UNIVERSITY OF PRIMORSKA, KOPER
APRIL 15, 2025

PETER KNEES
TU Wien, Faculty of Informatics, Austria

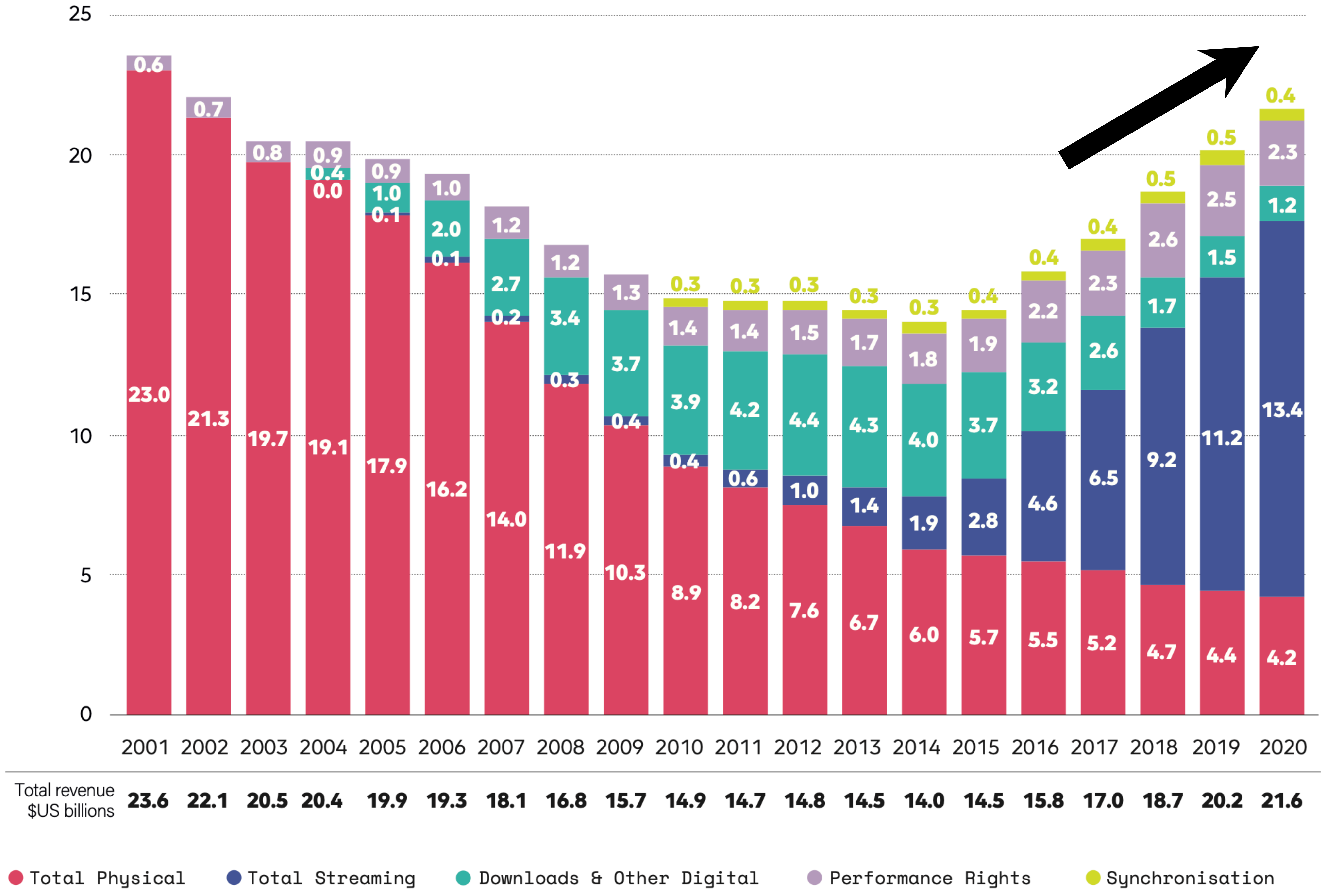
TU WIEN Informatics

MUSIC RECOMMENDER SYSTEMS

unesco Chair

MUSIC CONSUMPTION

GLOBAL RECORDED MUSIC INDUSTRY REVENUES 2001-2020 (US$ BILLIONS)



GLOBAL RECORDED MUSIC REVENUES BY SEGMENT 2020

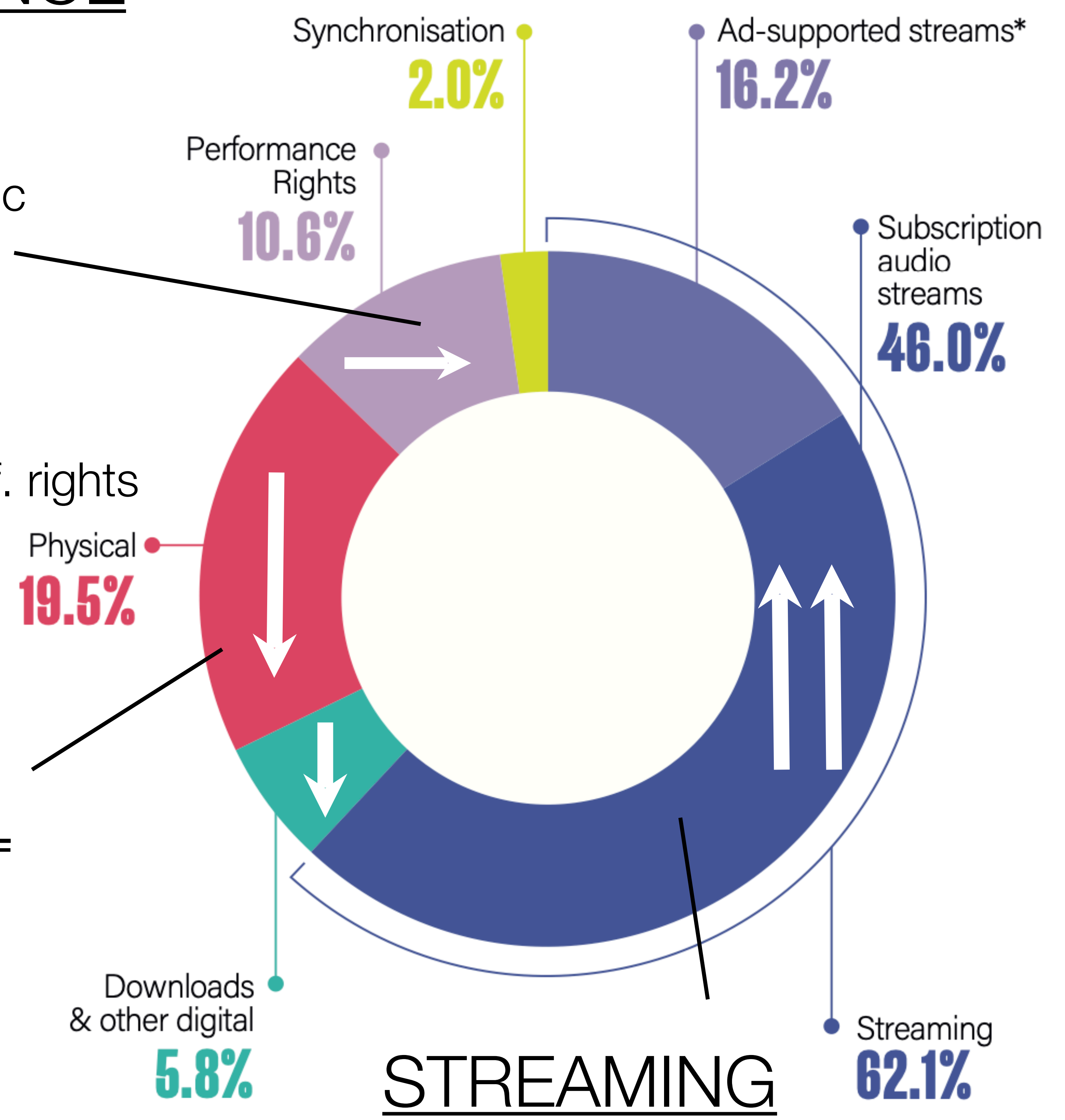
PERFORMANCE RIGHTS

Revenue from music reproduction:
- on AM/FM radio
- at public venues

(NB: Excluding perf. rights from Streaming)

PHYSICAL

e.g. CDs



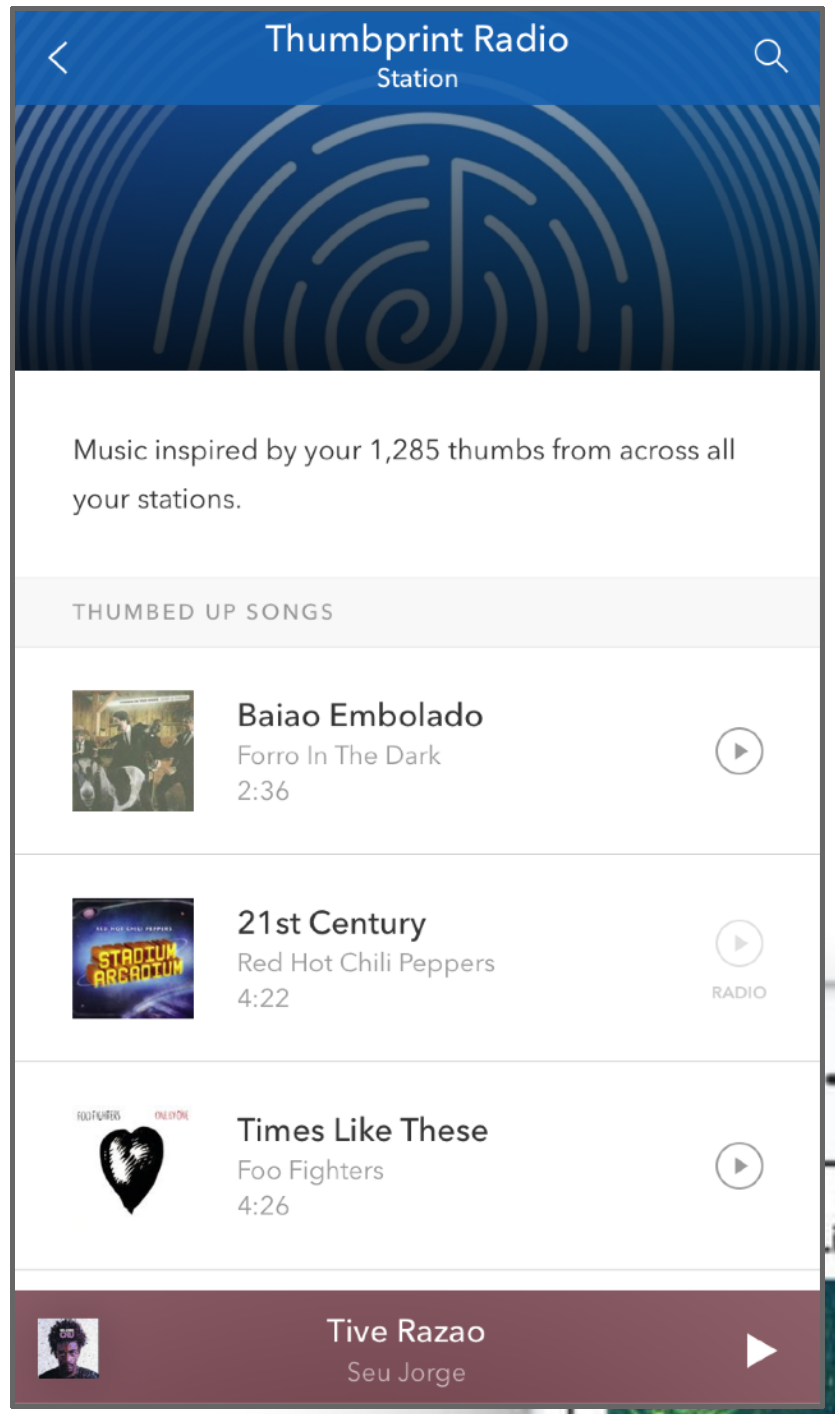
STREAMING

- Internet radio & on-demand
- Ad-supported & subscriptions

*INCLUDES AD-SUPPORTED STREAMS AND VIDEO STREAM REVENUES.

MUSIC DISCOVERY

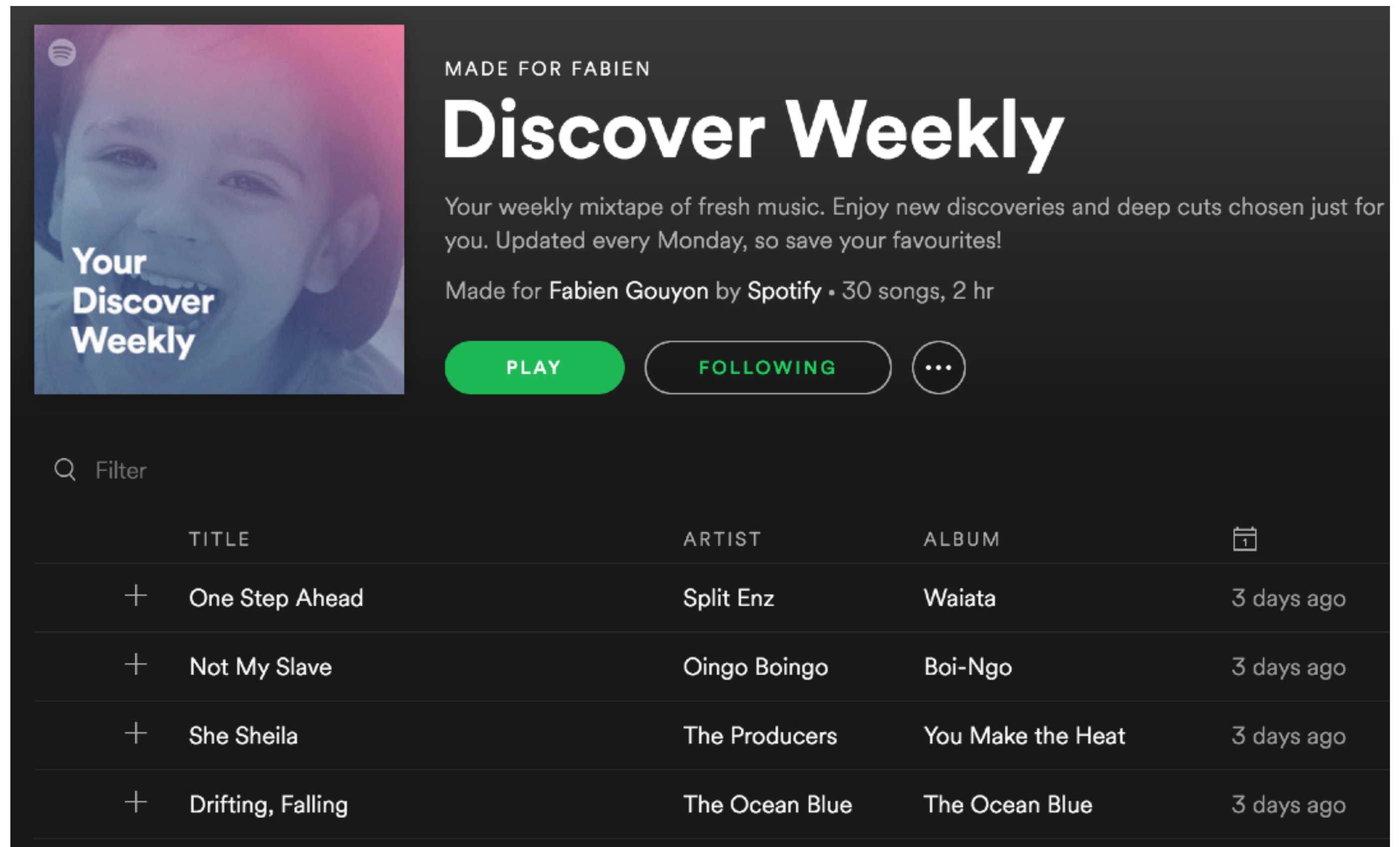
- Streaming "taking over" physical & downloads
- But competing with terrestrial radio, too

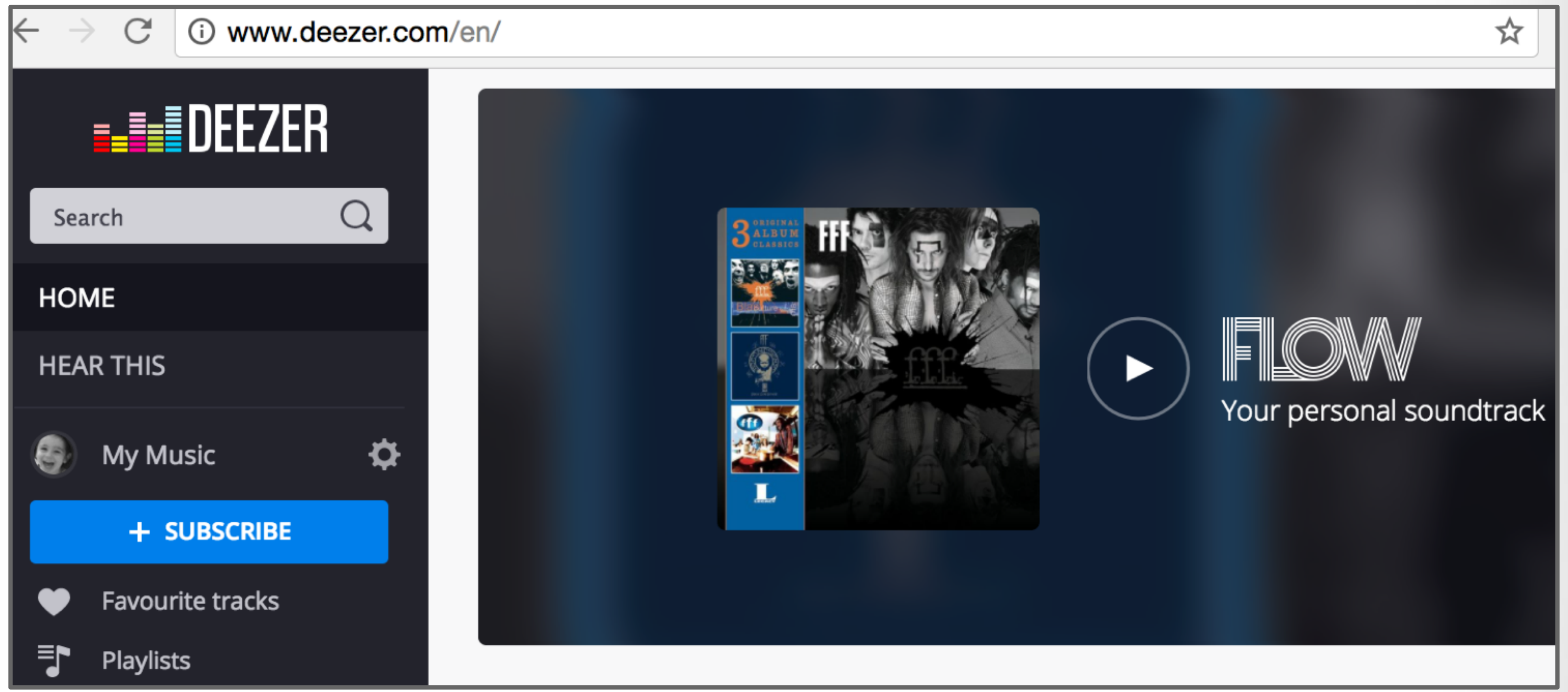


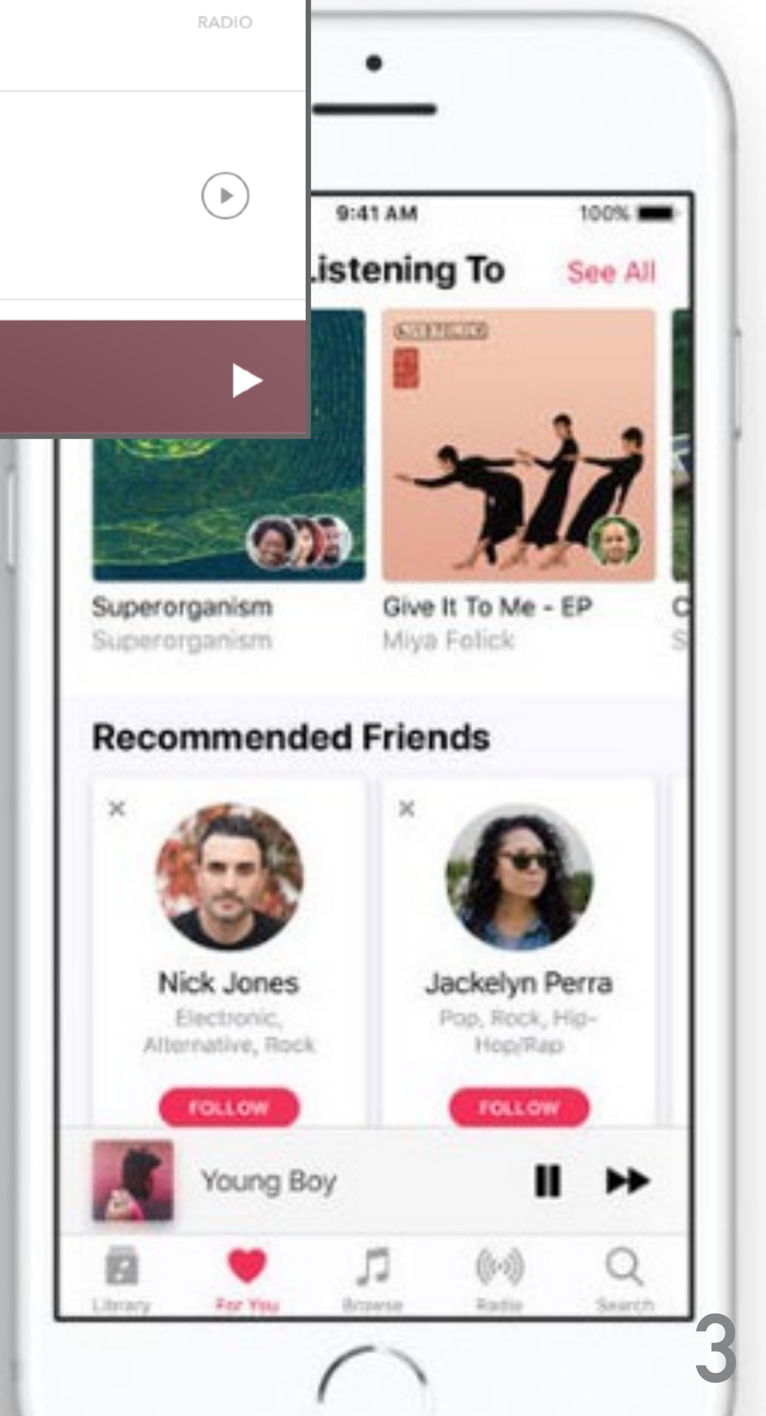
The Quest for "Discovery"

Ongoing quest for defining listening format calls for:

- Innovative Discovery features
- <u>Right balance between lean-in & lean-back experiences</u>







RECOMMENDER SYSTEMS

- Growing amounts of data artifacts available
(User generated + commercial)
- Providing tailored views onto massive data collections
➡ **Personalization**
- Recommender Systems are an <u>**Information Filter**</u>
- Provide right items (options, answers, …) at the right time
- Found in all areas, powers central services and platforms of digital economy

WHAT'S SPECIAL TO MUSIC RECOMMENDATION?

- Extremely relevant to the music industry due to **rise of streaming**
- Magnitude of available data items (millions) & data points (billions)
- **Various types of items** to recommend (songs, albums, artists, audio samples, concerts, venues, fans, etc.); **various tasks** (discovery, playlist building, etc.) subsumed under music recommendation
- Wide range of duration of items (2+ vs. 90+ minutes),
  Lower commitment, items more "disposable", **low item cost**
  → "bad" recommendations maybe not as severe
- Very often **consumed in sequence**
- **Re-recommendation** often appreciated (in contrast to e.g. movies)

WHAT'S SPECIAL TO MUSIC RECOMMENDATION?

- **Listener intent and context** are crucial
    - Different consumption locations/settings: static (e.g., via stereo at home) vs. variable (e.g., via headphones during exercise), alone vs. in group, etc.
    - Often consumed **passively** (while working, background music, etc.)
    - Importance of **social** component
    - Music often used for **self-expression**
    - Highly **emotionally connoted** (in contrast to products, e.g. home appliances)
- Recommendations relevant for various actors (listeners, producers, performers, etc.), **multi-stakeholder** recommendation scenarios
- **Diversity of modalities** (audio, user feedback, text, etc.)

DATA FUELS RECOMMENDERS

Interaction Data

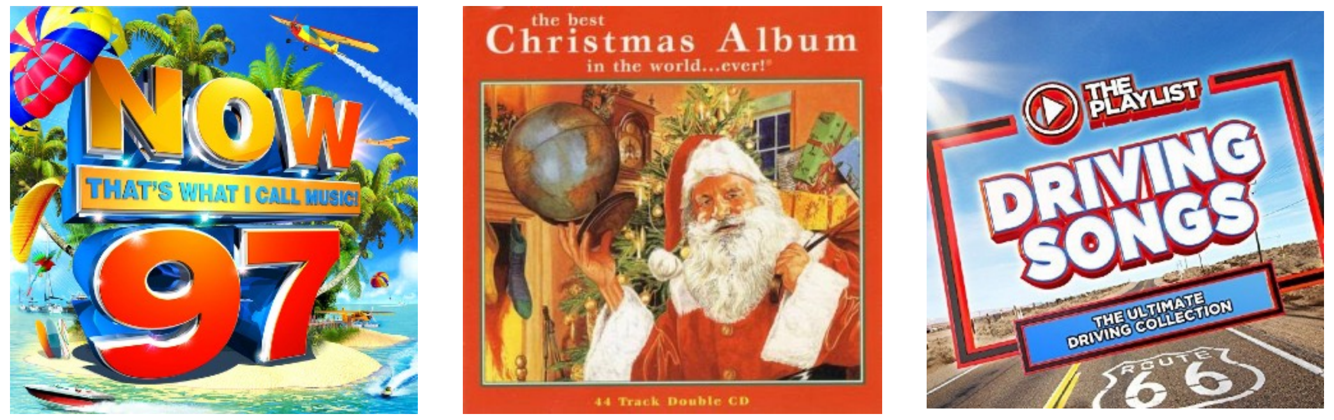
- Listening logs, listening histories
- Feedback ("thumbs"), purchases

User-generated

- Tags, reviews, stories

Curated collections

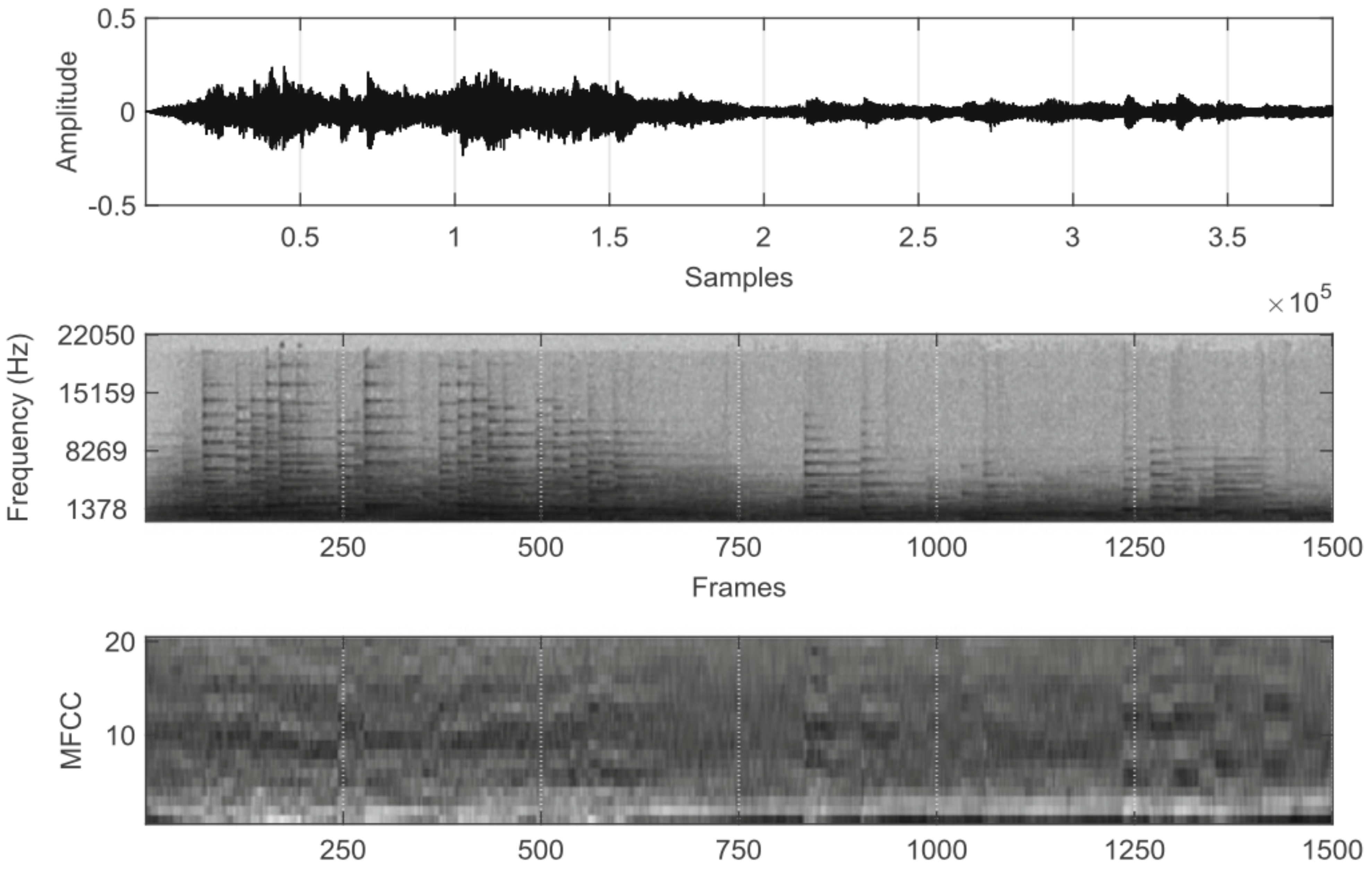
- Playlists, radio channels
- CD album compilations



DATA FUELS RECOMMENDERS

Content (audio, symbolic, lyrics)

- Machine listening/content analysis
- Human labelling



Meta-data

- Editorial
- Curatorial
- Multi-modal (album covers etc.)

AUDIO CONTENT ANALYSIS

➡️ CF. MUSIC-IR

- Finding "similar sounding" tracks
- Features can be extracted from any audio file
  → no other data or community necessary
  → no cultural biases (no popularity bias, no subjective ratings etc.)
- Manually crafted features or automatic via machine learning
- Representation learning and temporal modeling directly from signals (even as GenAI side products)

[Choi et al., 2017] *A Tutorial on Deep Learning for Music Information Retrieval*, arXiv:1709.04396.

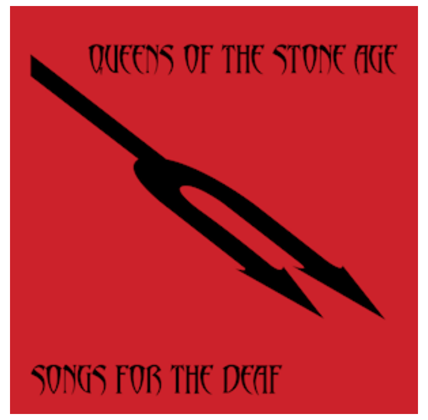
[Casey et al., 2008] *Content-based music information retrieval: Current directions and future challenges,* Proc IEEE 96 (4).

[Müller, 2015] *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*, Springer.

AUDIO CONTENT ANALYSIS: SELECTED FEATURES

Disturbed
The Sound of Silence

- Timbre (→ MFCCs)
  e.g. for genre classification,
  "more-of-this" recommendations



- Beat/downbeat → Tempo: 85 bpm

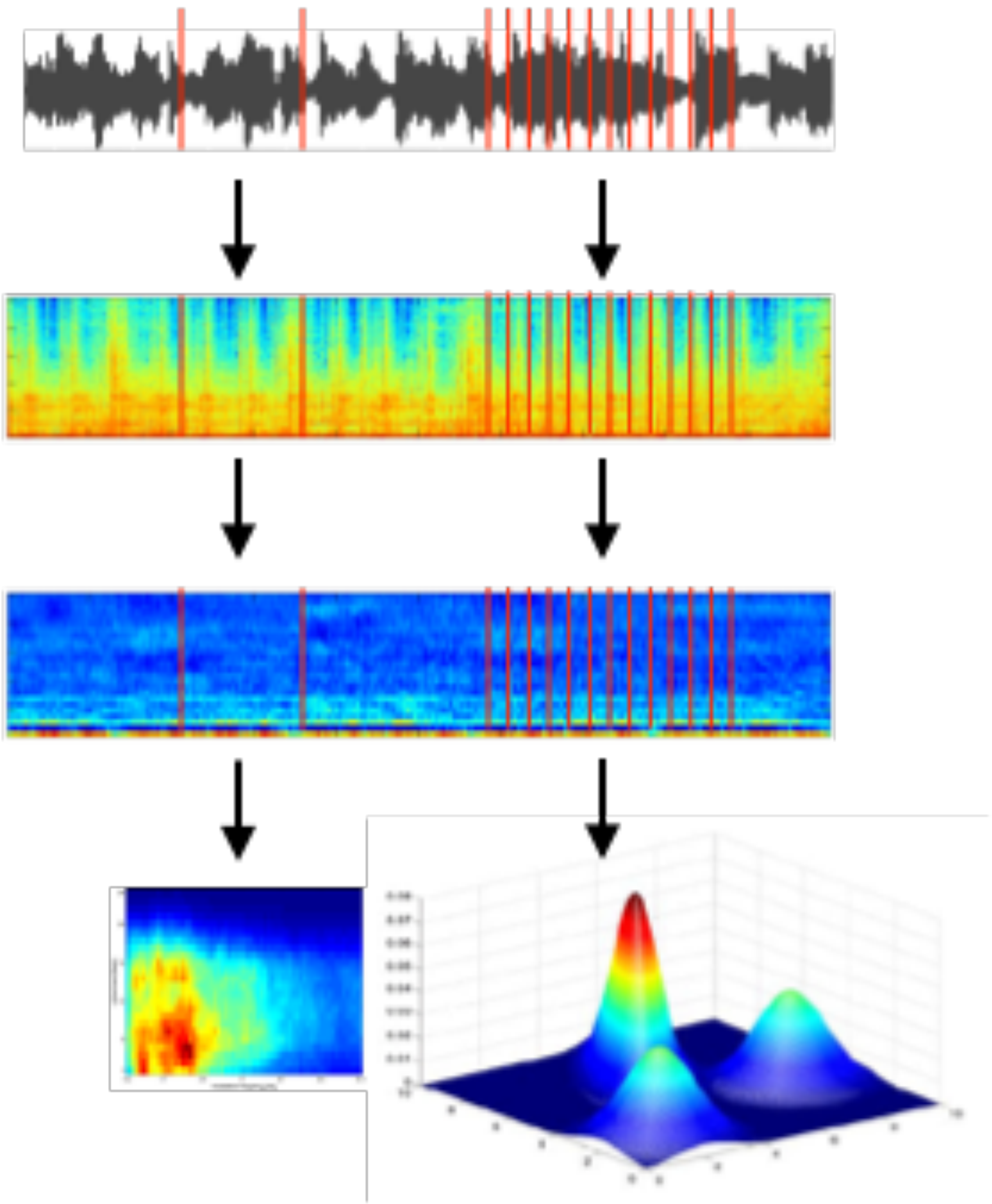
- Tonal features (→ Pitch-class profiles)
  e.g. for melody extraction,
  cover version identification

Different versions of this song:
Simon & Garfunkel - The Sound of Silence
Anni-Frid Lyngstad (ABBA) - En ton av tystnad
etc.

- Semantic categories via machine learning:
  `not_danceable, gender_male, mood_not_happy`

AUDIO FEATURES: BASIC PROCESSING STEPS

- Convert signal from time domain to frequency domain, e.g., using a Fast Fourier Transform (FFT)
- Psychoacoustic transformation (Mel-scale, Bark-scale, Cent-scale, …): mimics human listening process (not linear, but logarithmic!), removes aspects not perceived by humans, emphasizes low frequencies
- Extract features, e.g.,
  - Block-level (large time windows, e.g., 6 sec)
  - Frame-level (short time windows, e.g., 25 ms) needs model distribution of frames
- Calculate similarities between feature vectors/models



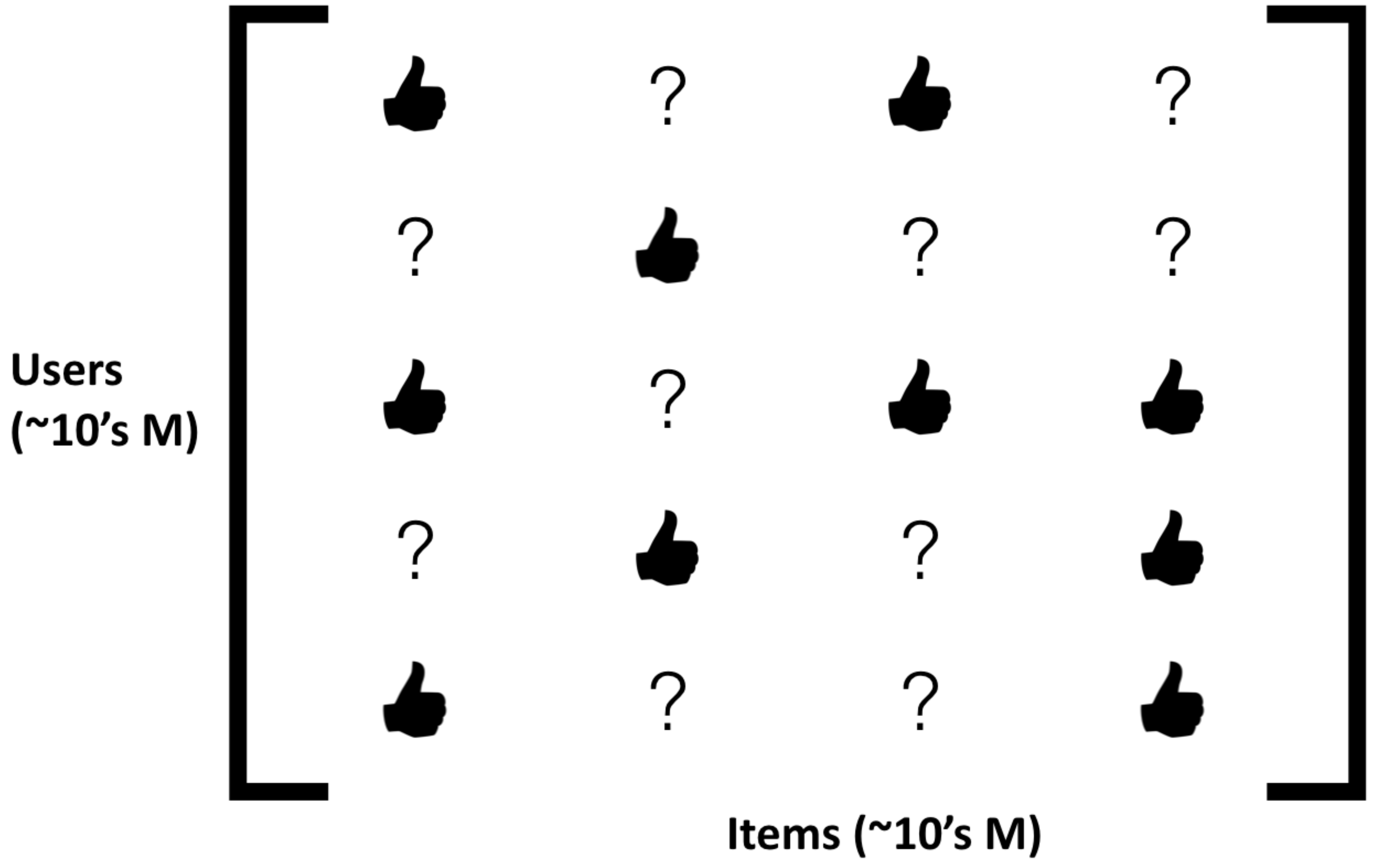
SO, GREAT … WHY IS THIS DIFFICULT THEN?

- “Objective” similarity measure of the “sound of music”
- Describes the output of the applied transformation as designed
- Works well for genre and mood classification …and that’s more or less it

- The resulting numbers represent a very narrow aspect of acoustic properties, describe no *musical* qualities (structure, development, time dependency, etc.)
- Which sound properties are important to whom and in which context?
- Lack of any personal preferences or experiences

COLLABORATIVE FILTERING (CF)

- Exploits interaction data
- *"People who listened to track A, also listened to track B"*
- Main underlying assumption: users that had similar taste in the past, will have similar taste in the future
- Stemming from "usage" of music
  → close to "what users want"
- Typical methods
  - Memory-based ("k-NN like")
  - Matrix factorization



EXAMPLE OF COLLABORATIVE FILTERING OUTPUT

People who liked **Disturbed — The Sound of Silence**, also liked…

1. Bad Wolves – Zombie

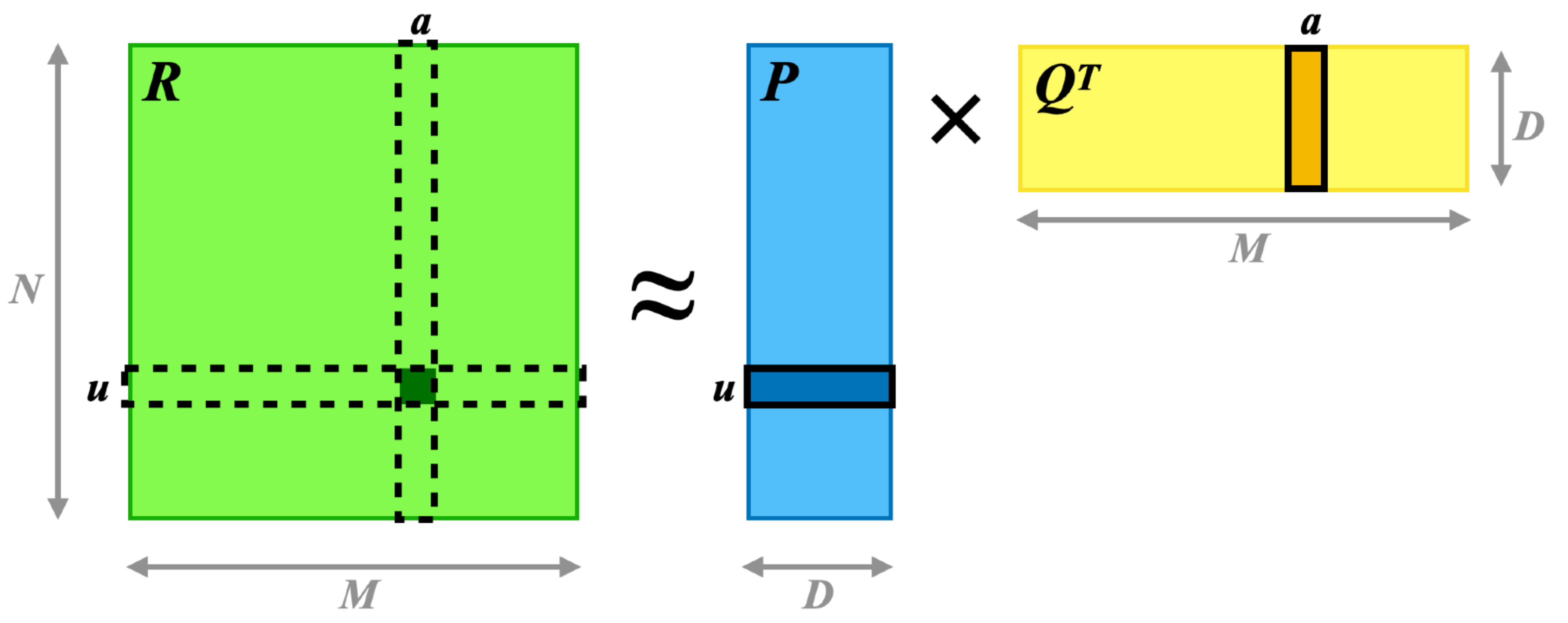
2. Five Finger Death Punch – Bad Company

3. Disturbed – The Light

4. Metallica – Nothing Else Matters

MATRIX FACTORIZATION

- Recommendation: a matrix completion problem
- Rationale: observed data are interactions of 2 factors: users and items
- Decompose interaction matrix $R \in \mathbb{N}^{N\times M}$ into user and item matrices P, Q of lower rank D: $P \in \mathbb{R}^{N\times D}$, $Q \in \mathbb{R}^{M\times D}$



- Objective: minimize reconstruction error

$$\min_{p_*,q_*} \sum_{r_{ua}\in R} (r_{ua} - p_u q_a^T)^2 + \boxed{\lambda(\|p_u\|^2 + \|q_a\|^2)}$$

regularization term

- Prediction of interaction: inner product of vectors of user u and item a

OPTIMIZATION THROUGH STOCHASTIC GRADIENT DESCENT

- Learn factors from R using stochastic gradient descent, cf. [Funk/Webb 2006]
- Gradient: calculating the direction of the steepest descent
- Need partial derivatives of the error function $\sum e_{ua}^2 = (r_{ua} - p_u q_a^T)^2$ wrt. each parameter p_{ud}, q_{ad} for all $u, a, d \quad d\ldots \textit{latent dimension}$
- Determine step length or use fixed length parameter as multiplier on the gradient ⇒ learning rate γ
- Update p_{ud} and q_{ad} using “backpropagation" (cf. neural networks)

$$p_{uD} \leftarrow p_{uD} - \gamma \frac{\partial e_{iA}^2}{\partial p_{uD}} = p_{uD} + \gamma 2(r_{uA} - r'_{uA}) q_{AD}$$

$$q_{aD} \leftarrow q_{aD} - \gamma \frac{\partial e_{iA}^2}{\partial q_{aD}} = q_{aD} + \gamma 2(r_{uA} - r'_{uA}) p_{uD}$$

- Note: factors not necessarily interpretable (just capture variance in data)

LIMITATIONS FOR MUSIC RECOMMENDATION

- Extreme sparsity in music interaction data
- Special case: Cold-start Problem
  - Recommending for new users: no prior interactions/preferences
  - Recommending new tracks: no prior usage
  - Mitigation: **incorporate additional data sources**
- Explicit feedback rare in music listening

  - Summing up listening events, often on artist level (combat sparsity)
  - **Factorization techniques for implicit data**

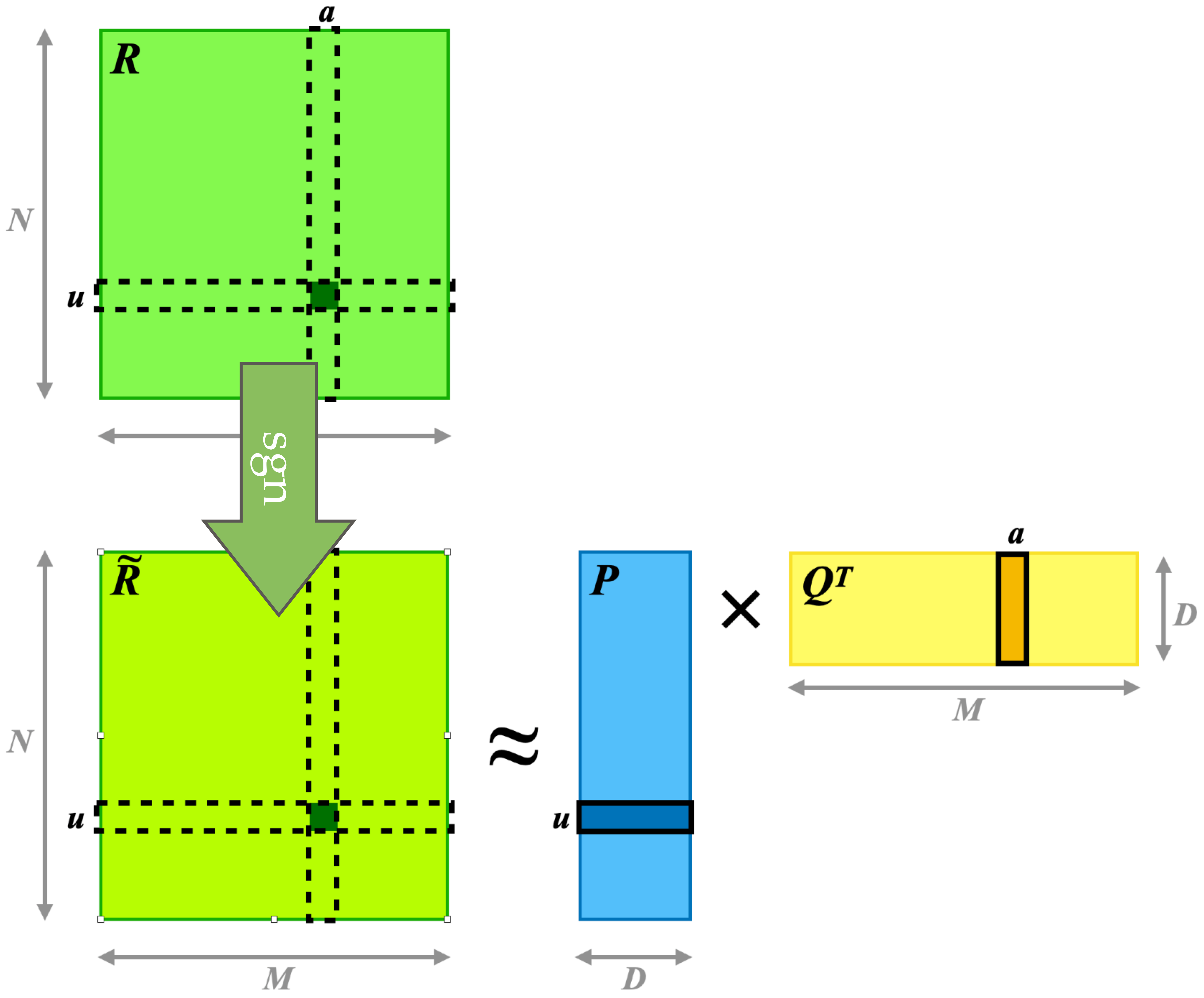
IMPLICIT FEEDBACK MATRIX FACTORIZATION (MF)

- Extension of Matrix Factorization for implicit data by Hu et al. [2008]
- R … aggregated play counts per artist
- $\tilde{R}_{ua} = \begin{cases} 1 & \text{if } R_{ua} > 0 \\ 0 & \text{if } R_{ua} = 0 \end{cases}$ $\quad w(\eta, x) = 1 + \eta \log(1 + x)$

  "preference" "confidence"
- Objective

$$\min_{P,Q} \sum_{ua \in R} w(\alpha, R_{ua})(\tilde{R}_{ua} - P_u Q_a^T)^2 + \lambda(\|P\|_F^2 + \|Q\|_F^2)$$

- λ, α determined by grid search



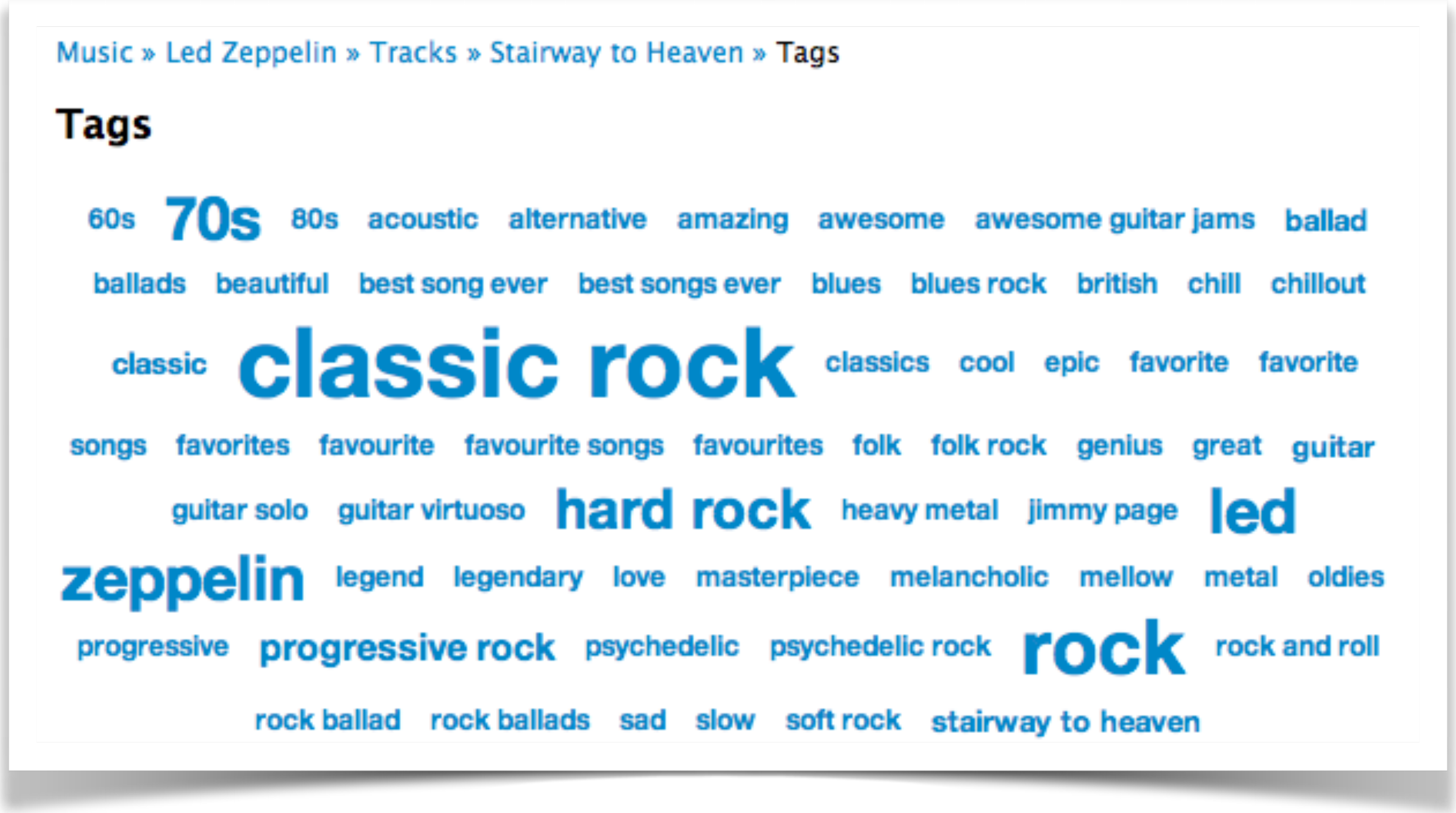
- Gradient updates

$$P_u = (Q^T W_R^{r_u} Q + \lambda I)^{-1}(Q^T W_R^{r_u} R_{r_u}^T)$$

$$Q_a = (P^T W_R^{c_a} P + \lambda I)^{-1}(P^T W_R^{c_a} R_{c_a}^T)$$

HYBRID RECOMMENDATION APPROACH

- Sparsity in interaction data
- Idea: incorporate information expressed in different types of data
- Tags: explicit semantic information assigned by users
- Combine implicit factorization with tagging data
- Extending model by Hu et al. (2008) by incorporating
  - users' tagging preferences (user-tag model)
  - how tracks have been tagged (track-tag model)



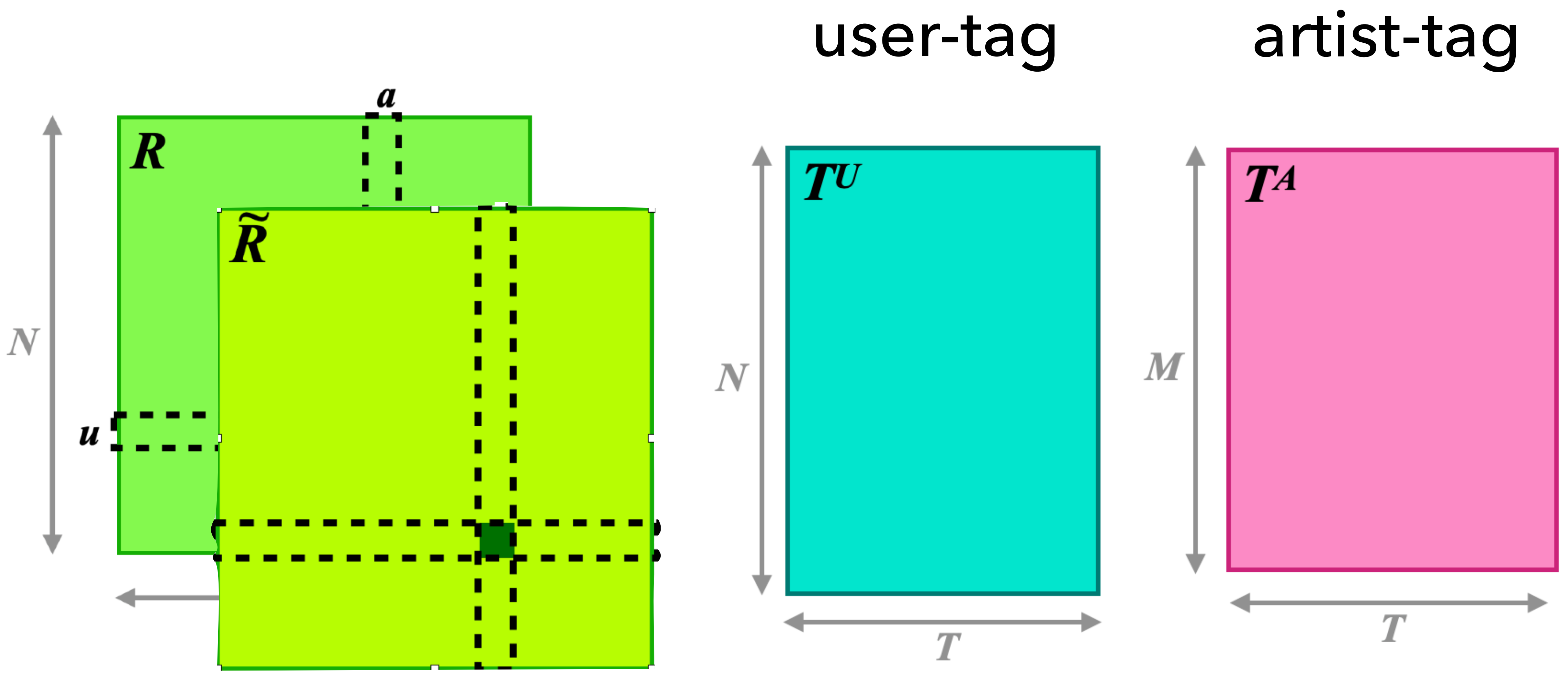
Improving Music Recommendations with a Weighted Factorization of the Tagging Activity.
A. Vall, M. Skowron, P. Knees, and M. Schedl. In: Proc 16th International Society for Music Information Retrieval Conference (ISMIR), 2015.

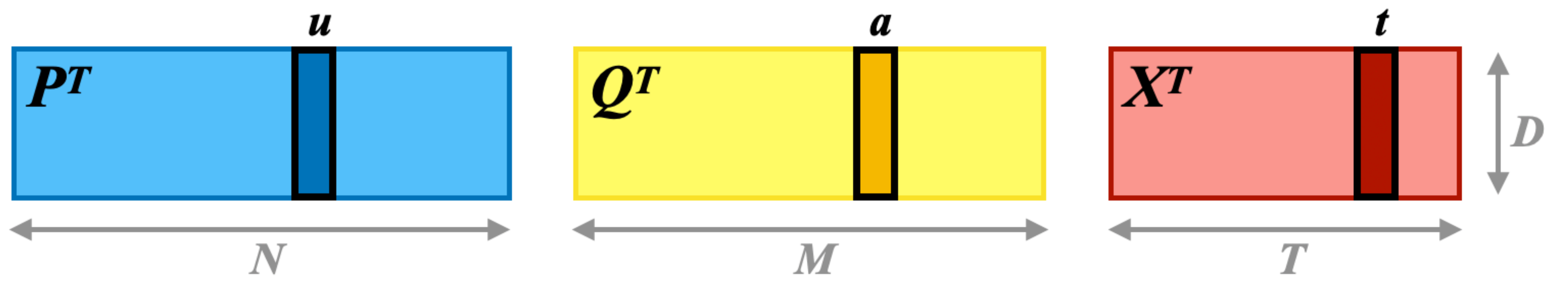
New Paths in Music Recommender Systems Research.
M. Schedl, P. Knees, and F. Gouyon. In: Proc 11th ACM Conference on Recommender Systems (RecSys), 2017.

IMPLICIT FEEDBACK MATRIX FACTORIZATION WITH TAGGING (TMF)

- Introducing T tags
- $T^U \in \mathbb{N}^{N \times T}$ … user-tag frequency matrix
- $T^A \in \mathbb{N}^{M \times T}$… artist-tag frequency matrix
- Latent factor matrix for tags $X \in \mathbb{R}^{T \times D}$
- Objective

$$
\begin{aligned}
\min_{P,Q,X} \; & \sum_{ua \in R} w(\alpha, R_{ua}) \left(\tilde{R}_{ua} - P_u Q_a^T\right)^2 \\
& + \mu_1 \sum_{ut \in T^U} \left(T^U_{ut} - P_u X_t^T\right)^2 \\
& + \mu_2 \sum_{at \in T^A} \left(T^A_{at} - Q_a X_t^T\right)^2 \\
& + \lambda \left(\|P\|_F^2 + \|Q\|_F^2 + \|X\|_F^2 \right).
\end{aligned}
$$





- Gradient updates

$$P_u = (Q^T W_R^{r_u} Q + \mu_1 X^T X + \lambda I)^{-1} (Q^T W_R^{r_u} R_{r_u}^T + \mu_1 X^T T_{r_a}^{UT})$$

$$Q_a = (P^T W_R^{c_a} P + \mu_2 X^T X + \lambda I)^{-1} (P^T W_R^{c_a} R_{c_a}^T + \mu_2 X^T T_{r_a}^{AT})$$

$$X_t = (\mu_1 P^T P + \mu_2 Q^T Q + \lambda)^{-1} (\mu_1 P^T T_{c_t}^{UT} + \mu_2 Q^T T_{c_t}^{AT})$$

IMPLICIT FEEDBACK MATRIX FACTORIZATION WITH WEIGHTED TAGGING (WTMF)

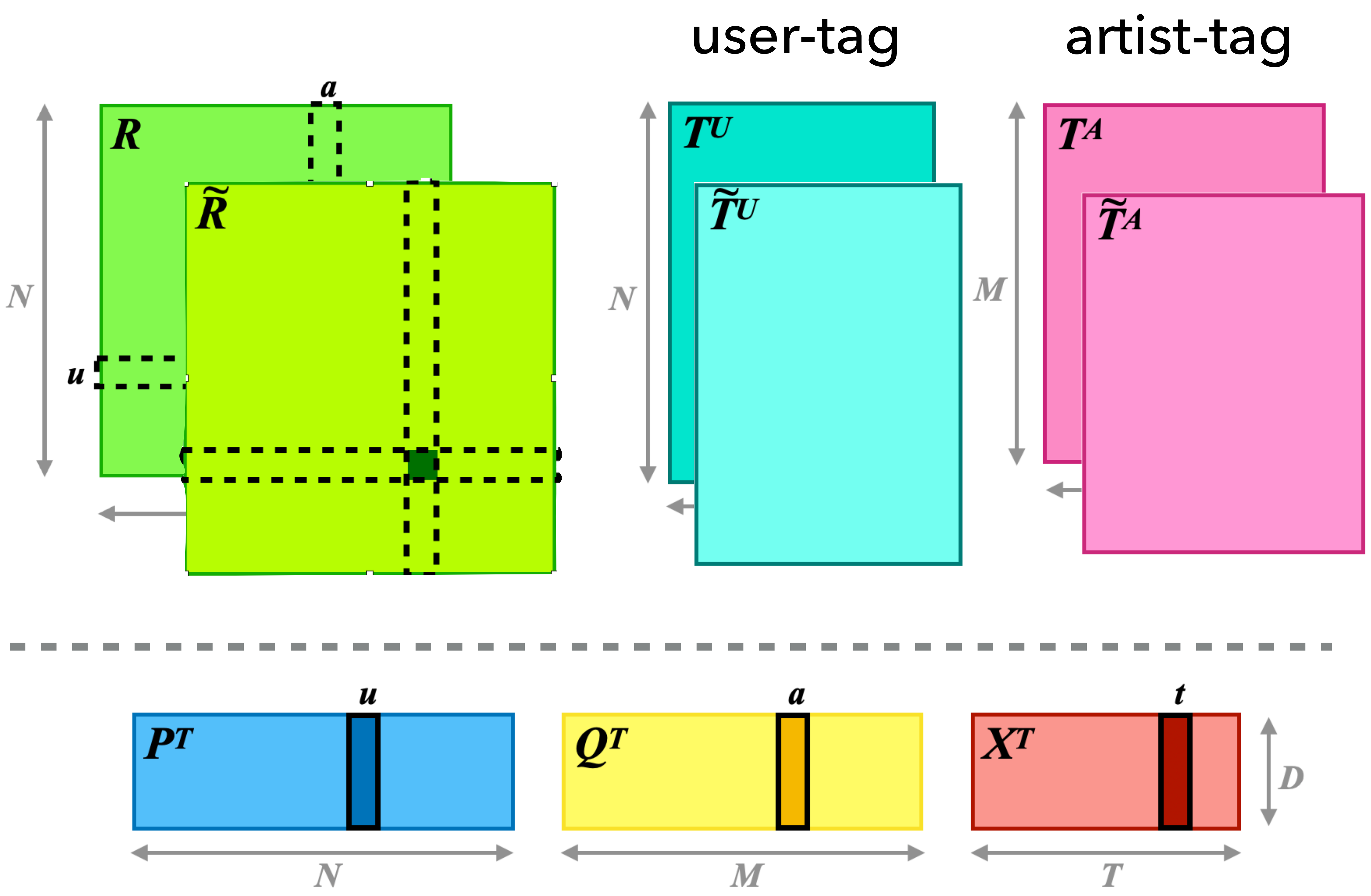
- Extending weighting to tag matrices

$$\tilde{T}^U_{ut} = \begin{cases} 1 & \text{if } T^U_{ut} > 0 \\ 0 & \text{if } T^U_{ut} = 0 \end{cases} \qquad \tilde{T}^A_{at} = \begin{cases} 1 & \text{if } T^A_{at} > 0 \\ 0 & \text{if } T^A_{at} = 0 \end{cases}$$

- Objective

$$\min_{P,Q,X} \sum_{ua \in R} w(\alpha, R_{ua}) \left(\tilde{R}_{ua} - P_u Q_a^T \right)^2 + \mu_1 \sum_{ut \in T^U} w(\beta, T^U_{ut}) \left(\tilde{T}^U_{ut} - P_u X_t^T \right)^2 + \mu_2 \sum_{at \in T^A} w(\gamma, T^A_{at}) \left(\tilde{T}^A_{at} - Q_a X_t^T \right)^2 + \lambda \left(\|P\|_F^2 + \|Q\|_F^2 + \|X\|_F^2 \right).$$

- Prediction: $Z = PQ^T$



- Gradient updates

$$P_u = (Q^T W_R^{r_u} Q + \mu_1 X^T W_{T^U}^{r_u} X + \lambda I)^{-1} (Q^T W_R^{r_u} R_{r_u}^T + \mu_1 X^T W_{T^U}^{r_u} T_{r_a}^{UT})$$

$$Q_a = (P^T W_R^{c_a} P + \mu_2 X^T W_{T^A}^{r_a} X + \lambda I)^{-1} (P^T W_R^{c_a} R_{c_a}^T + \mu_2 X^T W_{T^A}^{r_a} T_{r_a}^{AT})$$

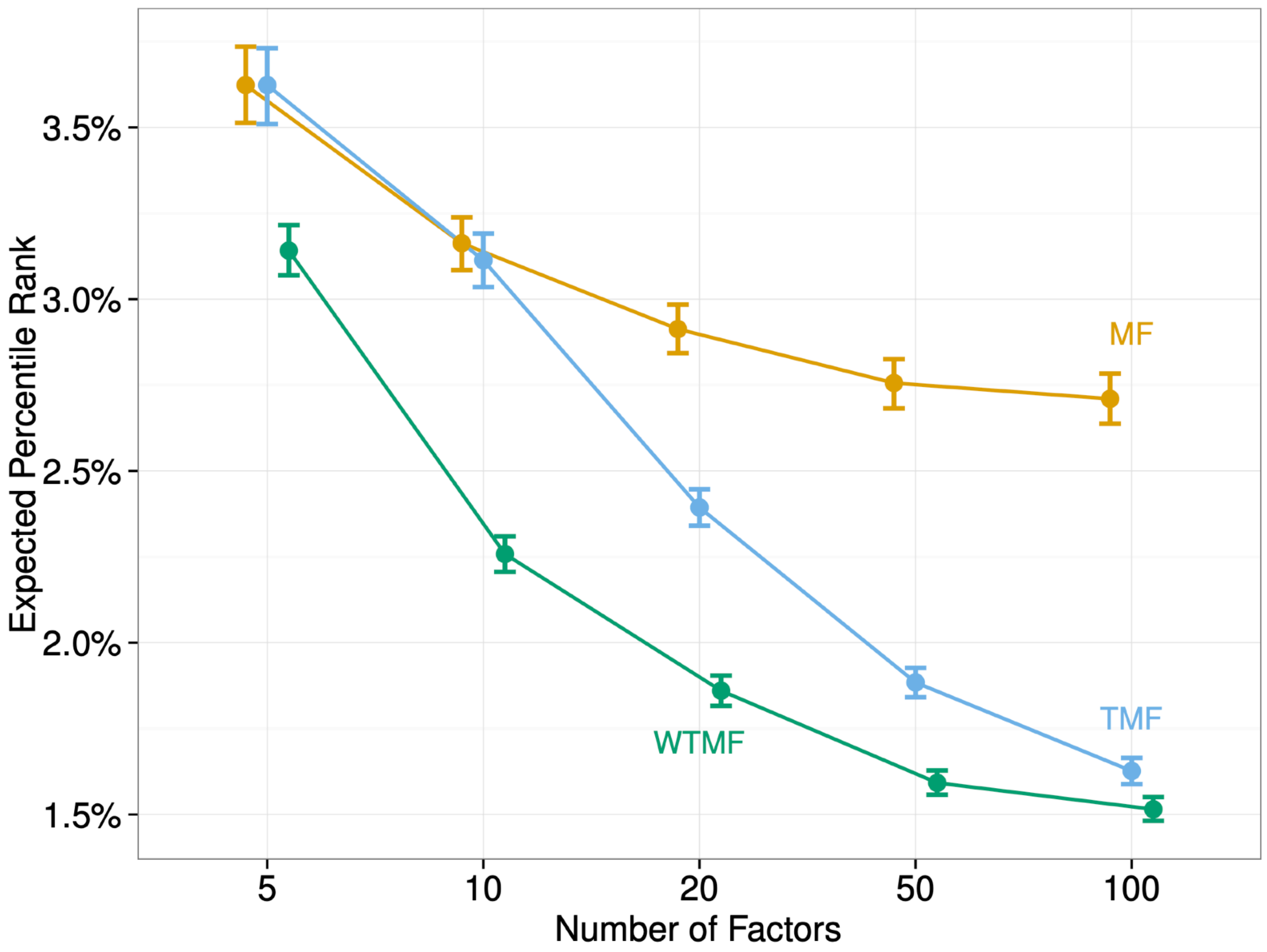
$$X_t = (\mu_1 P^T W_{T^U}^{c_t} P + \mu_2 Q^T W_{T^A}^{c_t} Q + \lambda)^{-1} (\mu_1 P^T W_{T^U}^{c_t} T_{c_t}^{UT} + \mu_2 Q^T W_{T^A}^{c_t} T_{c_t}^{AT})$$

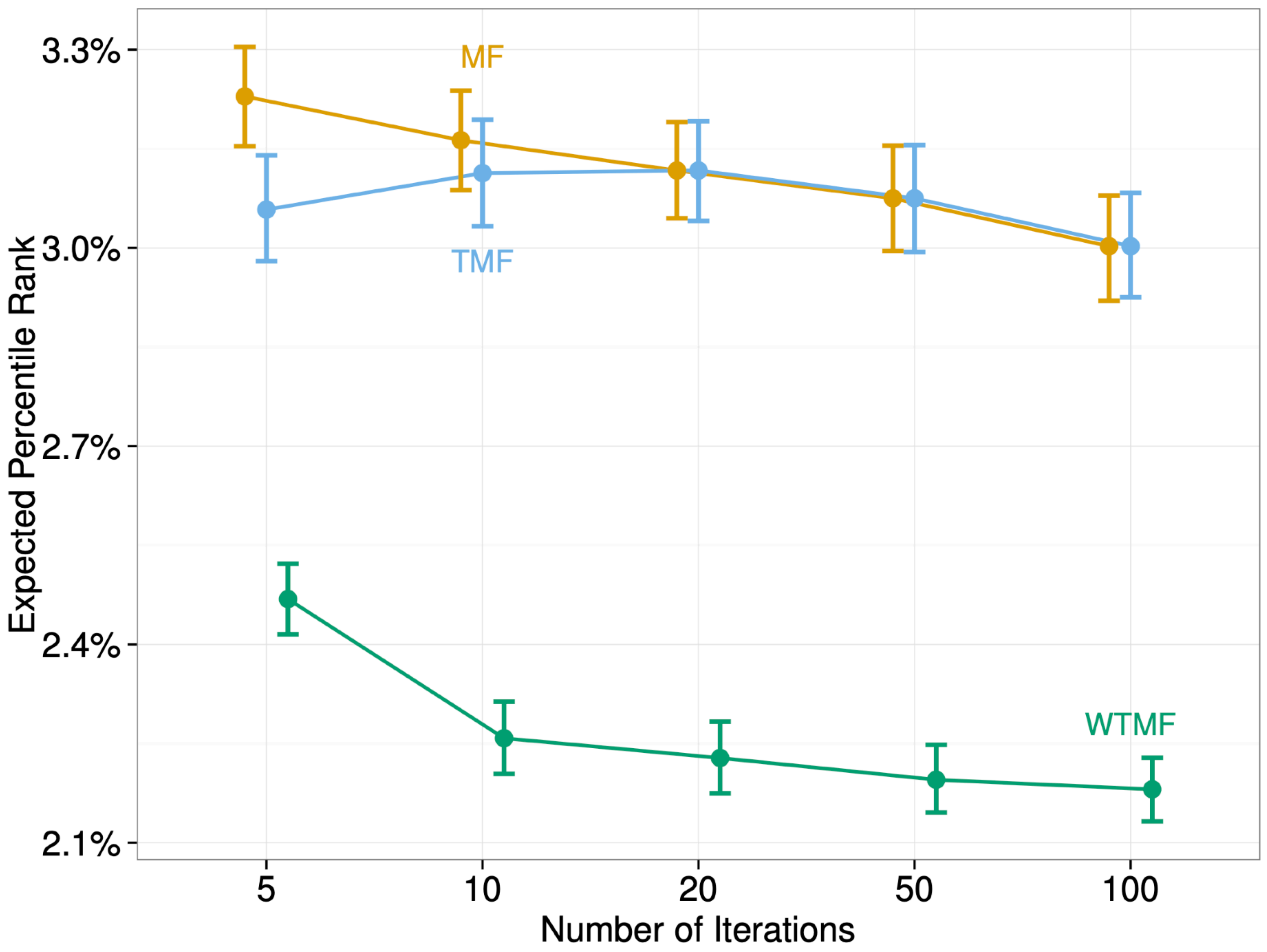
EXPERIMENTAL STUDY

- Dataset: Last.fm listening data + tagging data
  - ~22M listening events, ~3K user, ~71K artists
  - 687K matrix entries; 99.7% sparsity
  - ~500 tags; tags for nearly 100% of artists, only 20% of users
- Evaluation criteria: expected percentile rank
  - Insert artists to evaluate into ranking of randomly selected artists
  - Resulting percentile ranks averaged, weighted by frequency
  - Bootstrap confidence intervals (1K repetitions)

$$\overline{rank} = \frac{\sum_{ua \in R} R_{ua} rank_{ua}}{\sum_{ua \in R} R_{ua}}$$

EVALUATION RESULTS





TAKE AWAYS FROM THIS STUDY

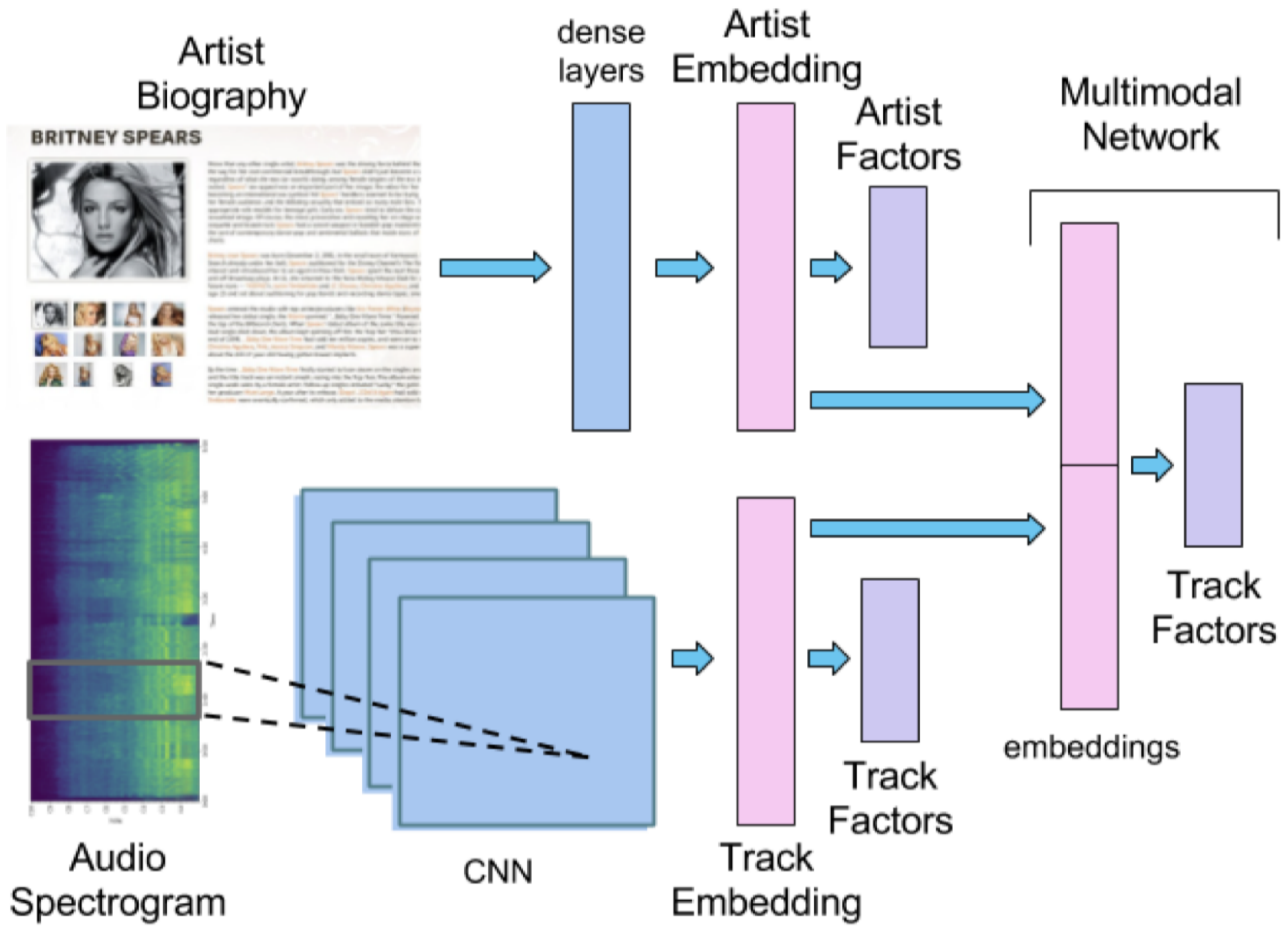
- Matrix factorization is well suited to build models for recommender systems
- Implicit matrix factorization already a strong baseline
- Cold start a central problem
- Additional information from external sources can lead to improved results, even when starting from high level
- Introduces tag cold start problem

OTHER MULTIMODAL APPROACHES

- Incorporation of different sources / complementary information
- Content to handle cold-start problem in CF

- E.g. combining artist biography text embeddings with CNN-trained track audio embeddings

[Oramas et al., 2017] *A Deep Multimodal Approach for Cold-start Music Recommendation.* RecSys DLRS workshop.

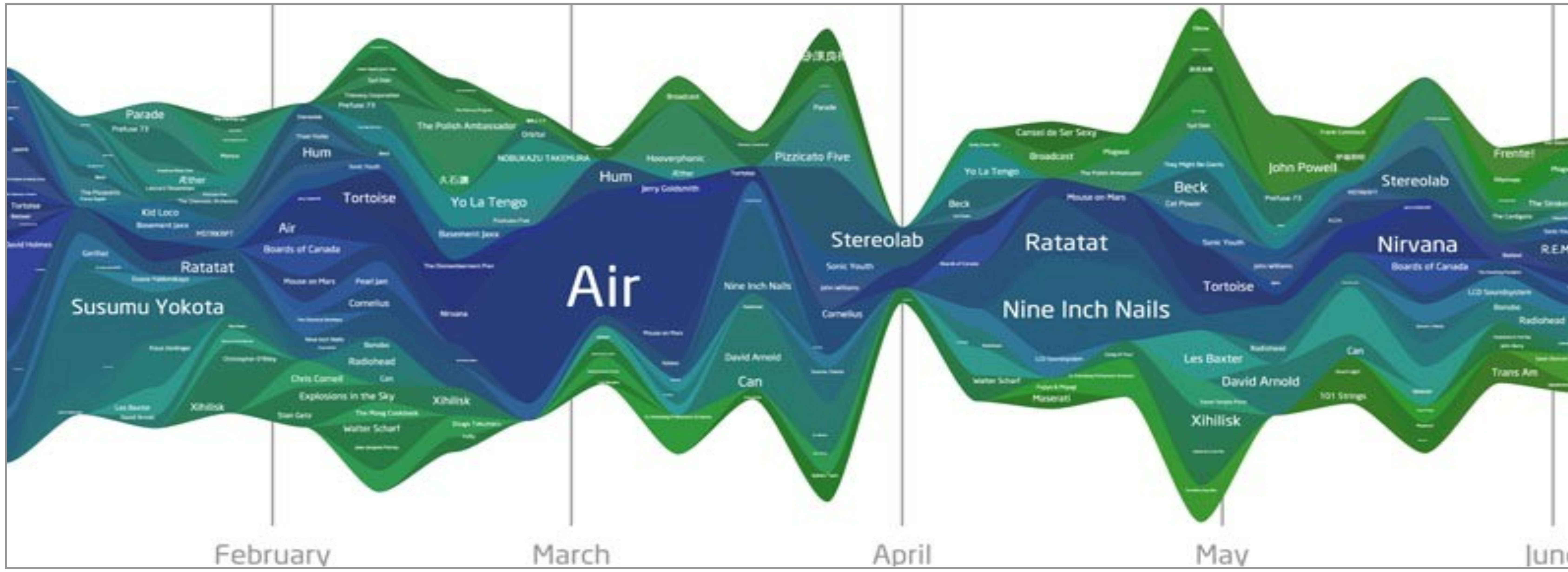


- E.g. fusing deep features from audio and image (album covers) and text

[Oramas et al., 2018] *Multimodal Deep Learning for Music Genre Classification.* TISMIR 1(1).

Sequential Listening and Playlists

WAIT, WHAT ABOUT TIME?



- Well, it's important
- "Music rotation rules" from AM/FM radio programming, e.g.:
  - Popularity categories: "Current", "Recurrent", "Gold"
  - Musical attributes: tempo, male vs. female vocals, danceability, major vs. minor, etc.
  - Sound attributes: synth vs. acoustic, intensity, etc.
  - Artist separation

[Price, 2015]: *After Zane Lowe: Five More Things Internet Radio Should Steal from Broadcast*, NewSlangMedia blog post

A "GOOD" RECOMMENDATION?

What makes a good recommendation:

- Accuracy
- Good balance of:
  - Novelty vs. familiarity / popularity
  - Diversity vs. similarity
- Transparency / Interpretability
- Listener Context

Influential factors:

- Listener
- Musical anchor
- Focus / Intent

It's about recommending a listening experience

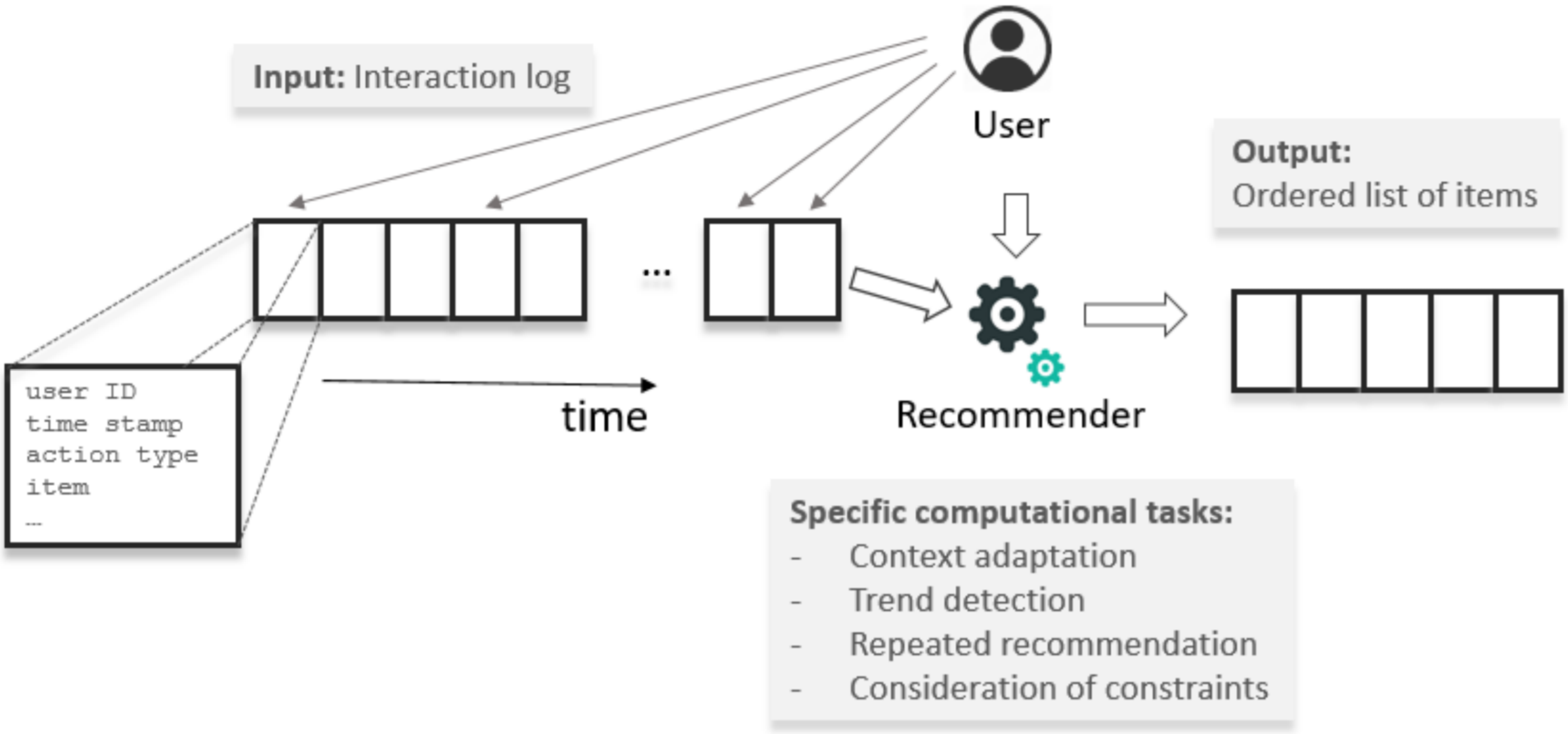
[Celma, 2010] *Music Recommendation and Discovery: The Long Tail, Long Fail, and Long Play in the Digital Music Space*, Springer

[Celma, Lamere, 2011] *Music Recommendation and Discovery Revisited*, ACM Conference on Recommender Systems

[Jannach, Adomavicius, 2016] *Recommendations with a Purpose*, RecSys

[Amatriain, Basilico, 2016] *Past, Present, and Future of Recommender Systems: An Industry Perspective*, RecSys

SEQUENCE-AWARE RECOMMENDATION - OVERVIEW



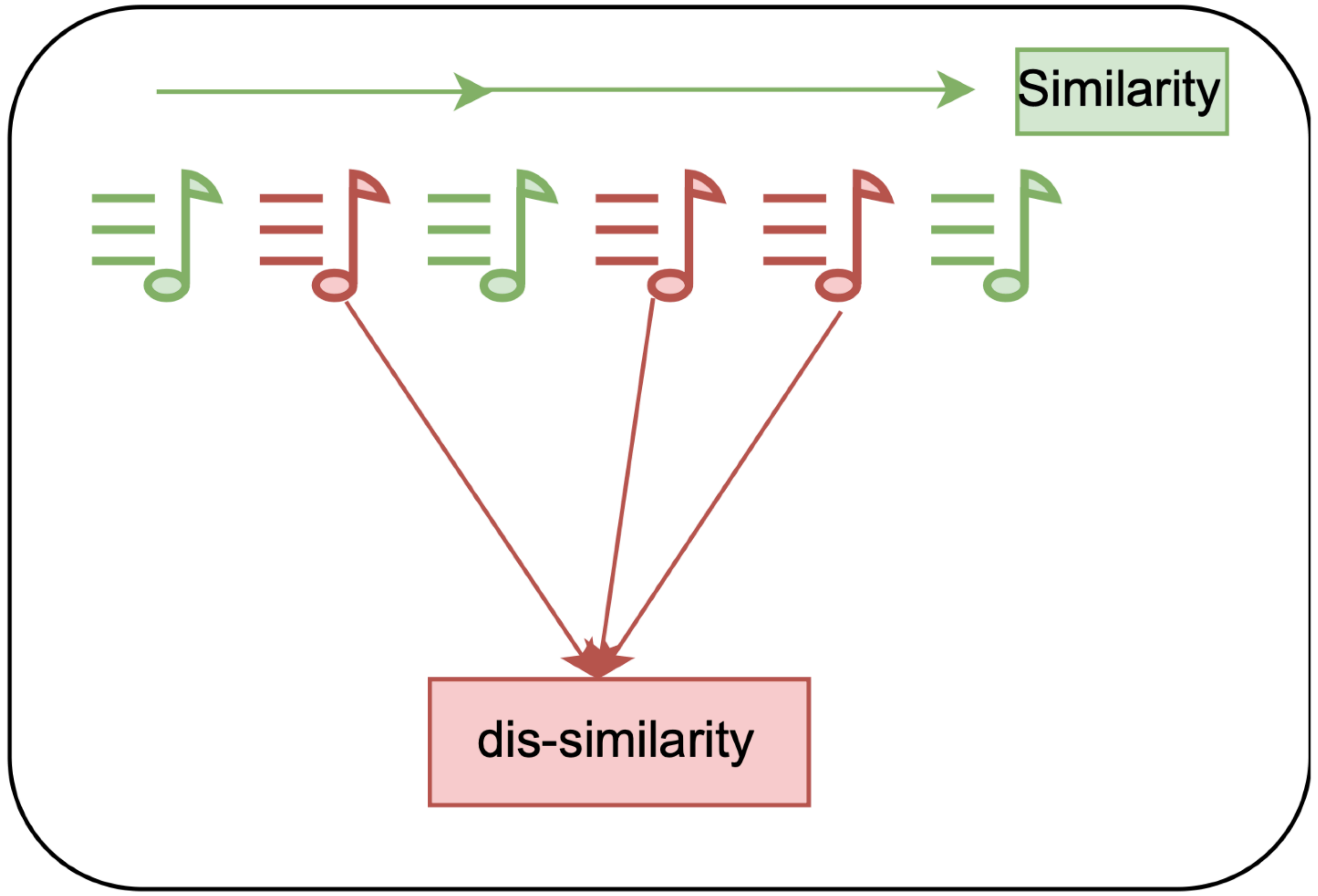
[Quadrana et al., 2018]: *Sequence-Aware Recommender Systems*, https://arxiv.org/abs/1802.08452

SEQUENTIAL MODELS

- In session-based recommendation:
  - user cold-start, (almost) no personalization
  - e.g., news recommendation, shopping basket completion, playlist continuation
- Exemplary Models:
  - **CASER**, **NextItRec**: Convolutional Neural Network-based
  - **GRU4Rec**: Recurrent Neural Network-based
  - **BERT4Rec**: Bi-directional attention mechanism
  - **SASRec**: Uni-directional attention mechanism, longer context

[Seshadri et al., 2024]: *Enhancing Sequential Music Recommendation with Negative Feedback-informed Contrastive Learning*, Proc. RecSys.

CASE STUDY



- Including also negative feedback
  - Skipped tracks, short listens
  - Should not only learn what goes together but what doesn't fit
- Contrastive Learning
  - Maximize similarity of latent factors of similar (positive) examples, while minimizing similarity of dissimilar (negative) examples
  - Changing the learning objective via loss function

[Seshadri et al., 2024]: *Enhancing Sequential Music Recommendation with Negative Feedback-informed Contrastive Learning*, Proc. RecSys.

RESULTS

Data	Metric	SASRec		BERT4Rec		GRU4Rec		CASER		WRMF			BPR		
		orig.	*ours*	*orig.*	*ours*	*orig.*	*ours*	*orig.*	*ours*	*orig.*	*-BL*	*-NR*	*orig.*	*-BL*	*-NR*
MSSD	HR@1	.377	**.410 (9%)**	.204	.355 **(74%)**	.210	.235 **(12%)**	.223	.251 **(13%)**	.203	.207	.211	.208	.213	.216
	HR@5	.615	**.628 (2%)**	.450	.608 **(35%)**	.398	.431 **(8%)**	.412	.455 **(10%)**	.400	.406	.411	.406	.413	.421
	HR@10	.696	**.706 (1%)**	.553	.693 **(25%)**	.491	.534 **(9%)**	.518	.530 **(2%)**	.488	.498	.505	.502	.510	.517
	HR@20	.767	**.774 (1%)**	.648	.767 **(18%)**	.600	.627 **(5%)**	.616	.649 **(5%)**	.597	.603	.608	.605	.609	.612
	MAP@10	.397	**.417 (5%)**	.185	.369 **(99%)**	.176	.193 **(10%)**	.195	.221 **(13%)**	.149	.156	.164	.160	.168	.173
LFM-2B	HR@1	.190	**.221 (16%)**	.101	.117 **(16%)**	.096	.102 **(6%)**	.102	.112 **(10%)**	.097	.098	.102	.098	.105	.107
	HR@5	.371	**.400 (8%)**	.227	.248 **(9%)**	.203	.221 **(9%)**	.197	.208 **(6%)**	.138	.142	.147	.143	.152	.169
	HR@10	.452	**.477 (6%)**	.292	.320 **(10%)**	.273	.291 **(7%)**	.281	.302 **(7%)**	.269	.273	.279	.276	.286	.294
	HR@20	.532	**.553 (4%)**	.366	.394 **(8%)**	.311	.342 **(10%)**	.326	.354 **(9%)**	.305	.311	.316	.314	.324	.348
	MAP@10	.188	**.219 (16%)**	.098	.110 **(12%)**	.078	.085 **(9%)**	.081	.088 **(9%)**	.062	.065	.067	.066	.072	.078
LFM-1K	HR@1	.152	**.181 (19%)**	.069	.086 **(25%)**	.048	.059 **(23%)**	.052	.071 **(37%)**	.042	.044	.047	.043	.050	.052
	HR@5	.301	**.330 (10%)**	.207	.230 **(11%)**	.182	.200 **(10%)**	.188	.198 **(5%)**	.139	.146	.177	.150	.153	.159
	HR@10	.392	**.421 (7%)**	.299	.320 **(7%)**	.261	.289 **(11%)**	.269	.293 **(9%)**	.270	.285	.294	.292	.298	.301
	HR@20	.478	**.491 (3%)**	.413	.433 **(5%)**	.388	.397 **(2%)**	.390	.404 **(4%)**	.346	.368	.375	.369	.374	.376
	MAP@10	.092	**.107 (16%)**	.049	.064 **(31%)**	.034	.042 **(24%)**	.037	.040 **(8%)**	.028	.031	.034	.032	.036	.038

Table 1: *Hit Ratio @ [1, 5, 10, 20]* and *Mean Average Precision @ 10* for the sequential models (SASRec, BERT4Rec, GRU4Rec, CASER) and the non-sequential baselines (WRMF, BPR) on the three datasets. Non *"orig."* models incorporate negative feedback. They are compared to their *"orig."* baselines which do not model negative feedback. Numbers in parentheses show the relative increase in the percentage of the approach over the respective baseline; bold entries mark the better performing approach between the baseline and negative feedback-informed approach. The overall best performance is also highlighted in bold (i.e., SASRec-ours).

MORE RESULTS

- Looking at the skipped tracks in SASRec and BERT4Rec
- Correspondingly, they should be ranked lower…
- Generally, they are

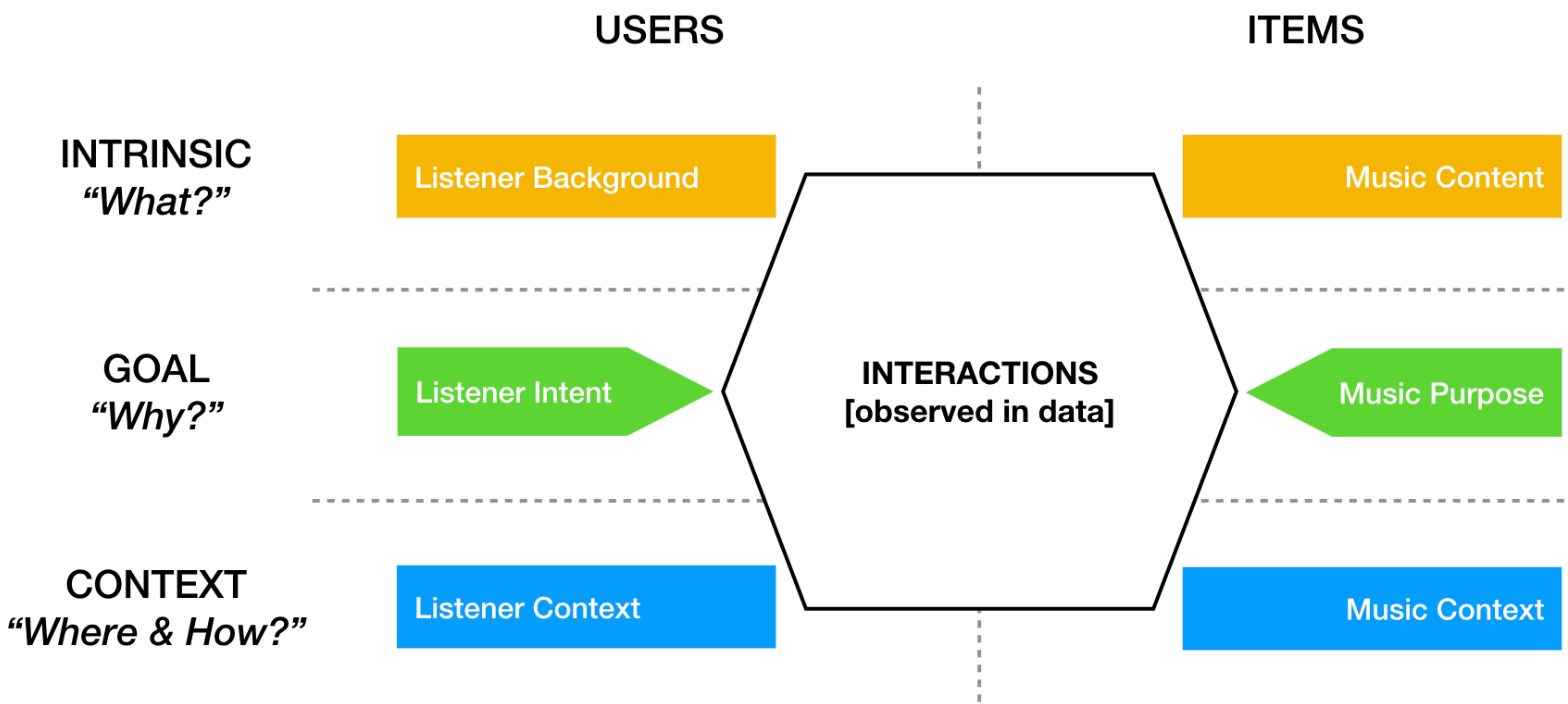
		SASRec		BERT4Rec	
Metric		*orig.*	*ours*	*orig.*	*ours*
MRR@10	**MSSD**	.960	.950 **(-1%)**	.840	.950 **(13%)**
MRR@10	**LFM-2B**	.969	.911 **(-6%)**	.953	.950 **(-0.3%)**
MRR@10	**LFM-1K**	.731	.731 **(0%)**	.540	.460 **(-15%)**

Table 2: *Mean Reciprocal Rank @ 10* on skip targets for the best sequential models (SASRec, BERT4Rec) on the three datasets. Lower values are better.

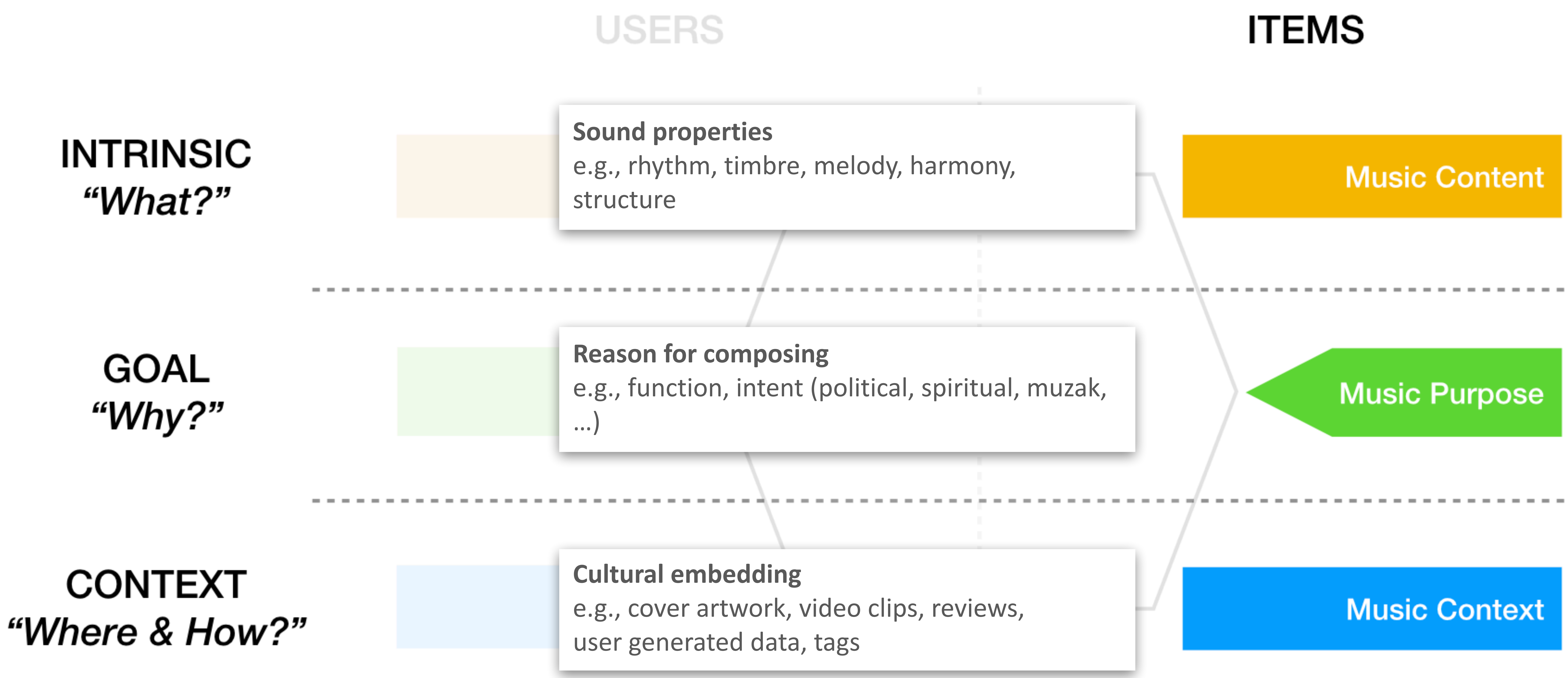
- But on MSSD, they are also ranked higher, i.e., positive and negative tracks get more likely recommended
- Hypothesis: BERT4Rec with negative feedback closer to Spotify recommendations

It’s even more complex

TWO FACTORS…


USERS
ITEMS
INTRINSIC
"What?"
Listener Background
Music Content
GOAL
"Why?"
Listener Intent
INTERACTIONS
[observed in data]
Music Purpose
CONTEXT
"Where & How?"
Listener Context
Music Context

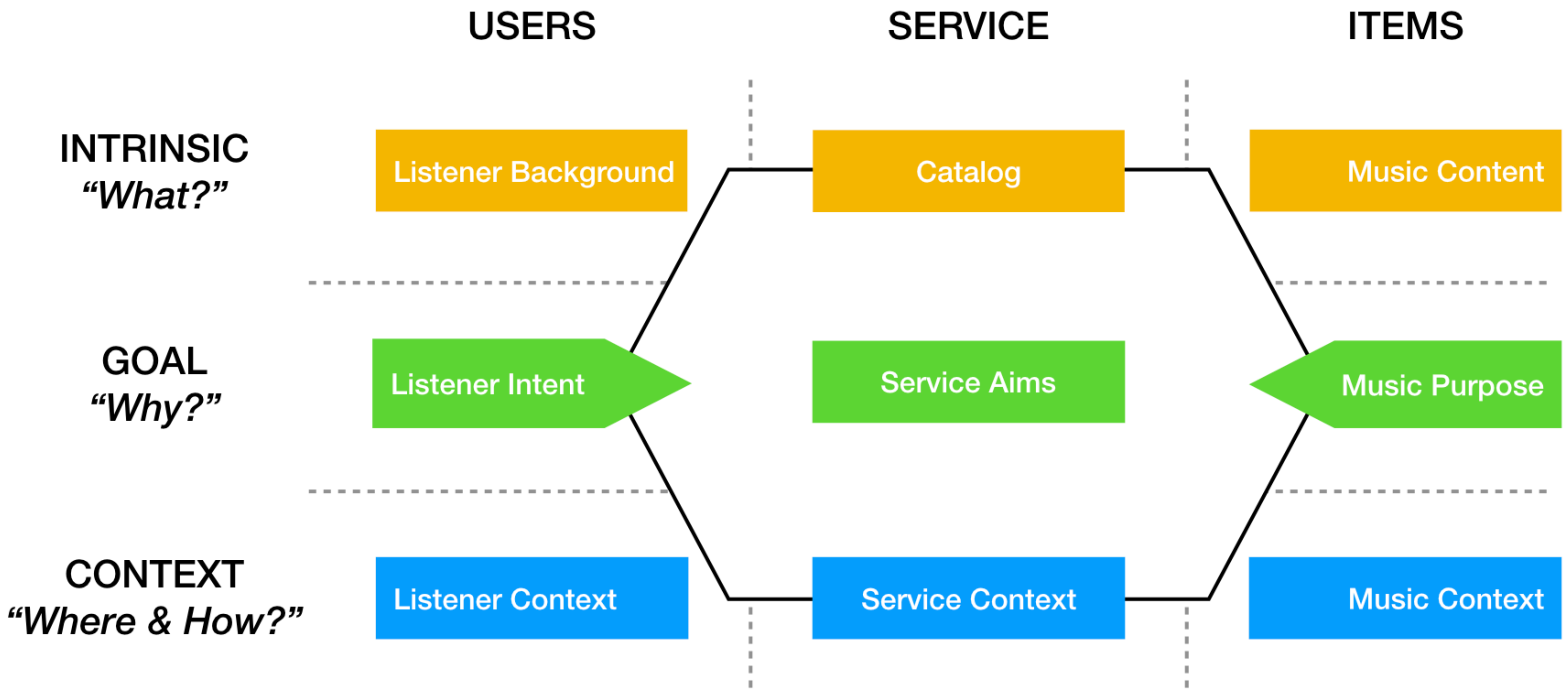
FACTORS HIDDEN IN THE DATA



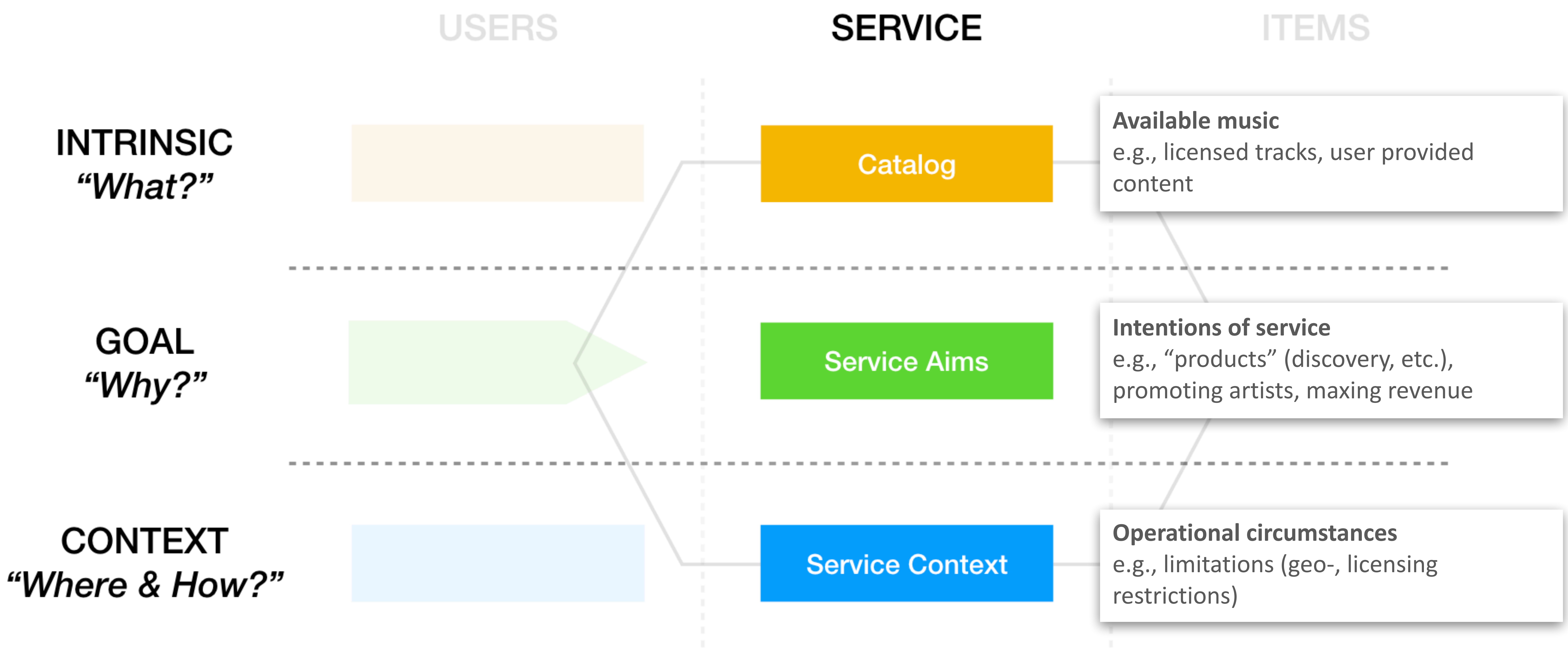
FACTORS HIDDEN IN THE DATA

	USERS		ITEMS
INTRINSIC *"What?"*	Listener Background	**Personal characteristics** e.g., music preference, experience, musical training, demographics	
GOAL *"Why?"*	Listener Intent	**Reason to select music/listen** e.g., self-regulation, emotion evocation, demonstration	
CONTEXT *"Where & How?"*	Listener Context	**Situation of listener** e.g., mood, activity, social context, spatio-temporal context	

FACTORING THE SERVICE INTO THE PICTURE



FACTORS HIDDEN IN THE DATA



THE SERVICE IMPACTS DATA COLLECTED

Catalog

- Which content is provided/recommended?
- e.g. Soundcloud recommends different content than Spotify

Service Aims

- Why is this service in place? What is the purpose/market niche?
- What are the identified use cases? (Radio? Exclusives? Quality?)
- Do they push their own content (cf. Netflix)?

Service Context

- How do catalog and service aims depend on context?
- Are there licensing issues/restrictions in particular countries?
- Is the service context-aware? (e.g. app vs desktop/browser)

THE SERVICE PLAYS A CENTRAL ROLE

- Data from one service not generalizable to others

SOUNDCLOUD ≠ bandcamp ≠ Spotify ≠ DEEZER ≠ pandora ≠ …

- Particularly for niche markets and different parts of the world

≠ IDAGIO ≠ qobuz ≠ KKbox ≠ SUPERPLAYER ≠ simfy africa ≠ …

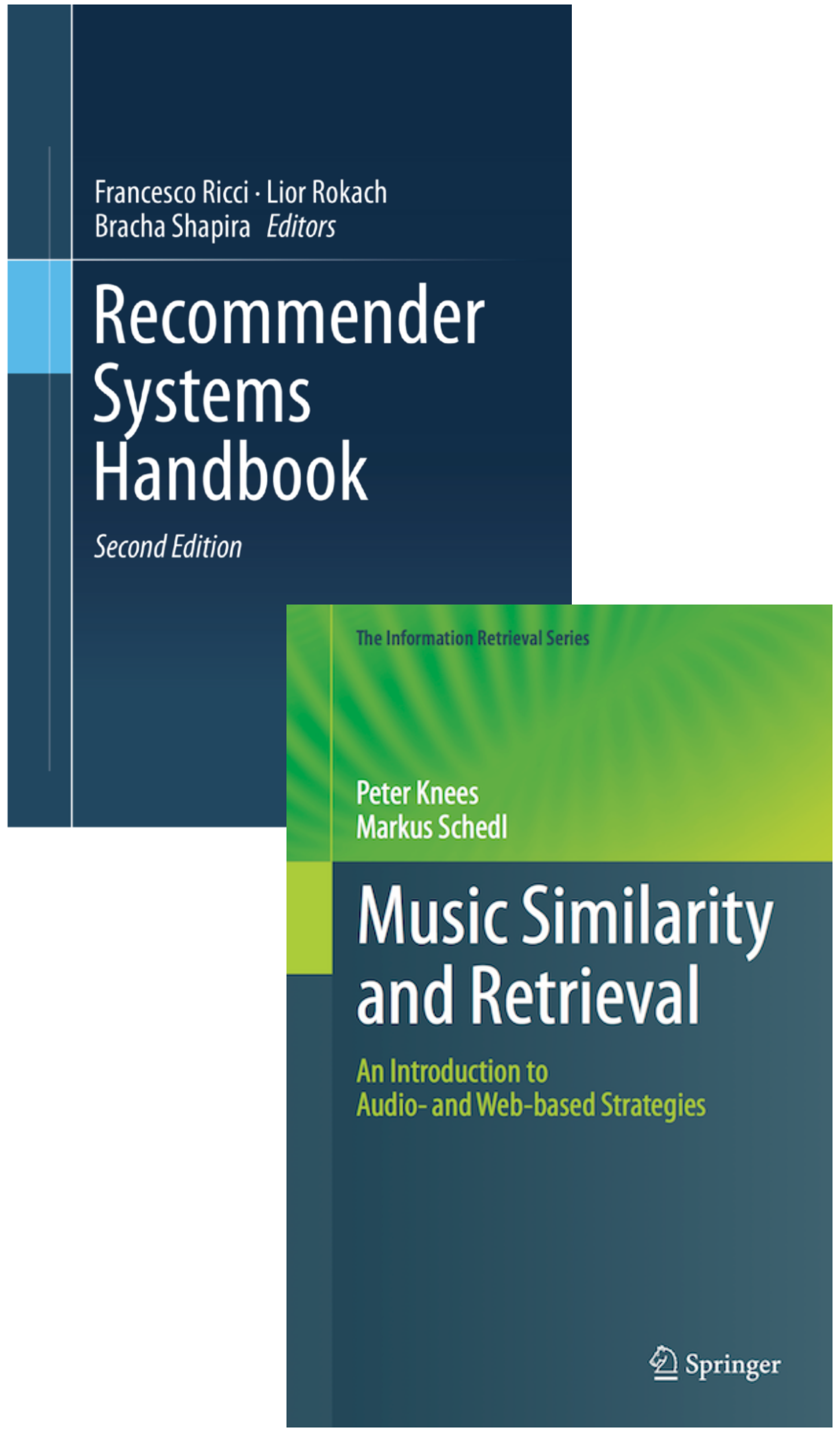
- **Service influences listening behavior**
- But there is no listening "in the wild"

cf. [Porcaro & Gomez, 2019] *20 years of Playlists: a Statistical Analysis on Popularity and Diversity",* ISMIR.

CONCLUSIONS

- Various music listening use cases, stakeholders, data
  - similarity-based retrieval methods
  - model-based recommendation techniques
  - exploitation of multi-modal data
- Real-world data not only noisy but potentially biased, exhibits interests of various stakeholders
- Academic research does not have the data of industry, but could focus on uncovering commercial biases and interests

FURTHER READING



- Recommender Systems Handbook (3rd ed.)
  Chapter: Music Recommendation Systems: Techniques, Use Cases, and Challenges
  by M. Schedl, P. Knees, B. McFee, D. Bogdanov. Springer, 2022.
- **Music Similarity and Retrieval: An Introduction to Audio and Web-based Strategies**
  by P. Knees and M. Schedl. Springer, 2016.
- **Tutorial: Overview and New Challenges of Music Recommendation Research in 2018**
  by M. Schedl, P. Knees, F. Gouyon. ISMIR'18.
  https://www.slideshare.net/FabienGouyon/music-recommendation-2018-116102609