

Project: This Is My Jam

Presented by Charlotte Friis Theisen,
Aymen Gannouni, and Antoine Wehenkel

25.01.2018



— A Network Tour of Data Science —



Agenda

1. Introduction
2. Data acquisition
3. Data exploration
4. Data exploitation
5. Conclusion

What is This Is My Jam?

- This Is My Jam is a social music network that existed between 2011 and 2015
- Users could post a favourite song as a jam
- Users could follow each other, like and comment on each other's jams
- The website went into archive mode in September 2015



Project Goals

- Collect the data and improve the quality
- Explore the networks built from the dataset
- Address the following research questions?
 - How is the evolution of jams over time?
 - Does the social network reveal communities that share the same music preferences?
 - If not, can we construct a better network which does?

Data Acquisition

- The dataset was collected on archive.org
- It contains 3 tsv files
 - jams.tsv
 - likes.tsv
 - followers.tsv
- The jams, likes and users were anonymized
- It has over 2 million jams posted by 101 thousand users
- Some songs have a corresponding spotify URI

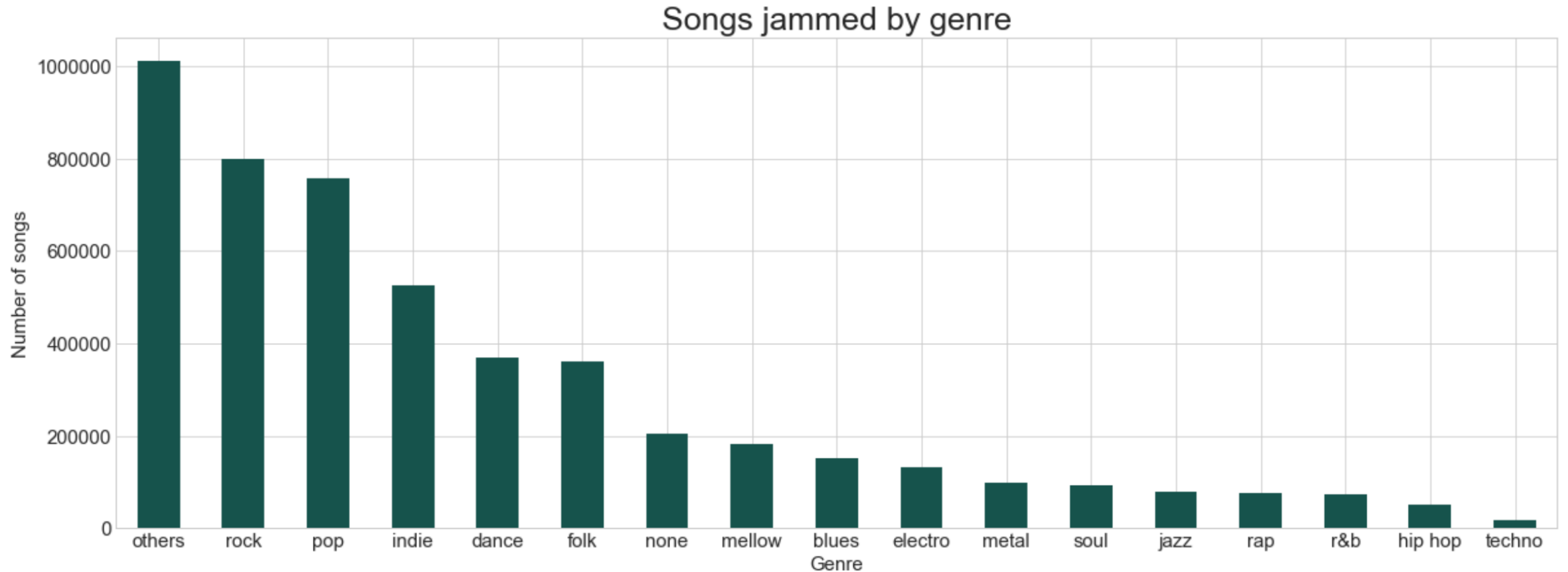


Data Cleaning and Enhancement

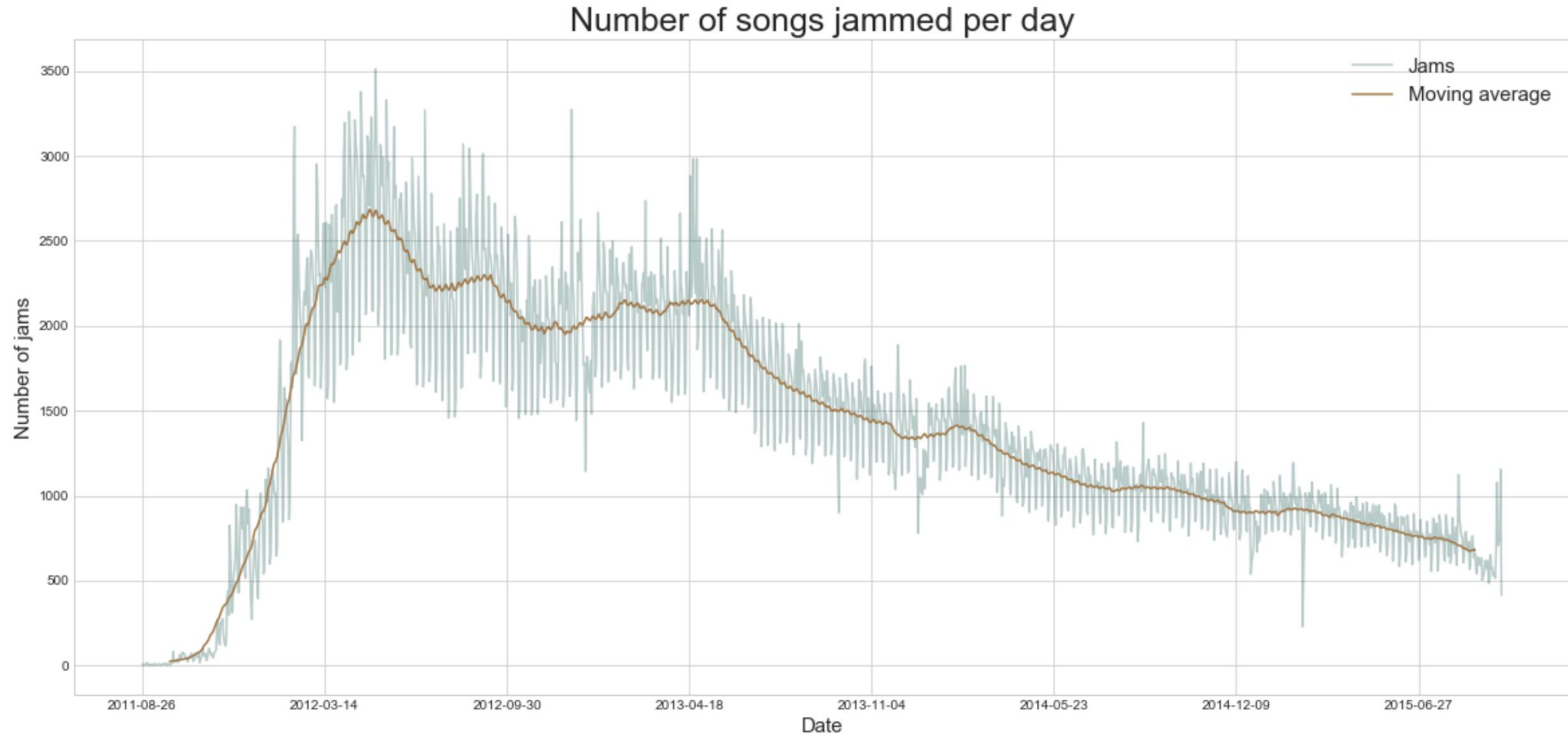
- Removed songs which could not be identified by the spotify API
- The API helped to correct wrong song titles and to fill in the missing values
- The API endpoints were also used to extract the genres, release year, and audio features



Descriptive Analysis

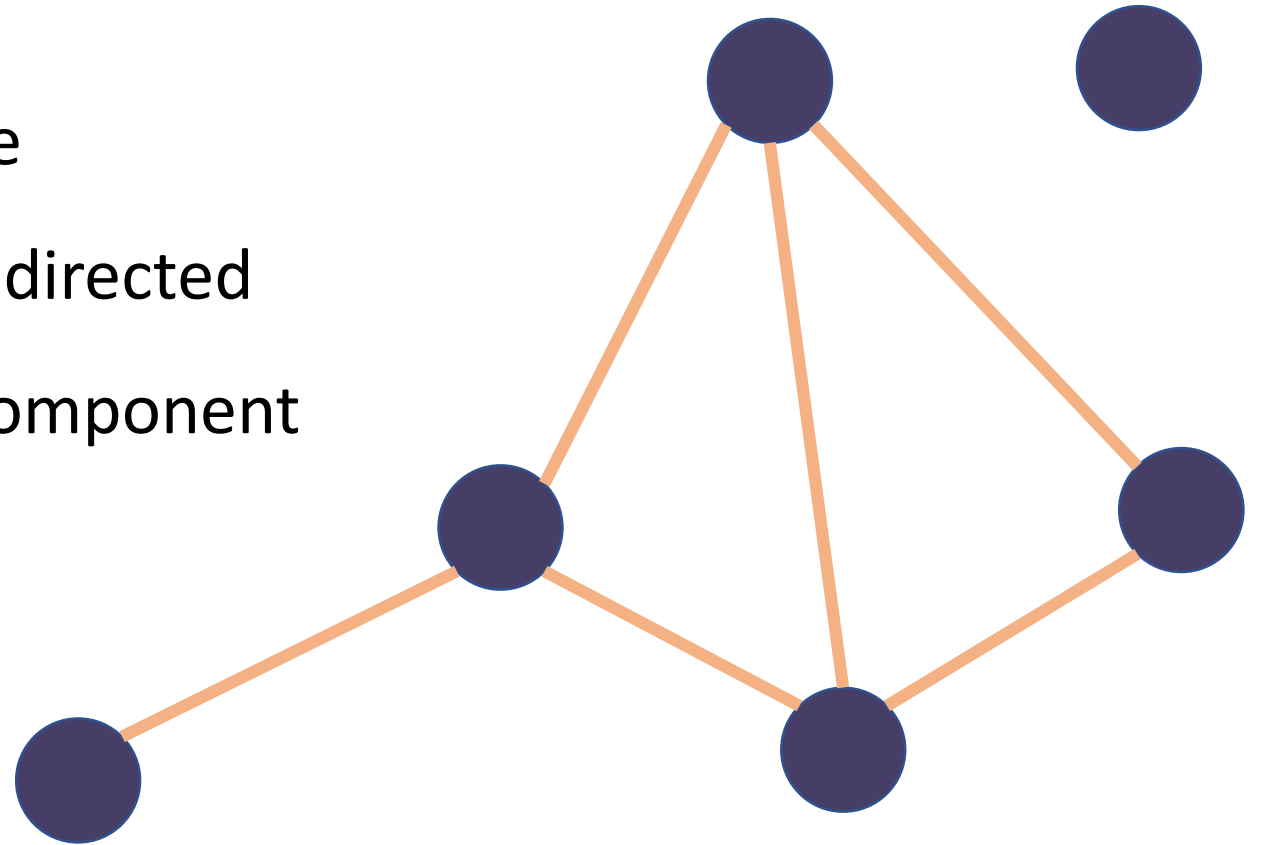


Descriptive Analysis Cont'd

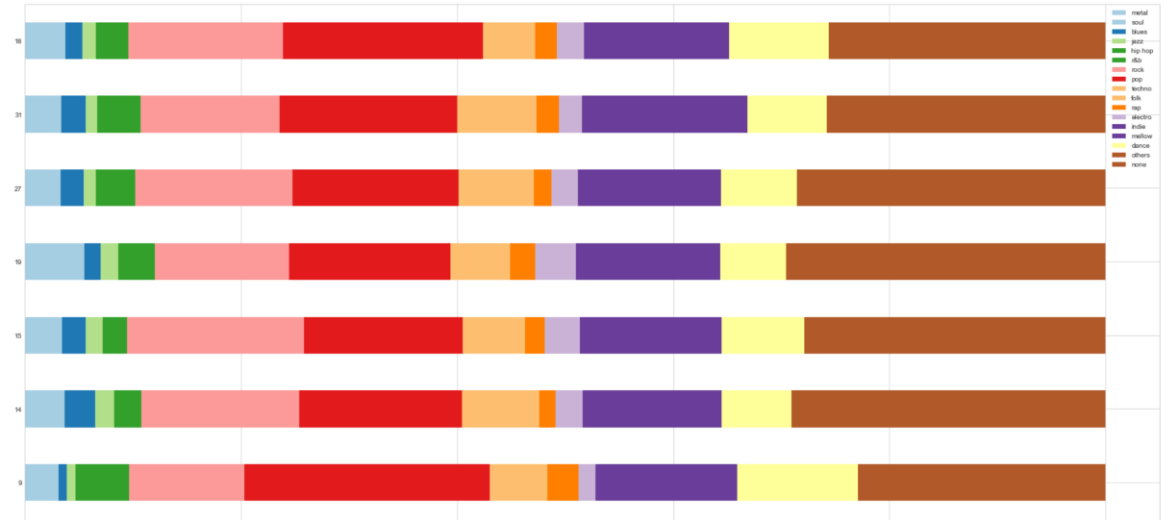
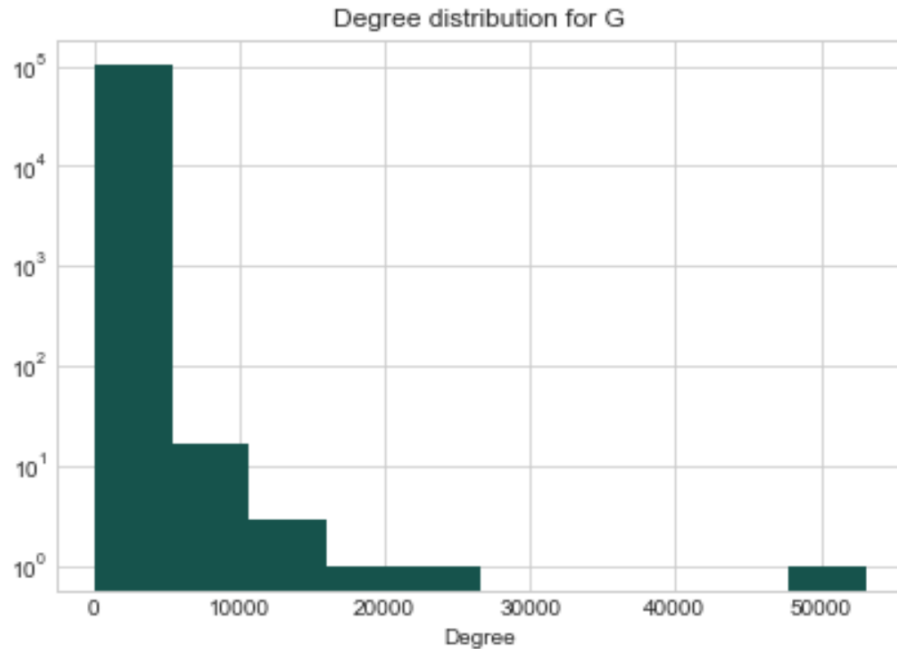


Building the Social Graphs

- Each user is represented as a node
- Edges are based on followers – undirected
- Extracted the largest connected component



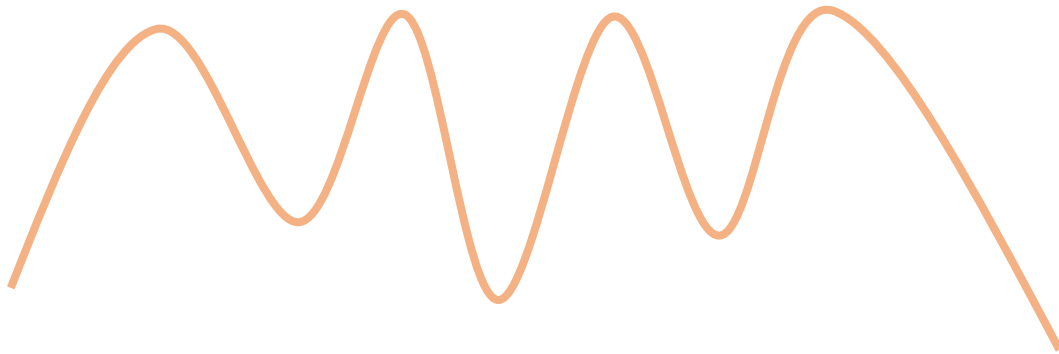
Unweighted Graph Characteristics



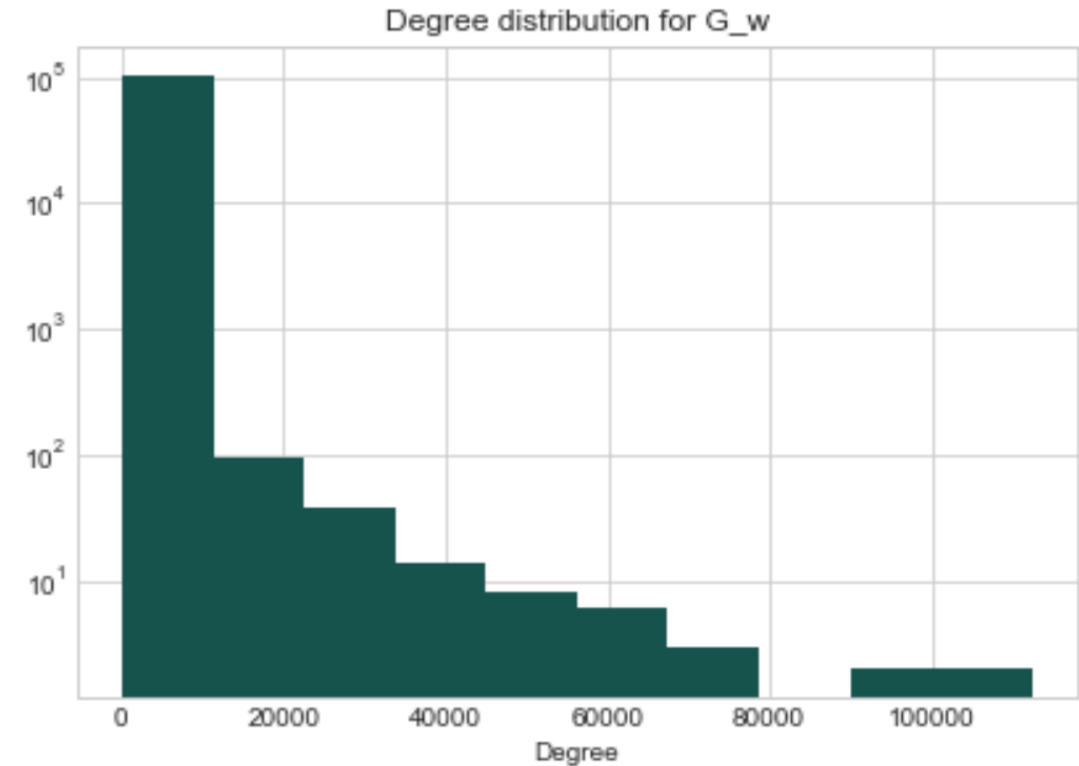
No connection between communities and their preferences in genre

Weighted Graph Characteristics

- Weighted by the jam likes
- **Nodes:** 100,518
Edges: 1,294,160
Average degree: 123.00

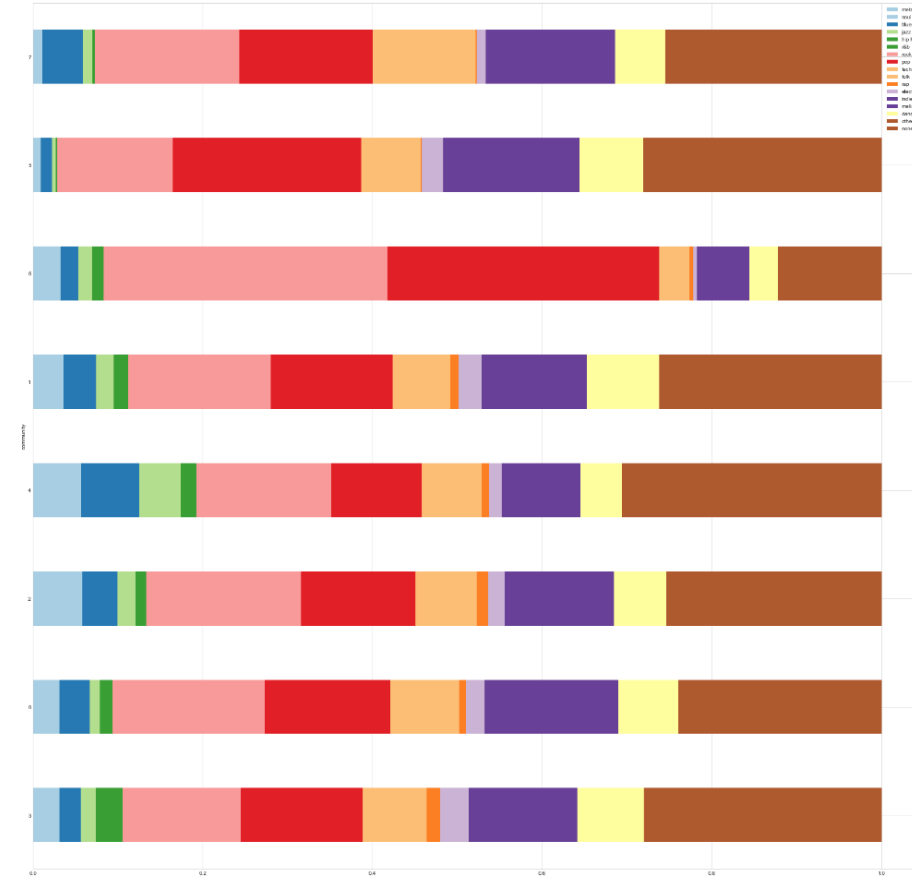
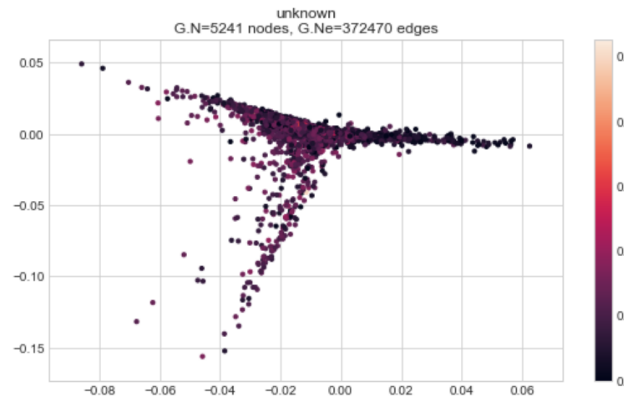
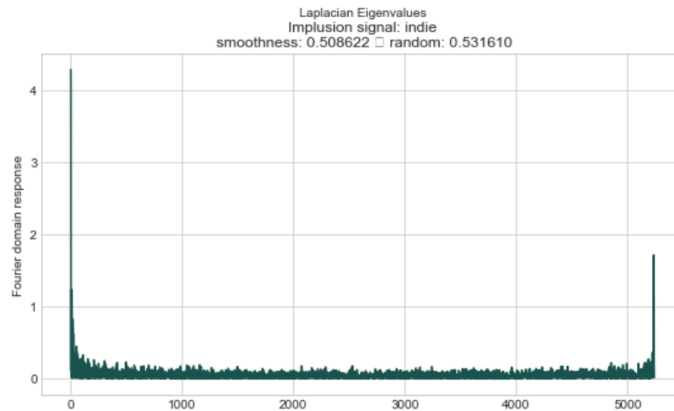


- Timestamp of jams as signal



Social Network Analysis

- Based on the weighted social network & pruning
- Communities have different music preferences
- However signal smoothness is not satisfying

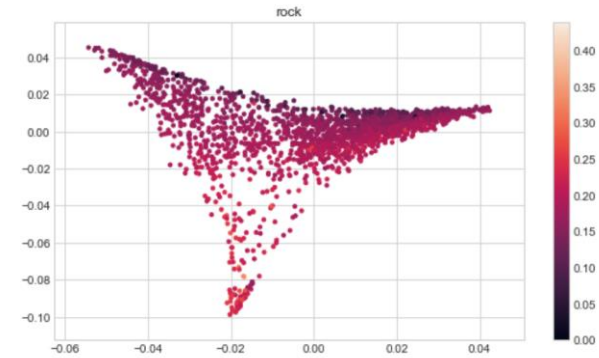
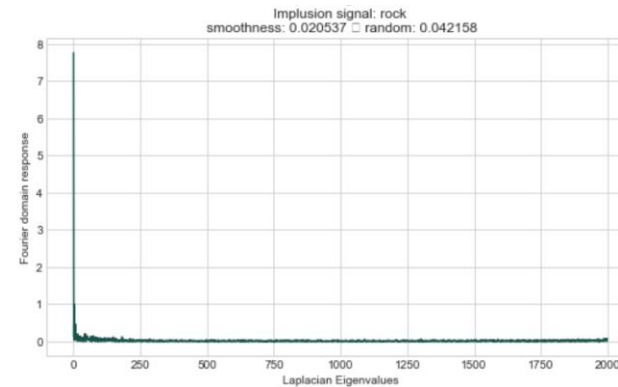
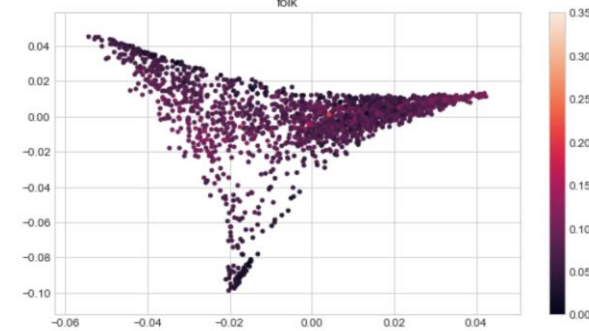
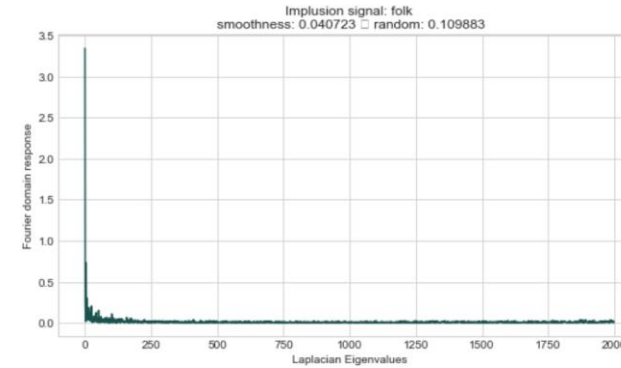


Like-Based Network

- New network to solve the encountered problems
 - Select the most active users (Likes and Jams)
 - Each user is defined by the genre of music he liked
 - Build the complete graph based on the similarity between users (cosine distance & Gaussian Kernel)
 - Pruning of the network

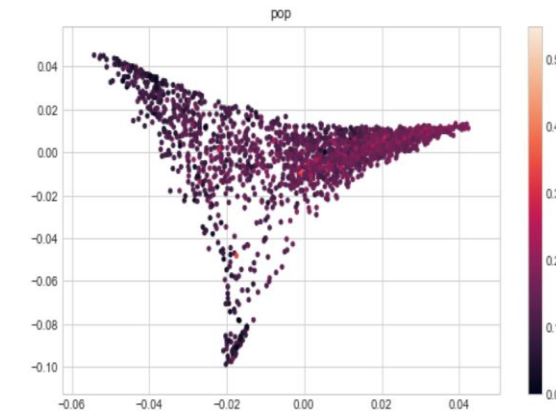
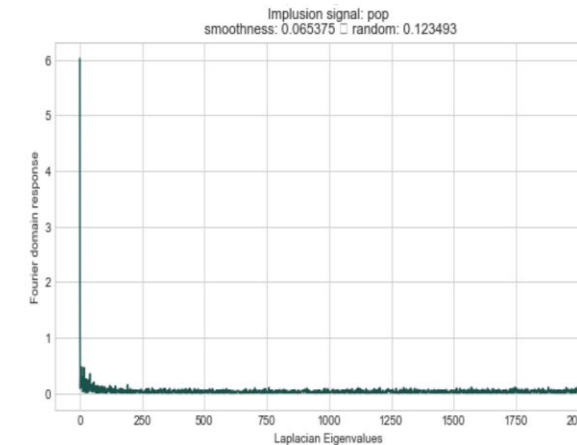
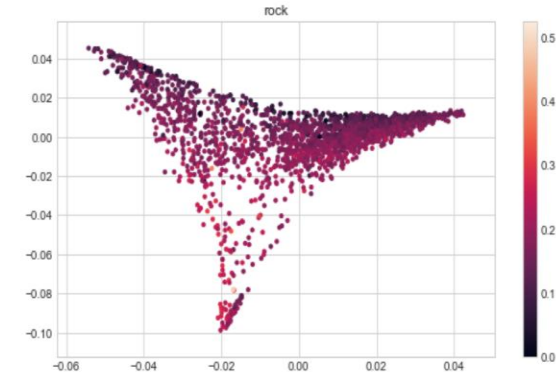
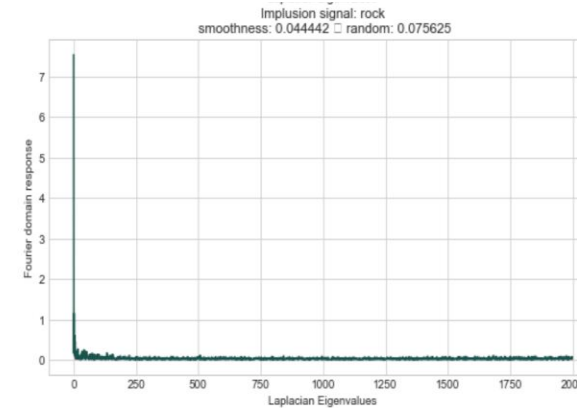
Likes Genre Signals

- 17 signals – Each of these corresponds to a genre (Rock, Indie, Others...)
- Not surprisingly, the signals propagate smoothly in the network



Jams Genre Signals

- Also 17 signals, but based on the genre of the jammed music
- Smoothness is kept
- People who like similar jams will also jam similar music
- This network can be built even for silent users, and recommend them which song they could jam



Conclusion

- Jams spread analysis is difficult
- Social network communities are not well linked to the music preferences
- However like-based communities are more related to music preferences
- A network based on genre of liked music represents well the genre of jammed music
- Future work: build a recommender system based on user likes

References

1. Shuman, David I., et al. "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains." IEEE Signal Processing Magazine 30.3 (2013): 83-98.
2. <https://archive.org/details/thisismyjam-datadump>
3. <https://www.thisismyjam.com/>

Thanks for your attention!

Questions