DSC250 | Winter 2025

Hybrid Graph-Based Content Filtering (HyG-Con) for Music Recommendation

Eliminating Filter Bubbles

Sarah Borsotto, John Driscoll, Taylor Martinez, Thuy Nguyen

Filter Bubbles

The phenomenon where Internet customization effectively isolates individuals from diverse opinions or materials

In music recommendation:

- Users are only recommended popular songs (collaborative bubble)
- Users are only recommended similar songs which often fall into the same genre (content bubble)

Spotify radio often suggests songs in your liked playlist or trending songs to maintain user engagement

Traditional Recommender Systems

Content Filtering

What is Content Filtering?

Relies on audio features such as tempo, instrumentalness, and danceability, to recommend songs similar to a user’s listening history

Limitations

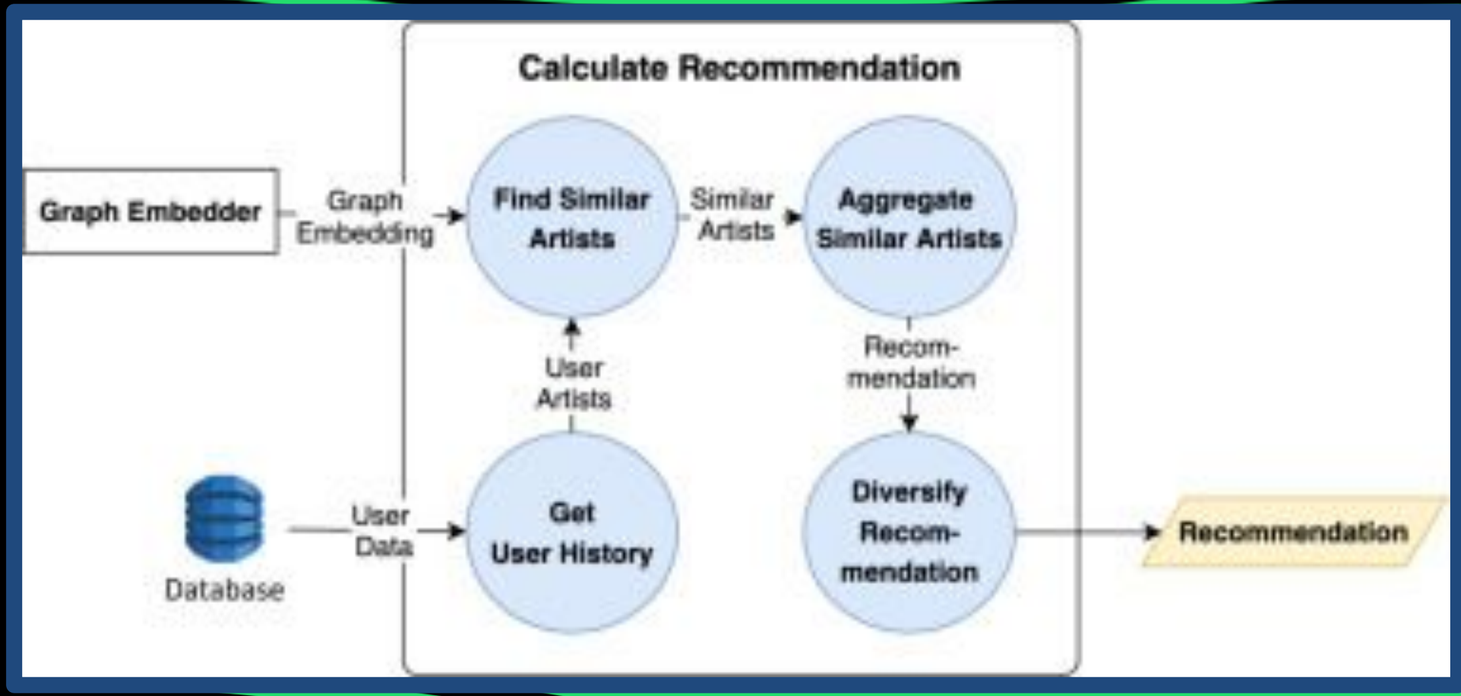
Lacks novelty, keeping users stuck in a "genre bubble" with limited exposure to new or diverse music

No genre diversity is incorporated

EARS

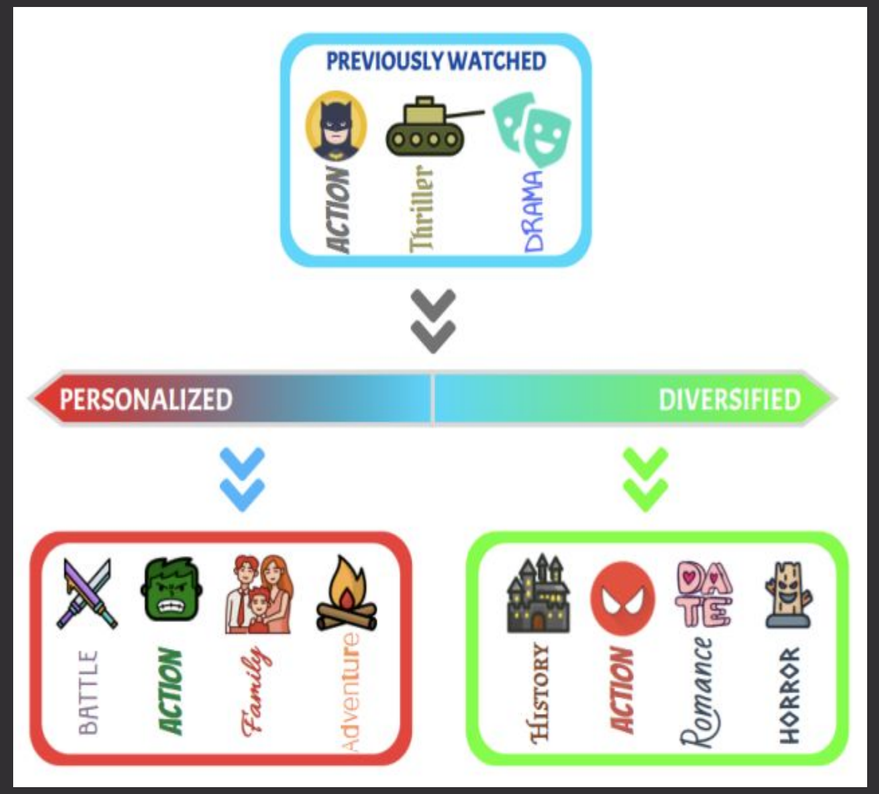
(Embedding-Based Artist Recommender System)

- Constructs collaborative knowledge graphs (hybrid)
- Use graph embeddings to generate accurate and diverse recommendations
- Struggles with capturing relationships between genres and songs; domain knowledge required



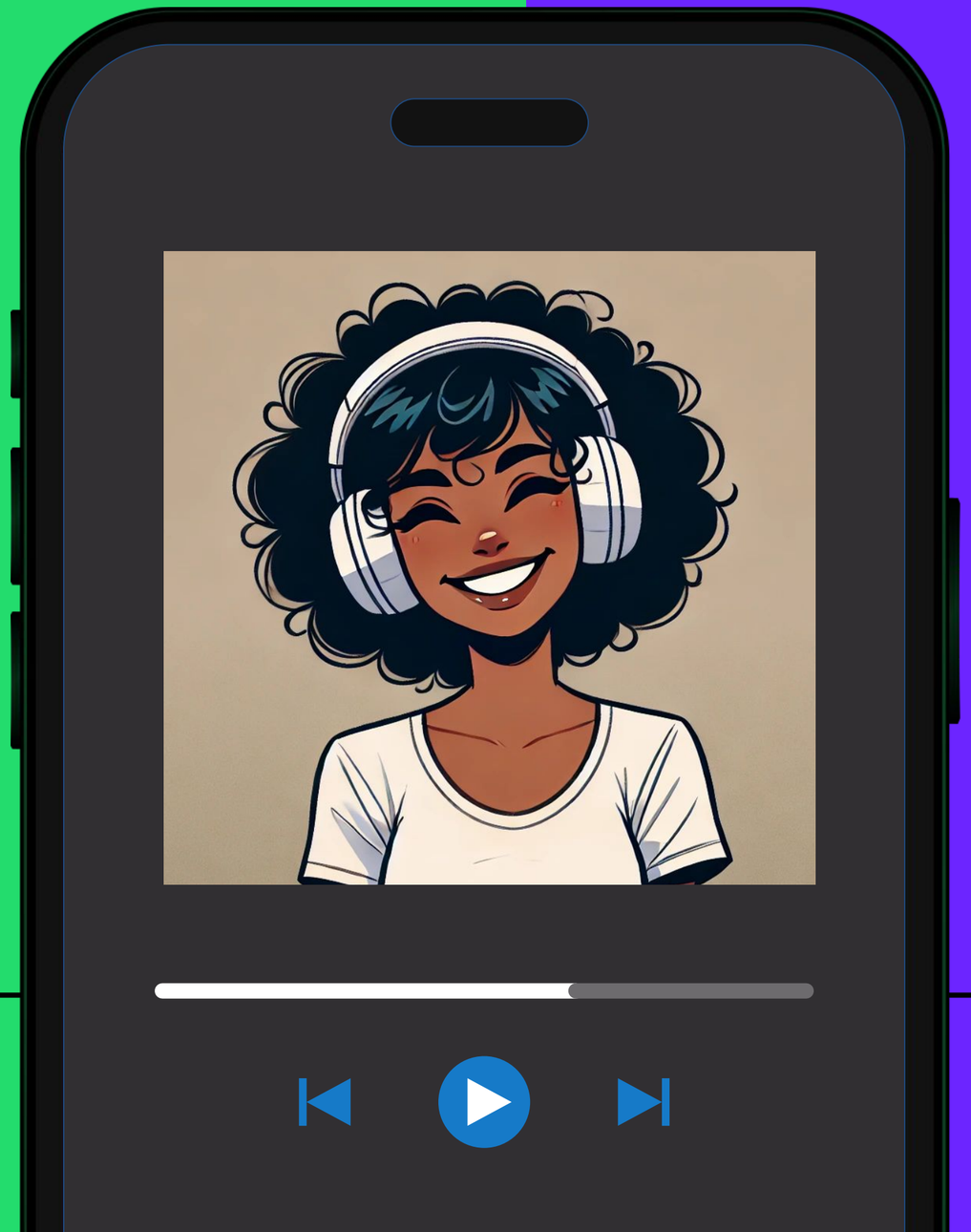
Limitations of Current Methods

- Restrict listeners' exposure to new genres and diverse musical content
- Listener behavior creates feedback loops (e.g., confirmation bias)
- Don't incorporate knowledge about how musical styles developed



Proposed Method:
Hy-G Con

A hybrid approach to music recommendation that combines content-based filtering with a domain-specific knowledge graph to promote cross-genre discovery while maintaining relevance



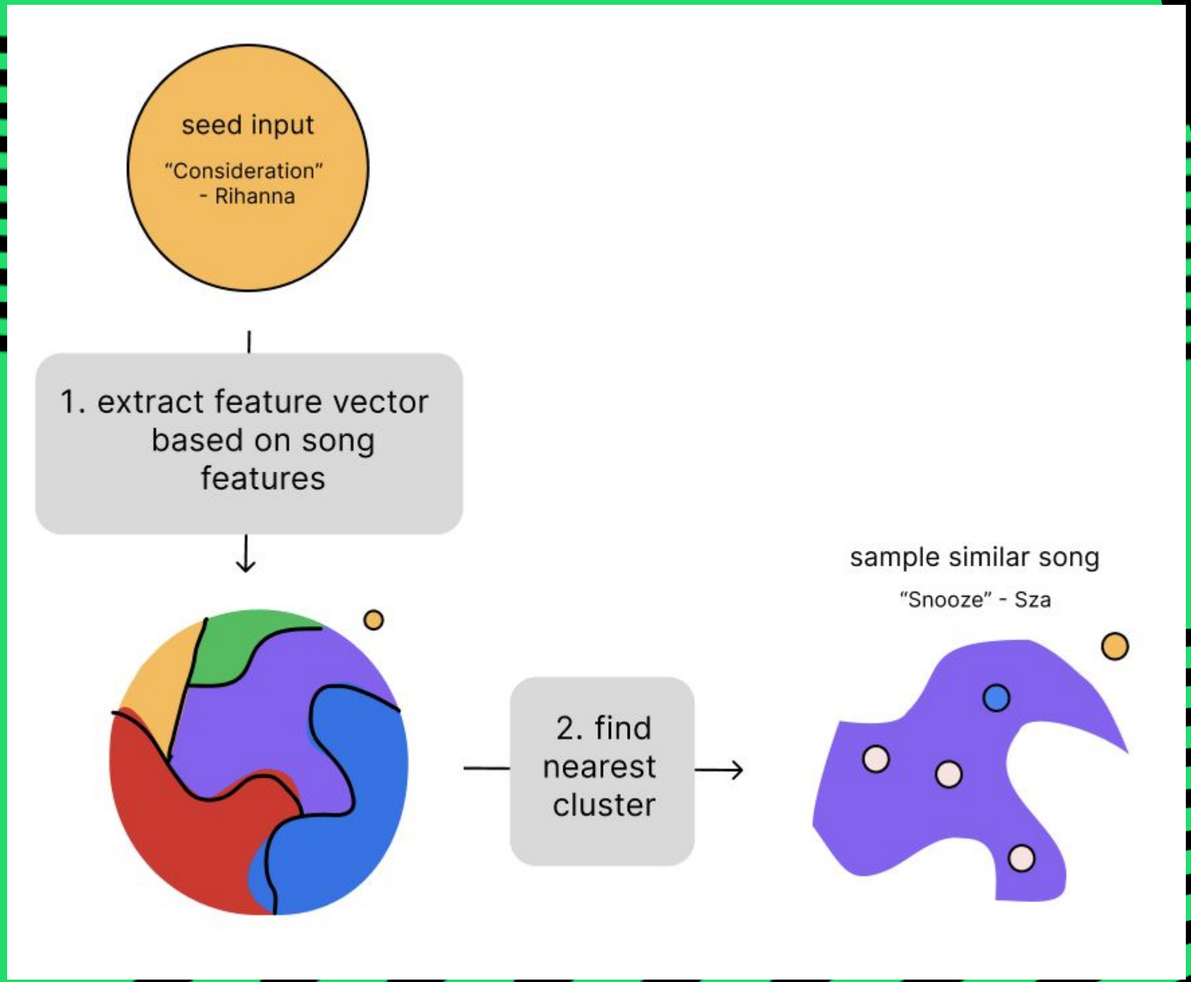
Spotify Tracks Dataset

- 114,000 tracks with relevant song features across 125 genres
- Tools: Spotify Web API and Python
- Cleaned and preprocessed

Music Map

- Expert-curate genealogy of music genres based on historical data
  - Removed 15 genres that are too niche or diverse within themselves: sleep, comedy, happy, sad, german, acoustic, turkish, iranian, malay, spanish, songwriter, british, french, swedish, indian

Content-Based Filtering Baseline



Model approach: MiniBatch K-Means.

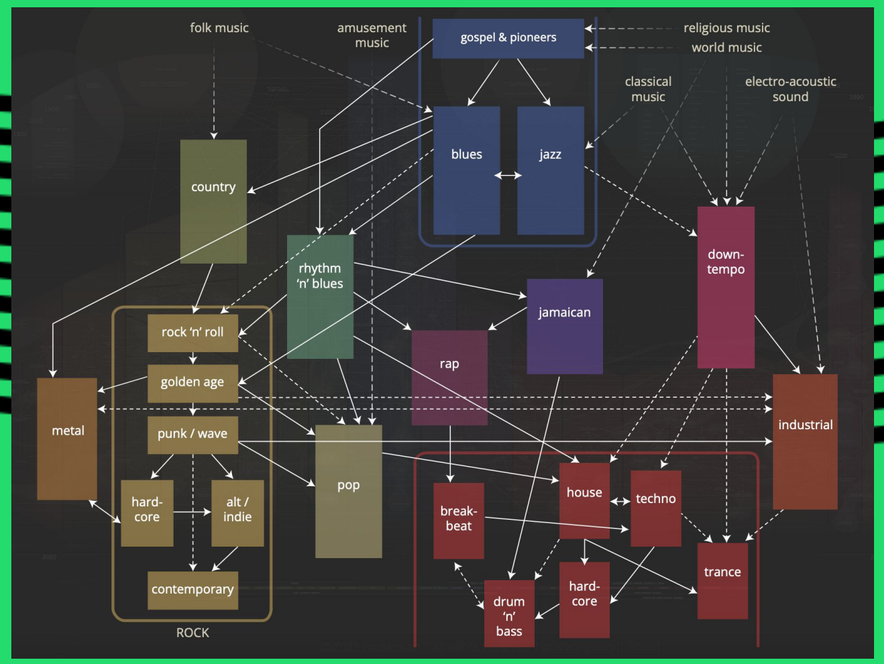
Goal: Create many small clusters (min. **20 songs per cluster**) to **preserve distinctiveness**.

- **Optimal cluster count: 500 clusters**
  a. Each cluster contains **50 to 500 songs**.
  b. Even distribution across the feature space.

Recommendation process:

- Find **closest cluster centroid**.
- Randomly sample **similar songs** from that cluster
- Since songs from the same genre tend to have similar features, there will be a lack of genre diversity within each cluster

Secondary Approach - Stochastic model: pick a random cluster nearby



Domain-Specific Knowledge Graph

Goal: Use historical semantic data to facilitate seamless transitions between related music genres.

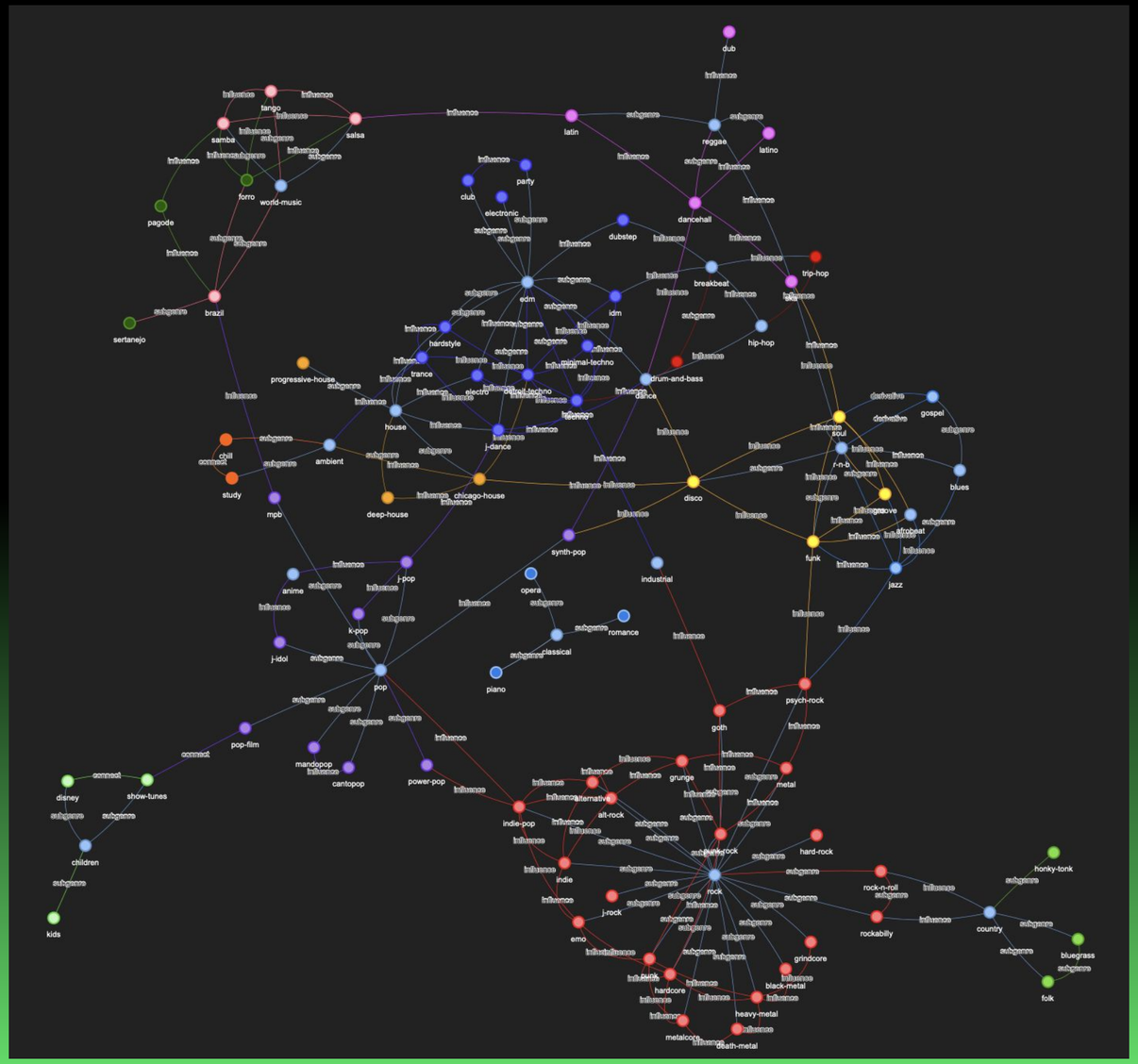
Construction: Manually encoded dataset genres to match Music Map (expert curated genre genealogy)

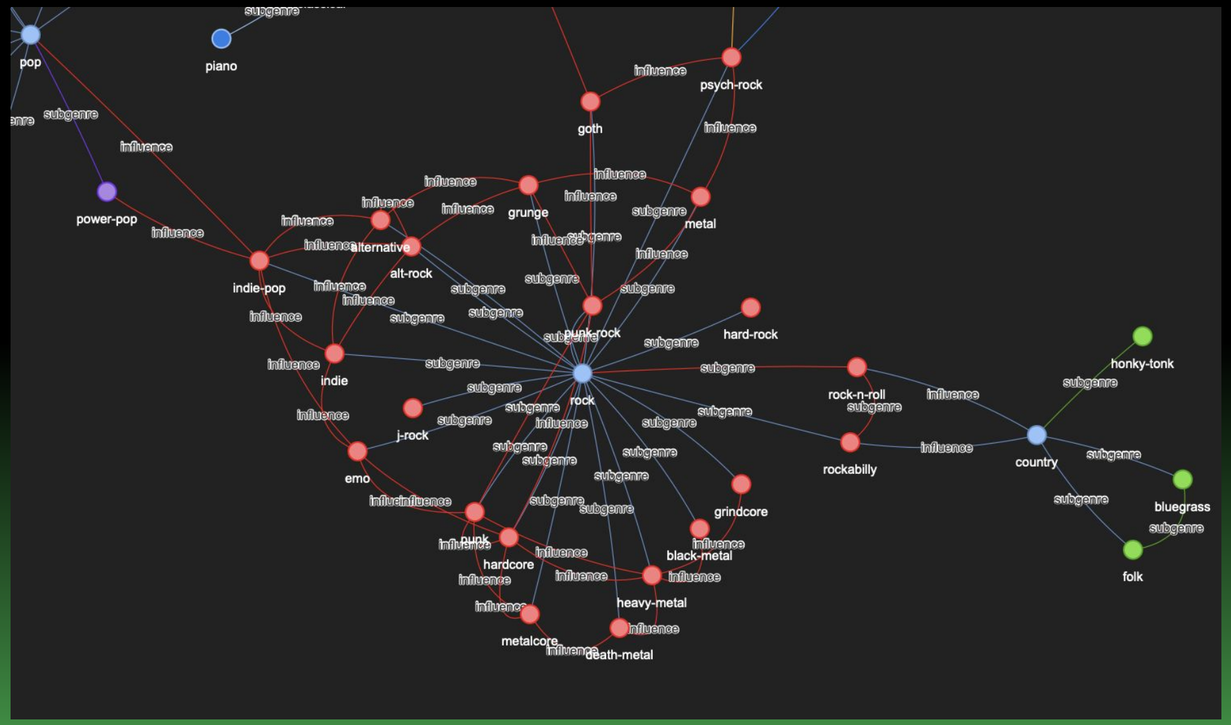
- Organized parent genres in the simplified model
- Identified detailed connections using the interactive graph

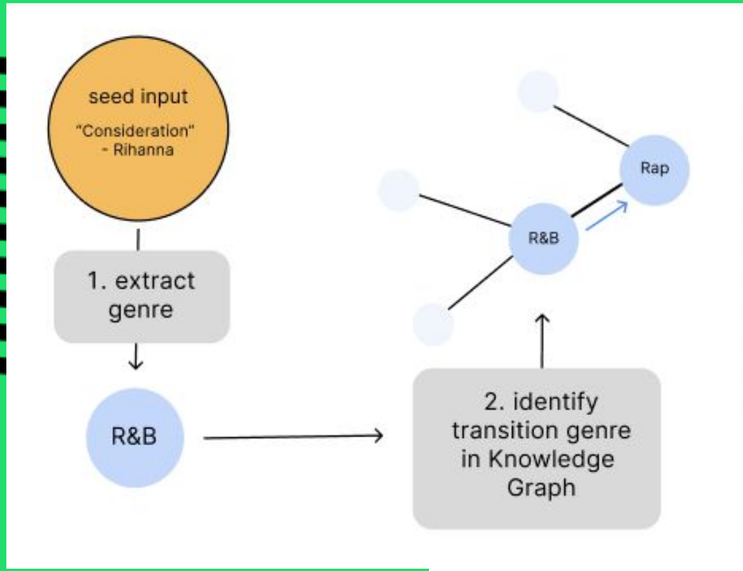
Determining Genre Transitions - Graph Traversal

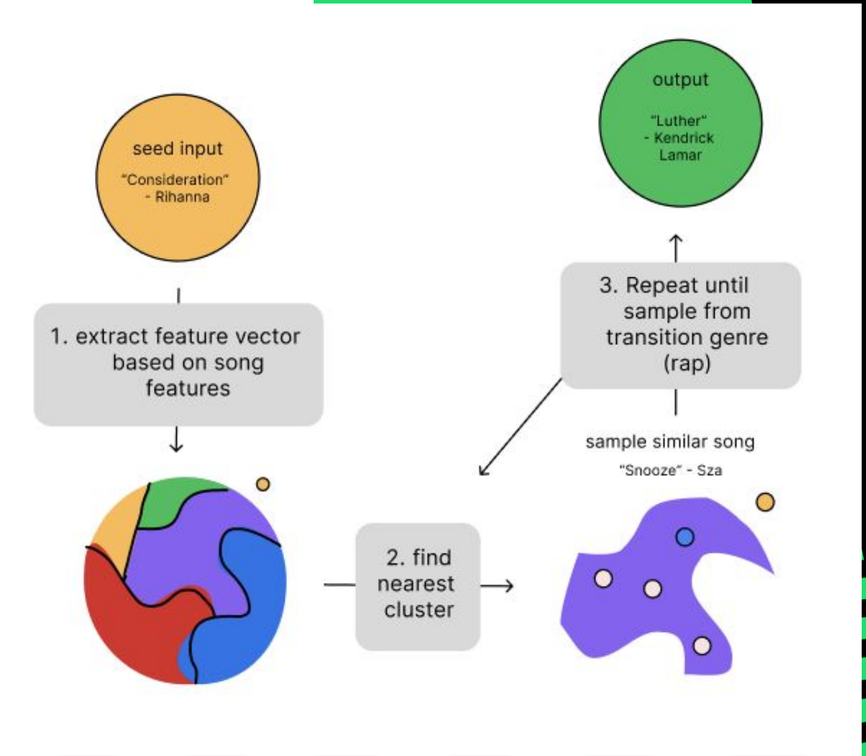
- Computed shortest paths from current genre to all others.
- Randomly selected the next genre, prioritizing:
  - Short or uncommon paths (dead-end paths).
  - Closer genres (limited depth traversal)

Knowledge Graph Implemented









Hybrid Model

Hybrid Model: Integrating Genre Transitions

- **Step 1**: Start with the seed song (e.g., *Consideration*).
- **Step 2**: Identify the next transition genre using knowledge graph (e.g., R&B → Rap).
- **Step 3:** Find a nearby cluster and sample a song from that cluster that matches the transition genre
  - If no match is found, move to a nearby cluster.
- **Step 4**: Repeat the transition process, progressing through genres until 15 songs are recommended.

Compare this to baseline & stochastic!

Evaluation

Metrics

Intra-list Diversity

Diversity of entire playlist

$$ILD@k = \left(\sum_{i=1}^{k+1}\sum_{j=1}^{k+1} d(i,j)\right)/(k*(k-1))$$

Genre Coverage

Total unique genres

Feature Variance

Average variance of features

Models

Content Filtering

Baseline clustering

Stochastic CF

Baseline + randomness to sample different cluster

Hy-G Con

Knowledge graph + content filtering

Content-Based Filtering Results

- Songs are grouped into clusters based on the predefined features
- Computes transition probabilities between clusters based on the model's centroids
- The content-based filtering graph has promising results for generating playlists, but underperforms alone
- Can be inconsistent with this as the only recommender system
- Metrics on a randomly generated playlist:

Intra-List Diversity: 0.5476030109814571

Genre Coverage: 5.71%

Feature Variance Score: 0.04309619654899438

Content-Based Filtering Results

```
==================================================
Generated Playlist (15 songs)
==================================================
1. Hey Lady - WANIMA - punk-rock
   Features: danceability: 0.46, energy: 0.94, acousticness: 0.00, instrumentalness: 0.00, valence: 0.49, tempo: 0.47
2. Fenix - Tronic - punk-rock
   Features: danceability: 0.47, energy: 0.94, acousticness: 0.00, instrumentalness: 0.00, valence: 0.43, tempo: 0.57
3. Endlessly - The Cab - punk
   Features: danceability: 0.51, energy: 0.86, acousticness: 0.05, instrumentalness: 0.00, valence: 0.52, tempo: 0.53
4. Para Pau - GIT - punk-rock
   Features: danceability: 0.61, energy: 0.31, acousticness: 0.05, instrumentalness: 0.00, valence: 0.19, tempo: 0.49
5. Mr. Soul - Buffalo Springfield - psych-rock
   Features: danceability: 0.68, energy: 0.71, acousticness: 0.11, instrumentalness: 0.00, valence: 0.97, tempo: 0.55
6. Every Day - Eric Prydz - progressive-house
   Features: danceability: 0.48, energy: 0.73, acousticness: 0.00, instrumentalness: 0.26, valence: 0.03, tempo: 0.52
7. Brilla Tu Luz Para Mí - Sumo - punk-rock
   Features: danceability: 0.61, energy: 0.71, acousticness: 0.00, instrumentalness: 0.00, valence: 0.70, tempo: 0.43
8. Si Supieras - PXNDX - punk-rock
   Features: danceability: 0.49, energy: 0.96, acousticness: 0.02, instrumentalness: 0.00, valence: 0.43, tempo: 0.42
9. Unwrap You at Christmas - The Monkees - psych-rock
   Features: danceability: 0.68, energy: 0.80, acousticness: 0.25, instrumentalness: 0.00, valence: 0.48, tempo: 0.53
10. Wonderful Christmastime - The Ready Set - punk
   Features: danceability: 0.60, energy: 0.49, acousticness: 0.16, instrumentalness: 0.00, valence: 0.53, tempo: 0.39
11. Therapy - All Time Low - punk-rock
   Features: danceability: 0.32, energy: 0.71, acousticness: 0.03, instrumentalness: 0.00, valence: 0.20, tempo: 0.70
12. Anywhere With You - Afrojack;Lucas & Steve;DubVision - progressive-house
   Features: danceability: 0.61, energy: 0.74, acousticness: 0.07, instrumentalness: 0.00, valence: 0.53, tempo: 0.52
13. All Day And All Of The Night - The Kinks - psych-rock
   Features: danceability: 0.56, energy: 0.96, acousticness: 0.51, instrumentalness: 0.00, valence: 0.71, tempo: 0.56
14. Me Without Us (with Justin Jesso) - Matisse & Sadko;Raiden;Justin Jesso - progressive-house
   Features: danceability: 0.52, energy: 0.61, acousticness: 0.22, instrumentalness: 0.00, valence: 0.42, tempo: 0.60
15. The Happiest Days of Our Lives - Pink Floyd - psych-rock
   Features: danceability: 0.75, energy: 0.23, acousticness: 0.55, instrumentalness: 0.75, valence: 0.55, tempo: 0.43
==================================================
```

Progressive house and **punk rock** are fundamentally different genres with distinct characteristics and origins

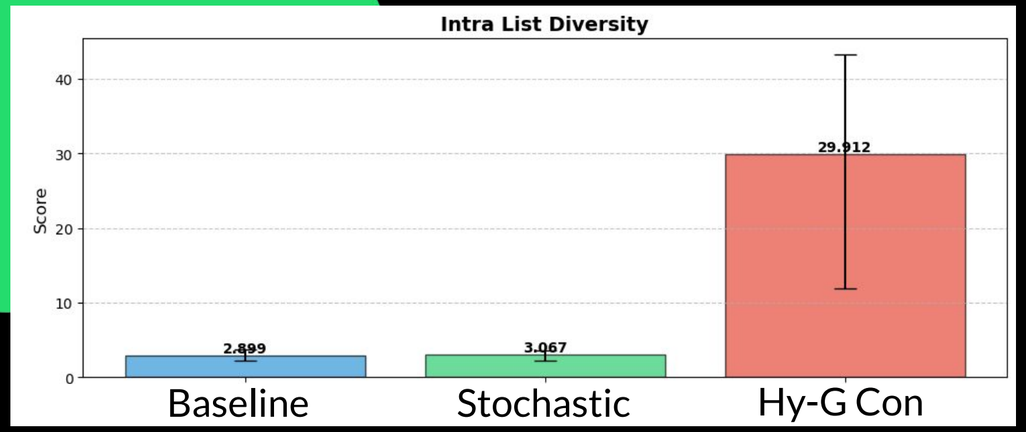
Knowledge Graph Results

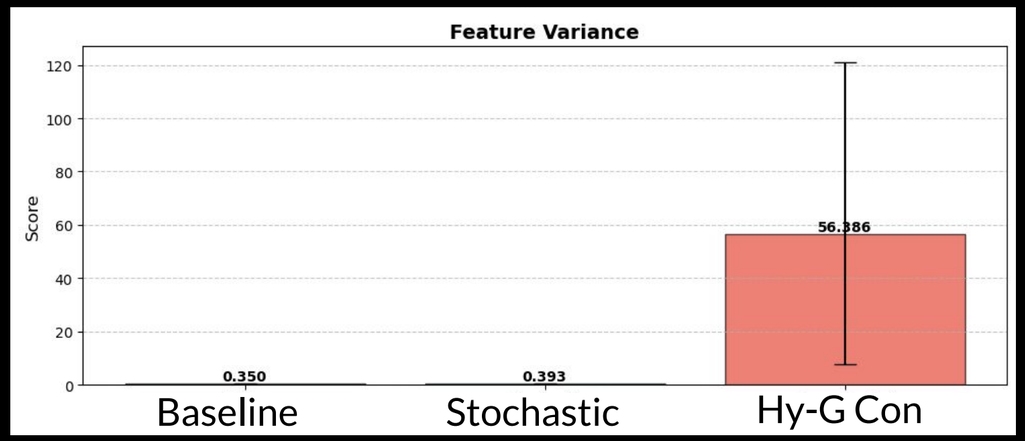
- The knowledge graph's role is to make sure our recommendations explore related genres
- Have predefined genre maps to be able to traverse through all 110 genres
- Results show the recommendations are able to explore multiple similar genres smoothly

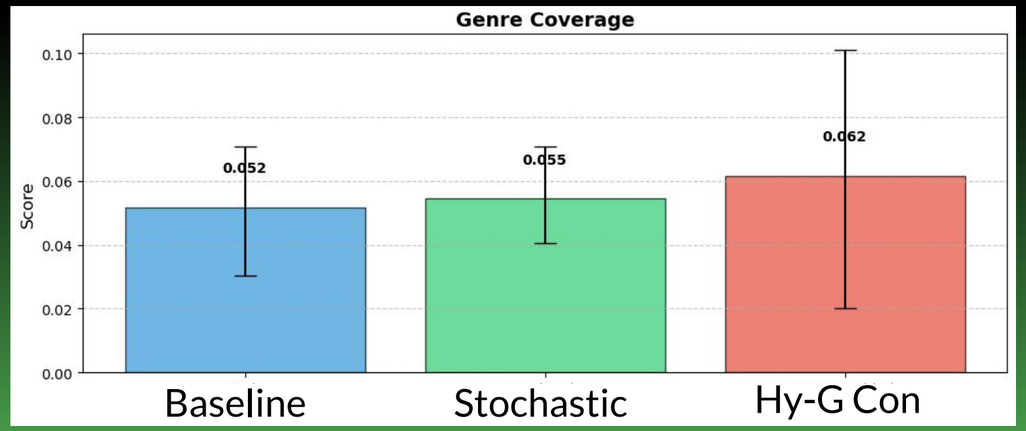
```
['punk-rock',
 'hardcore',
 'rock',
 'punk-rock',
 'goth',
 'punk-rock',
 'hardcore',
 'punk-rock',
 'hardcore',
 'punk-rock',
 'metal',
 'punk-rock',
 'rock',
 'punk-rock',
 'metal']
```

Ex. Results Starting from 'punk-rock'

Preliminary Results







Results over 30 playlists of 15 songs each

Each playlist trio generated from same start

Find trade off between diversity & variance

Potential Next Steps

- An additional metric to evaluate satisfaction
- Consider extending hybrid approach to our clustering method
- Evaluate effects of playlist length on metrics

Thank You

★★★★★

Metrics

Deciding how to analyze playlists in the context of diversity and relevance

Top Artists

1 Artist Name

2 Artist Name

3 Artist Name

4 Artist Name

5 Artist Name

DATA

COLLECTIO

DATA

OLLECTIO

DATA

Traditional Recommender Systems

Memory- vs Model-Based

Memory-Based

Create a utility matrix to store pairwise information
→ consider the preferences of an item's neighbourhood.
Updating requires recomputing pairs.

Model-Based

Pre-train a model to quickly recommend a group of items
→ clustering-based techniques.
We adopt this approach for scalability.

1234 mins

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

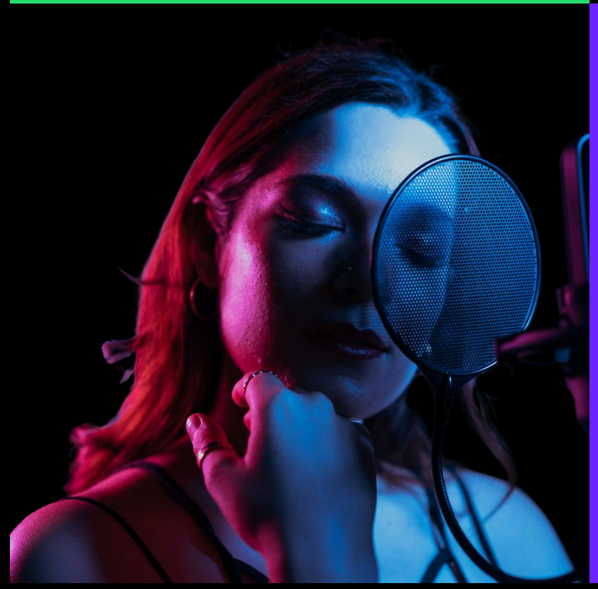
5678 songs

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Baseline Results

Progress on experiments ran for content based filtering graph and knowledge graph

14.5M
Streams



6.2K
Listeners

Hybrid Graph Results

- We wanted to leverage the benefits of content-based filtering with a knowledge graph to improve recommendations
- Combination of the models into the HyG-Con model gave promising results

All | Music | Podcasts | Audiobooks

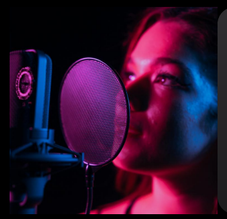
Most Played Tracks

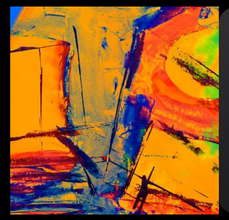
From top tracks to unforgettable moments, here's a snapshot of your musical journey.

Album #1

Album #2

Album #3


Album #4


Album #5

Album #6

Album #7

Album #8

01	Artist Name
02	Artist Name
03	Artist Name
04	Artist Name
05	Artist Name

Your Most Loved Artists

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

01

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt.

02

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt.

03

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt.

04

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt.

RESOURCE PAGE

This presentation template uses the following free fonts:

Titles :
League Spartan

Headers :
Lato

Body Copy :
Lato

You can find these fonts online too.

COLORS

#23DC6D #5D19E8