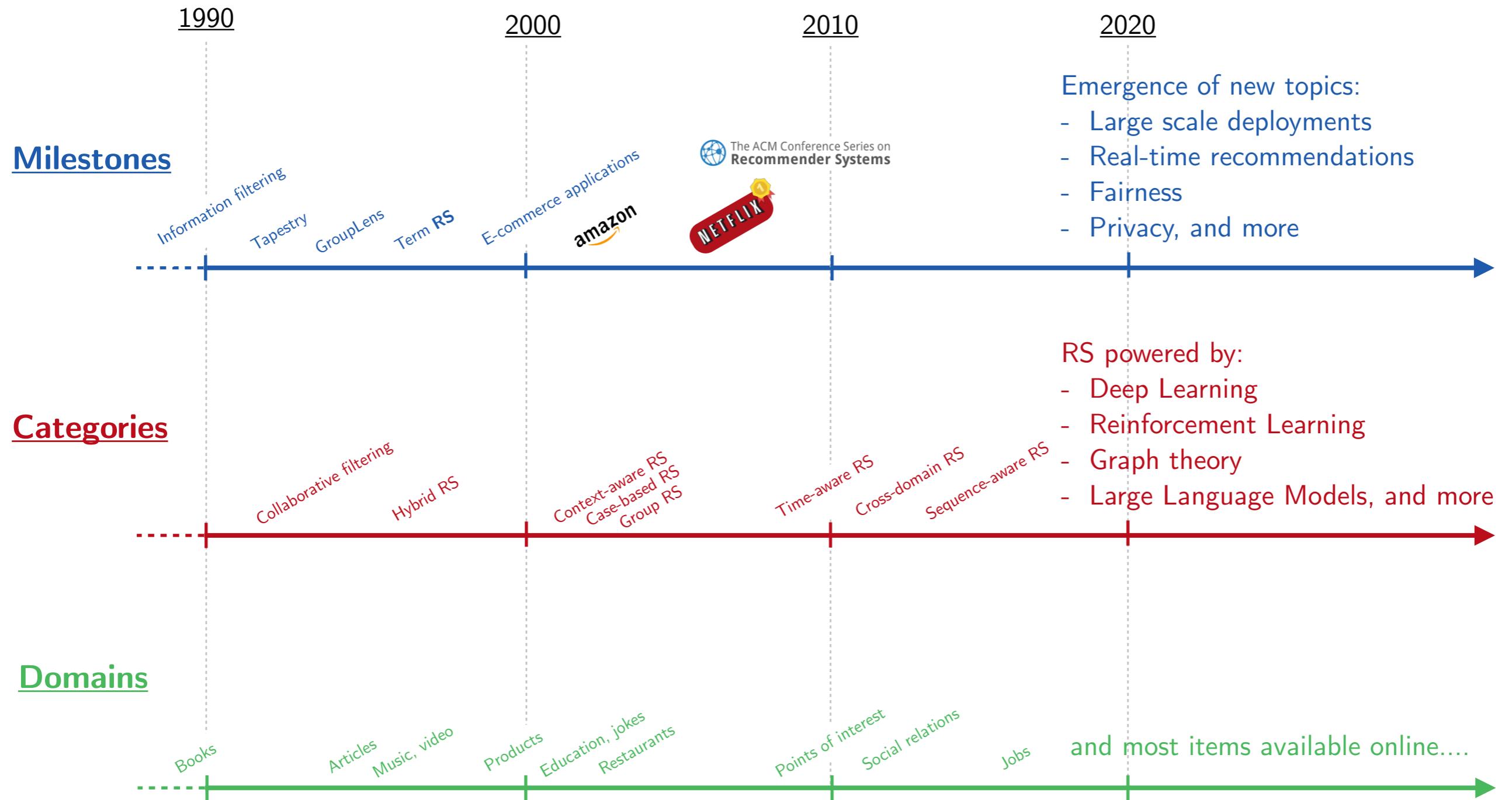
# Building a recommender system



# Historical perspective on recommender systems



# Historical perspective on recommender systems



# The recommendation problem

# The Netflix Prize (2006)

- **End-goal:** Recommend movies to Netflix users
- Competition problem:
  - ▶ Predict movie ratings on a 5-star scale
  - ▶ Dataset of 100M ratings from 480k users on 18k movies
  - ▶ \$1M prize for 10% improvement over accuracy of existing solution



★★★★★	New favorite!!!
★★★★☆	Enjoyed it!
★★★★☆	Mixed feelings.
★★☆☆☆	Did not like it.
★★☆☆☆	Hated it!

Ratings

James Bennett and Stan Lanning. *The Netflix Prize*. In Proc. KDD Cup and Workshop, 2007.

# The Netflix Prize (2006)

- **End-goal:** Recommend movies to Netflix users
  - Competition problem:
    - ▶ Predict movie ratings on a 5-star scale
    - ▶ Dataset of 100M ratings from 480k users on 18k movies
    - ▶ \$1M prize for 10% improvement over accuracy of existing solution
- Recommendation problem formulated as **rating prediction**



★★★★★	New favorite!!!
★★★★☆	Enjoyed it!
★★★★☆	Mixed feelings.
★★☆☆☆	Did not like it.
★★☆☆☆	Hated it!

Ratings

# The recommendation problem

## Rating prediction

- The task of the RS is to predict the rating that a user will give to an item he did not rate.
  - Objective: Minimize the error between predicted and observed ratings
  - Recommendation: Items having the highest ratings for a user

# The recommendation problem

## Rating prediction

- The task of the RS is to predict the rating that a user will give to an item he did not rate.
  - Objective: Minimize the error between predicted and observed ratings
  - Recommendation: Items having the highest ratings for a user

Limitation: In real-world RS, it is more valuable to provide relevant top recommendations than accurately predicting ratings for all items, including those that are at the bottom list.

# The recommendation problem

## Rating prediction

- The task of the RS is to predict the rating that a user will give to an item he did not rate.
  - Objective: Minimize the error between predicted and observed ratings
  - Recommendation: Items having the highest ratings for a user

Limitation: In real-world RS, it is more valuable to provide relevant top recommendations than accurately predicting ratings for all items, including those that are at the bottom list.

## Top-N recommendation

- The task of the RS is to predict the relevance of an item for a user, i.e., predict whether the user will choose an item or not.

# General formulation

- Let  $\mathcal{U}$  be the set of users and  $\mathcal{I}$  the set of items.
- Let  $f$  be a utility function that measures the usefulness, e.g., rating or relevance, of an item for a user:  $f : \mathcal{U} \times \mathcal{I} \rightarrow \mathcal{R}$ , where  $\mathcal{R}$  is a totally ordered set.
- For each user  $u \in \mathcal{U}$ , we recommend the items maximizing  $f$ :

$$\forall u \in \mathcal{U}, i^* = \arg \max_{i \in \mathcal{I}} f(u, i)$$

$f$  can represent the rating or the relevance of an item

# A recommendation algorithm

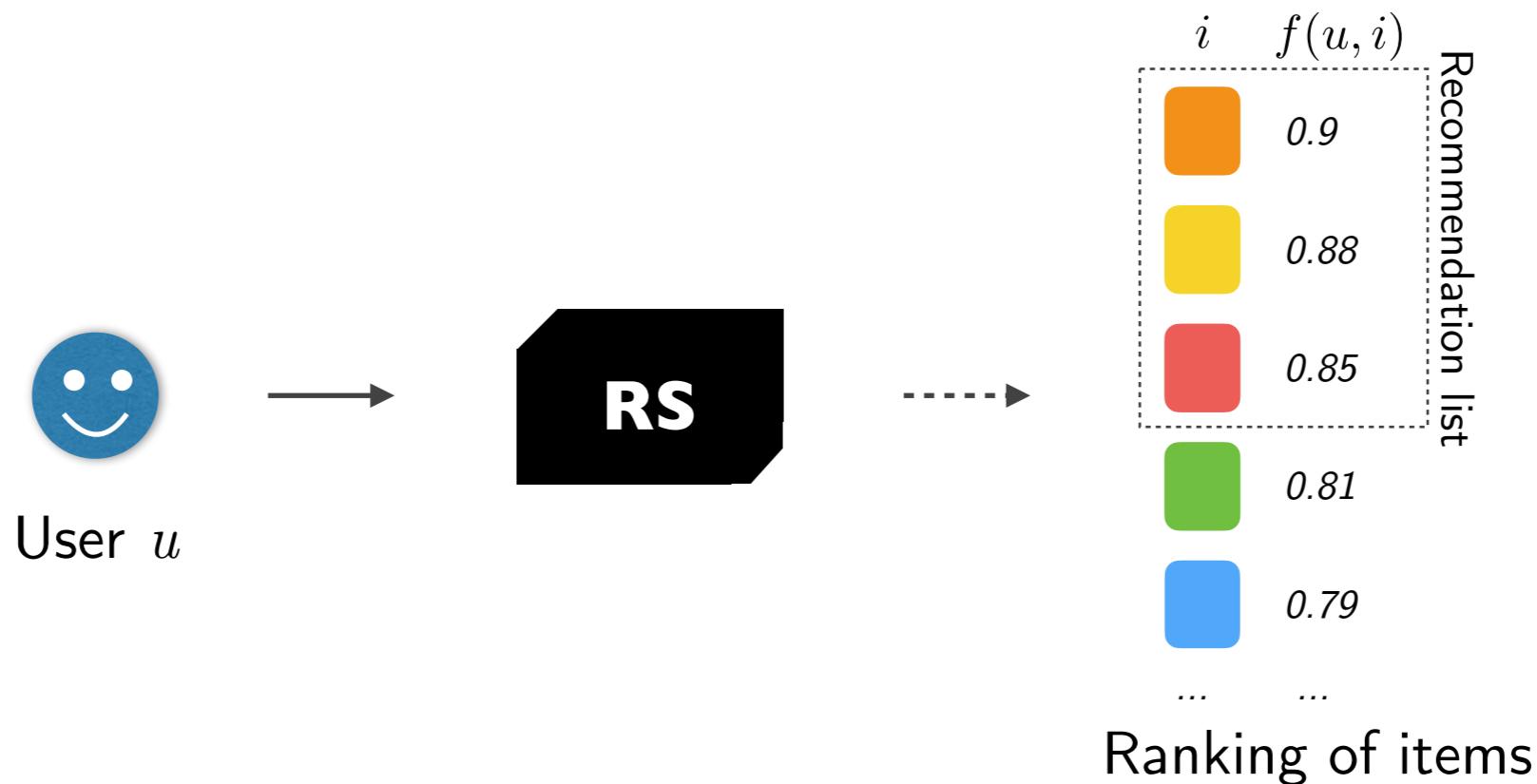
---

**Algorithm 1** A recommendation algorithm

---

**Input:** user  $u$ , number of items to recommend  $N$

- 1 **for**  $i$  in  $\mathcal{I} \setminus \mathcal{I}_u$  **do**
  - 2     Predict  $\hat{r}_{ui}$
  - 3 **end for**
  - 4 Create list of  $\mathcal{I} \setminus \mathcal{I}_u$  items ordered by decreasing order of  $\hat{r}_{ui}$
  - 5 Return  $N$  first items of the list
- 

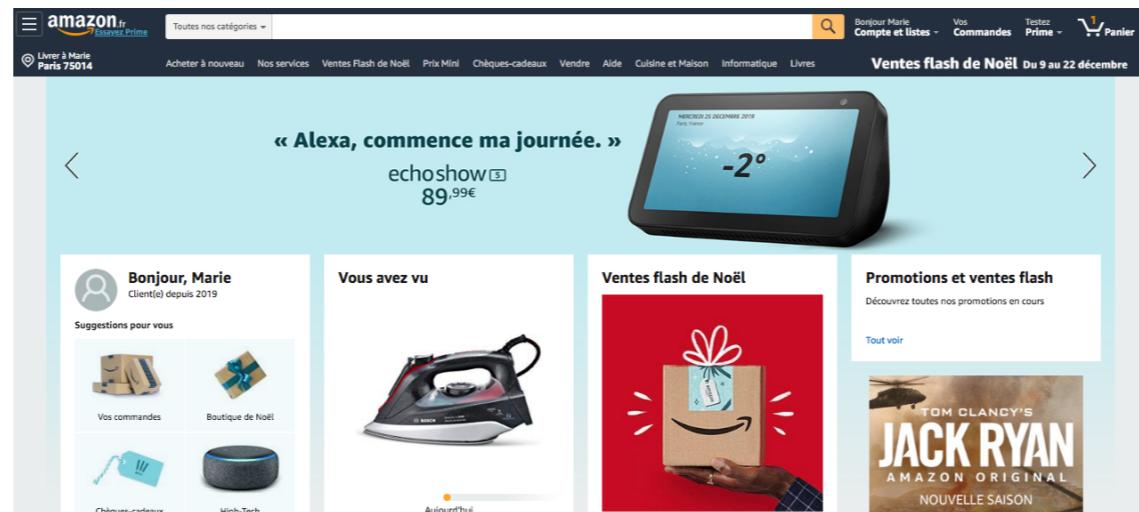


# User feedback

# What is a recommender system?

- Knowledge of the **user**
  - ▶ Historical data - interactions with items
    - Purchases, clicks, views, ratings,...
  - ▶ Demographics
    - Age, gender, nationality, political views, income,...
  - ▶ Context
    - Device, location, time of day/week/year, alone or in a group,...
- Knowledge of the **item**
  - ▶ Content information
    - Features, tags, text, images, multimedia signals,...

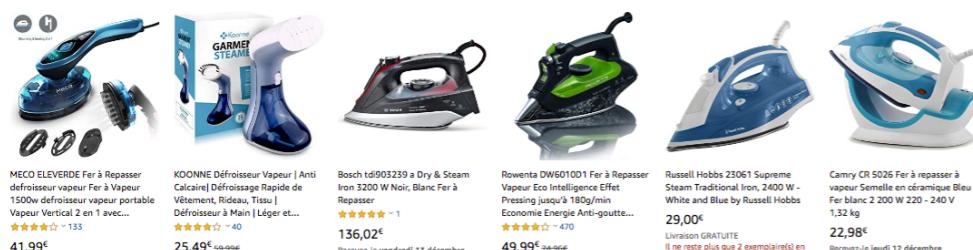
# Browsing on Amazon...



Click on card #2



Search with keywords “w<sub>1</sub> w<sub>2</sub>”



Impression with items #1, 2, 3, 4, 5, 6

## Informations sur le produit

Marque	Bosch
Numéro de modèle	TDI903239A
Couleur	Noir, Blanc
Poids de l'article	2,12 Kg
Dimensions du colis	33,8 x 20,2 x 12,4 cm
Capacité	0,4 litres
Puissance	3200 Watts
Matériau	Aluminium

Read description of item #3



Add item #3 to cart



Click on item #3



★★★★★ Parfait

29 novembre 2017

Achat vérifié

Produit très très qualitatif !

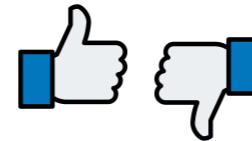
Rate and review item #3

# Explicit feedback

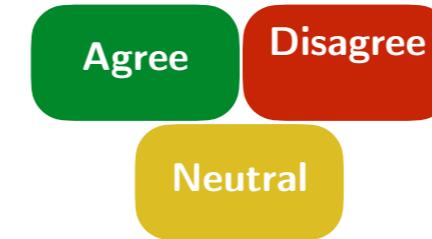
- Explicit feedback is a form of feedback directly reported by the user to the system.



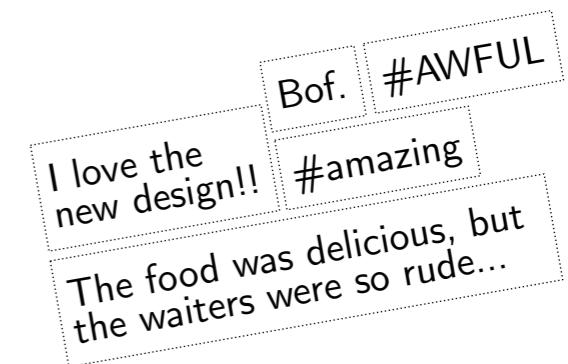
Ratings, 5-star scale



Like/Dislike



Ordinal scale



Textual reviews

# Explicit feedback

- Explicit feedback is a form of feedback directly reported by the user to the system.

# Explicit feedback

- Explicit feedback is a form of feedback directly reported by the user to the system.
- Limitations:
  - ▶ Users do **not always provide** explicit feedback.

# Explicit feedback

- Explicit feedback is a form of feedback directly reported by the user to the system.
- Limitations:
  - Users do **not always provide** explicit feedback.
  - The system may be **unable to collect** explicit feedback, e.g., cannot collect rating of people for LinkedIn's "*People You May Know*".

# Explicit feedback

- Explicit feedback is a form of feedback directly reported by the user to the system.
- Limitations:
  - ▶ Users do **not always provide** explicit feedback.
  - ▶ The system may be **unable to collect** explicit feedback, e.g., cannot collect rating of people for LinkedIn's "*People You May Know*".
  - ▶ Is the rating revealing the **true opinion** of the user?

# Explicit feedback

- Explicit feedback is a form of feedback directly reported by the user to the system.
- Limitations:
  - Users do **not always provide** explicit feedback.
  - The system may be **unable to collect** explicit feedback, e.g., cannot collect rating of people for LinkedIn's "*People You May Know*".
  - Is the rating revealing the **true opinion** of the user?
  - Is the user being **influenced** by his social circle when rating an item?

# Explicit feedback

- Explicit feedback is a form of feedback directly reported by the user to the system.
- Limitations:
  - Users do **not always provide** explicit feedback.
  - The system may be **unable to collect** explicit feedback, e.g., cannot collect rating of people for LinkedIn's "*People You May Know*".
  - Is the rating revealing the **true opinion** of the user?
  - Is the user being **influenced** by his social circle when rating an item?
  - Is the user rating the **complete package** or is there an item's **detail** affecting the overall rating?

# Explicit feedback

- Explicit feedback is a form of feedback directly reported by the user to the system.
- Limitations:
  - Users do **not always provide** explicit feedback.
  - The system may be **unable to collect** explicit feedback, e.g., cannot collect rating of people for LinkedIn's "*People You May Know*".
  - Is the rating revealing the **true opinion** of the user?
  - Is the user being **influenced** by his social circle when rating an item?
  - Is the user rating the **complete package** or is there an item's **detail** affecting the overall rating?
  - Explicit feedback exhibit some **biases**, e.g., some users tend to give higher ratings than others for a similar level of appreciation.

# Implicit feedback

- Implicit feedback is collected by the system without the intervention of the user.



Bosch tdi903239 a Dry & Steam Iron 3200 W Noir, Blanc Fer à Repasser de Bosch  
 ★★★★ 1 évaluation  
 Prix : 136,02 € Livraison GRATUITE en France métropolitaine. Détails  
 Tous les prix incluent la TVA.  
 Payez : 34,00 € x 4 (+0,00 € de frais inclus) Voir conditions et plus de facilités de paiement  
 Message promotionnel Economisez € 17,99 sur Echo Dot lorsque vous achetez ... 1 promotion +  
 Assistance produit Amazon gratuite incluse -  
 Livraison GRATUITE (0,01€ pour les livres) en point retrait. Détails  
 Neufs & occasions (5) \$2,04 € + Livraison GRATUITE  
 • 2 unités de cet article soldée(s) à partir du 10 janvier 2018 8h (uniquement sur les unités vendues et expédiées par Amazon)  
 Comparer avec des articles similaires  
 Signaler des informations incorrectes sur les produits  
 Nos prix incluent l'éco-participation sur tous les produits concernés. Vous voulez recycler votre appareil électrique ou électronique gratuitement ? En savoir plus ici.

**Clicks**

## Informations sur le produit

Marque	Bosch
Numéro de modèle	TDI903239A
Couleur	Noir, Blanc
Poids de l'article	2,12 Kg
Dimensions du colis	33,8 x 20,2 x 12,4 cm
Capacité	0,4 litres
Puissance	3200 Watts
Matériau	Aluminium

**Views of item's description**

Acheter cet article

**Purchases**

Ajouter à votre liste

**Add to Wishlist**

# Implicit feedback

- Implicit feedback is collected by the system without the intervention of the user.



Bosch tdi903239 a Dry & Steam Iron 3200 W Noir, Blanc Fer à Repasser de Bosch  
★★★☆☆ 1 évaluation

Prix : 136,02 € Livraison GRATUITE en France métropolitaine. Détails  
 Tous les prix incluent la TVA.  
 Payez : 34,00 € x 4 (+0,00 € de frais inclus) Voir conditions et plus de facilités de paiement

**Message promotionnel** Economisez € 17,99 sur Echo Dot lorsque vous achetez ... 1 promotion +  
 Assistance produit Amazon gratuite incluse ~  
 Livraison GRATUITE (0,01€ pour les livres) en point retrait. Détails

Neufs & occasions (5) \$2,04 € + Livraison GRATUITE  
 • 2 unités de cet article soldée(s) à partir du 10 janvier 2018 8h (uniquement sur les unités vendues et expédiées par Amazon)

Comparer avec des articles similaires  
 Signaler des informations incorrectes sur les produits

Nos prix incluent l'éco-participation sur tous les produits concernés. Vous voulez recycler votre appareil électrique ou électronique gratuitement ? En savoir plus ici.

Clicks

## Informations sur le produit

Marque	Bosch
Numéro de modèle	TDI903239A
Couleur	Noir, Blanc
Poids de l'article	2,12 Kg
Dimensions du colis	33,8 x 20,2 x 12,4 cm
Capacité	0,4 litres
Puissance	3200 Watts
Matériau	Aluminium

Acheter cet article

Purchases

Ajouter à votre liste

Add to Wishlist

→ More abundant than explicit feedback

# Implicit feedback

- Implicit feedback is collected by the system without the intervention of the user.

# Implicit feedback

- Implicit feedback is collected by the system without the intervention of the user.
- Challenges:
  - Implicit feedback is **inherently noisy**.
    - Every *positive item*, i.e., item the user has interacted with, is not necessarily an item liked by the user.

# Implicit feedback

- Implicit feedback is collected by the system without the intervention of the user.
- Challenges:
  - Implicit feedback is **inherently noisy**.
    - Every *positive item*, i.e., item the user has interacted with, is not necessarily an item liked by the user.
  - There is **no negative feedback**.
    - Negative items are a mixture of unliked items and unknown items.

# Feedback matrix




# Feedback matrix - Explicit feedback



4

5

1



5

4

3



3

1

5



2

2

5

4



5

2

# Feedback matrix - Implicit feedback (Binary)



1		1	1		
			1	1	1
	1	1	1		
1	1			1	1
		1	1		



# Feedback matrix - Implicit feedback (Weights)



2

2.8

0.8



3.2

2.4

1



1.2

1

2



0.4

0.9

2.1

1.3



3

0

# Challenges

# Rating data sparsity

- Users only interact with a small number items selected from a very large catalog of available items.

# Rating data sparsity

- Users only interact with a small number items selected from a very large catalog of available items.

Examples of datasets:

 <b>1</b>
480k users
18k movies
100M ratings
Period of ~7 years

**Density: 1.1%**

James Bennett and Stan Lanning. *The Netflix Prize*.  
In Proc. KDD Cup and Workshop, 2007.


7.8M users
4.5k hotels
34.7M bookings
Period of ~4 years

**Density: 0.1%**

Marie Al-Ghossein et al. *Exploiting Contextual and External Data for Hotel Recommendation*. In Proc. UMAP, 2018.


20.98M users
9.35M items
82.83M ratings
Period of ~18 years

**Density: 0.00004%**

[https://cseweb.ucsd.edu/~jmcauley/datasets.html#amazon\\_reviews](https://cseweb.ucsd.edu/~jmcauley/datasets.html#amazon_reviews)

# Cold-start

- The *cold-start* problem designates the problems of:
  - Recommending items to **new users** where the historical data is missing,
    - Recommendations cannot be computed before the user has rated a certain number of items, but the user expects to receive suggestions before that.
  - Recommending **new items** to existing users.

# Cold-start

- The *cold-start* problem designates the problems of:
  - Recommending items to **new users** where the historical data is missing,
    - Recommendations cannot be computed before the user has rated a certain number of items, but the user expects to receive suggestions before that.
  - Recommending **new items** to existing users.
- A related problem is the *gray sheep* problem:
  - Recommending items to users who have very **unique tastes**, such that no other user (or very few users) has interacted with the same items.

# Overspecialization

- Recommendation algorithms tend to recommend items that are **too similar** to what the user has **already experienced**, resulting in overspecialized recommendations.

# Overspecialization

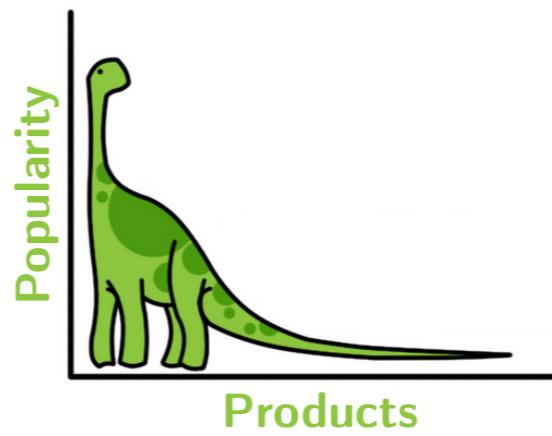
- Recommendation algorithms tend to recommend items that are **too similar** to what the user has **already experienced**, resulting in overspecialized recommendations.

Is this recommendation list useful?



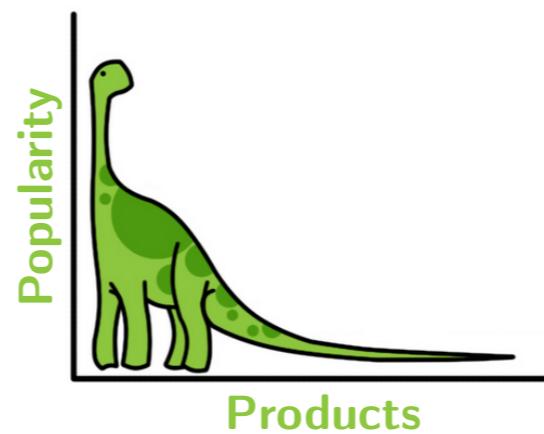
# Popularity bias

- Most item catalogs exhibit the *long tail* effect,



# Popularity bias

- Most item catalogs exhibit the *long tail* effect,



which raises two issues:

- Risk of recommending popular items to everyone,

Is this recommendation list useful?



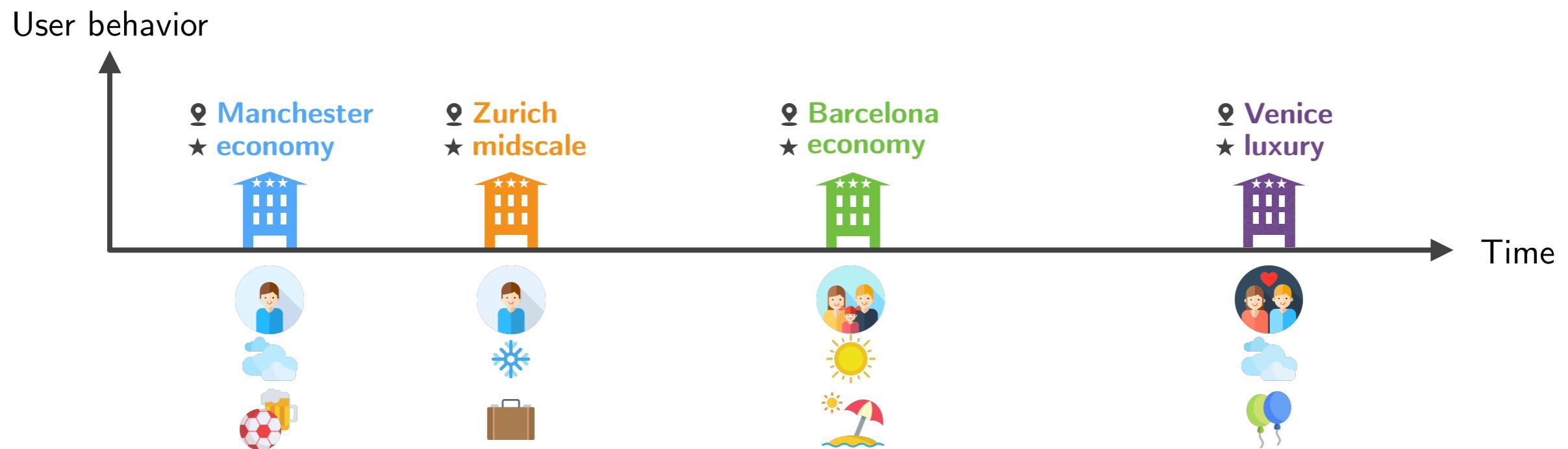
- Difficulty to promote long tail items with little feedback.

# Temporal dynamics

- Users preferences and item perceptions are changing over time
  - ▶ Reasons: emergence of new items, seasonality factors, etc.

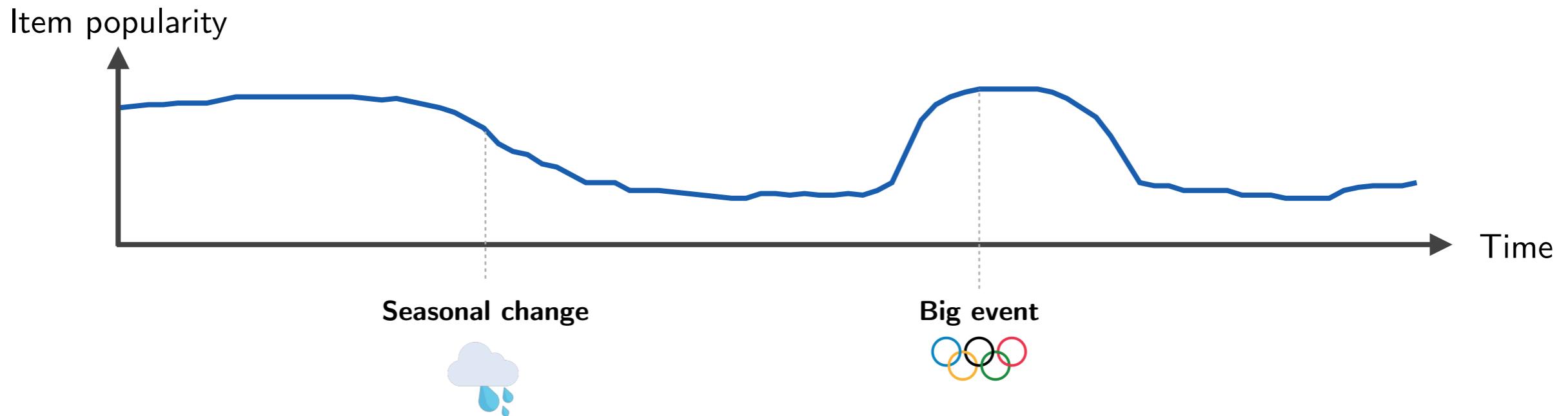
# Temporal dynamics

- Users preferences and item perceptions are changing over time
  - Reasons: emergence of new items, seasonality factors, etc.



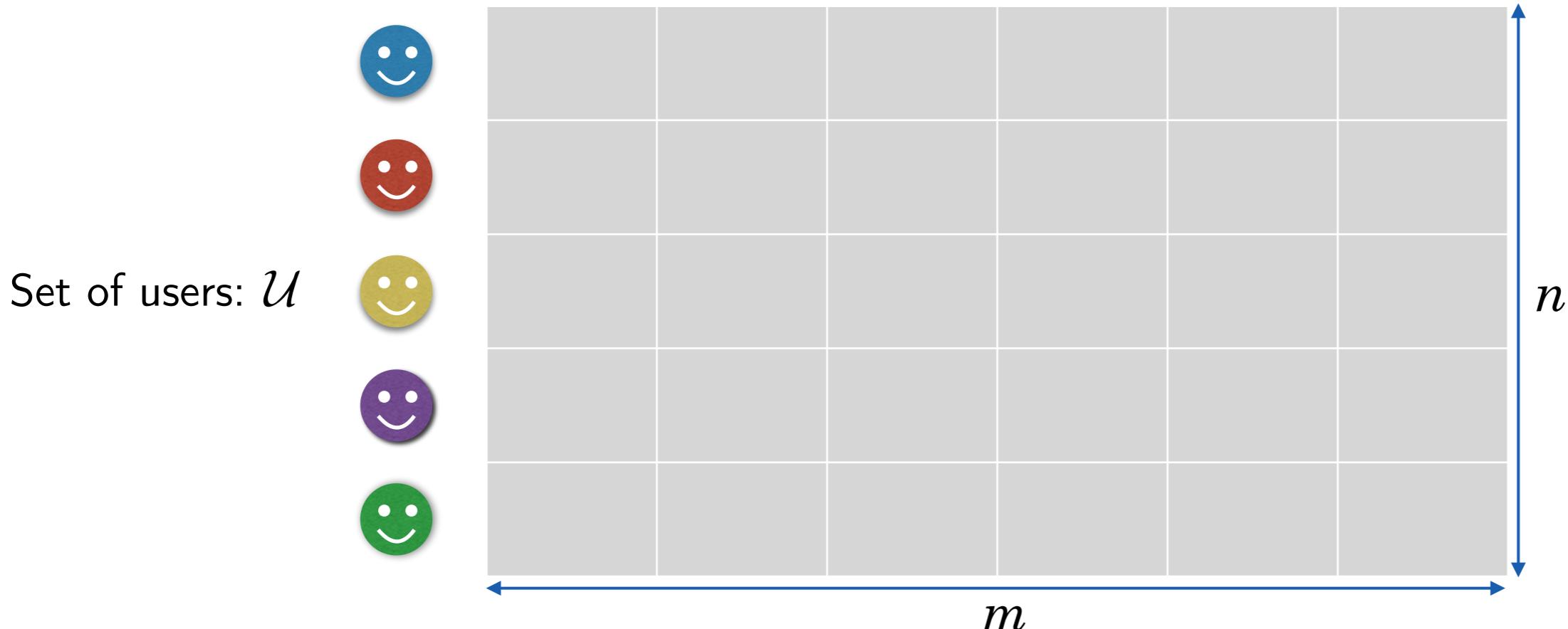
# Temporal dynamics

- Users preferences and item perceptions are changing over time
  - Reasons: emergence of new items, seasonality factors, etc.



# Feedback matrix - Notations

Set of items:  $\mathcal{I}$ , set of items rated by user  $u$ :  $\mathcal{I}_u$



Feedback matrix:  $R$

Rating:  $r_{ui}$ , predicted rating:  $\hat{r}_{ui}$

# Evaluation of RS

# Evaluation methods

- Offline evaluation
  - Performed on previously collected datasets and simulate the user behavior

# Evaluation methods

- Offline evaluation
  - Performed on previously collected datasets and simulate the user behavior
  - **Pros:** Does not require an interaction with real users, allows replicability,

# Evaluation methods

- Offline evaluation
  - Performed on previously collected datasets and simulate the user behavior
  - **Pros:** Does not require an interaction with real users, allows replicability,
  - **Cons:** Cannot measure the influence of the RS on the user behavior

# Evaluation methods

- Offline evaluation
  - Performed on previously collected datasets and simulate the user behavior
  - **Pros:** Does not require an interaction with real users, allows replicability,
  - **Cons:** Cannot measure the influence of the RS on the user behavior
- User study
  - Recruit a small number of users and ask them to interact with the RS

# Evaluation methods

- Offline evaluation
  - Performed on previously collected datasets and simulate the user behavior
  - **Pros:** Does not require an interaction with real users, allows replicability,
  - **Cons:** Cannot measure the influence of the RS on the user behavior
- User study
  - Recruit a small number of users and ask them to interact with the RS
  - **Pros:** Provides more insights than offline evaluation,

# Evaluation methods

- Offline evaluation
  - Performed on previously collected datasets and simulate the user behavior
  - Pros: Does not require an interaction with real users, allows replicability,
  - Cons: Cannot measure the influence of the RS on the user behavior
- User study
  - Recruit a small number of users and ask them to interact with the RS
  - Pros: Provides more insights than offline evaluation,
  - Cons: May introduce biases in the collected data

# Evaluation methods

- Offline evaluation
  - Performed on previously collected datasets and simulate the user behavior
  - Pros: Does not require an interaction with real users, allows replicability,
  - Cons: Cannot measure the influence of the RS on the user behavior
- User study
  - Recruit a small number of users and ask them to interact with the RS
  - Pros: Provides more insights than offline evaluation,
  - Cons: May introduce biases in the collected data
- Online evaluation
  - A/B testing on a real RS deployed online

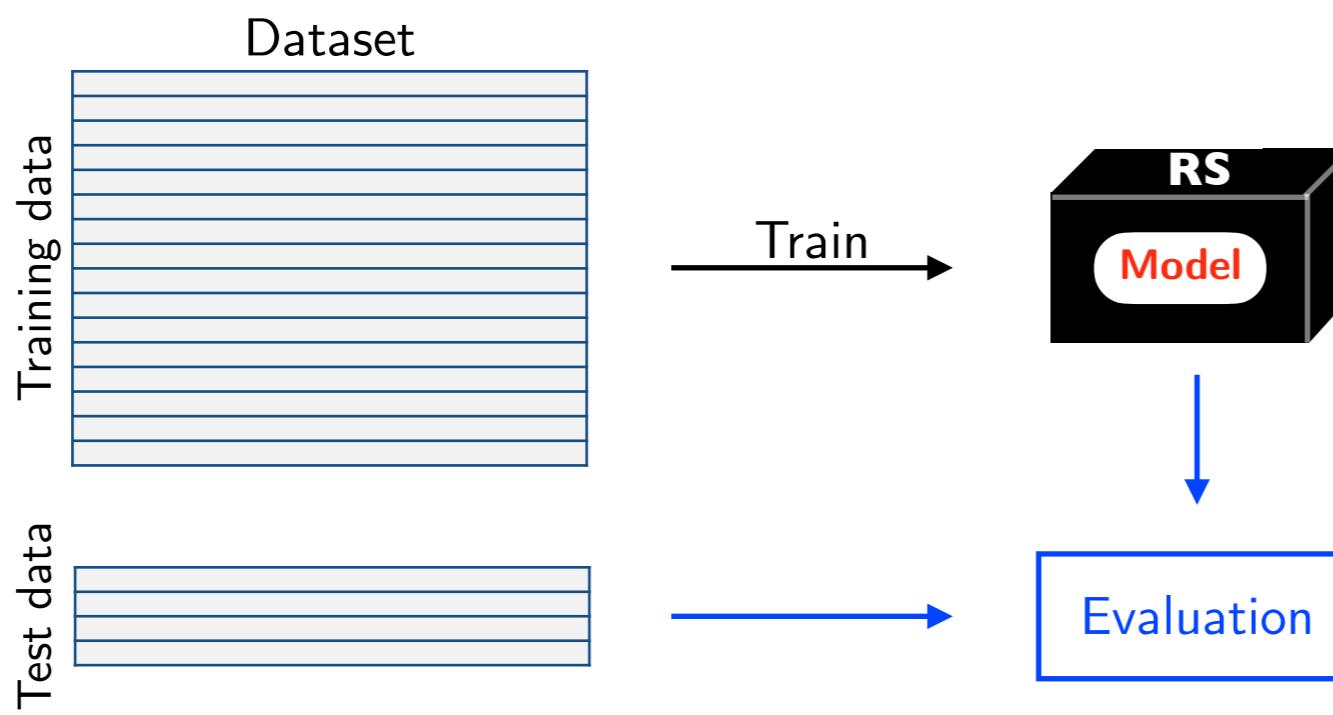
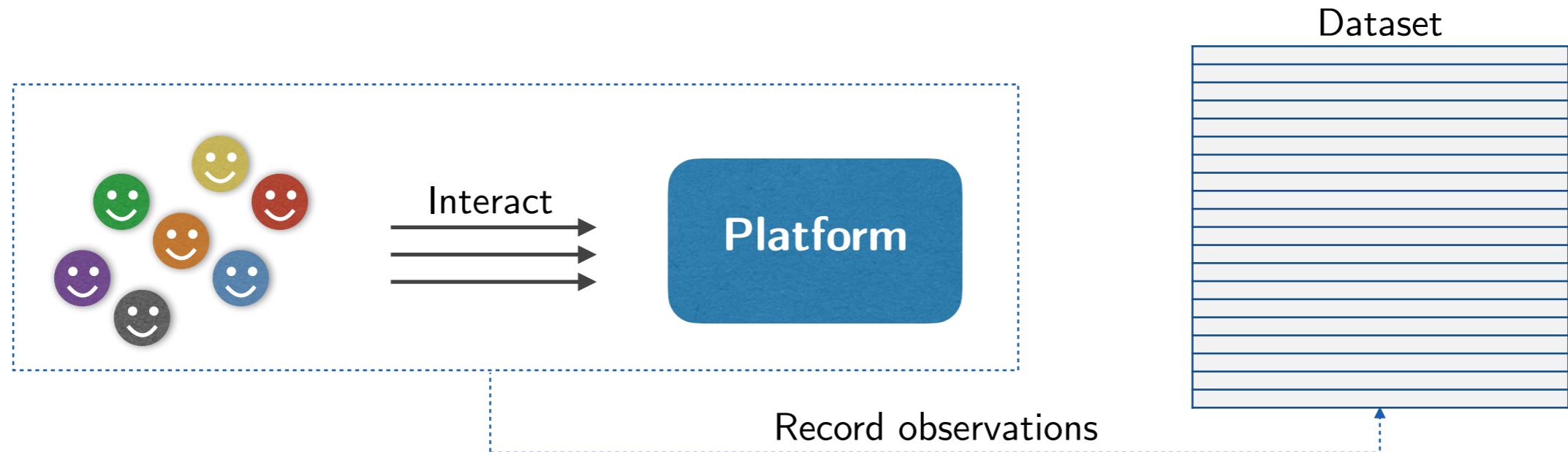
# Evaluation methods

- Offline evaluation
  - Performed on previously collected datasets and simulate the user behavior
  - Pros: Does not require an interaction with real users, allows replicability,
  - Cons: Cannot measure the influence of the RS on the user behavior
- User study
  - Recruit a small number of users and ask them to interact with the RS
  - Pros: Provides more insights than offline evaluation,
  - Cons: May introduce biases in the collected data
- Online evaluation
  - A/B testing on a real RS deployed online
  - Pros: Provides the strongest evidence,

# Evaluation methods

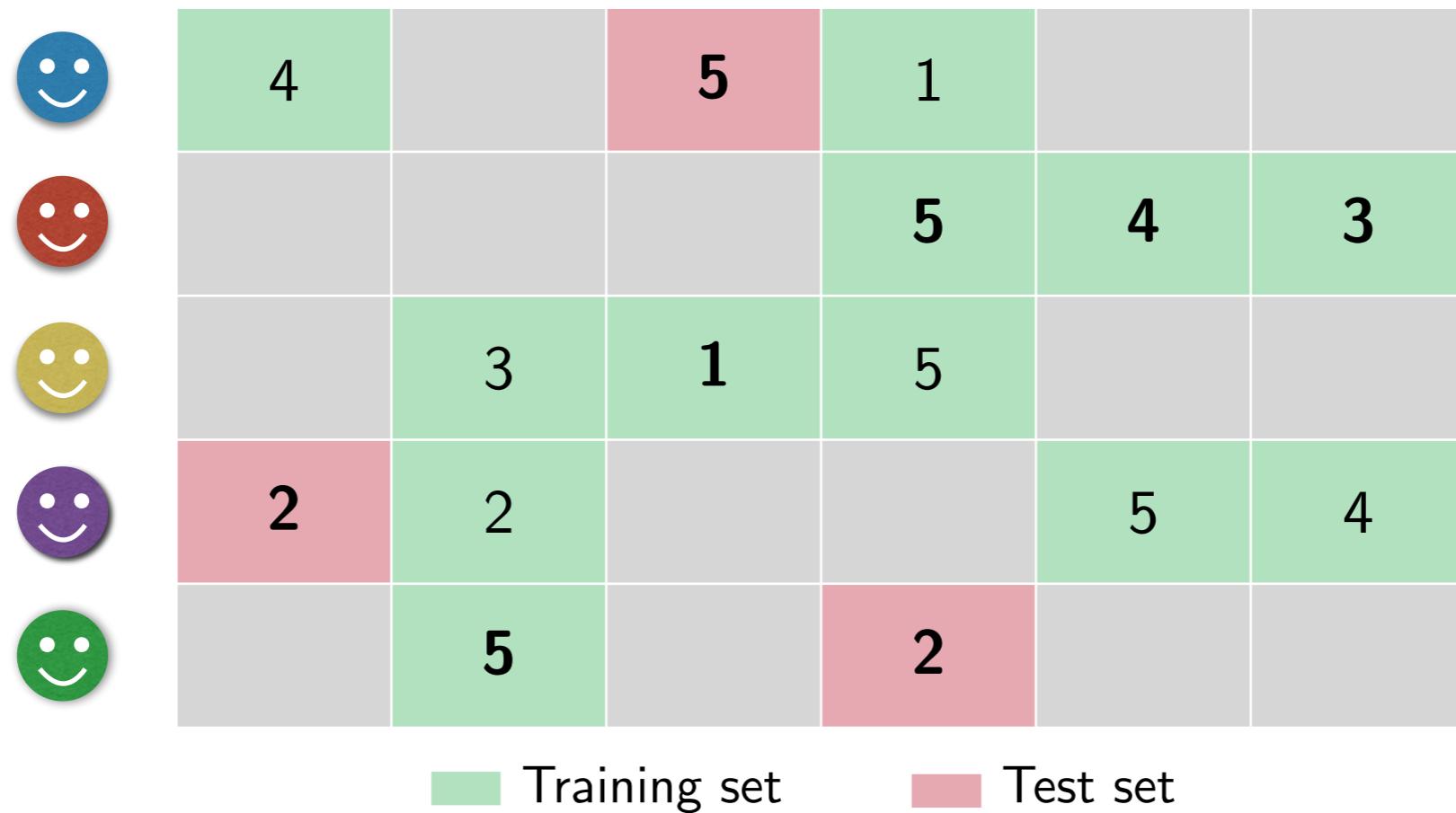
- Offline evaluation
  - Performed on previously collected datasets and simulate the user behavior
  - Pros: Does not require an interaction with real users, allows replicability,
  - Cons: Cannot measure the influence of the RS on the user behavior
- User study
  - Recruit a small number of users and ask them to interact with the RS
  - Pros: Provides more insights than offline evaluation,
  - Cons: May introduce biases in the collected data
- Online evaluation
  - A/B testing on a real RS deployed online
  - Pros: Provides the strongest evidence,
  - Cons: Risk of negatively affecting the real users's experience

# Offline evaluation



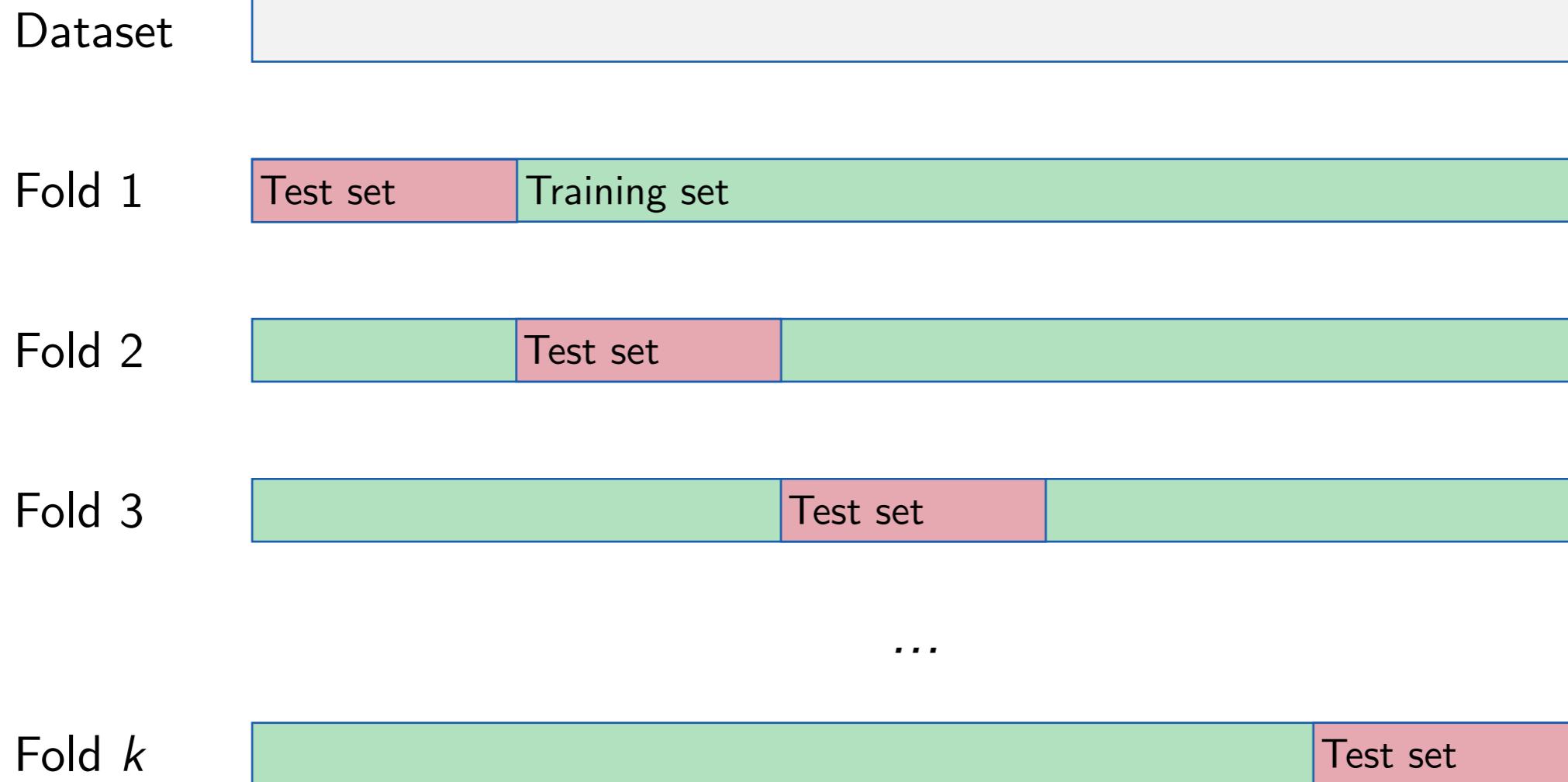
# Offline evaluation

- Random split



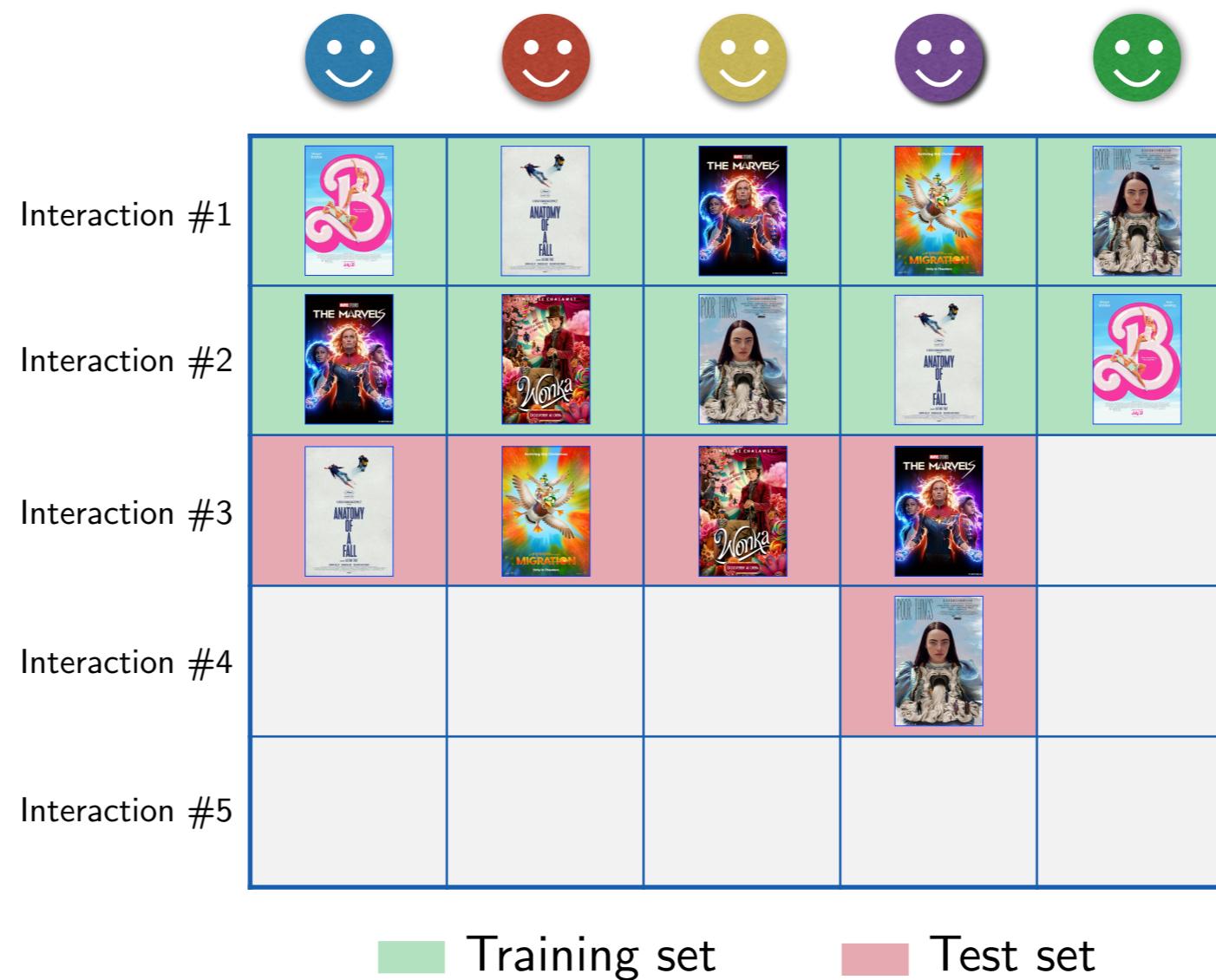
# Offline evaluation

- Random split,  $k$ -fold cross validation



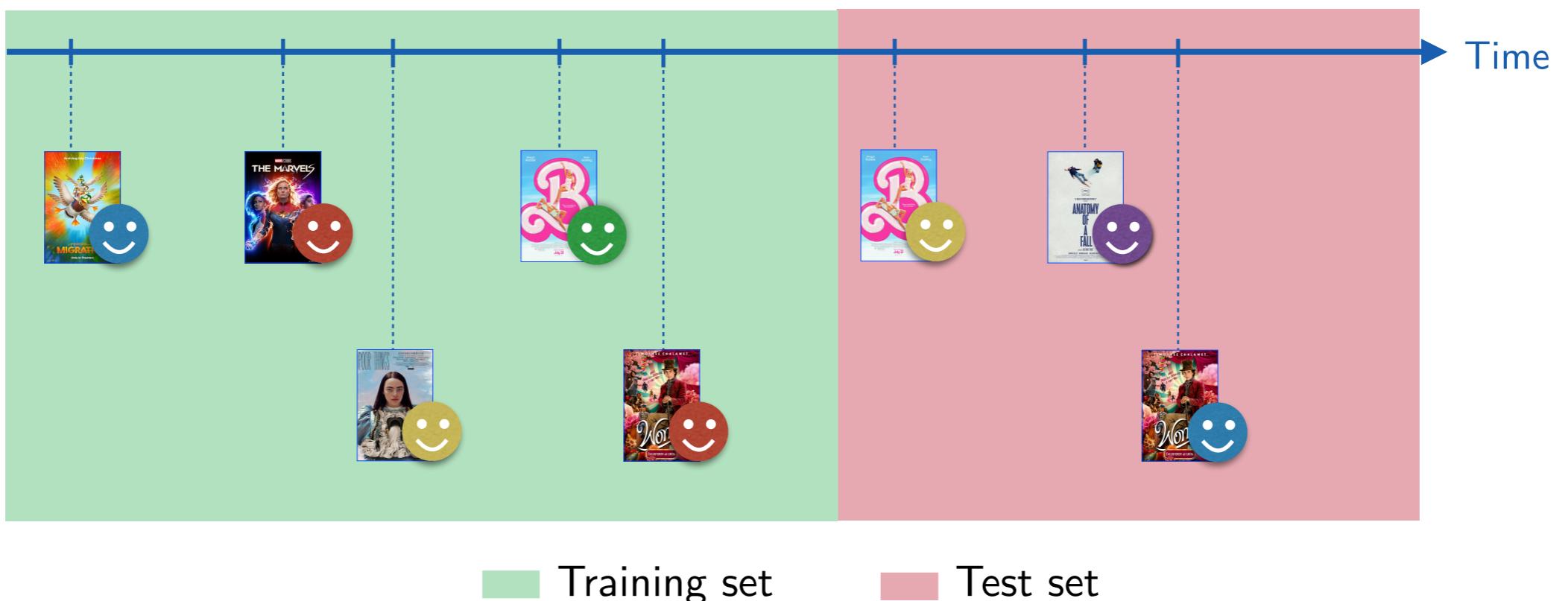
# Offline evaluation

- Given- $n$  split



# Offline evaluation

- Chronological split



# Evaluation criteria

- Accuracy
  - ▶ How accurate are the predictions of utility of items?

# Evaluation criteria

- Accuracy
  - How accurate are the predictions of utility of items?
- Coverage
  - How many items from the catalog are the recommendations covering?

# Evaluation criteria

- Accuracy
  - How accurate are the predictions of utility of items?
- Coverage
  - How many items from the catalog are the recommendations covering?
- Diversity
  - How diverse are the computed recommendations?

# Evaluation criteria

- Accuracy
  - How accurate are the predictions of utility of items?
- Coverage
  - How many items from the catalog are the recommendations covering?
- Diversity
  - How diverse are the computed recommendations?
- Serendipity
  - How surprising are the relevant computed recommendations?

# Evaluation criteria

- Accuracy
  - How accurate are the predictions of utility of items?
- Coverage
  - How many items from the catalog are the recommendations covering?
- Diversity
  - How diverse are the computed recommendations?
- Serendipity
  - How surprising are the relevant computed recommendations?
- Explainability
  - Are we able to explain and justify the recommendations?

# Evaluation metrics for accuracy

Measuring the accuracy of the predicted ratings.

- Mean Absolute Error (MAE)

$$MAE = \frac{1}{|\mathcal{T}|} \sum_{(u,i) \in \mathcal{T}} |r_{ui} - \hat{r}_{ui}|$$

# Evaluation metrics for accuracy

Measuring the accuracy of the predicted ratings.

- Mean Absolute Error (MAE)

$$MAE = \frac{1}{|\mathcal{T}|} \sum_{(u,i) \in \mathcal{T}} |r_{ui} - \hat{r}_{ui}|$$

User	Actual rating	Predicted rating	Error
	★★★★★	★★★★★	★
	★★★★★	★★★★★	★★★
	★★★★★	★★★★★	★★
	★★★★★	★★★★★	
	★★★★★	★★★★★	★

$$MAE = 1.4$$

# Evaluation metrics for accuracy

Measuring the accuracy of the predicted ratings.

- Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{|\mathcal{T}|} \sum_{(u,i) \in \mathcal{T}} (r_{ui} - \hat{r}_{ui})^2}$$

# Evaluation metrics for accuracy

Measuring the accuracy of the predicted ratings.

- Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{|\mathcal{T}|} \sum_{(u,i) \in \mathcal{T}} (r_{ui} - \hat{r}_{ui})^2}$$

User	Actual rating	Predicted rating	Error
	★★★★★	★★★★★	★
	★★★★★	★★★★★	★★★
	★★★★★	★★★★★	★★
	★★★★★	★★★★★	★★
	★★★★★	★★★★★	★

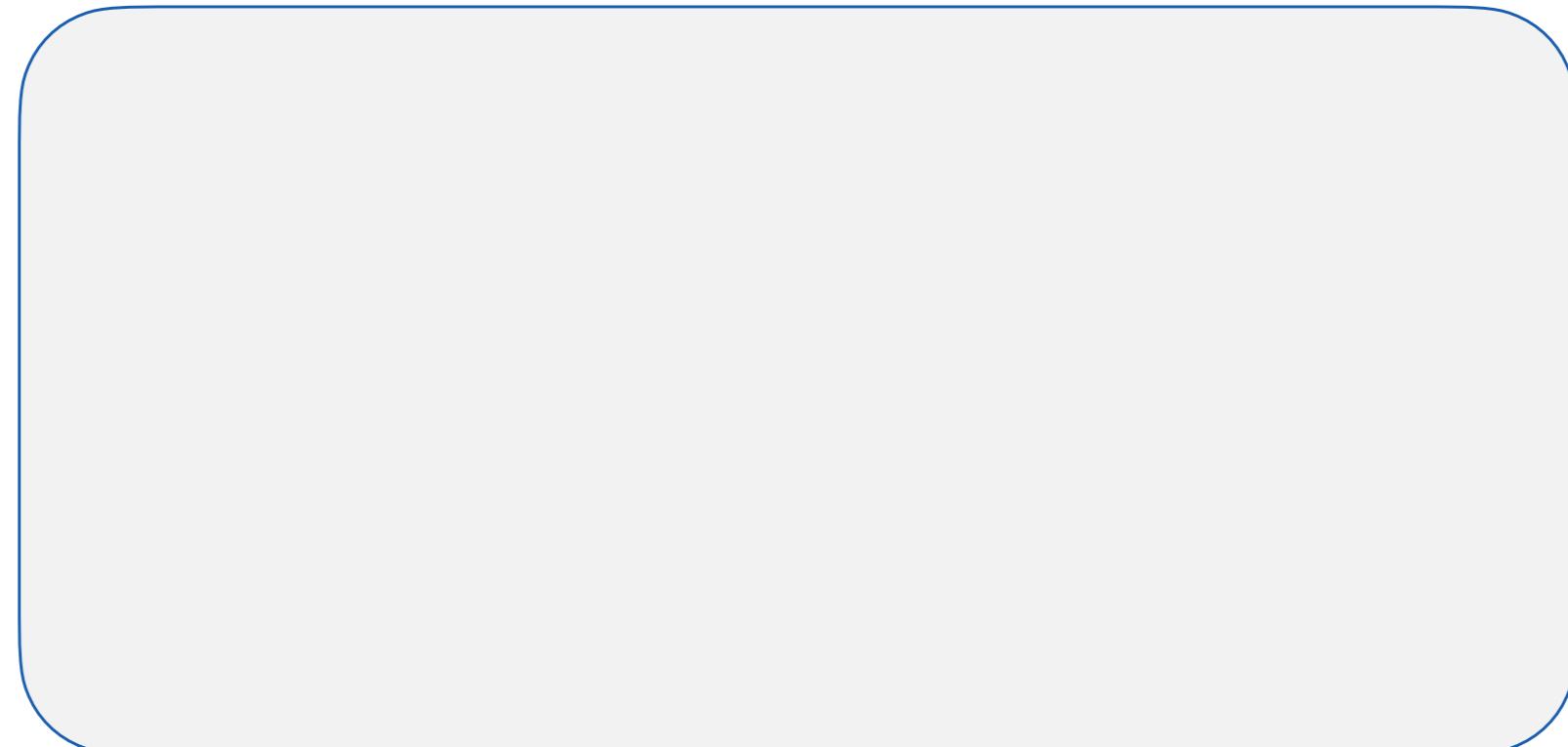
$$RMSE = 3$$

# Evaluation metrics for accuracy

Measuring the accuracy of top- $N$  recommendations

# Evaluation metrics for accuracy

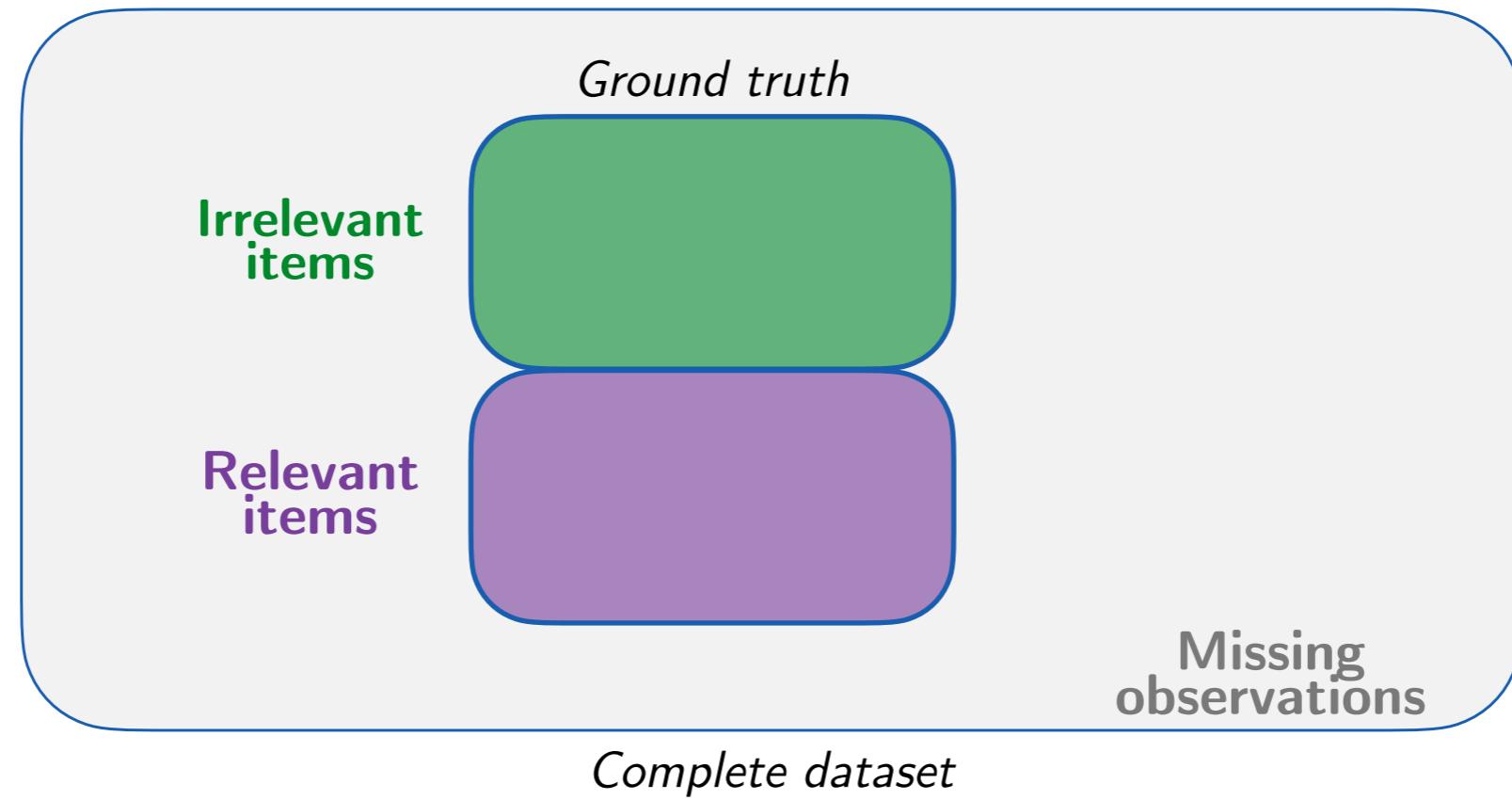
Measuring the accuracy of top- $N$  recommendations



*Complete dataset*

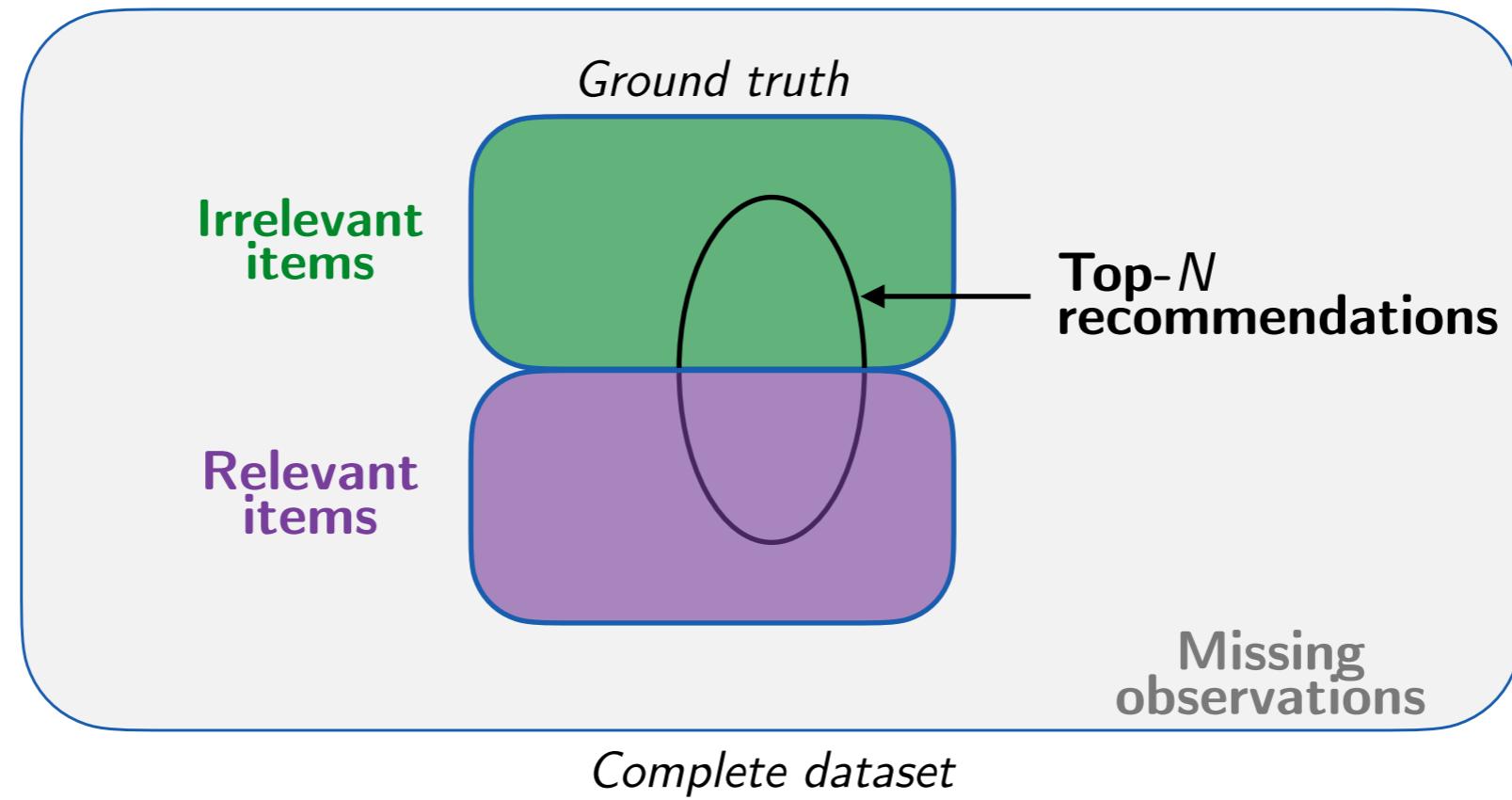
# Evaluation metrics for accuracy

Measuring the accuracy of top- $N$  recommendations



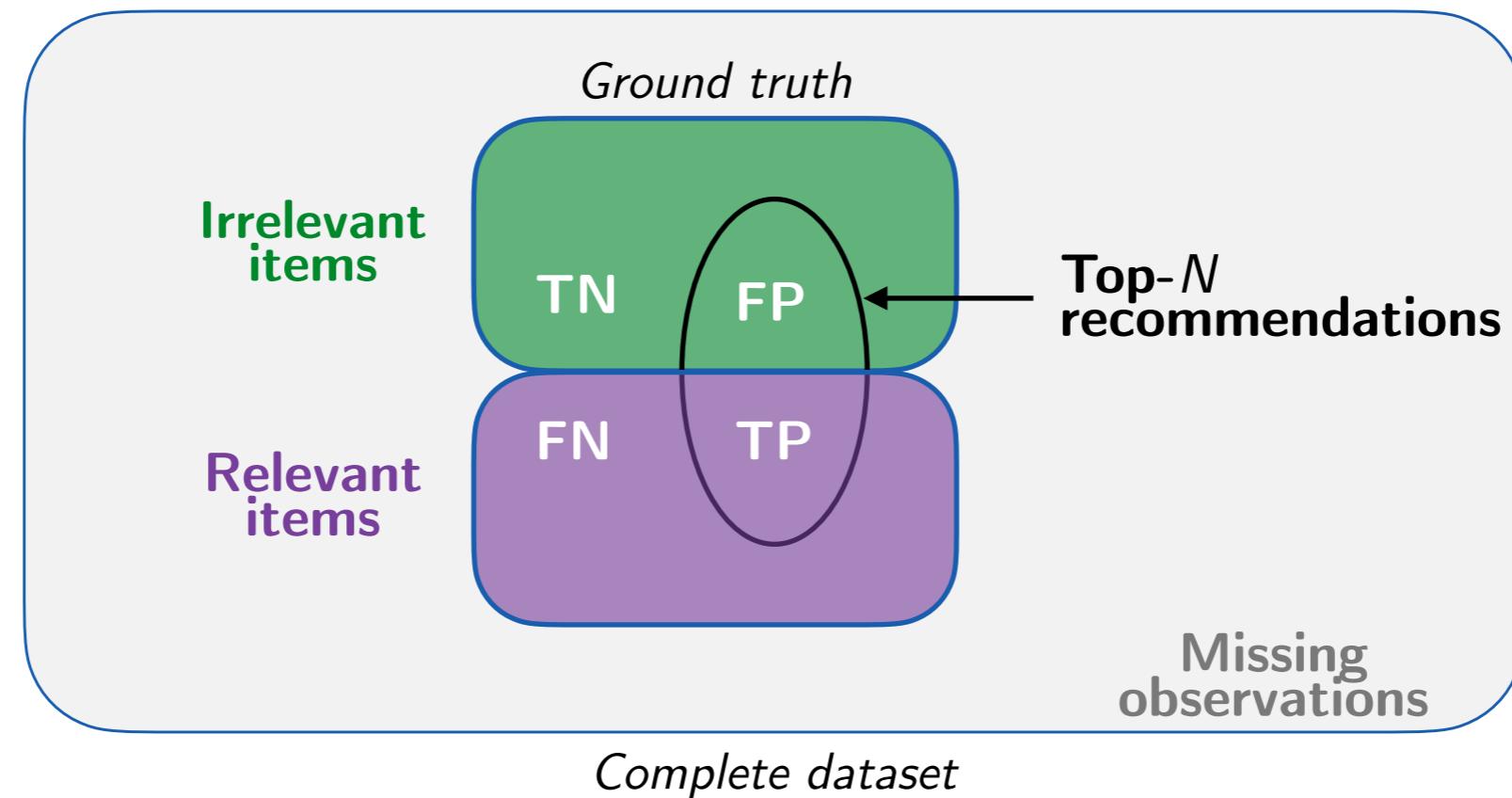
# Evaluation metrics for accuracy

Measuring the accuracy of top- $N$  recommendations



# Evaluation metrics for accuracy

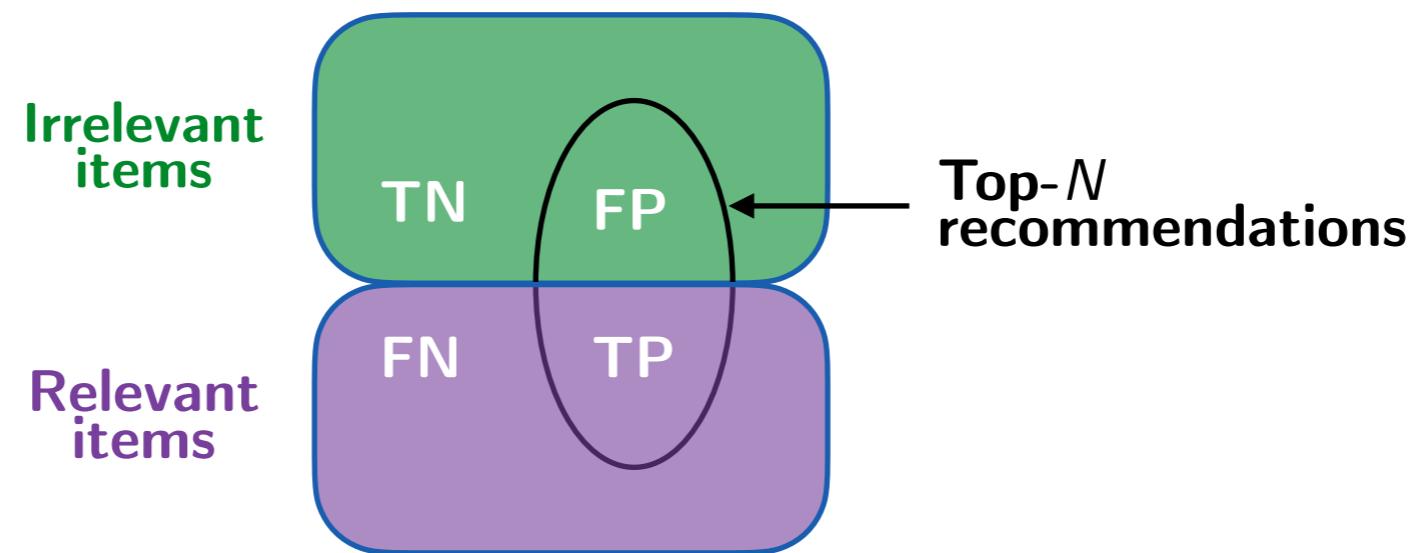
Measuring the accuracy of top- $N$  recommendations



# Evaluation metrics for accuracy

Measuring the accuracy of top- $N$  recommendations

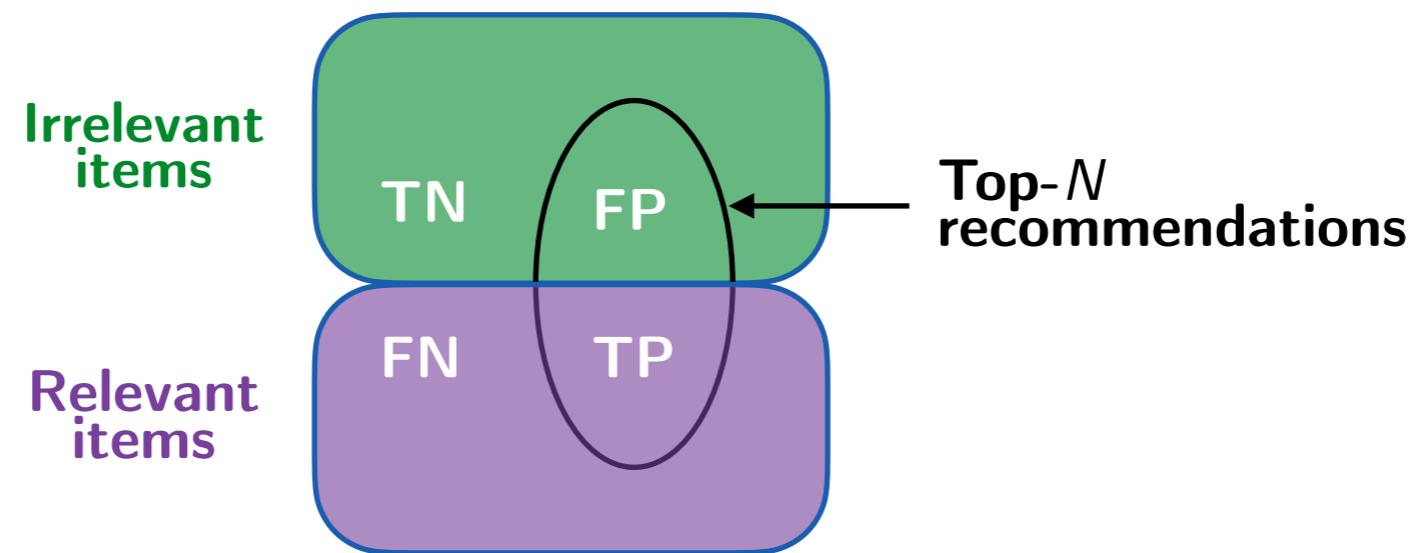
$$Recall_u@N = \frac{TP}{FN + TP} = \frac{\# \text{relevant recommended items}}{\# \text{relevant items}}$$



# Evaluation metrics for accuracy

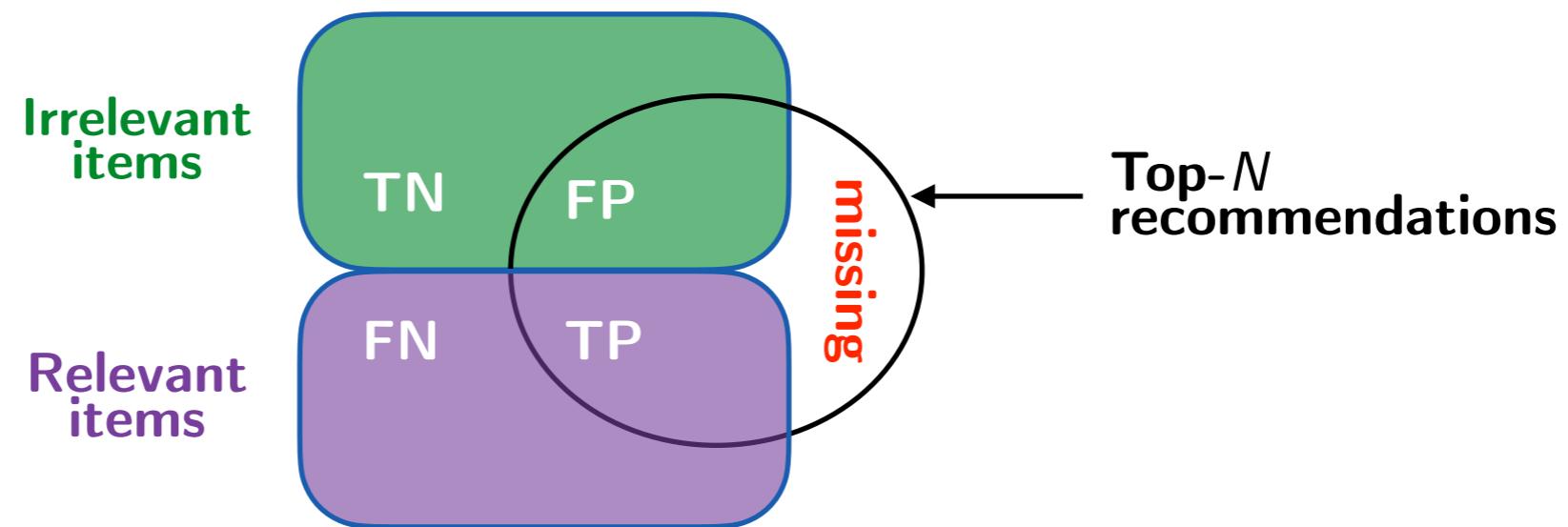
Measuring the accuracy of top- $N$  recommendations

$$Precision_u@N = \frac{TP}{FP + TP} = \frac{\# \text{relevant recommended items}}{\# \text{recommended items}}$$



# Evaluation metrics for accuracy

- Recall: *Missing Not As Random* assumption
  - Non-relevant items are more likely to be missing than relevant items
- Precision: *All Missing As Negative* assumption
  - All missing ratings are irrelevant



# Evaluation metrics for ranking

Measuring the quality of ranking of top- $N$  recommendations

- Discounted Cumulative Gain (DCG):

$$DCG_u@N = \sum_{i=1}^N \frac{rel_{ui}}{\log_2(i+1)}$$

where  $rel_{ui} = 1$  if item at rank  $i$  is relevant, 0 otherwise.

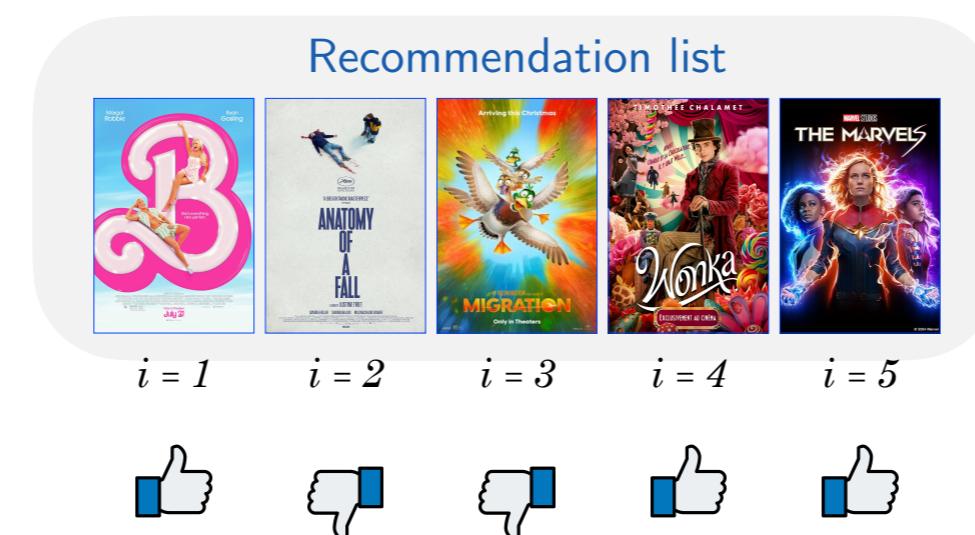
# Evaluation metrics for ranking

Measuring the quality of ranking of top- $N$  recommendations

- Discounted Cumulative Gain (DCG):

$$DCG_u@N = \sum_{i=1}^N \frac{rel_{ui}}{\log_2(i + 1)}$$

where  $rel_{ui} = 1$  if item at rank  $i$  is relevant, 0 otherwise.



$$DCG_u \quad 1 \quad + \quad 0 \quad + \quad 0 \quad + 0.43 + 0.38$$

# Evaluation metrics for ranking

Measuring the quality of ranking of top- $N$  recommendations

- Normalized Discounted Cumulative Gain (DCG):

$$NDCG@N = \frac{1}{|\mathcal{T}_u|} \sum_{u \in \mathcal{T}_u} \frac{DCG_u@N}{DCG_u^*@N}$$

where  $DCG_u^*$  is the best possible  $DCG_u$  that can be obtained.

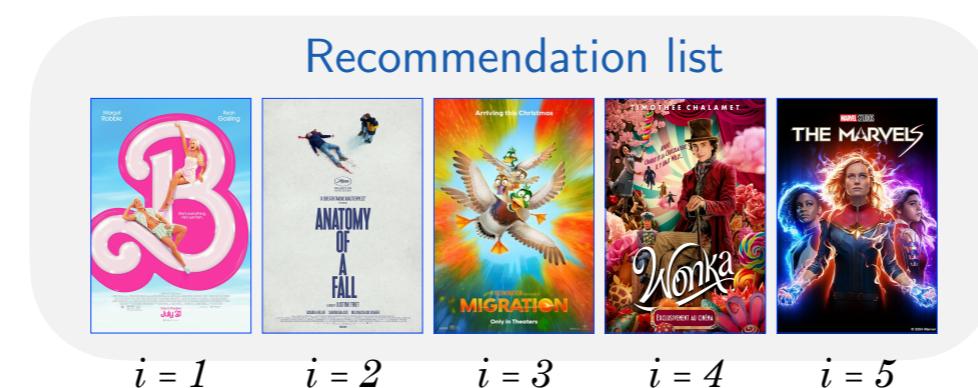
# Evaluation metrics for ranking

Measuring the quality of ranking of top- $N$  recommendations

- Normalized Discounted Cumulative Gain (DCG):

$$NDCG@N = \frac{1}{|\mathcal{T}_u|} \sum_{u \in \mathcal{T}_u} \frac{DCG_u@N}{DCG^*_u@N}$$

where  $DCG^*_u$  is the best possible  $DCG_u$  that can be obtained.



$$DCG^*_u$$

$$1 + 0.63 + 0.5 + 0.43 + 0.38$$

# List of other used metrics

- Mean Average Precision (MAP)
- Mean Reciprocal Rank (MRR)
- Area under the ROC Curve (AUC)
- Hit-Rank (HR)
- Average Reciprocal Hit-Rank (ARHR)

...

# **Non-personalized recommendations**

# Non-personalized recommendations

- Personalized recommendations
  - Tailor the recommendation list to a specific individual
  - Using information about the target user and similar users

vs.

- Non-personalized recommendations
  - Do not require to know anything about the individual user
  - Using aggregated information about users

# Most popular item recommendation

**Top 50**

**Global Top 50**

Your daily update of the most played tracks right now.

3 New Entries • Last Updated: 19 hours ago

FOLLOWERS 14,478,257

PLAY

#	TITLE	ARTIST	DAILY PLAYS
1	Dance Monkey	Tones and I	6,024,896
2	Lucid Dreams	Juice WRLD	4,260,964
3	ROXANNE	Arizona Zervas	4,181,885
4	Circles	Post Malone	3,805,122
5	Memories	Maroon 5	3,655,212
6	Señorita	Shawn Mendes, Camila Cabello	3,404,481
7	Don't Start Now	Dua Lipa	3,272,428
8	everything i wanted	Billie Eilish	3,229,442

**Top destinations**

AMSTERDAM 55 things to do →

VIENNA 59 things to do →

LONDON 72 things to do →

Barcelona 77 things to do →

Berlin 35 things to do →

Milan 39 things to do →

**Les meilleures ventes**

Nos produits les plus populaires selon les ventes. Mises à jour chaque heure.

Les meilleures ventes en Jeux vidéo

#1 Luigi's Mansion 3 Nintendo	#2 Ring Fit Adventure pour Nintendo Switch	#3 PokéMón Epée	#4 Mario Kart 8 Deluxe Nintendo	#5 Mario & Sonic at the Olympic Games Tokyo 2020 Nintendo	#6 PokéMón Bouclier	#7 Minecraft switch standard Nintendo
★★★★★ 406	★★★★★ 297	★★★★★ 262	★★★★★ 808	★★★★★ 33	★★★★★ 140	★★★★★ 306
44,99 € prime	44,99 € prime	44,99 € prime	44,99 € prime	44,99 € prime	44,99 € prime	23,99 € prime
#8 The Legend of Zelda: Link's Awakening Nintendo	#9 New Super Mario Bros. U Deluxe Nintendo	#10 Nintendo Switch avec paire de Joy-Con Rouge...	#11 Super Mario Maker 2 Nintendo	#12 Ubisoft Just Dance 2020 - Switch	#13 Super Mario Party Switch	#14 Sony Manette PlayStation 4 Officielle...
★★★★★ 531	★★★★★ 325	★★★★★ 137	★★★★★ 257	★★★★★ 33	★★★★★ 188	★★★★★ 122
44,49 € prime	44,49 € prime	299,99 € prime	44,99 € prime	7 offres à partir de EUR 49,90	44,49 € prime	39,99 € prime

**TIFFANY & CO.**

My Account

Jewelry Love & Engagement Watches Home & Accessories Fragrance Men's Gifts

**Most Popular Gifts**

New

# Association rules

## Customers who viewed this item also viewed



Nespresso by De'Longhi  
ENV155TAE VertuoPlus  
Deluxe Coffee and  
Espresso Machine...  
★★★★★ 229  
\$124.99



Nespresso by De'Longhi  
ENV155BAE VertuoPlus  
Deluxe Coffee and  
Espresso Machine...  
★★★★★ 229  
\$149.99



Nespresso by De'Longhi  
ENV135T Vertuo Evoluo  
Coffee and Espresso  
Machine by De'Longhi,...  
★★★★★ 937  
\$139.30



Nespresso by De'Longhi  
ENV155T VertuoPlus  
Deluxe Coffee and  
Espresso Machine by...  
★★★★★ 229  
\$138.96



Nespresso by De'Longhi  
ENV155B VertuoPlus  
Deluxe Coffee and  
Espresso Machine by...  
★★★★★ 229  
\$119.99



Breville-Nespresso USA  
BNV420BLK1BUC1  
VertuoPlus Coffee and  
Espresso Machine, Black  
★★★★★ 343  
\$119.99

## Frequently bought together



# Association rules



Confidence:  $c(i \rightarrow j) = \frac{|\mathcal{U}_i \cap \mathcal{U}_j|}{|\mathcal{U}_i|}$

Support:  $S(i \rightarrow j) = \frac{|\mathcal{U}_i \cap \mathcal{U}_j|}{|\mathcal{U}|}$

# Towards personalized recommendations...

- Recommendations approaches
  - ▶ Collaborative Filtering approaches
  - ▶ Content-Based Filtering approaches
  - ▶ Hybrid approaches
  - ▶ Context-aware approaches