Universitat Pompeu Fabra Barcelona

MTG Music Technology Group

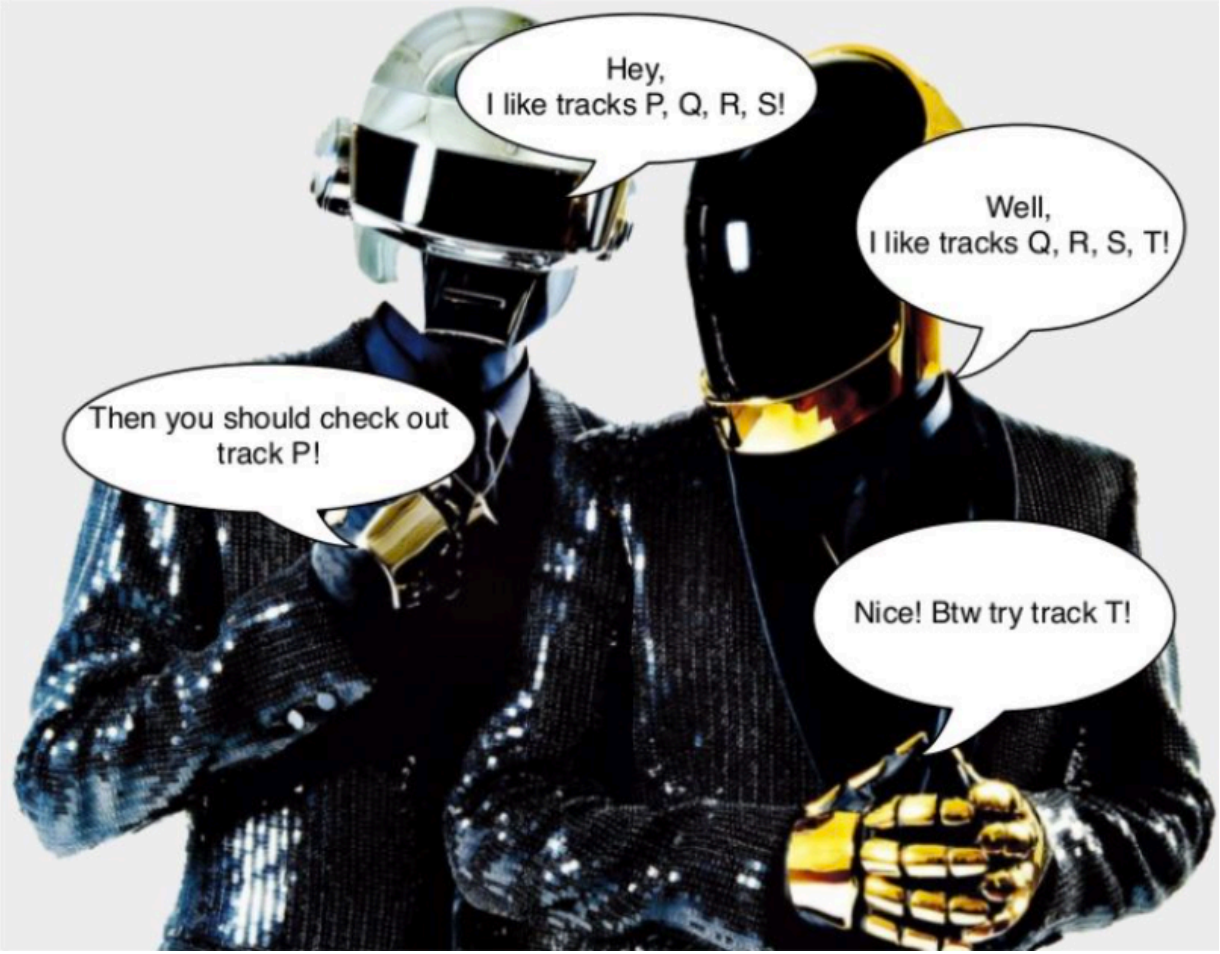
Knowledge Extraction and Feature Learning for Music Recommendation in the Long Tail

Sergio Oramas

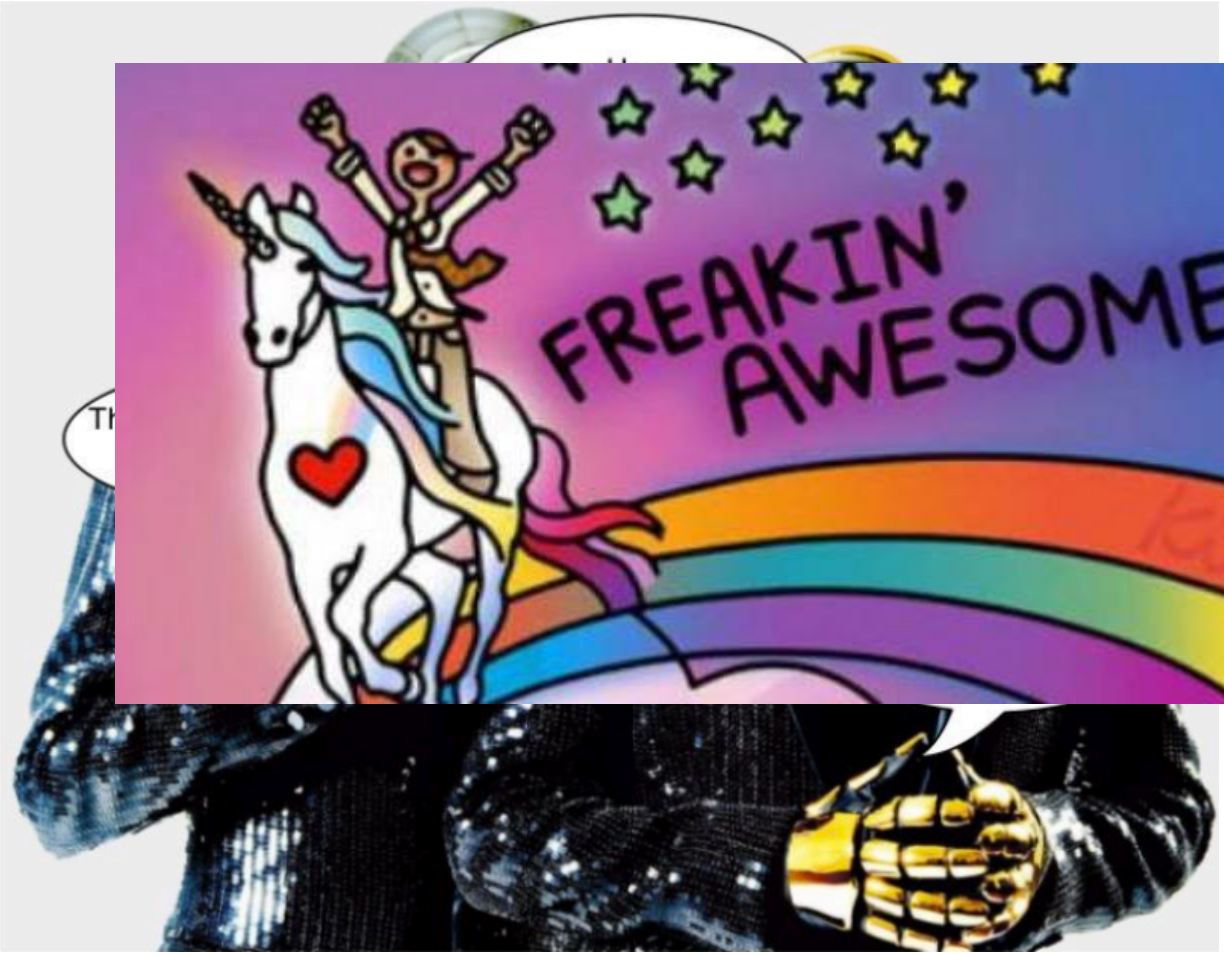
Collaborators: Oriol Nieto, Vito Claudio Ostuni. Tommaso Di Noia, Mohamed Sordo, Xavier Serra

Politecnico di Bari

Collaborative Filtering

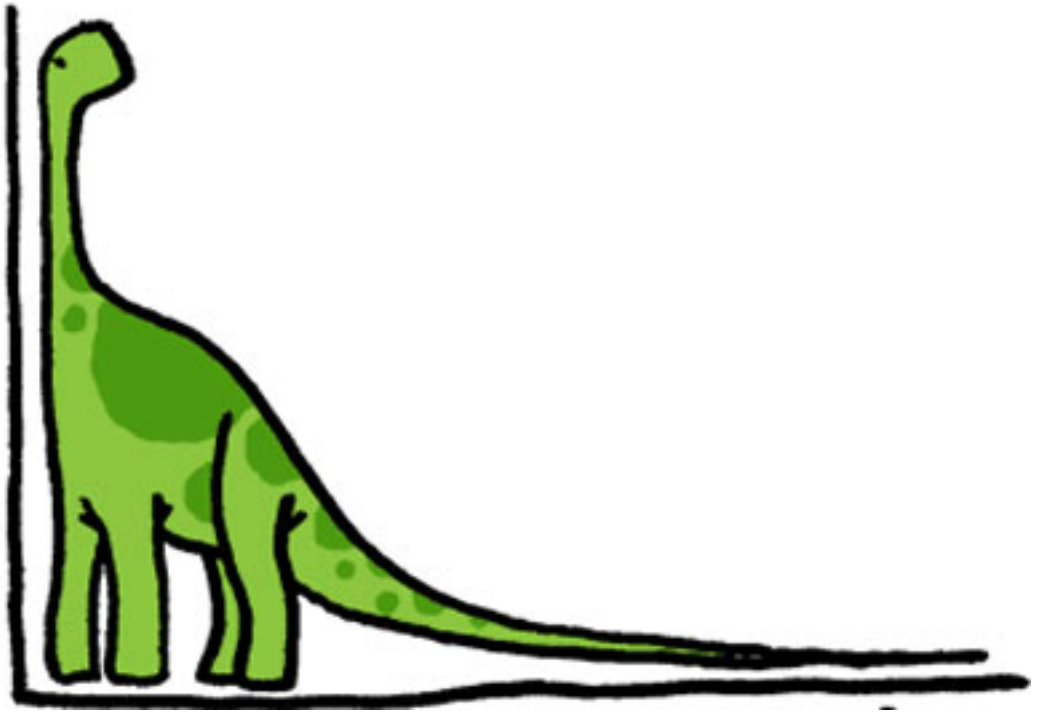


Collaborative Filtering

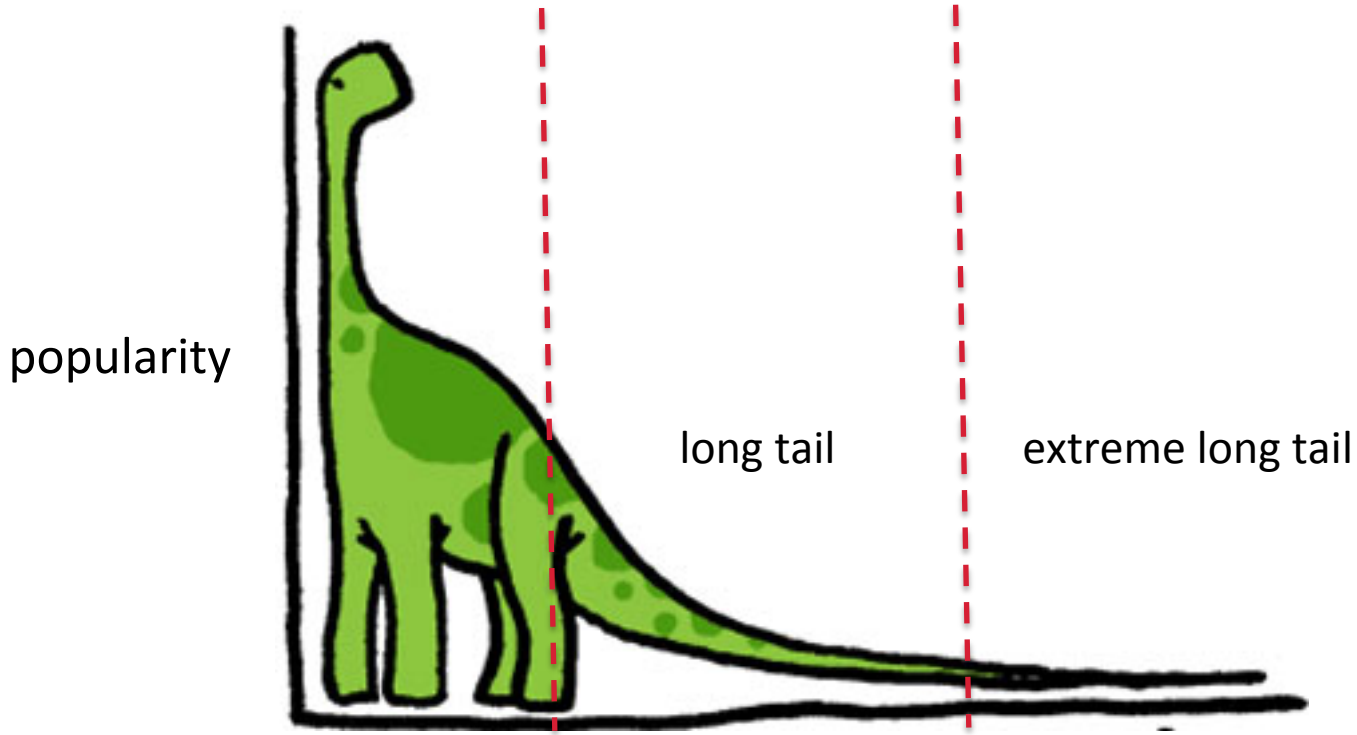


Long Tail

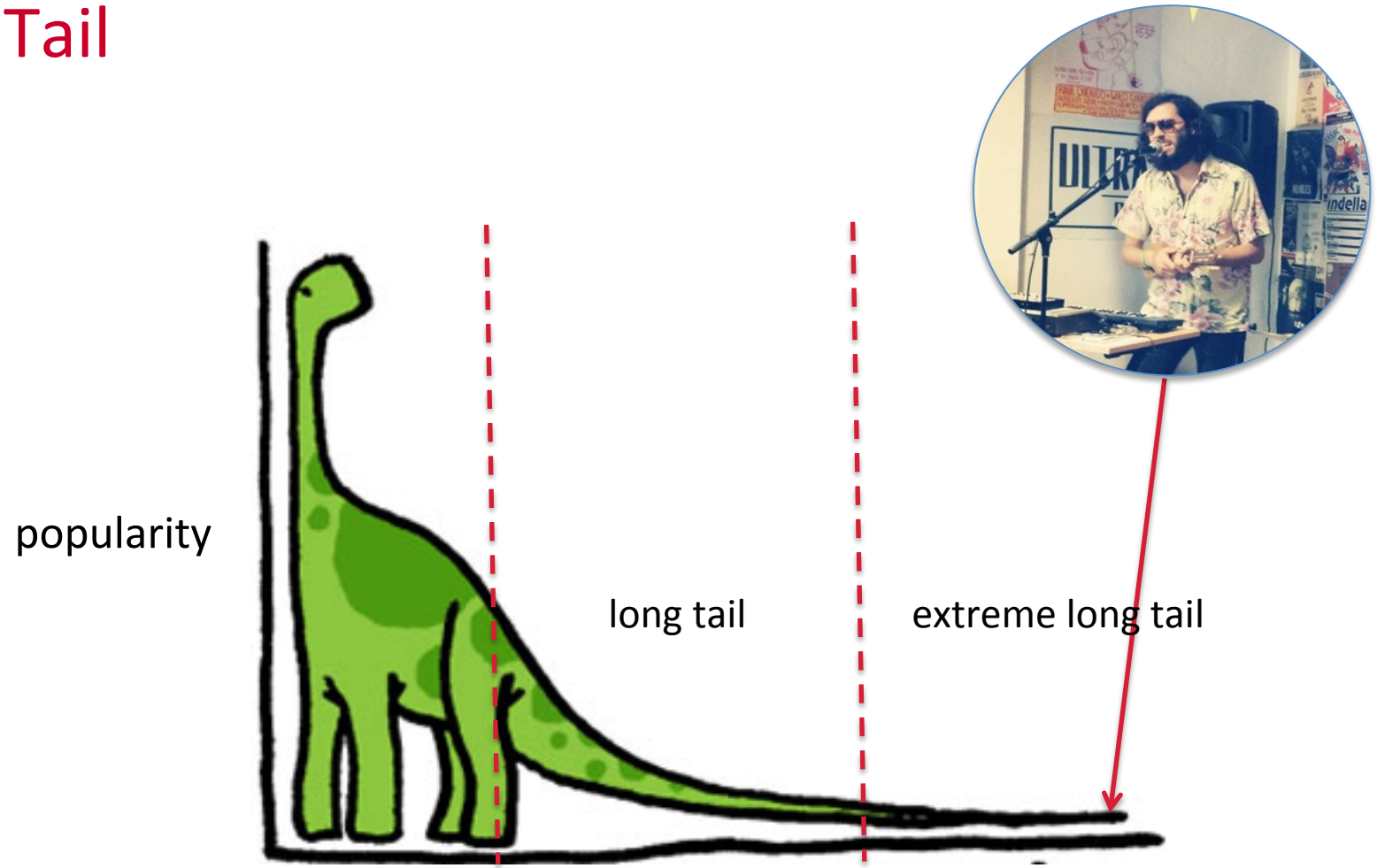
popularity



Long Tail



Long Tail



Cold-start

Cold-start Problem: New Music Releases

- Online Streaming Services daily ingest new releases from:
  - Existing Artists
  - **Novel Artists**

Cold-start Problem: Ingest Items in Catalog

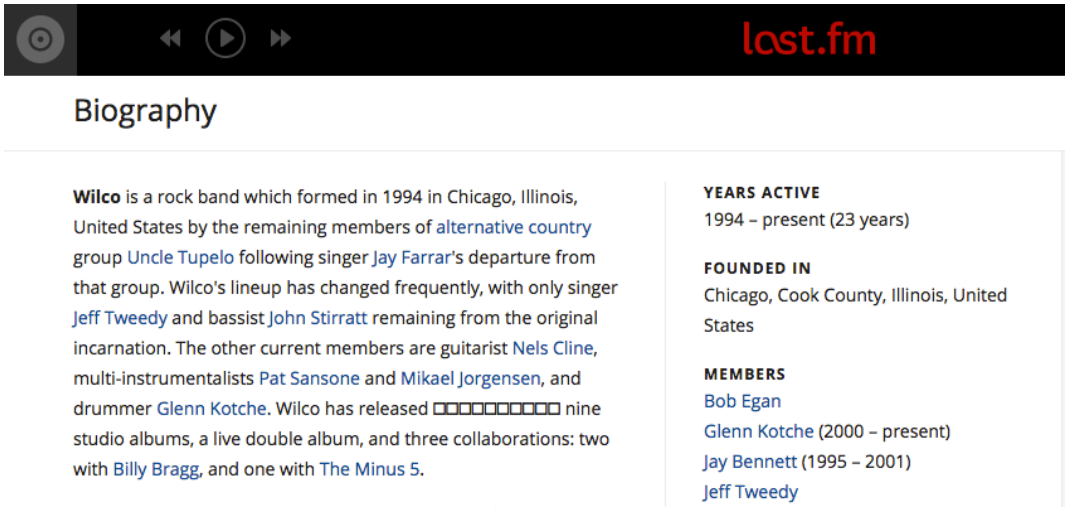
- The Pandora case:
  - New on-demand service
  - From millions to tens of millions

Content features

- **Tags**
  - Widely used
  - Need a curation process (experts or a crowd)
  - Rich contextual info: genres, places, relations, dates

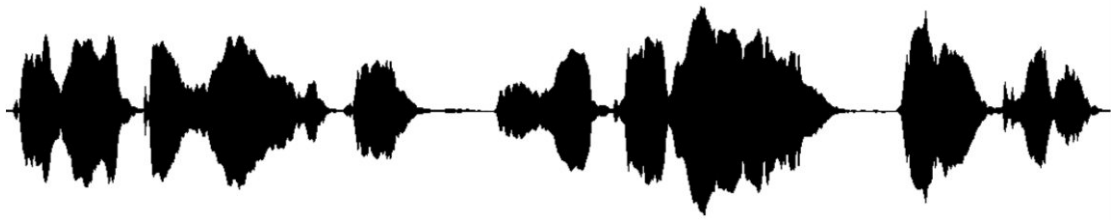
Content features

- **Artist biographies or press releases**
  - Barely used
  - Often self-made
  - Rich contextual info: genres, places, relations, dates
  - Noisy

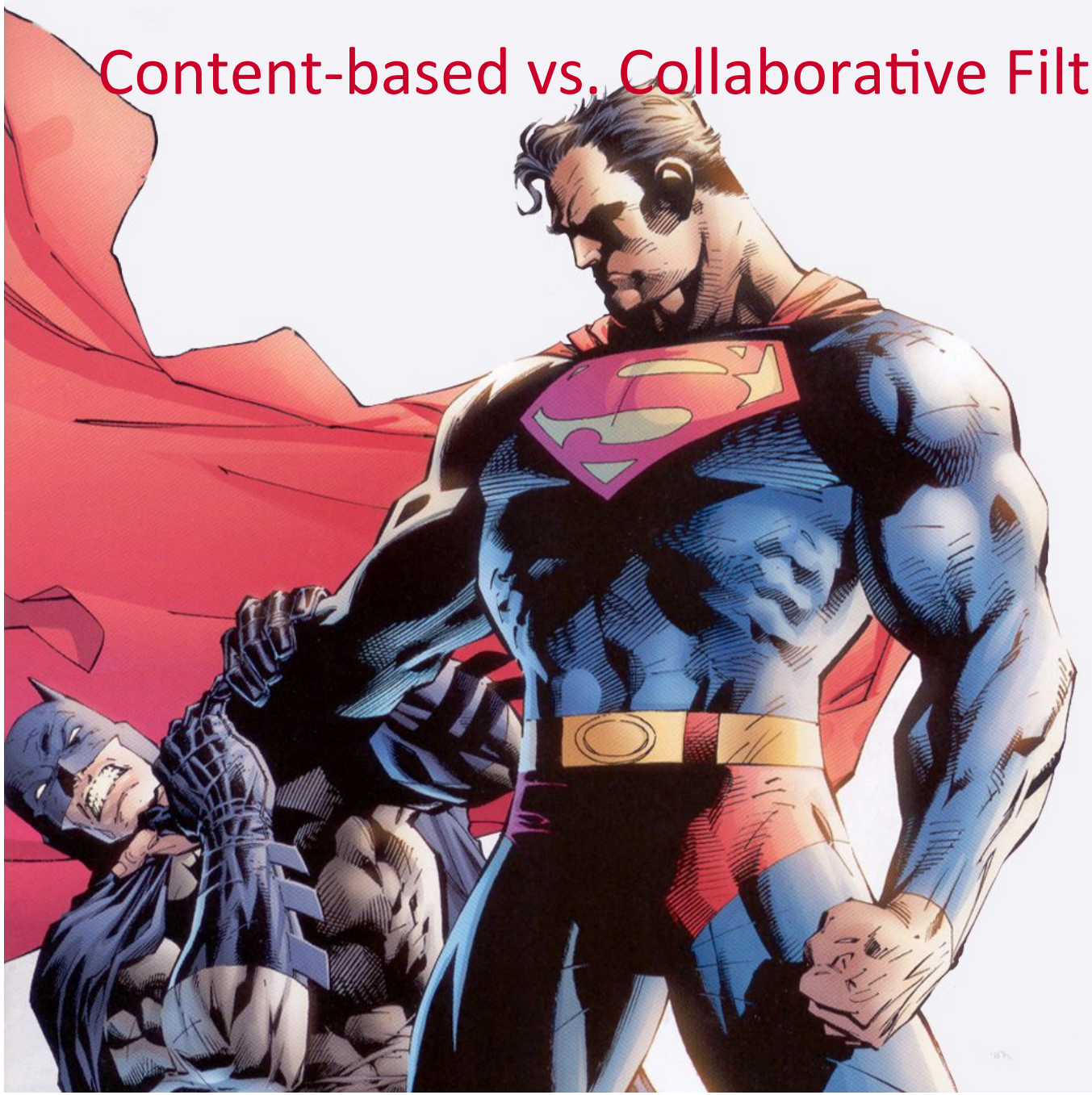


Content features

- **Audio**
  - Slightly used
  - Always available
  - Rich musical info: genres, instruments, rhythms, timber
  - Semantic gap

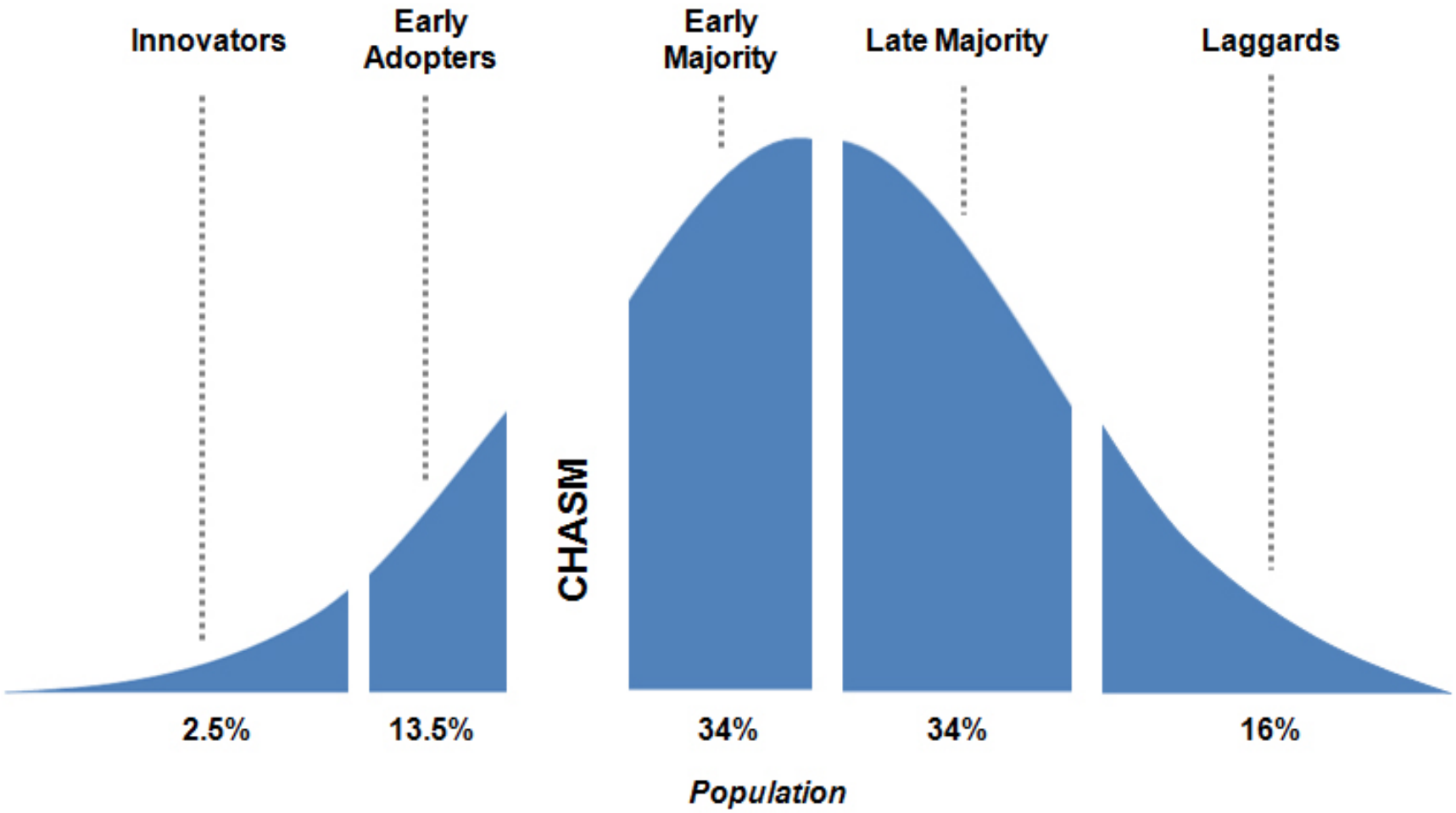


Content-based vs. Collaborative Filtering

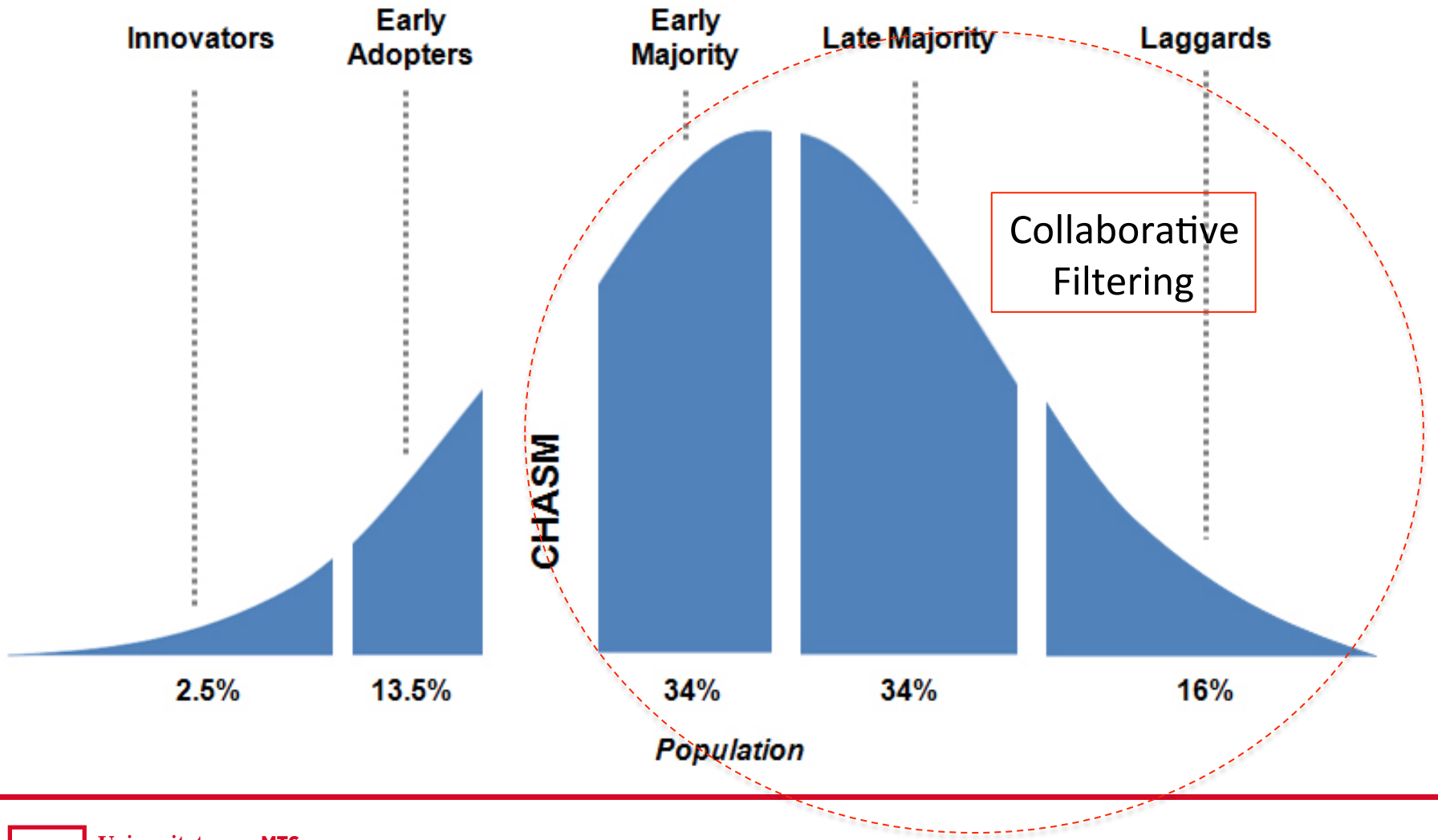


Hybrid approaches

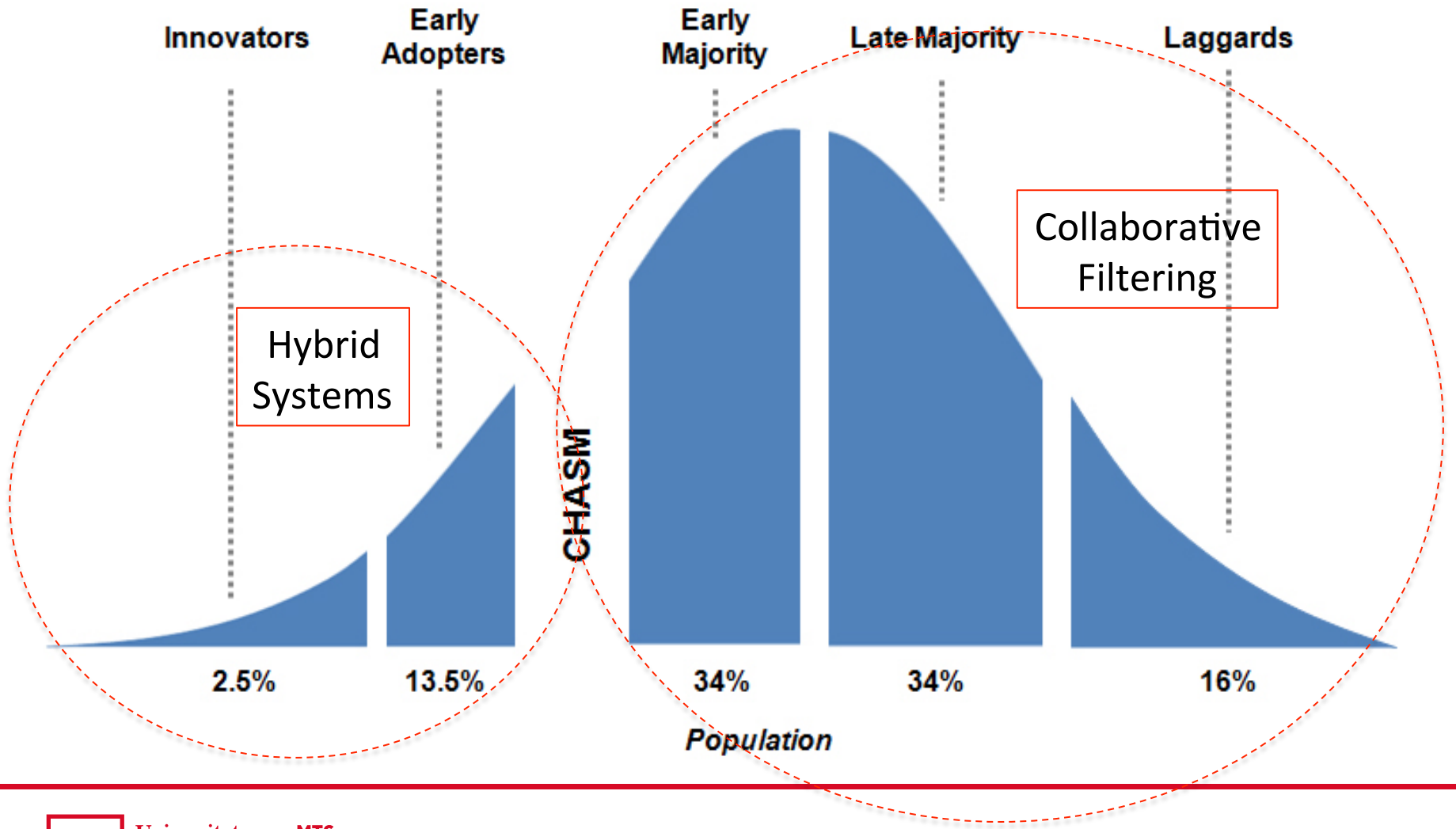
Users


Innovators
Early Adopters
Early Majority
Late Majority
Laggards
CHASM
2.5%
13.5%
34%
34%
16%
Population

Users


Innovators
Early Adopters
Early Majority
Late Majority
Laggards
CHASM
Collaborative Filtering
2.5%
13.5%
34%
34%
16%
Population

Users


Innovators
Early Adopters
Early Majority
Late Majority
Laggards
Hybrid Systems
Collaborative Filtering
CHASM
2.5%
13.5%
34%
34%
16%
Population

Knowledge Extraction

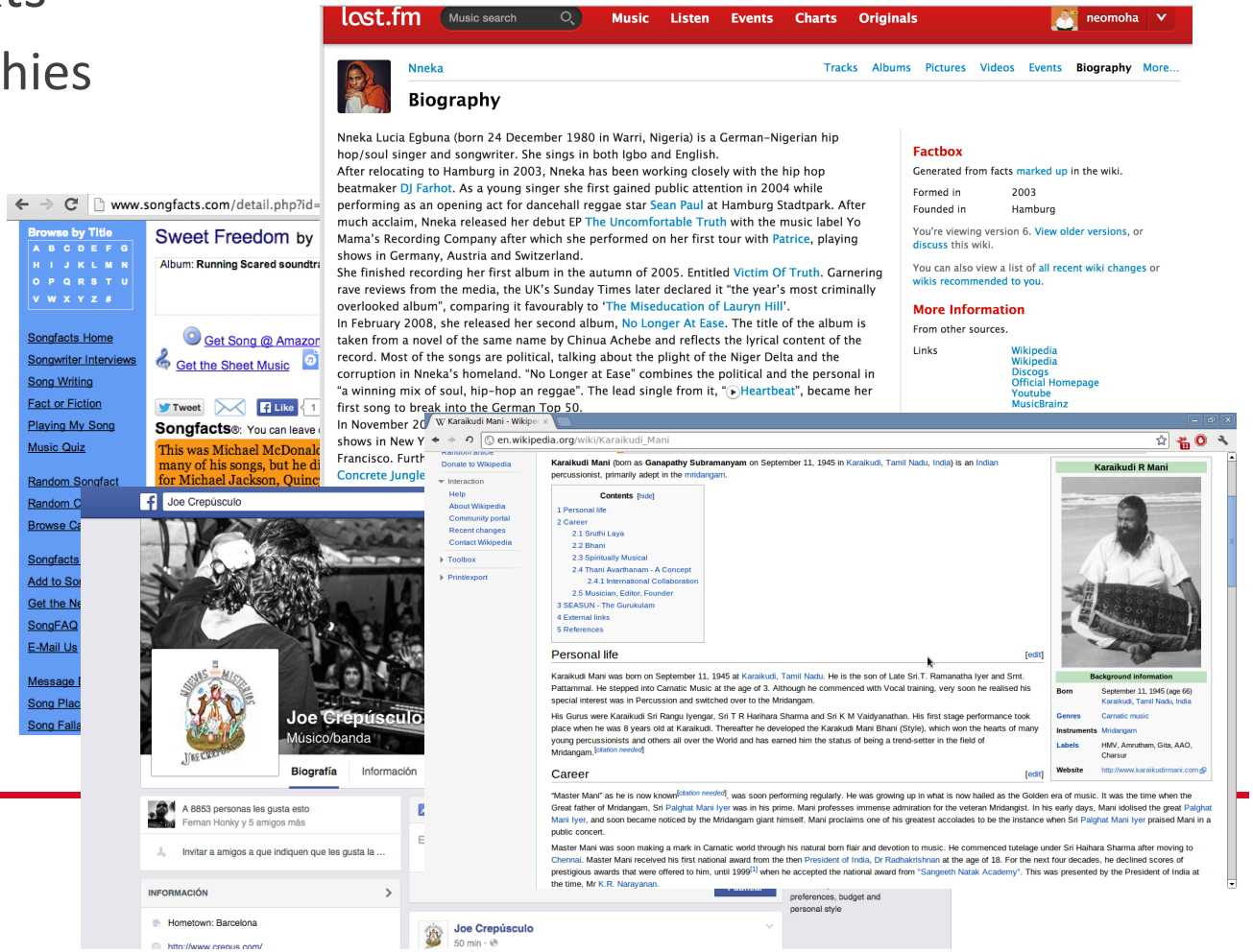
upf. Universitat Pompeu Fabra Barcelona
MTG Music Technology Group

Structured Data Sources

- Structured information about music is incomplete
  - Only popular artists and western music
  - Only editorial and some biographical information

Unstructured Data Sources

- Huge amount of music information remains implicit in unstructured texts
  - Artists biographies
  - Articles
  - Reviews
  - Web pages
  - User's posts



Entity Linking

Identify and link entity mentions in text to a Knowledge Base

Entity Linking

Songs in Sky Blue Sky were written by Wilco frontman , Jeff Tweedy.

Entity Linking

Songs in Sky Blue Sky were written by Wilco frontman , Jeff Tweedy.

Entity Linking

https://dbpedia.org/resource/Wilco

Songs in Sky Blue Sky were written by Wilco frontman , Jeff Tweedy.

https://dbpedia.org/resource/Sky_Blue_Sky

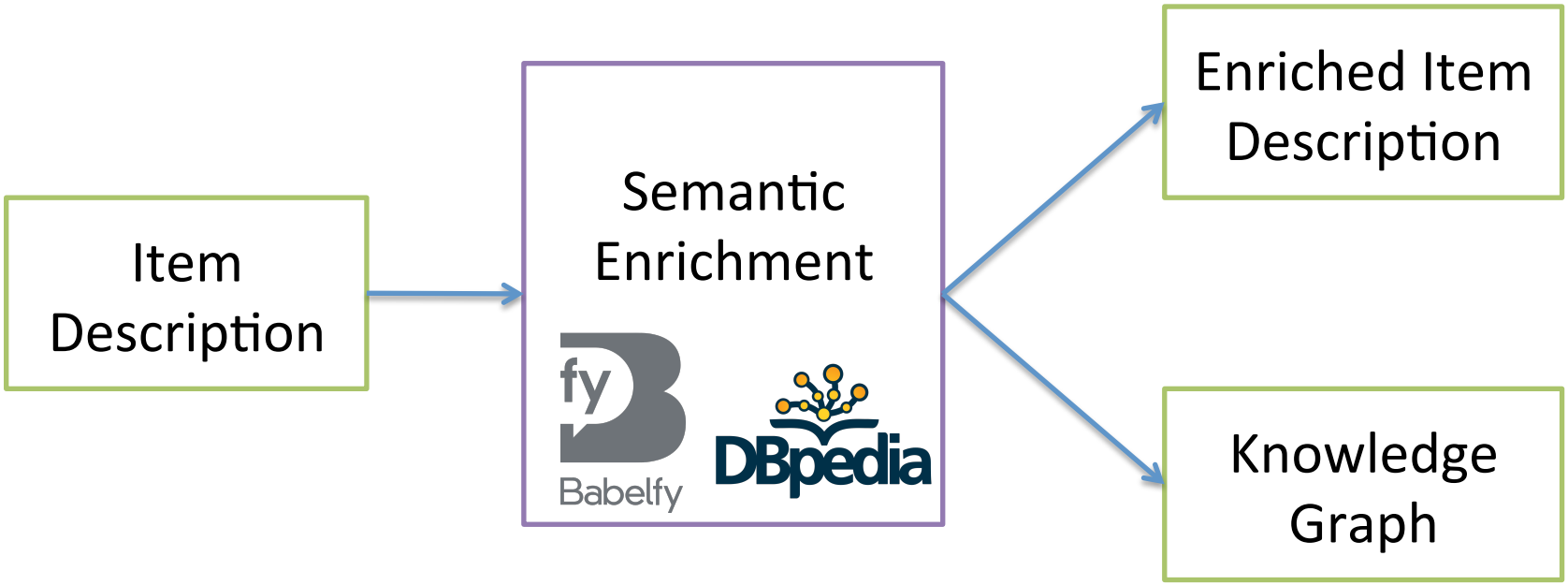
http://dbpedia.org/resource/Jeff_Tweedy

Semantic Information

- Wilco

dbo:genre	▪ dbr:Art_rock ▪ dbr:Experimental_rock ▪ dbr:Alternative_country ▪ dbr:Alternative_rock ▪ dbr:Indie_rock
dbo:hometown	▪ dbr:United_States ▪ dbr:Chicago ▪ dbr:Illinois
dbo:recordLabel	▪ dbr:Reprise_Records ▪ dbr:Nonesuch_Records ▪ dbr:Dbpm_records

Semantic Enrichment via Entity Linking



Enriched Item Description

Dies ist ein Blindtext. An ihm lässt sich vieles über die Schrift ablesen, in der er gesetzt ist. Auf den ersten Blick wird der Grauwert der Schriftfläche sichtbar. Dann kann man prüfen, wie gut die Schrift zu lesen ist und wie sie auf den Leser wirkt. Dies ist ein Blindtext. An ihm lässt sich vieles über die Schrift ablesen, in der er gesetzt ist. Auf den ersten Blick wird der Grauwert der Schriftfläche sichtbar. Dann kann man prüfen, wie gut die Schrift zu lesen ist und wie sie auf den Leser wirkt.

+

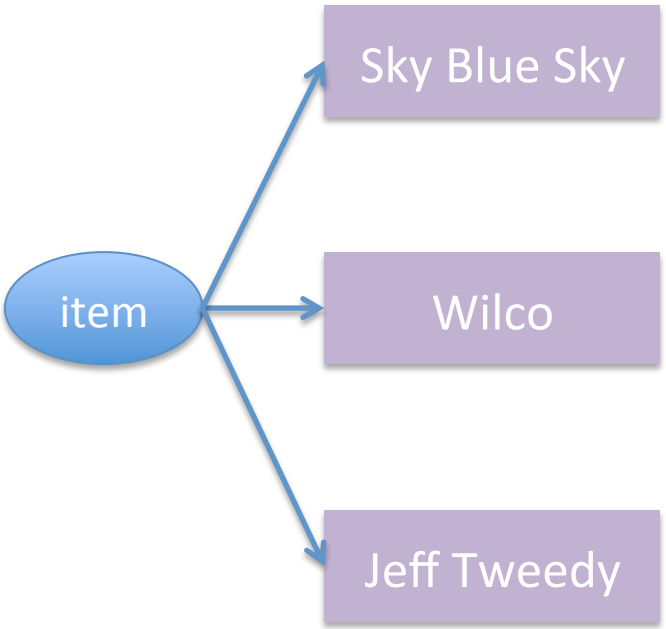
dbr:Art_rock dbr:Experimental_rock dbr:Alternative_country dbr:Alternative_rock dbr:Indie_rock dbr:United_States dbr:Chicago dbr:Illinois dbr:Reprise_Records dbr:Nonesuch_Records dbr:Dbpm_records

Biography text

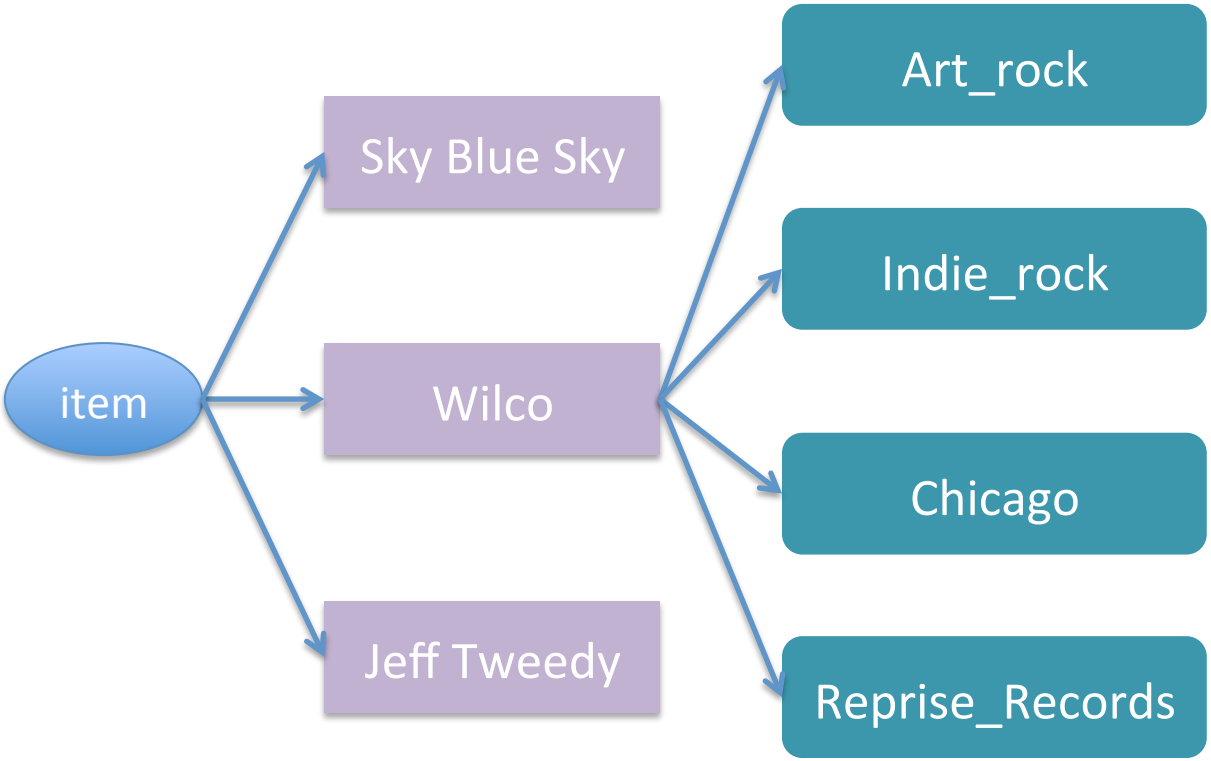
Semantic data

VSM
tf-idf

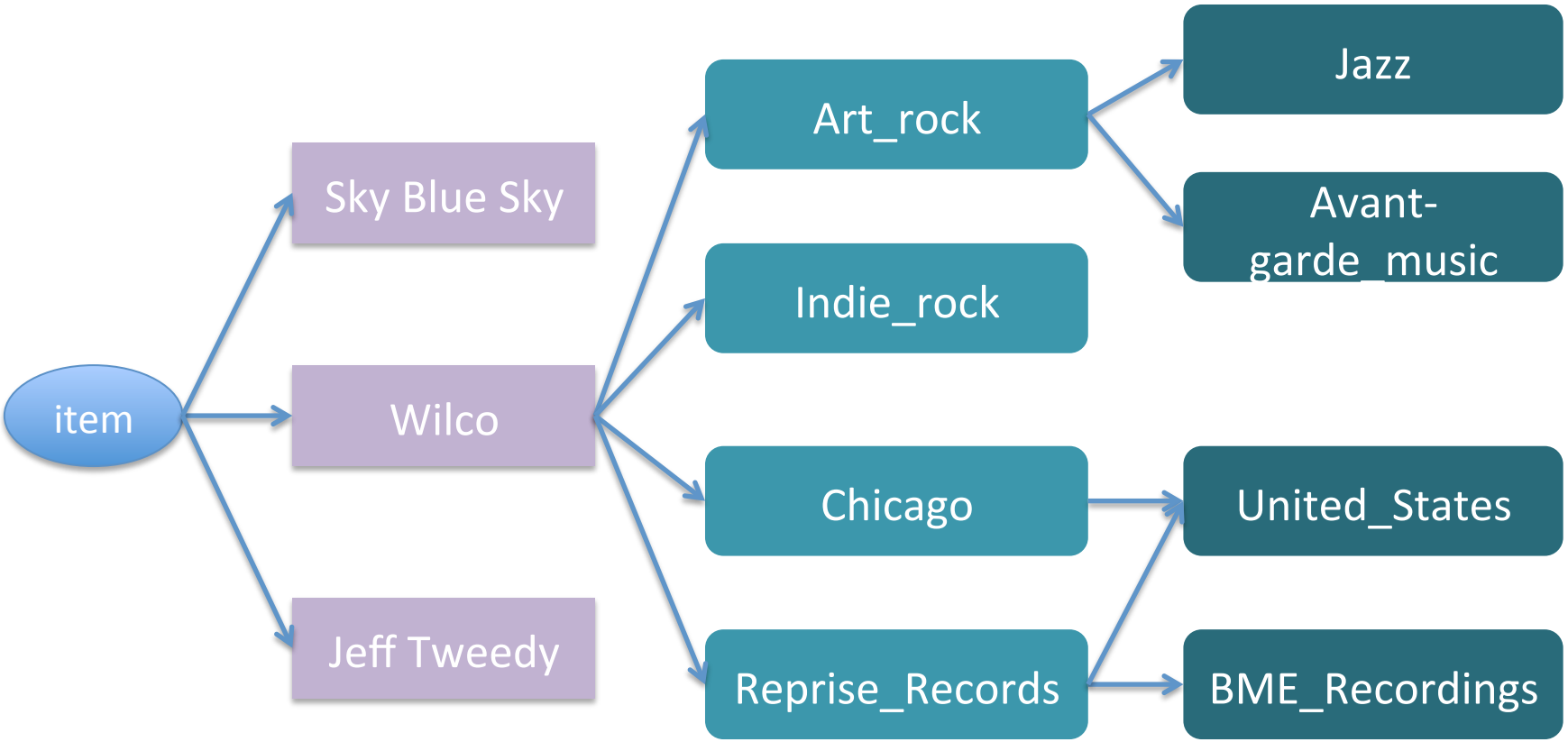
Knowledge Graph



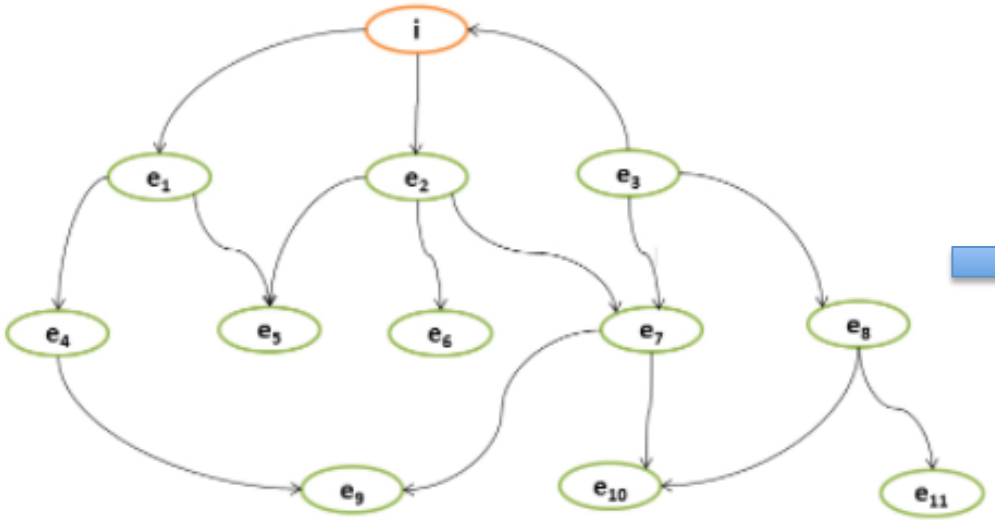
Knowledge Graph



Knowledge Graph



Graph Embedding



$$\varphi_i = (w_i,e_1 , w_i,e_2 , \ldots, w_i,e_t)$$

Long Tail Recommendations

Oramas S., Ostuni V.C., Di Noia T. et al. (2016). **Sound and Music Recommendation with Knowledge Graphs**. ACM-TIST

- Hybrid feature-combination approach
  - Knowledge Graph embedding
  - Collaborative features (user implicit feedback)

Experiments

- Sounds Recommendation
  - Freesound tags and descriptions
  - Implicit feedback (downloads)
  - 21,552 items and 20,000 users
- Music Recommendation
  - Last.fm tags and Songfacts descriptions
  - Implicit feedback (Last.fm listening habits)
  - 8,640 items and 5,199 users

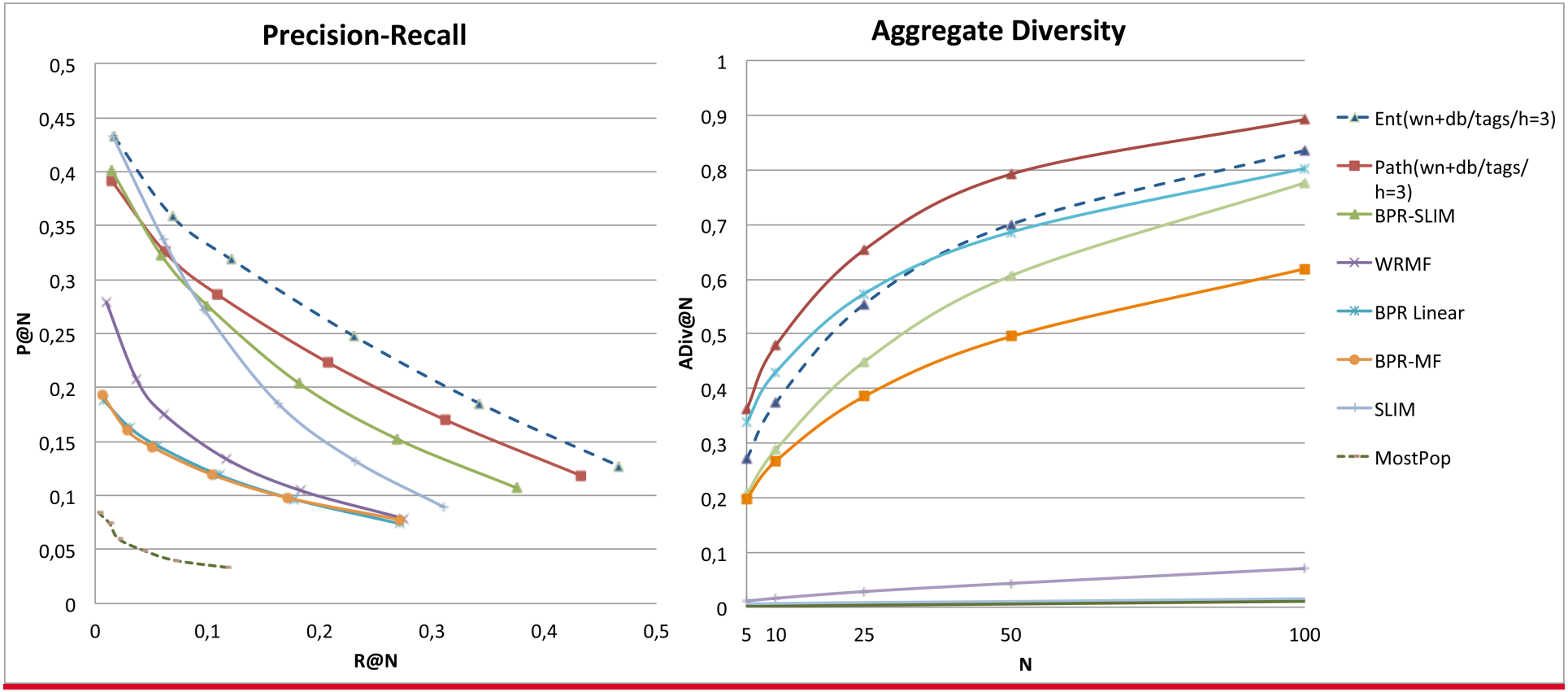
Results

- Semantic Enrichment in hybrid approach
  - Keeping high accuracy improve novelty and diversity
  - Better explore long tail items

Approach	Accuracy	Novelty	Diversity
Collab			
Collab + Tags			
Collab + Sem			
Sem only			

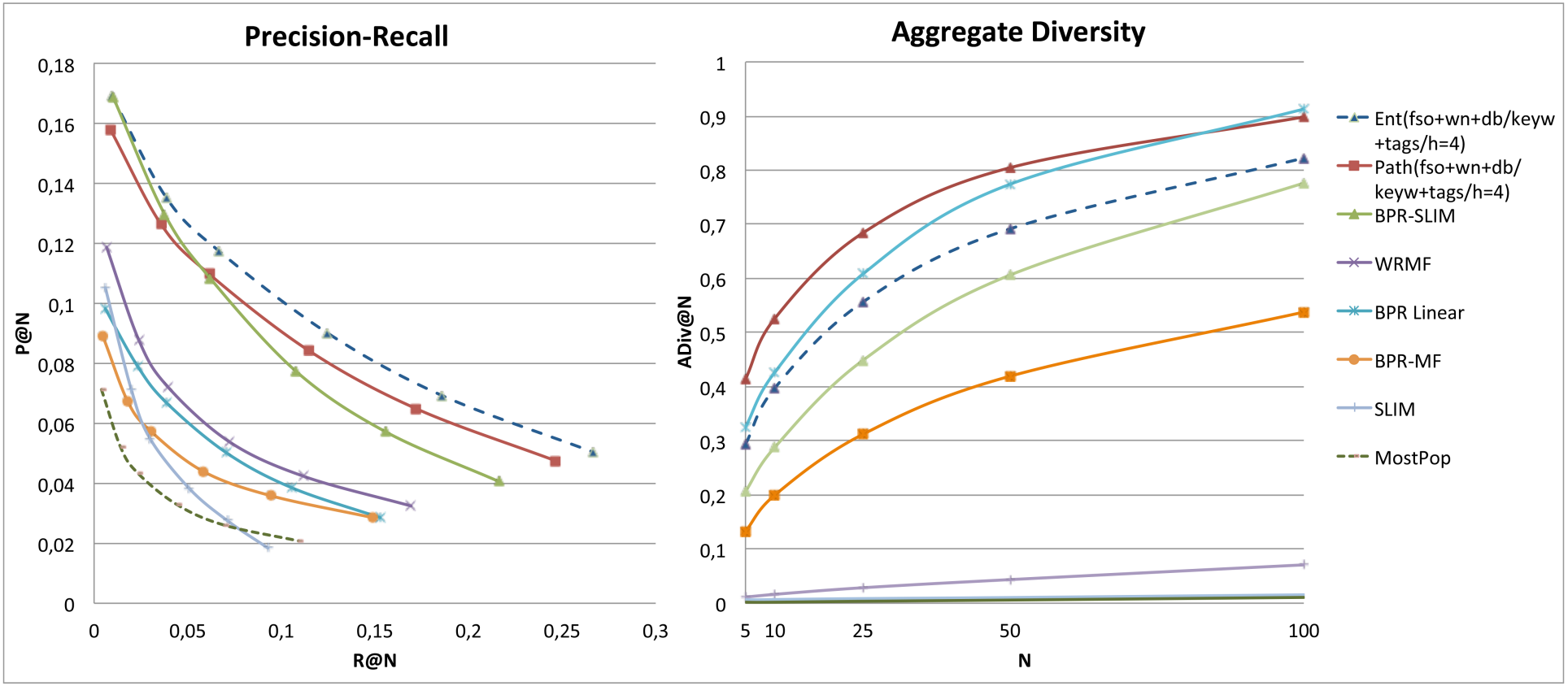
Results

- Music Recommendation



Results

- Sound Recommendation



Feature Learning

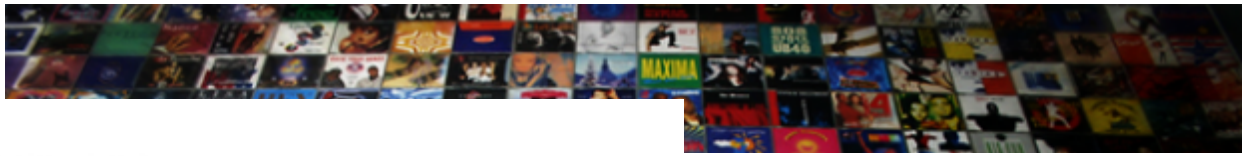
upf. Universitat Pompeu Fabra Barcelona
MTG Music Technology Group

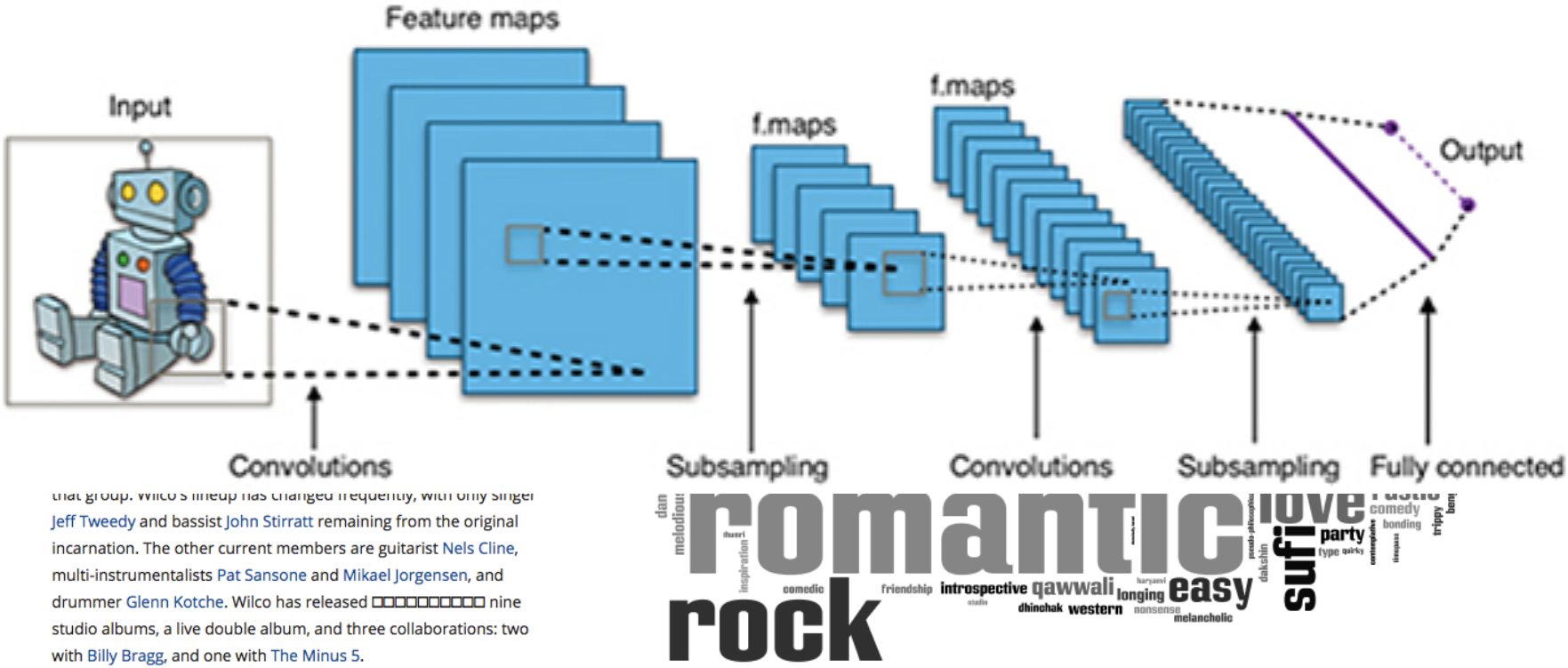
Multimodal Data

Biography

Wilco is a rock band which formed in 1994 in Chicago, Illinois, United States by the remaining members of alternative country group Uncle Tupelo following singer Jay Farrar's departure from that group. Wilco's lineup has changed frequently, with only singer Jeff Tweedy and bassist John Stirratt remaining from the original incarnation. The other current members are guitarist Nels Cline, multi-instrumentalists Pat Sansone and Mikael Jorgensen, and drummer Glenn Kotche. Wilco has released □□□□□□□□□□ nine studio albums, a live double album, and three collaborations: two with Billy Bragg, and one with The Minus 5.

Multimodal Data





that group. Wilco's lineup has changed frequently, with only singer Jeff Tweedy and bassist John Stirratt remaining from the original incarnation. The other current members are guitarist Nels Cline, multi-instrumentalists Pat Sansone and Mikael Jorgensen, and drummer Glenn Kotche. Wilco has released □□□□□□□□□□ nine studio albums, a live double album, and three collaborations: two with Billy Bragg, and one with The Minus 5.

Cold-start Recommendation

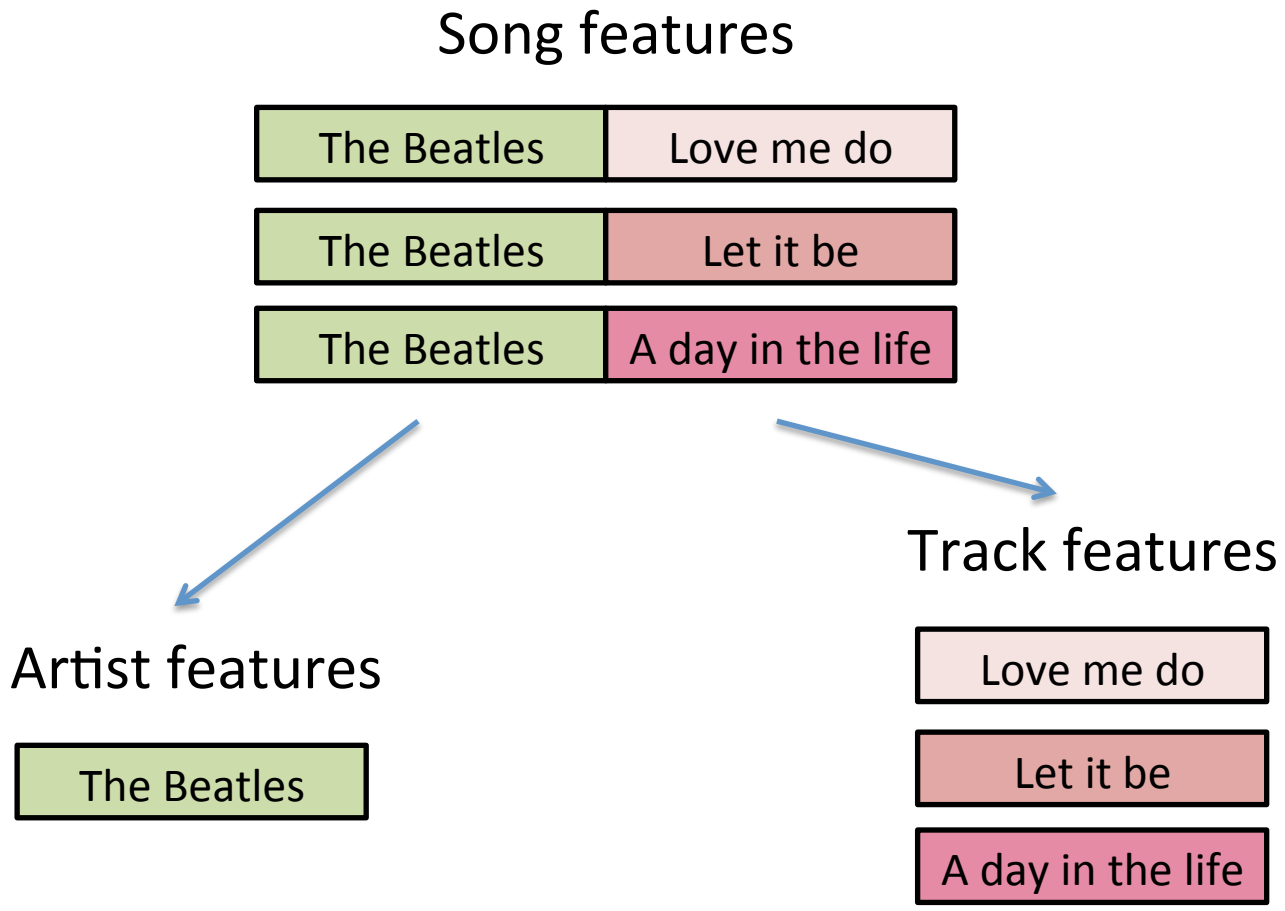
Sergio Oramas, Oriol Nieto, Mohamed Sordo, Xavier Serra (2017) **A Deep Multimodal Approach for Cold-start Music Recommendation**. DLRS Workshop, RecSys 2017.

Divide & Conquer

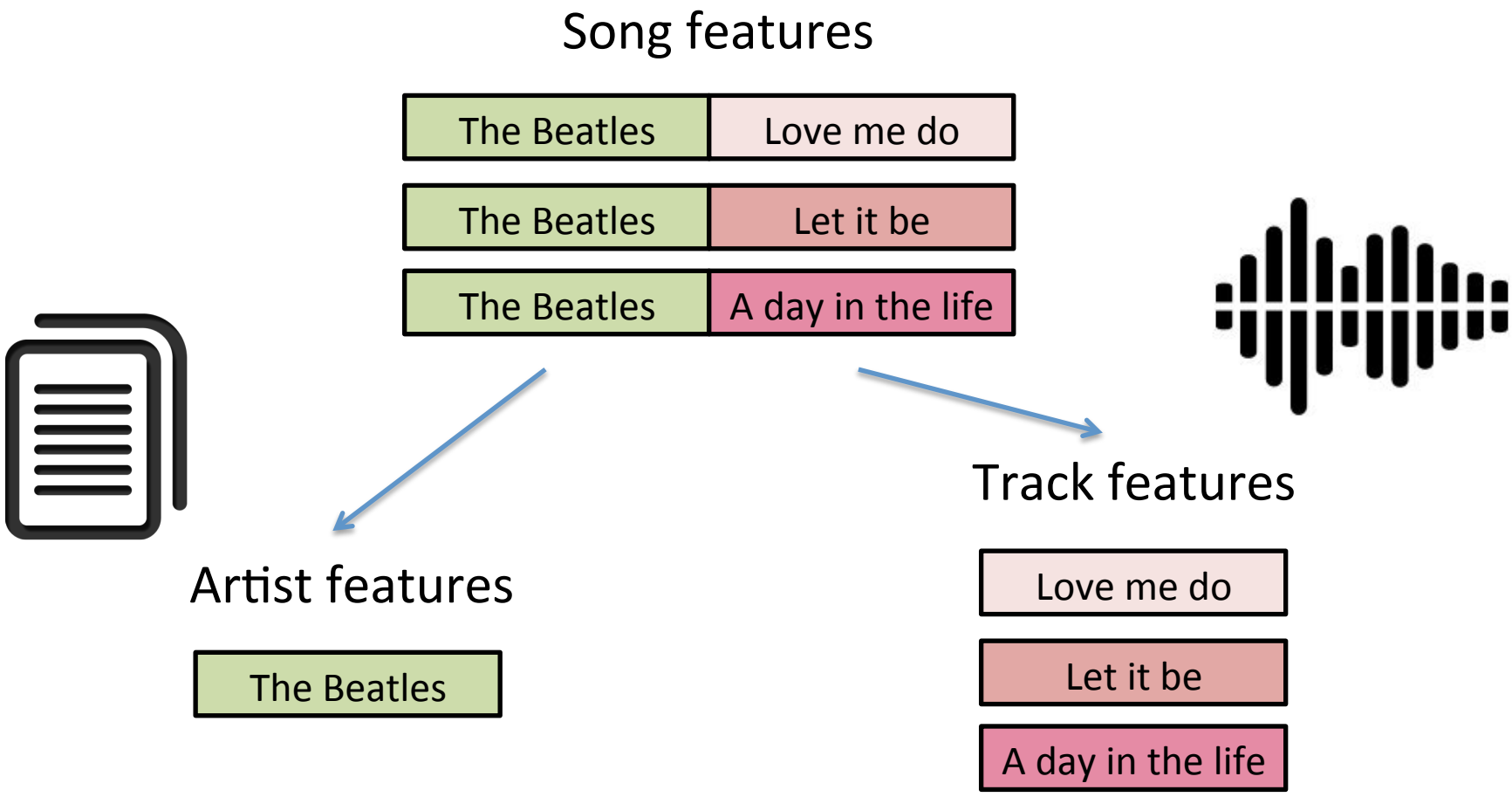
Song features

The Beatles	Love me do
The Beatles	Let it be
The Beatles	A day in the life

Divide & Conquer



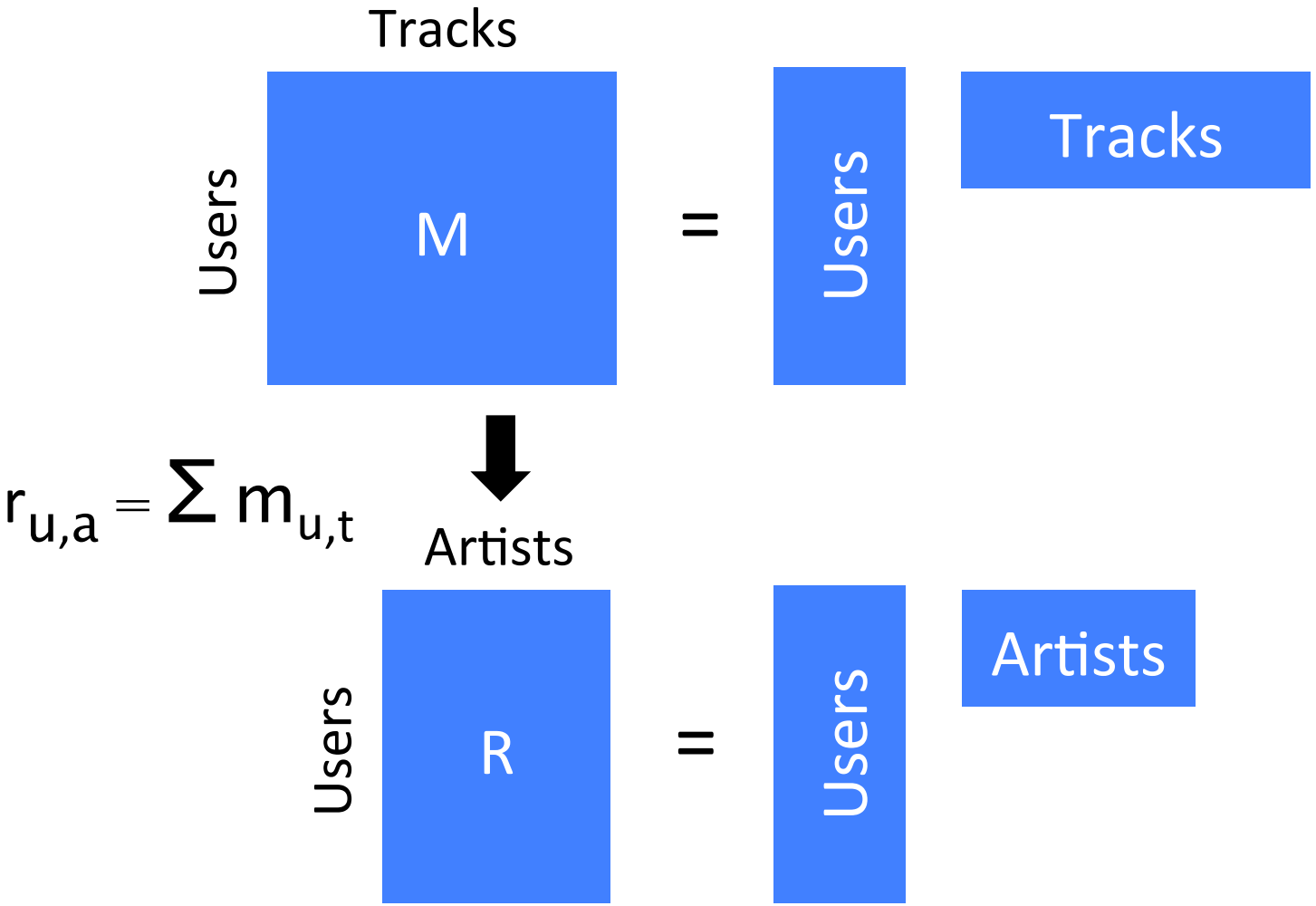
Divide & Conquer



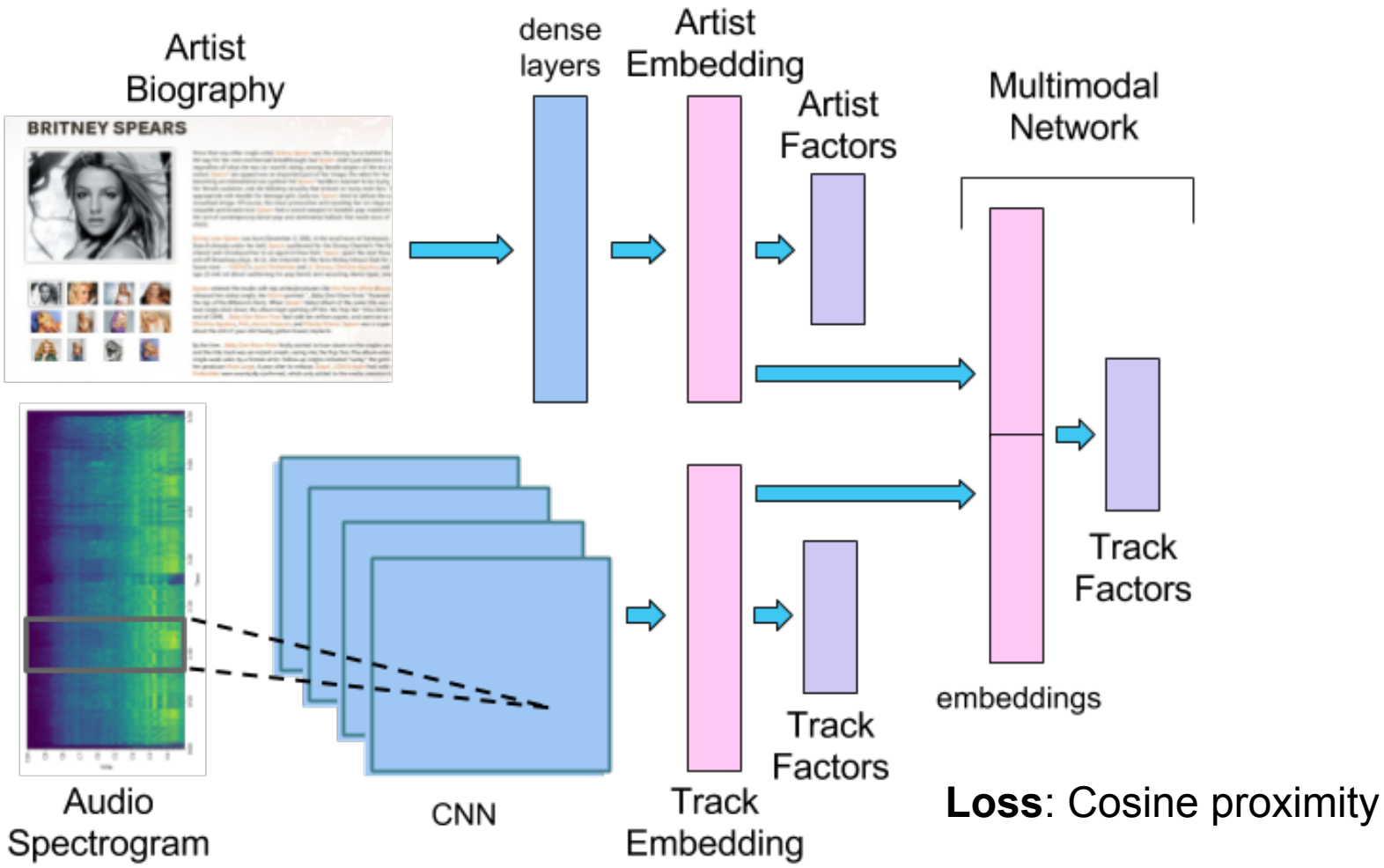
Recommendation Approach

1. Aggregate feedback data by artist
2. Learn artist feature embeddings from text
3. Learn track feature embeddings from audio
4. Fusion of feature embeddings

Matrix Factorization (WMF)



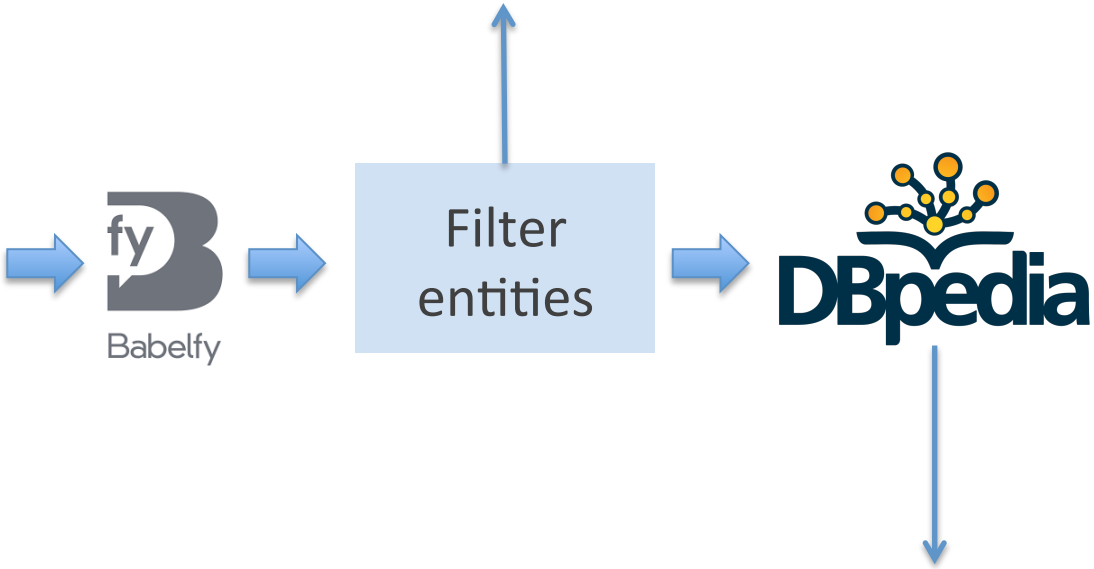
Recommendation Approach



Semantic Enrichment

Dies ist ein Blindtext. An ihm lässt sich vieles über die Schrift ablesen, in der er gesetzt ist. Auf den ersten Blick wird der Grauwert der Schriftfläche sichtbar. Dann kann man prüfen, wie gut die Schrift zu lesen ist und wie sie auf den Leser wirkt. Dies ist ein Blindtext. An ihm lässt sich vieles über die Schrift ablesen, in der er gesetzt ist. Auf den ersten Blick wird der Grauwert der Schriftfläche sichtbar. Dann kann man prüfen, wie gut die Schrift zu lesen ist und wie sie auf den Leser wirkt.

Types: MusicalArtist, Band, MusicGenre, MusicalWork, RecordLabel, Instrument, Engineer, and Place



Properties: homeTown, instrument, genre, associatedBand, writer, producer, recordedIn, etc.

Artist Text Embeddings: Semantic Enrichment

Dies ist ein Blindtext. An ihm lässt sich vieles über die Schrift ablesen, in der er gesetzt ist. Auf den ersten Blick wird der Grauwert der Schriftfläche sichtbar. Dann kann man prüfen, wie gut die Schrift zu lesen ist und wie sie auf den Leser wirkt. Dies ist ein Blindtext. An ihm lässt sich vieles über die Schrift ablesen, in der er gesetzt ist. Auf den ersten Blick wird der Grauwert der Schriftfläche sichtbar. Dann kann man prüfen, wie gut die Schrift zu lesen ist und wie sie auf den Leser wirkt.

+

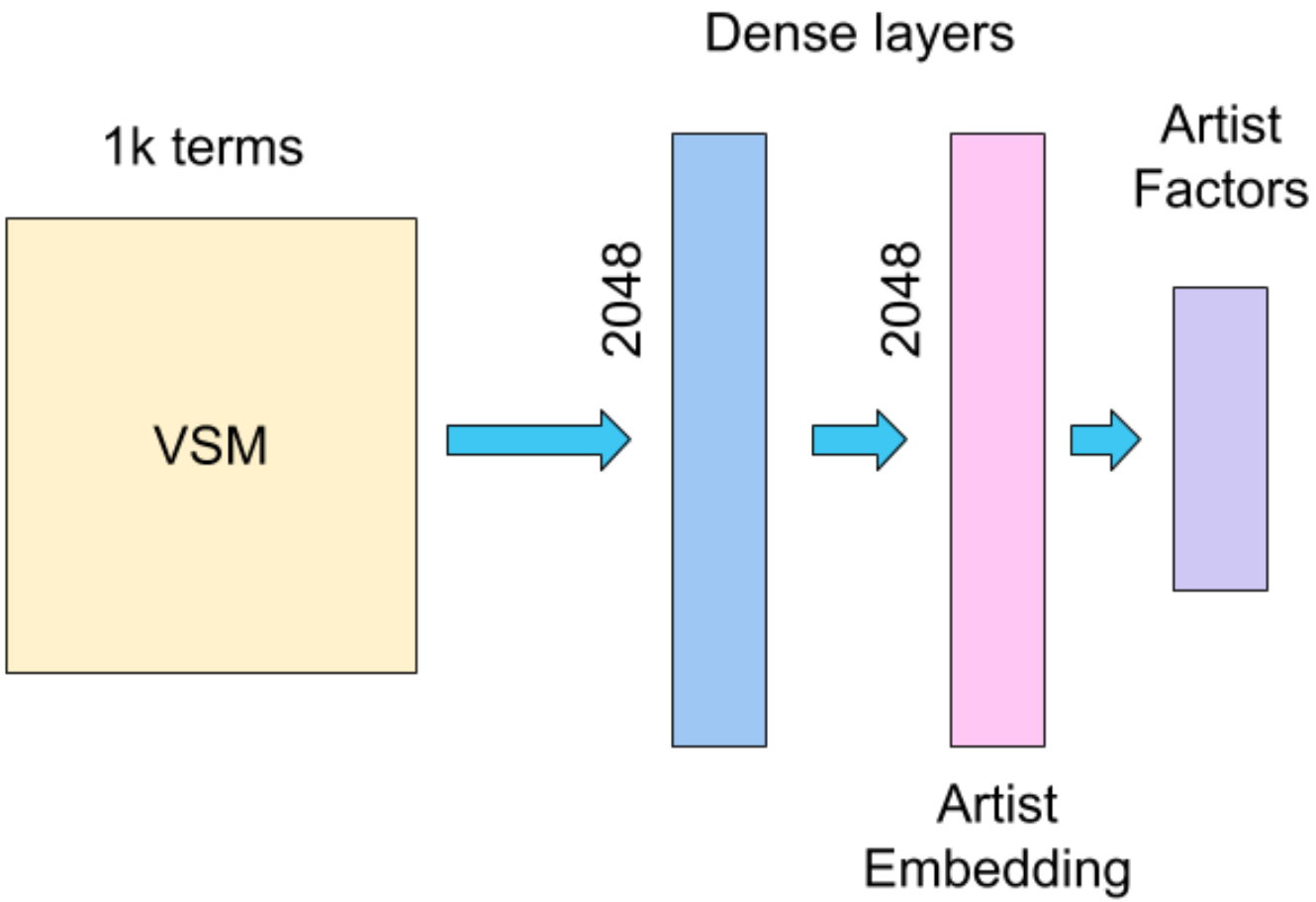
dbr:Art_rock dbr:Experimental_rock
dbr:Alternative_country
dbr:Alternative_rock dbr:Indie_rock
dbr:United_States dbr:Chicago
dbr:Illinois dbr:Reprise_Records
dbr:Nonesuch_Records
dbr:Dbpm_records

Biography text

Semantic data

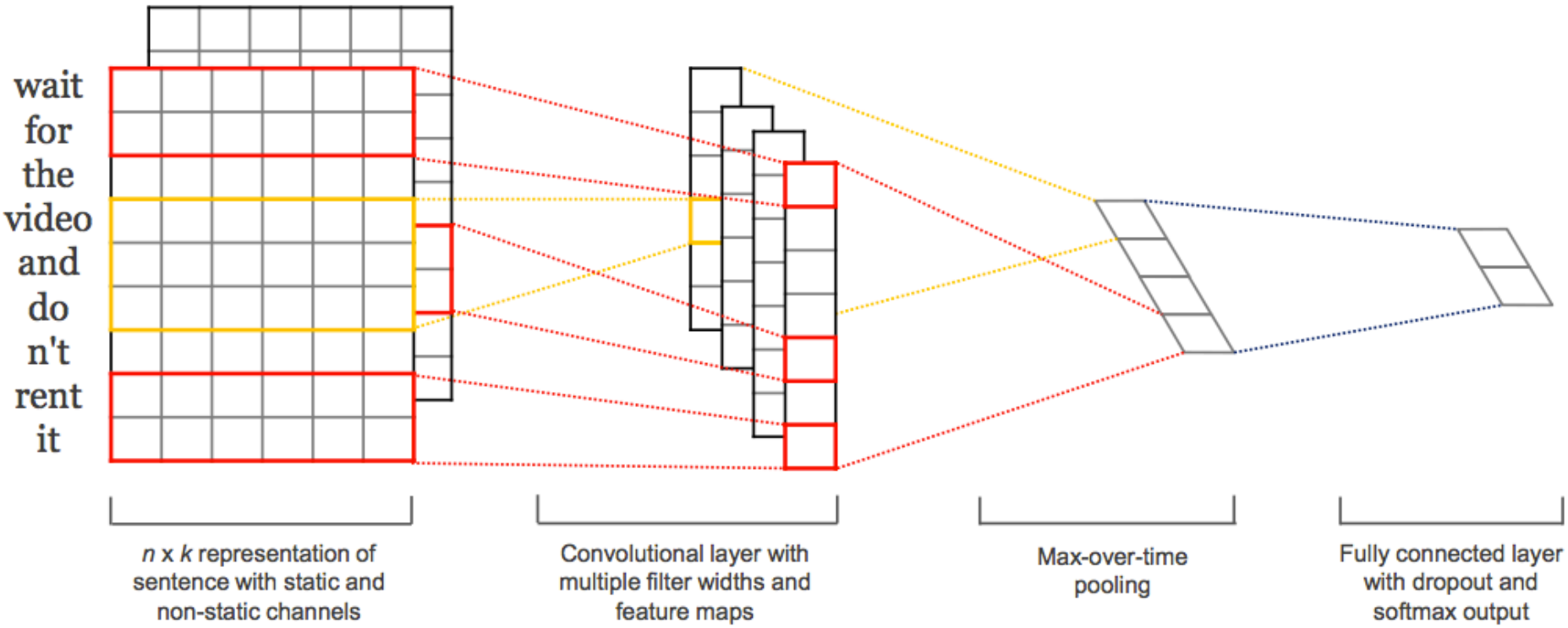
VSM
tf-idf

Artist Text Embeddings: Semantic Enrichment



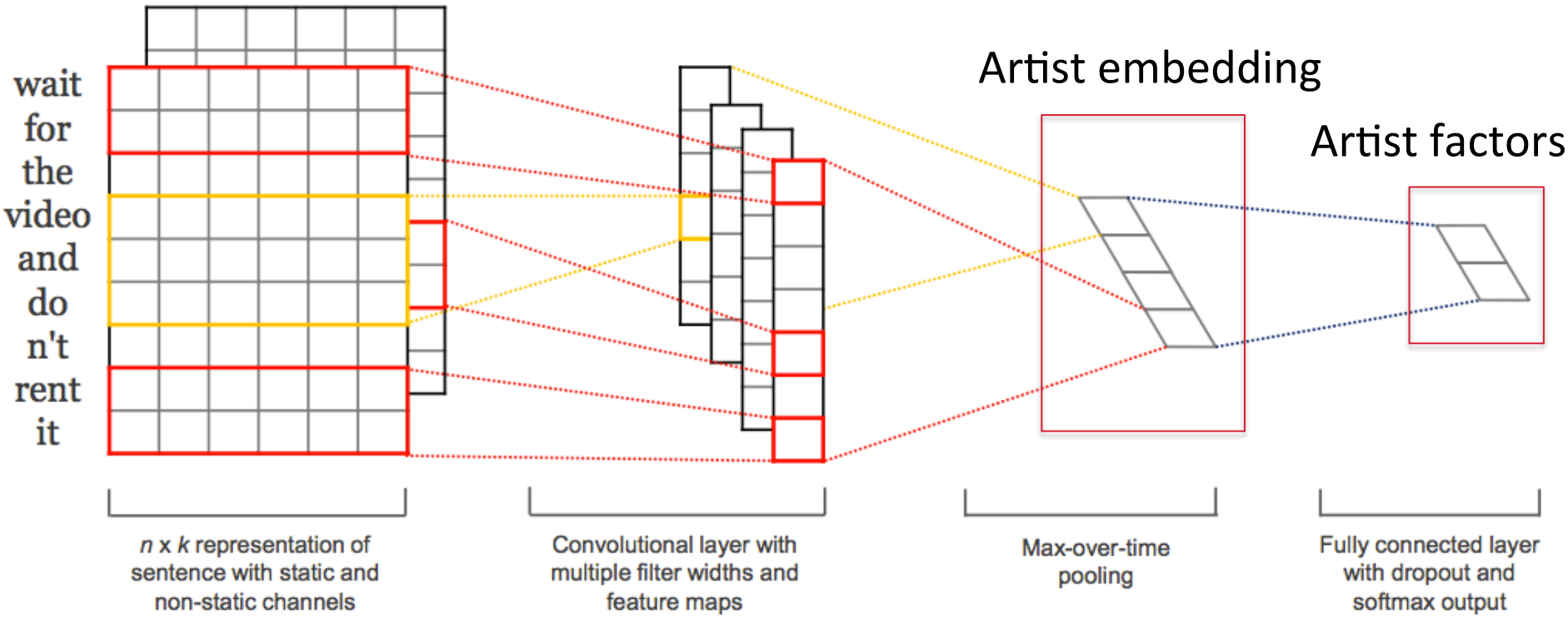
Artist Text Embeddings: Word Vectors

Kim, Y. (2014). Convolutional neural networks for sentence classification.



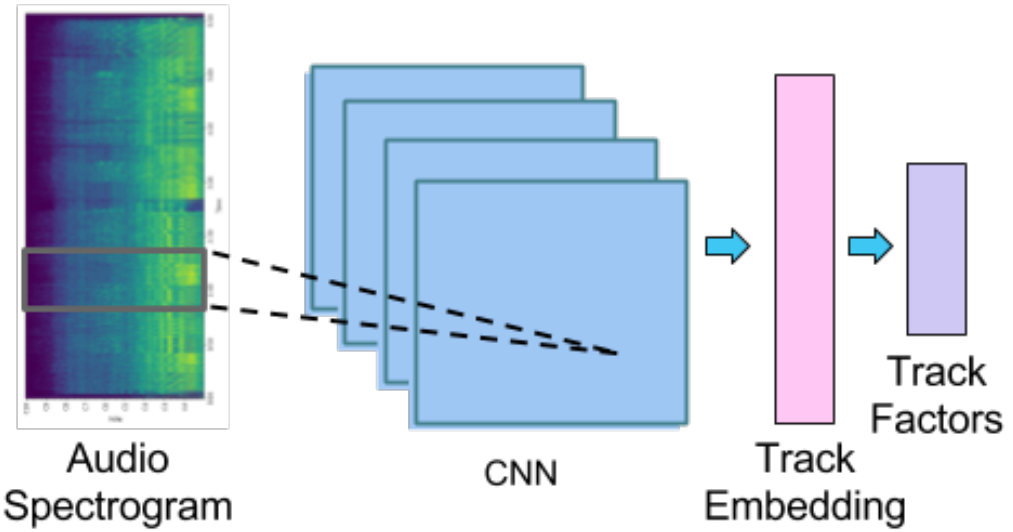
Artist Text Embeddings: Word Vectors

Kim, Y. (2014). Convolutional neural networks for sentence classification.



Track Audio Embeddings

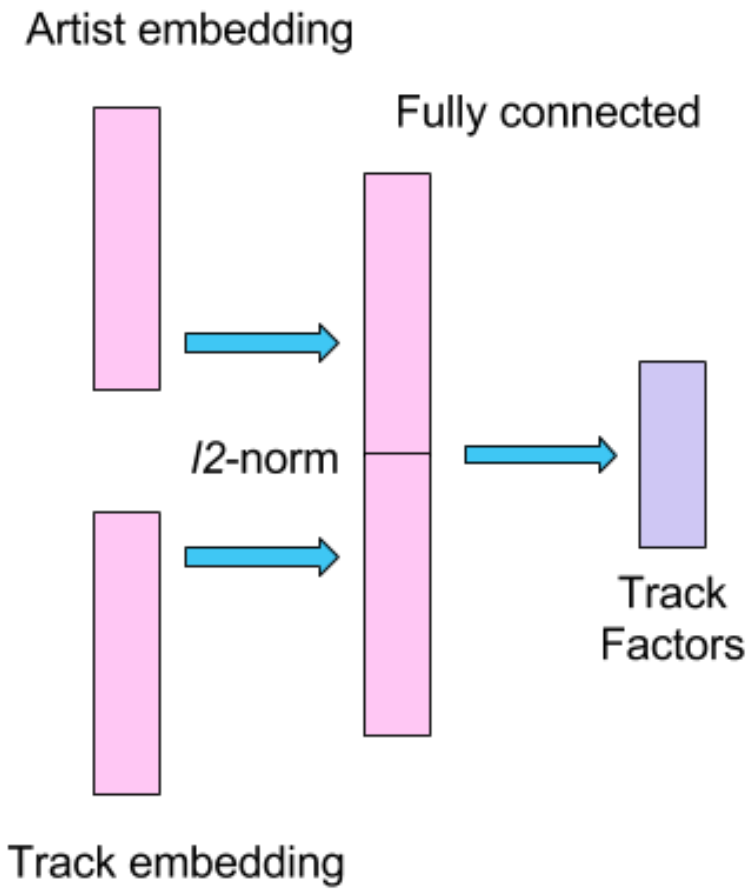
- CQT-spectrograms 96 bins
- 15 sec. patches
- 4 convolutional layers
- Time domain filters
- No dense layers
- Dropout 0.5
- Adam optimizer



Van den Oord et al. (2013) Deep content-based music recommendation

Multimodal Fusion

- *l2*-norm
- Concatenate embeddings
- Dropout 0.7



Dataset

- 328,821 tracks
- 24,043 artists
- ~1M users

+ last.fm

Taste Profile

Artists biographies and tags

Experiments

- Artist Recommendation
  - Text-based approaches

- Song Recommendation
  - Audio-based and multimodal approaches

- Splits: 80 train - 10 validation − 10 test
- **Different artists in each subset**
- Evaluation metric: MAP@500

Artist Recommendation: Text approaches

Aproach	Input	Data model	Arch	MAP
A-TEXT	Bio	VSM	FF	0.0161
A-SEM	**Sem Bio**	**VSM**	**FF**	**0.0201**
A-W2V-GOO	Bio	w2v-pretrain	CNN	0.0119
A-W2V	Bio	w2v-trained	CNN	0.0145
A-TAGS	Tags	VSM	FF	0.0314
TAGS-ITEMKNN	Tags	-	itemKnn	0.0161
TEXT-RF	Bio	VSM	RF	0.0089
RANDOM	-	-	-	0.0014
UPPER-BOUND	-	-	-	0.5528

Artist Recommendation: Text approaches

Aproach	Input	Data model	Arch	MAP
A-TEXT	Bio	VSM	FF	0.0161
A-SEM	**Sem Bio**	**VSM**	**FF**	**0.0201**
A-W2V-GOO	Bio	w2v-pretrain	CNN	0.0119
A-W2V	Bio	w2v-trained	CNN	0.0145
A-TAGS	Tags	VSM	FF	0.0314
TAGS-ITEMKNN	Tags	-	itemKnn	0.0161
TEXT-RF	Bio	VSM	RF	0.0089
RANDOM	-	-	-	0.0014
UPPER-BOUND	-	-	-	0.5528

Artist Recommendation: Tags vs. Bios

Aproach	Input	Data model	Arch	MAP
A-TEXT	Bio	VSM	FF	0.0161
A-SEM	**Sem Bio**	**VSM**	**FF**	**0.0201**
A-W2V-GOO	Bio	w2v-pretrain	CNN	0.0119
A-W2V	Bio	w2v-trained	CNN	0.0145
A-TAGS	Tags	VSM	FF	0.0314
TAGS-ITEMKNN	Tags	-	itemKnn	0.0161
TEXT-RF	Bio	VSM	RF	0.0089
RANDOM	-	-	-	0.0014
UPPER-BOUND	-	-	-	0.5528

Artist Recommendation: Competitor

Aproach	Input	Data model	Arch	MAP
A-TEXT	Bio	VSM	FF	0.0161
A-SEM	**Sem Bio**	**VSM**	**FF**	**0.0201**
A-W2V-GOO	Bio	w2v-pretrain	CNN	0.0119
A-W2V	Bio	w2v-trained	CNN	0.0145
A-TAGS	Tags	VSM	FF	0.0314
TAGS-ITEMKNN	Tags	-	itemKnn	0.0161
TEXT-RF	Bio	VSM	RF	0.0089
RANDOM	-	-	-	0.0014
UPPER-BOUND	-	-	-	0.5528

Artist Recommendation: Deep Learning vs. RF

Aproach	Input	Data model	Arch	MAP
A-TEXT	Bio	VSM	FF	0.0161
A-SEM	**Sem Bio**	**VSM**	**FF**	**0.0201**
A-W2V-GOO	Bio	w2v-pretrain	CNN	0.0119
A-W2V	Bio	w2v-trained	CNN	0.0145
A-TAGS	Tags	VSM	FF	0.0314
TAGS-ITEMKNN	Tags	-	itemKnn	0.0161
TEXT-RF	Bio	VSM	RF	0.0089
RANDOM	-	-	-	0.0014
UPPER-BOUND	-	-	-	0.5528

Song Recommendation: Audio approach

Approach	Artist Input	Track Input	Arch	MAP
AUDIO	-	audio spec	CNN	0.0015
SEM-VSM	Sem Bio	-	FF	0.0032
SEM-EMB	A-SEM	-	FF	0.0034
MM-LF-LIN	**A-SEM**	**AUDIO emb**	**MLP**	**0.0036**
MM-LF-H1	**A-SEM**	AUDIO emb	MLP	0.0035
MM	Sem Bio	audio spec	CNN	0.0014
TAGS-VSM	Tags	-	FF	0.0043
TAGS-EMB	A-TAGS	-	FF	0.0049
RANDOM	rnd emb	-	FF	0.0002
UPPER-BOUND	-	-	-	0.1649

Song Recommendation: Embeddings vs. VSM

Approach	Artist Input	Track Input	Arch	MAP
AUDIO	-	audio spec	CNN	0.0015
SEM-VSM	Sem Bio	-	FF	0.0032
SEM-EMB	A-SEM	-	FF	0.0034
MM-LF-LIN	**A-SEM**	**AUDIO emb**	**MLP**	**0.0036**
MM-LF-H1	**A-SEM**	AUDIO emb	MLP	0.0035
MM	Sem Bio	audio spec	CNN	0.0014
TAGS-VSM	Tags	-	FF	0.0043
TAGS-EMB	A-TAGS	-	FF	0.0049
RANDOM	rnd emb	-	FF	0.0002
UPPER-BOUND	-	-	-	0.1649

Song Recommendation: Late fusion vs. simultaneous

Approach	Artist Input	Track Input	Arch	MAP
AUDIO	-	audio spec	CNN	0.0015
SEM-VSM	Sem Bio	-	FF	0.0032
SEM-EMB	A-SEM	-	FF	0.0034
MM-LF-LIN	**A-SEM**	**AUDIO emb**	**MLP**	**0.0036**
MM-LF-H1	**A-SEM**	AUDIO emb	MLP	0.0035
MM	Sem Bio	audio spec	CNN	0.0014
TAGS-VSM	Tags	-	FF	0.0043
TAGS-EMB	A-TAGS	-	FF	0.0049
RANDOM	rnd emb	-	FF	0.0002
UPPER-BOUND	-	-	-	0.1649

Song Recommendation: Multimodal vs. single mod.

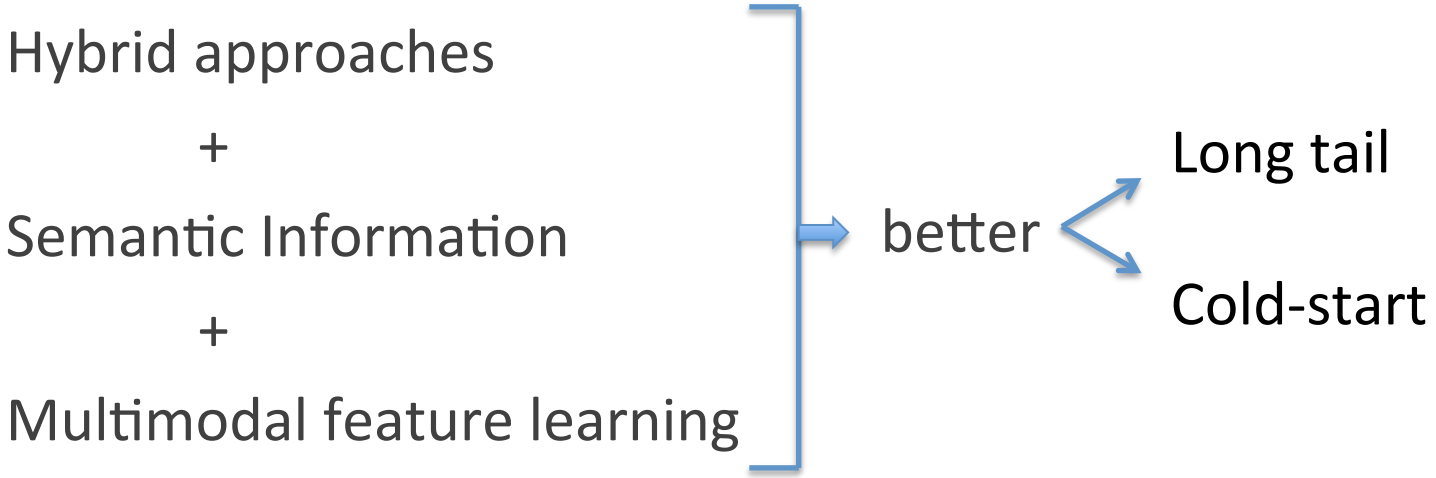
Approach	Artist Input	Track Input	Arch	MAP
AUDIO	-	audio spec	CNN	0.0015
SEM-VSM	Sem Bio	-	FF	0.0032
SEM-EMB	A-SEM	-	FF	0.0034
MM-LF-LIN	**A-SEM**	**AUDIO emb**	**MLP**	**0.0036**
MM-LF-H1	**A-SEM**	AUDIO emb	MLP	0.0035
MM	Sem Bio	audio spec	CNN	0.0014
TAGS-VSM	Tags	-	FF	0.0043
TAGS-EMB	A-TAGS	-	FF	0.0049
RANDOM	rnd emb	-	FF	0.0002
UPPER-BOUND	-	-	-	0.1649

Results

- Splitting the problem between artists and songs ✓
- Learning artist feature embedding separately ✓
- Use of artist biographies ✓
- Semantic enrichment via Entity Linking ✓
- Late fusion of multimodal feature embeddings ✓
- Multimodal better than single modality ✓

Conclusions

- Knowledge from semantic repositories incorporated via entity linking improves item profiles -> better diversity
- Learning deep feature representations from multimodal data and combining them improves cold-start recommendations



Reproducibility

- Datasets

  MSD-A: http://www.upf.edu/web/mtg/msd-a

  KG-REC: https://www.upf.edu/web/mtg/kgrec

- Source code

  TARTARUS: https://github.com/sergiooramas/tartarus

  lodreclib: https://github.com/sisinflab/lodreclib

Thanks!

@sergiooramas

Universitat Pompeu Fabra Barcelona

MTG Music Technology Group