Transitional playlists for new musical discoveries

A Network Tour of Data Science

Team 36:
Martial Bernard-Michel, Ludovic Coullery
Victor Faraut, Christophe Muller

Introduction

- Dataset: Free Music Archive → 100’000 music tracks
  - Artist, genre, album, ...
  - categorical/temporal features
  - Metadata curated in .csv files (github.com/mdeff/fma)
- Goal of the project:

  “Create a playlist connecting two tracks based on different features”

- Main libraries: Matplotlib, Numpy, Pandas, networkX

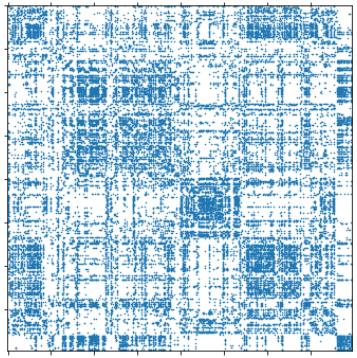
Pre-processing

- Using only the tracks in the top 10% of popularity for each genre
- Selecting genres that we want (Rock, Jazz, Classical, ...)

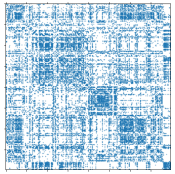
Features used:

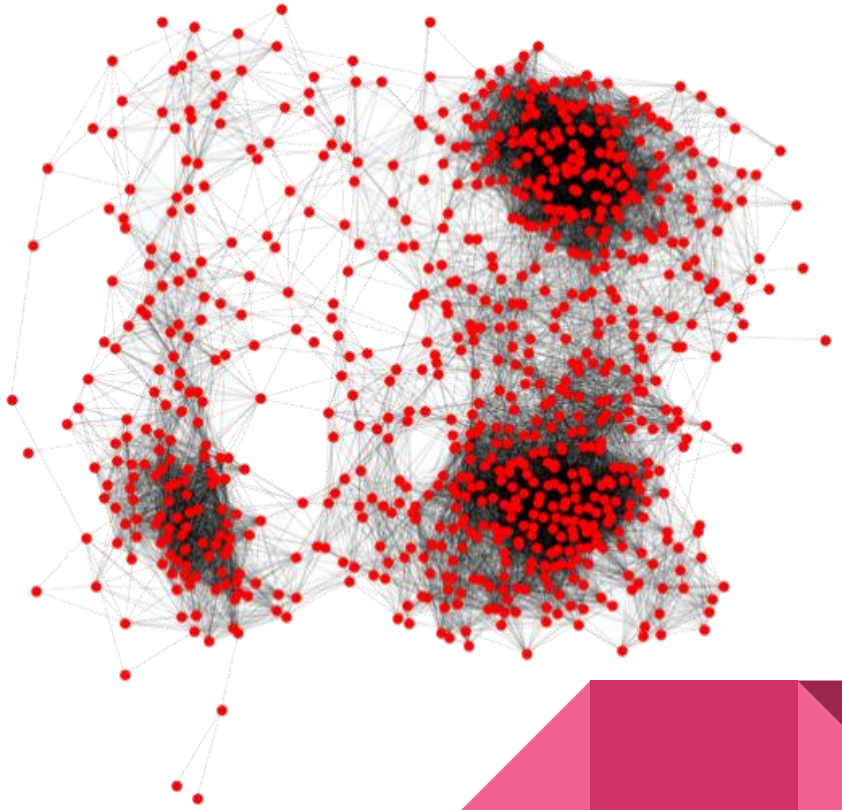
- Danceability
- Acousticness
- Instrumentalness
- Artist discovery

Compute the euclidean distance



Graph





Features used:

- Danceability
- Acousticness
- Instrumentalness
- Artist discovery

Weights threshold : 88.5%

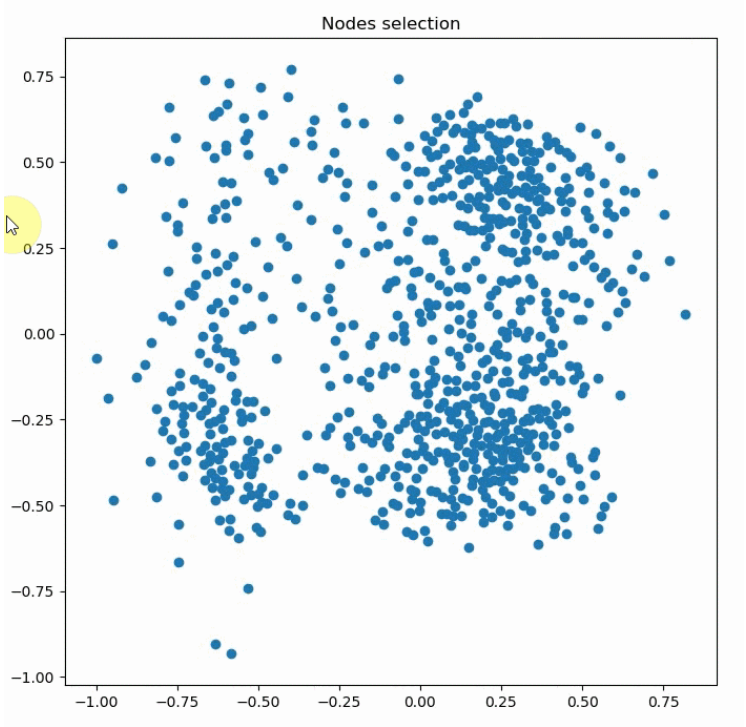
Biggest connected component:

- 795 nodes
- 16’786 edges
- 18% connectivity

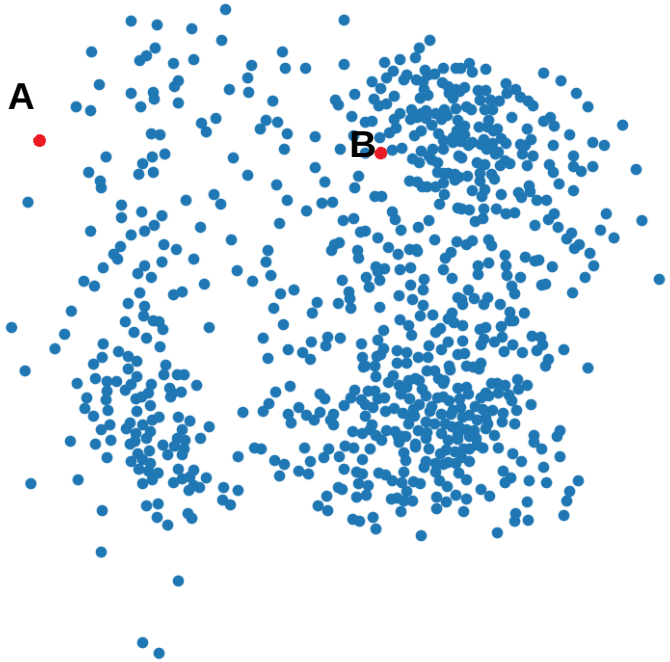
Layout : Kamada kawai

Node selection

- Interactive node selection using the Matplotlib library and callback
- Choose the song/artist/genre you like !



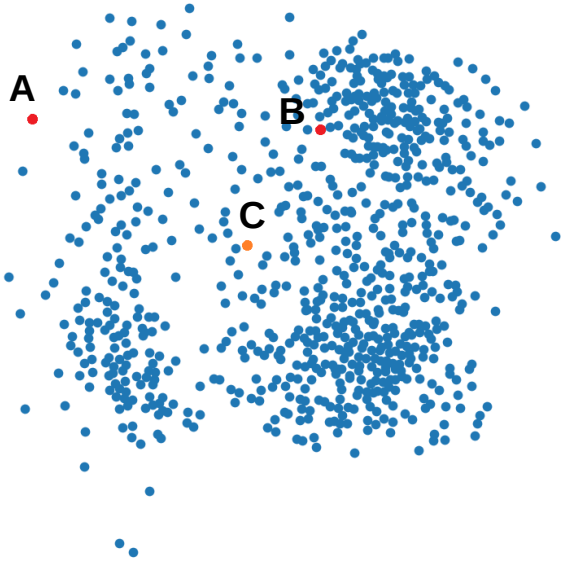
Node selection



A : Tombs : To Cross the Land (Rock)
B : Audible intelligence : Steppin’ In (Hip-Hop)

Path computation

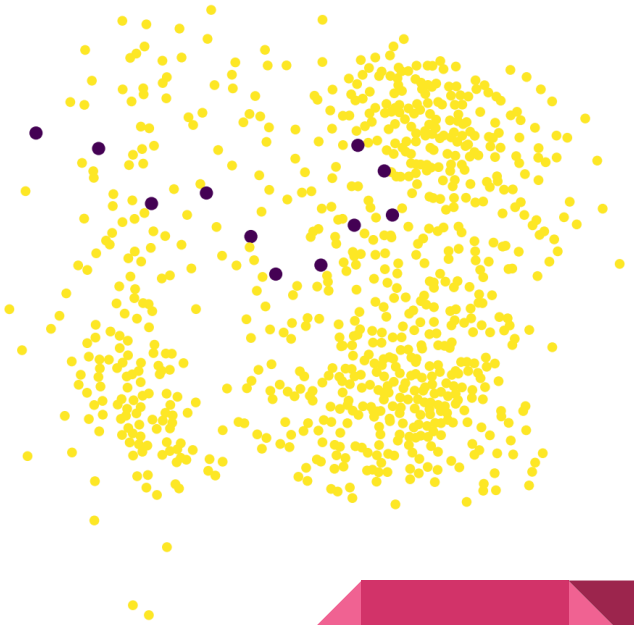
- High number of edges
  - High **number of possible paths** between two nodes
  - Extremely high **computation time**
  - First limitation: Graph diameter of 10



- Need reduction
  - Playlist size ∝ distance between start (A) & end (B) nodes
  - Two half-paths instead of a big one
  - “Random” midpoint selection (C)

Smoothest path

- Smoothness computed on different signals (selectable):
  - temporal features: variance/mean
  - categorical features: tempo, energy, instrumentalness, accousticness…



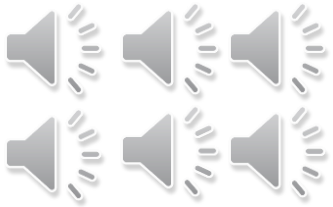
- Smoothness of a path = sum of the smoothness for all signals over the path

$$f^\intercal L f = \|\nabla_{\mathcal{G}} f\|_2^2 = \sum_{i \sim j} W_{ij}(f_j - f_i)^2$$

Playlist generation

- Transition with different genres → Goal reached !
- We used the average and the standard deviation of the temporal features as signals

artist	title	genre	standard_dev	average
Tombs	To Cross the Land	Rock	208.852634	44.013989
Bardo Pond	Absence	Rock	187.787273	36.666708
Throwing Muses	Say Goodbye	Rock	165.135419	35.728926
Lorenzo's Music	Indian Summer	Rock	178.305285	39.406680
Ultimate Painting	Winter In Your Heart	Rock	171.859449	35.945745
Anitek	Broke Fashion	Electronic	275.776705	55.938209
Ultimate Painting	Ten Street	Rock	272.117378	53.033893
El Niño del Parking	Los Dolores de Juana	International	223.700492	52.655255
Lasswell	Dig Deep Mix	Hip-Hop	203.046682	50.888559
2NRO8OT	Crash Of The First Interstellar Channel	Electronic	101.837313	32.836562
Audible Intelligence	Steppin' In	Hip-Hop	291.898491	66.613125



Errata: report

In section 5 (Results) we said we used “the average and standard deviation of the temporal features as metrics”.

A manipulation error in the notebook had the consequence of not taking into account the average (only for the results shown in the report).

Thank you for your attention!

Questions?