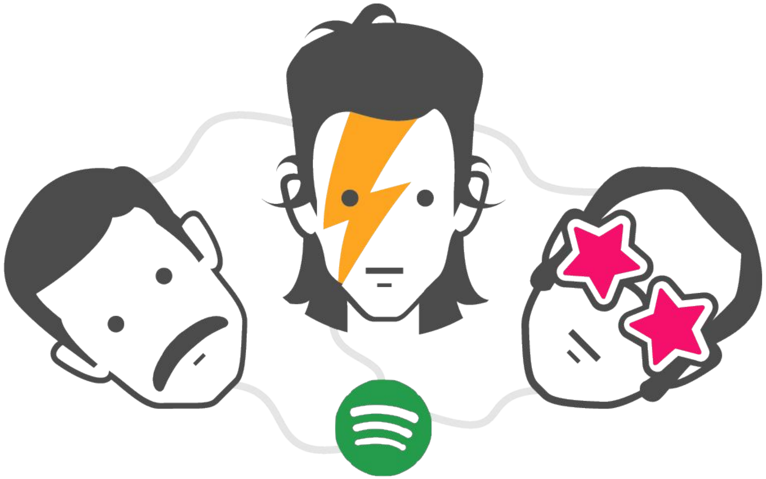


FRIENDS WILL BE FRIENDS

A NETWORK TOUR OF MUSICAL FRIENDSHIP

Team 33: Tobias Barblan / Laura Bujouves / Liana Mehrabyan / Jeremy Wanner

Some are happy, some are sad, oh we got to let the music play

Our starting questions:

- Are friends in real life more likely to share similar musical tastes?
- Can we be friends without listening to the same music?
- Are my tastes unique or do I have somebody with whom I could share?

What data can we use to analyze these musical relationships?

- An existing database:
  - Free Music Archive
  - MusicBrainz
  - The metabase Million Song Dataset
  - ...

The idea:

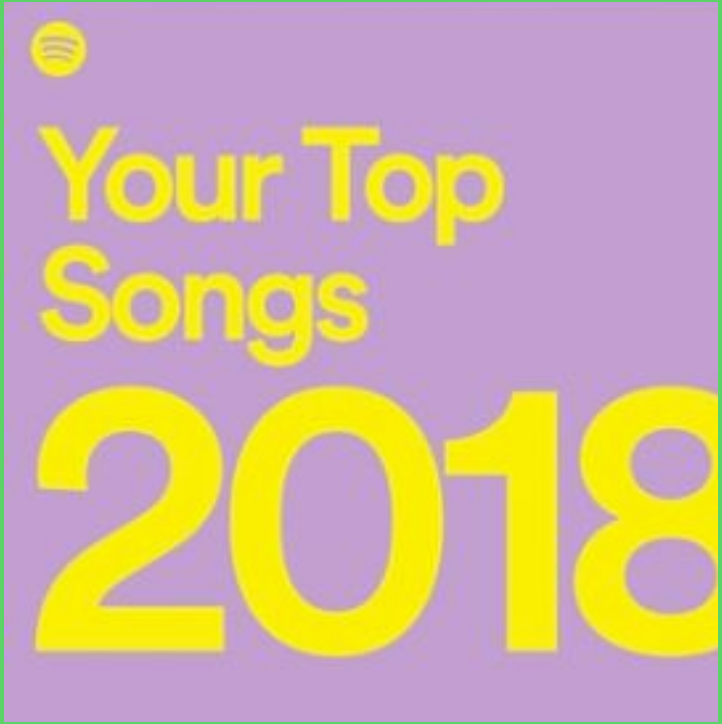
Spotify API

+

Spotipy

=

Top100_18 Database



https://spotifywrapped.com

The Top100_18 Database

25 friends

1274 artists

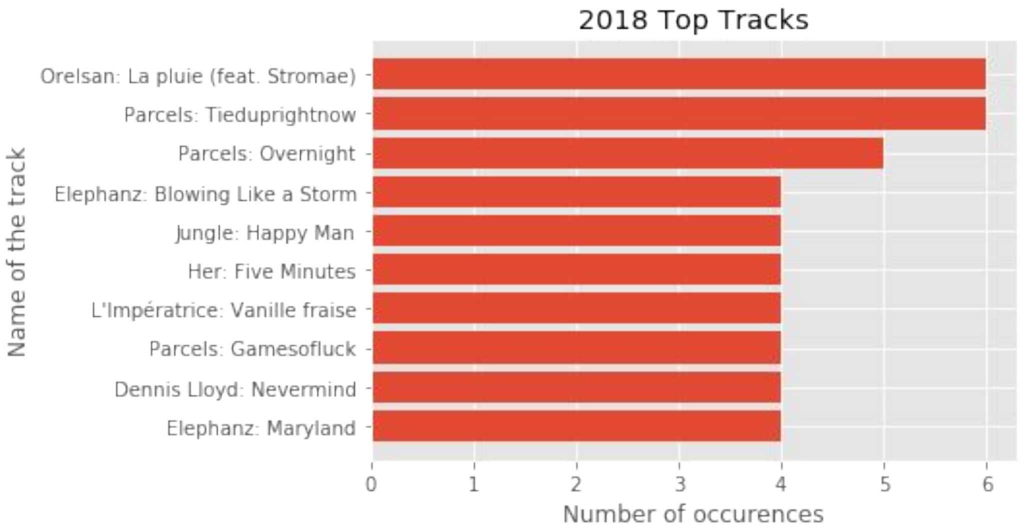
2207 songs

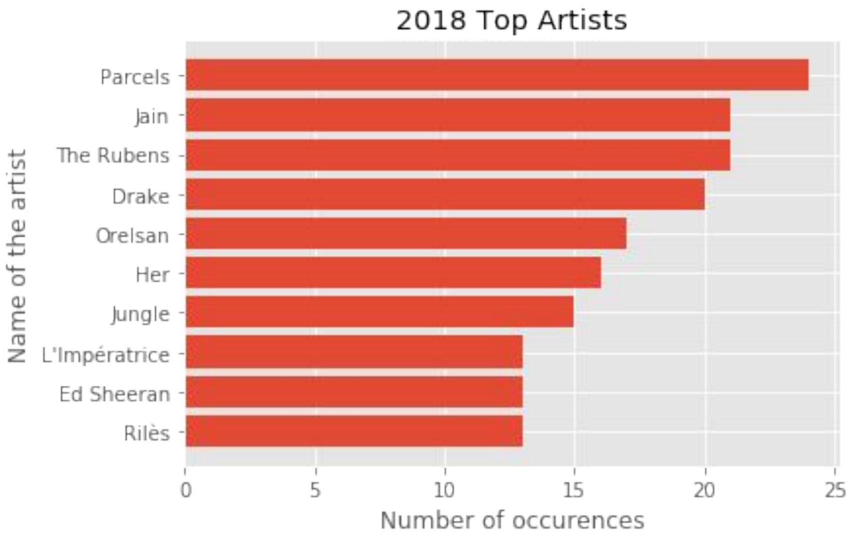
Artist	Key
Track Name	Loudness
Album Name	Mode
Album	Speechiness
Release Date	Acousticness
Track Number	Instrumentalness
Track Popularity	Liveness
Track Duration	Valence
Danceability	Tempo
Energy	Genres

•••

“You can’t always get what you want, but if you try sometimes well you might find”

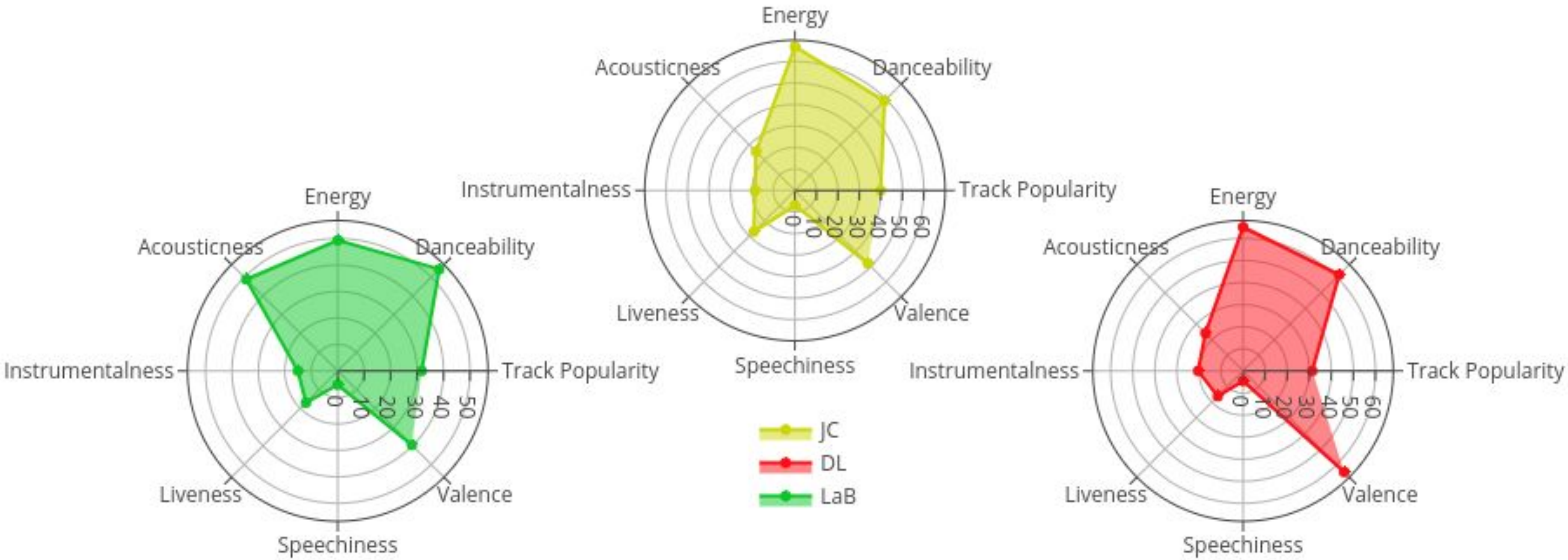
Just because a record has a groove doesn't make it in the groove





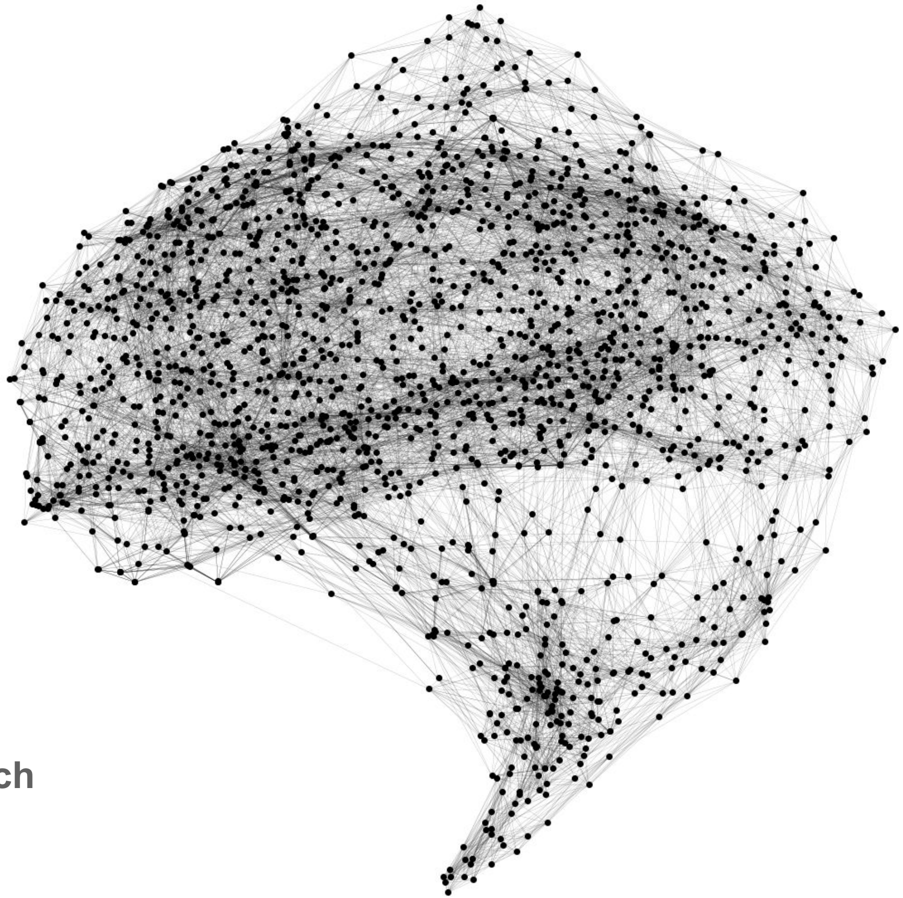
How does it change at the individual level?

Here are some of music's pioneers that time will not allow us to forget



Does it relate at another level?

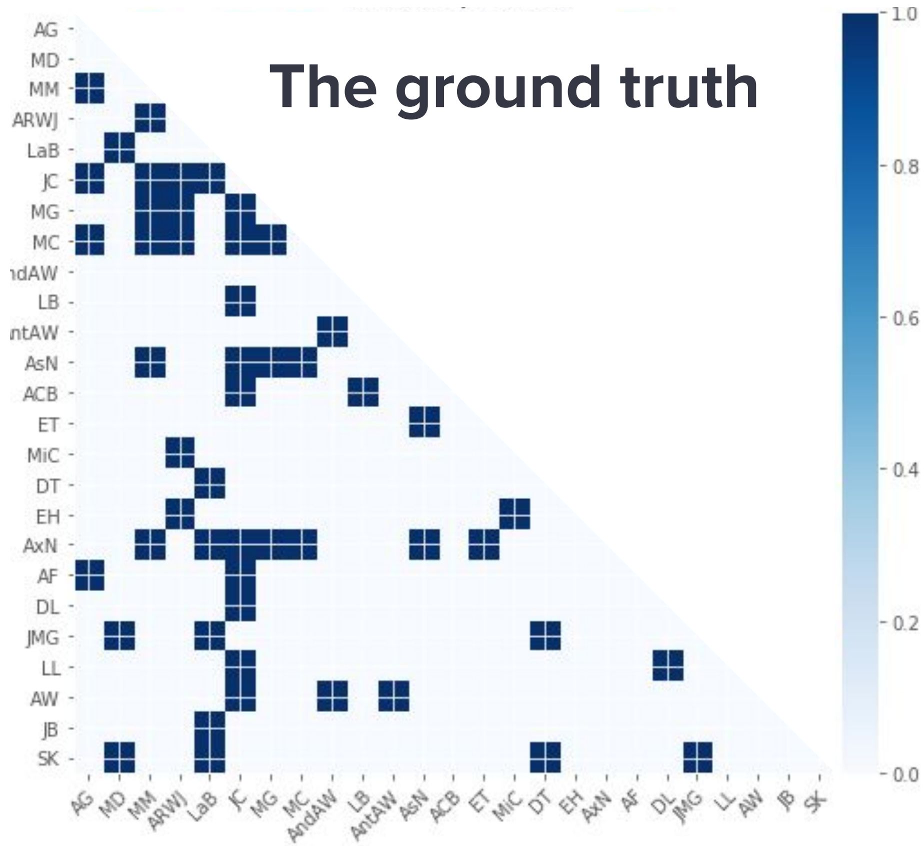
The Songs network

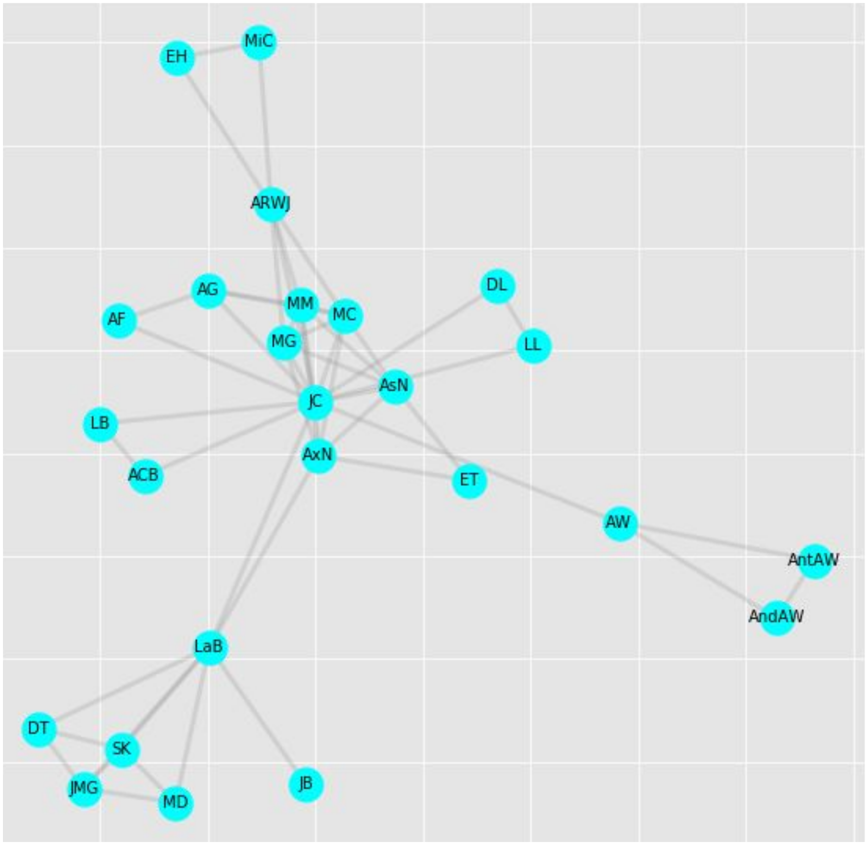


All musical tracks in the world are similar to each other, but some pairs match better than others.

Make it last forever, friendship never ends

The ground truth





The Spotify Graph

A tripartite graph:

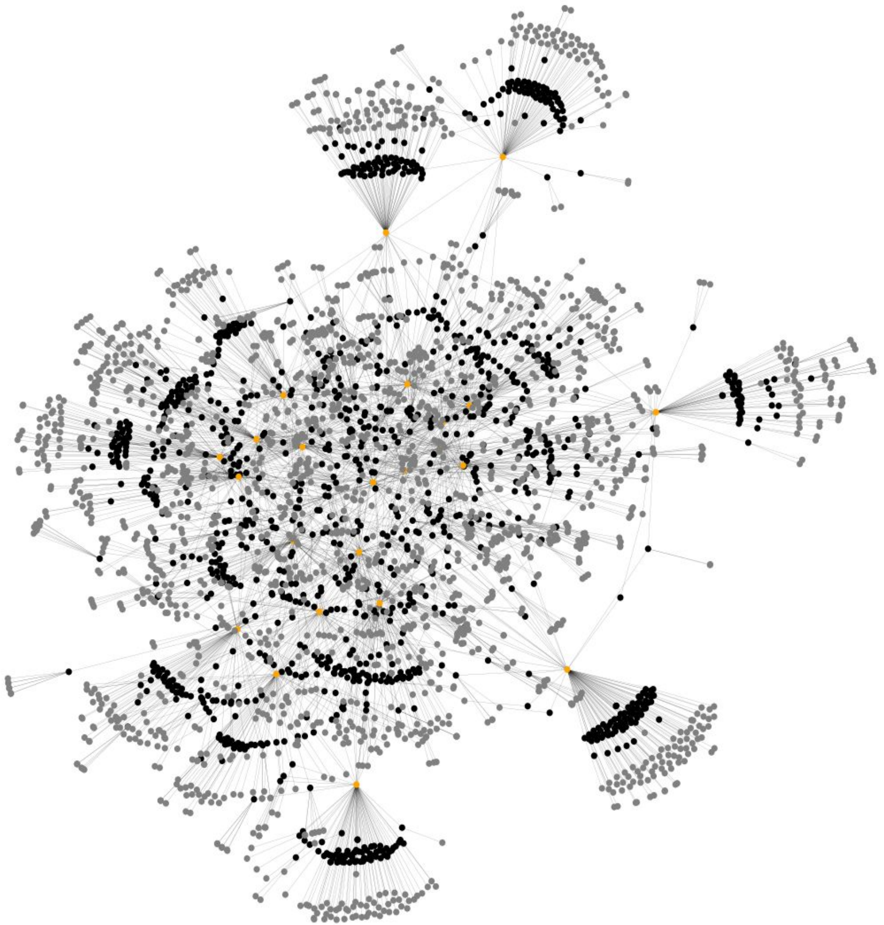
Number of nodes: 3499

Number of edges: 3926

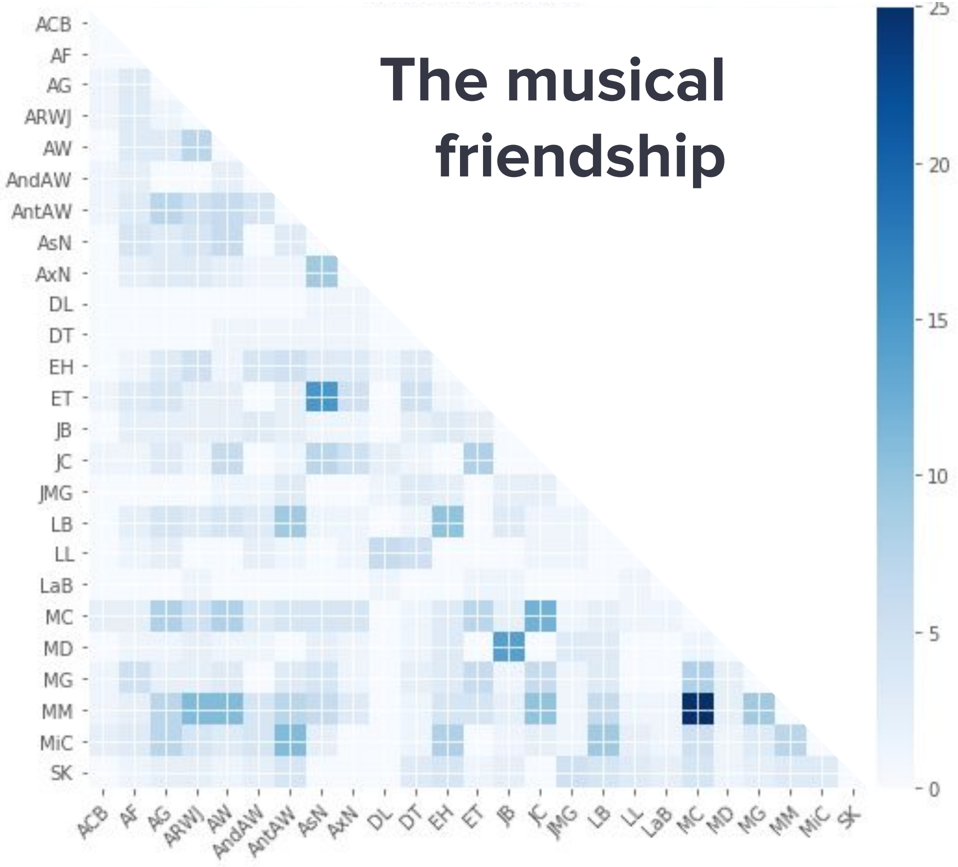
Idea:

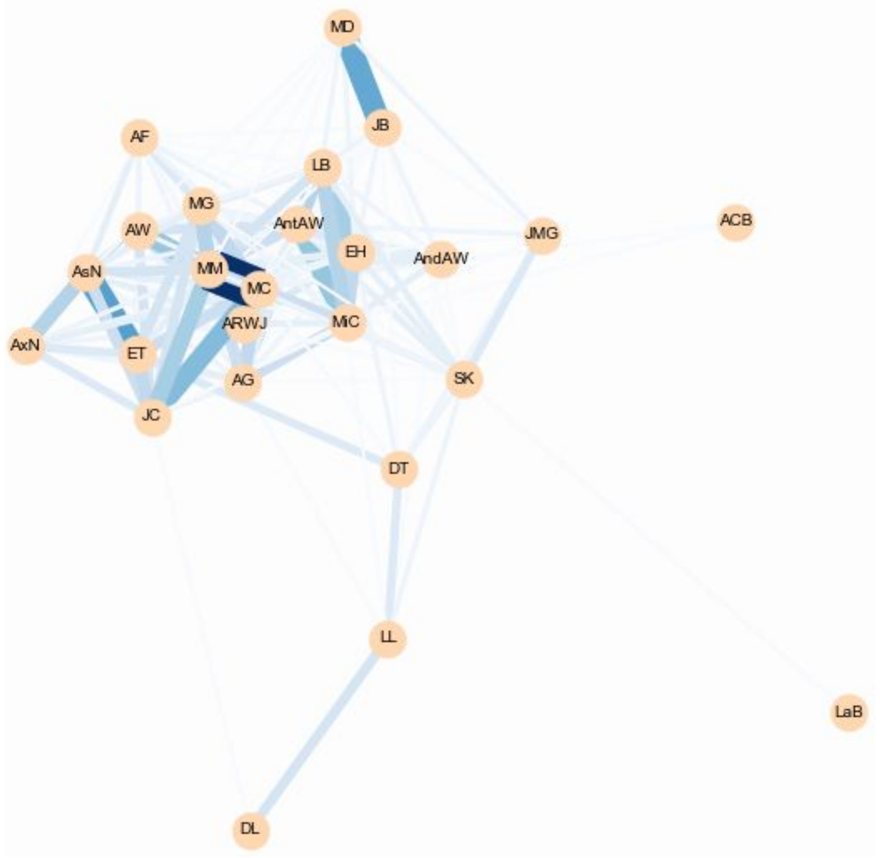
Use the shortest path length to compute the similarities between users

The higher number of shortest path of minimal length, the closer you are with the other user

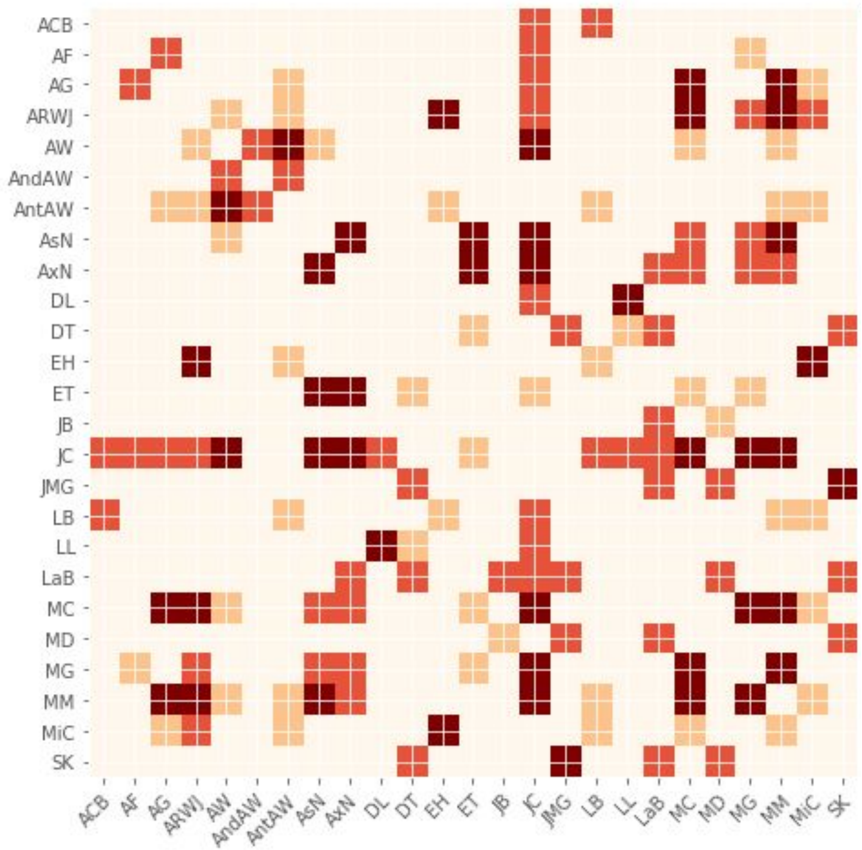


Friends will be friends?





Friends will be friends, right 'til the end!



22 both sides matches

23 reality only matches

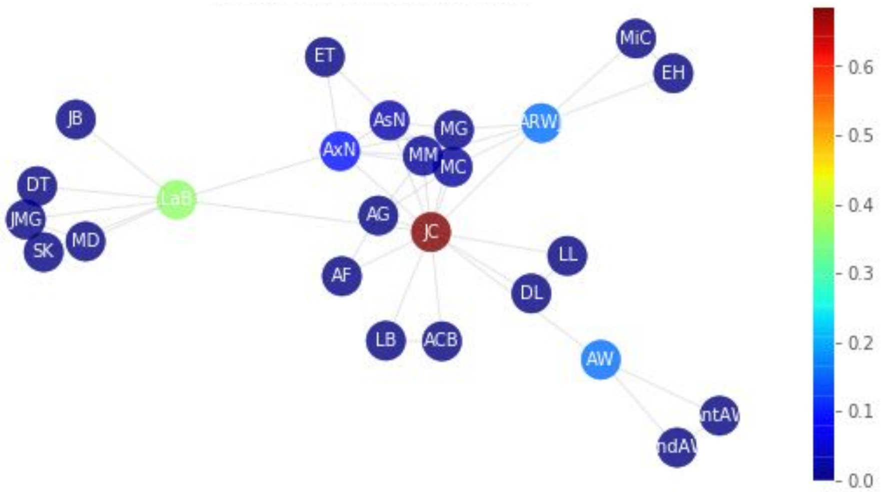
29 music only matches

&

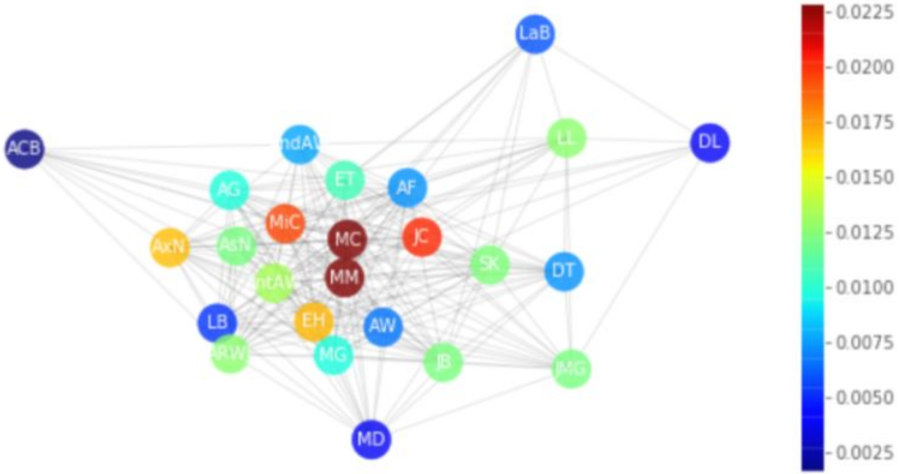
251 still possible relations!

Far across the distances and spaces between us

Betweenness: the sum of the fraction of all-pairs shortest paths that pass through the given node



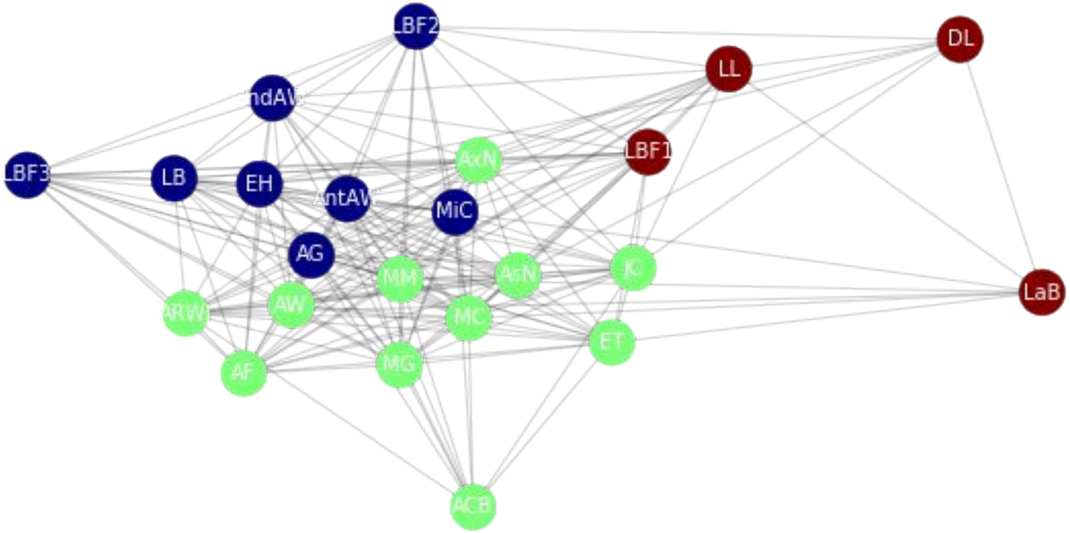
Ground truth



Musical Friendship

Because we are your friend, you'll never be alone again

Clustering using the **Louvain method**: optimisation of the modularity that measures the density of links inside communities compared to links between communities.



Brown: **Users based outside Switzerland**

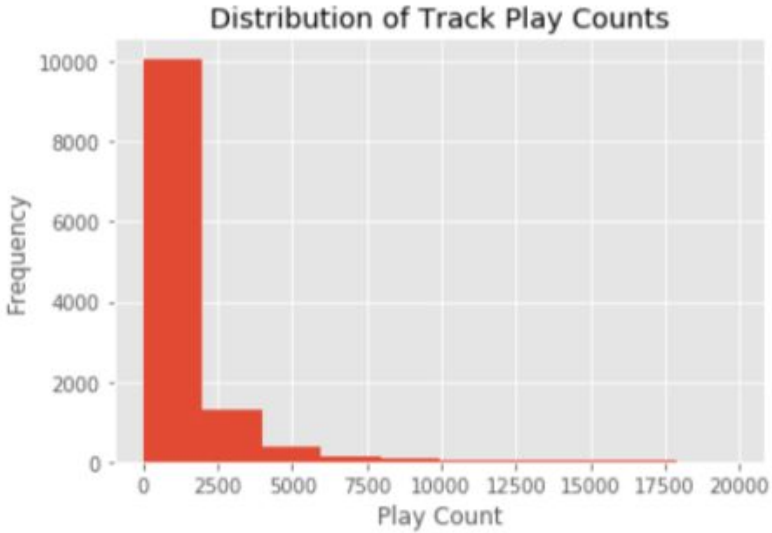
Blue: **Non-EPFL users**

Green: **EPFL users**

Data-Driven Tools for Artists and Producers

Predicting Song Popularity

Variable of interest: Track Play Count Problem



Transformation to a Classification

Main Idea:

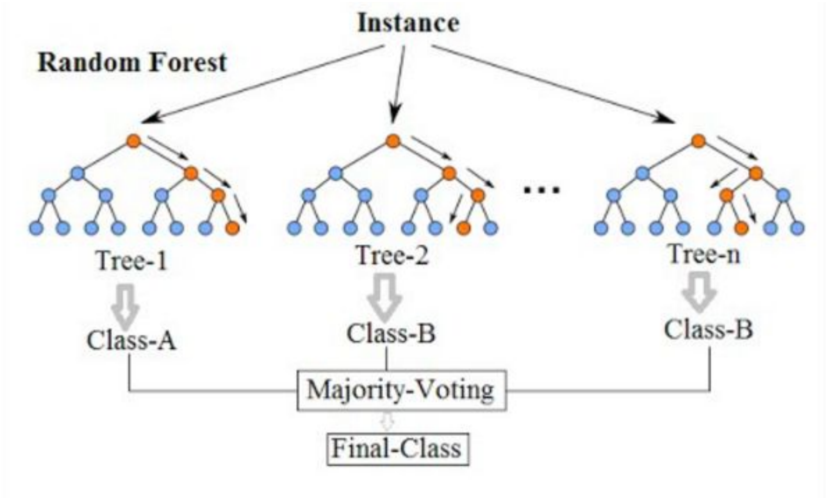
- Divide the ordered population into five subsets using quintiles.
- Assign scores 1-5 based on the quintile subset a track falls into.

Outcome:

- Multiclass classification problem.

Song Popularity Prediction: Random Forest Classifiers

RF with all features to assess feature importance

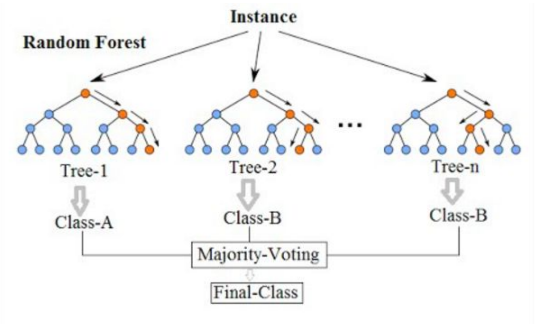


Feature	Importance Score
Interest	0.3683
Acousticness	0.0486
Artist Hotness	0.0471
Track Duration	0.0468
Artist Popularity	0.0466

Song Popularity Prediction: RF and KNN

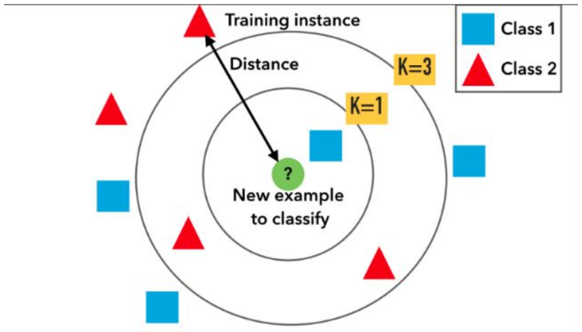
Random Forest Classifiers and K-Nearest Neighbour Algorithm with 5 important features

Random Forest



Accuracy: 80%

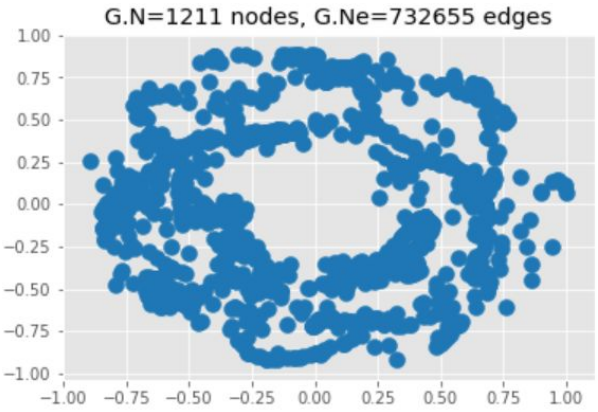
KNN (k=6)



Accuracy: 76.4%

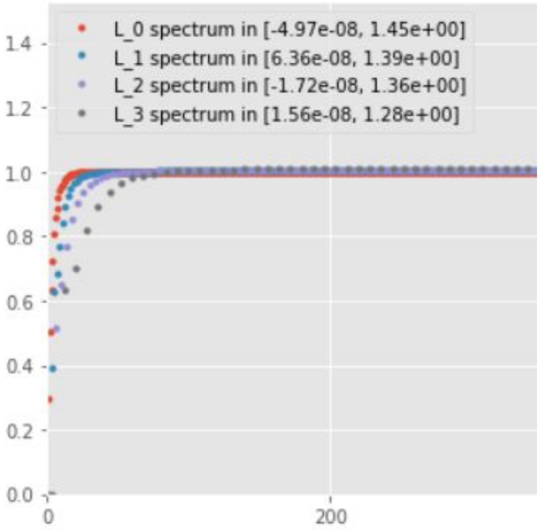
Using ConvNet to Predict Popularity

The Data: reduced features (top 5 based on importance), 10% of dataset



Adjacency matrix of songs created using euclidean distance

High Dimensional Irregular Domains Represented as a Graph



validation accuracy: peak = 53.31, mean = 51.86

Defferrard M., Bresson X., Vandergheynst P.,
Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering, arXiv:1606.09375 (2017)

The show must go on: what's next?

- Improve predictions using full dataset and expanded features such as librosa audio features
- Generate new meta data using Natural Language Processing on genre tags and incorporate it in the learning algorithms described above
- Start with one track and see how many artist hops are needed to reach another friend on Spotify (Stanley Milgram's small-world experiment)

- Reduce the number of favourite tracks per user from 100 to 10 or less and see how the new results compare

-

You really got me?

Questions

Teachers, leave them kids alone!

References

1. Defferrard M., Bresson X., Vandergheynst P., Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering, arXiv:1606.09375 (2017)
2. Klusowski Jason M., Complete Analysis of a Random Forest Model, arXiv:1805.02587v5 (2018)
3. Newman, Mark EJ, Power laws, Pareto distributions and Zipf's law. Journal of Contemporary physics vol. 46 no. 5, pp. 323–351 (2005)
4. Spotify 'Top 100 Songs of 2018' playlist taken from spotifywrapped.com (2018)