Network Tour of Data Science

Project Proposal

Charlotte Friis Theisen   *Aymen Gannouni*   *Antoine Wehenkel*

November 28, 2017

1 Project Description

Our project aims at analyzing a social music network called "This is My Jam ", which was active between 2011 and 2015, it can be accessed here. What motivates us the most is the opportunity we have to highlight the social interactions between people in the network and observe the spread of jams over time. In the course of this project, we will address a couple of research questions which are defined below. At the end of the project, a complete data story will be presented along with all results and conclusions that were obtained from the network analysis.

2 The Dataset

In thisismyjam social network, users could tell what their jam was in a certain week. In total, there are over 533,266 unique jams, which are songs that either derive from video platforms or music streaming services. Furthermore, users could follow each other, like and comment on each other's jams. The dataset contains 2 tab-seperated values(.tsv) files with a total size of 452 Mb. The structure of these is illustrated as follows:

- jams.tsv

  - jam_id
  - user_id
  - artist - As provided by the user
  - title - As provided by the user
  - creation_date - In the format YYYY-MM-DD
  - link - The link to the external source of the jam
  - spotify_uri - Best guess URI as provided by the Spotify API

- followers.tsv

  - followed_user_id
  - follower_user_id

3 Research Questions

Firstly, the project will study the properties of the network built on top of the "follow" links between users. It will put these properties in perspective with the spread of Jam over time in the network.

Secondly, considering the features provided by the dataset and possibly supplementary informations provided by the Spotify API, another network explaining at best the spreading of jams over itself will be built. It will be also studied with spectral graph theory.

Finally, based on the knowledge of the dataset acquired during the analysis of both graphs, a recommender system for users of the network will be built based on graph theory.

4 Project Plan

FIGURE 1 presents the main tasks along with a timeline. All milestones are here marked as the dark green vertical lines. These include handing in the project proposal, handing in the final report and giving the project presentation. The tasks will briefly be described below.

Data acquisition: The first task will involve to get the dataset on hand. It will also require to perform quick hand on the Spotify API in order to be ready if it reveals of need later.

Data exploration: It involves data pre-processing such as cleaning the data. Furthermore, we will here look at the descriptive statistics of the dataset and start creating the network along with the computing key features. A generic code to analyze network will be built.

Data exploitation: In this part of the project the research questions will be analyzed and answered. This will most likely lead to visualizations as well. Tools built in the previous step will be used for the second research question.

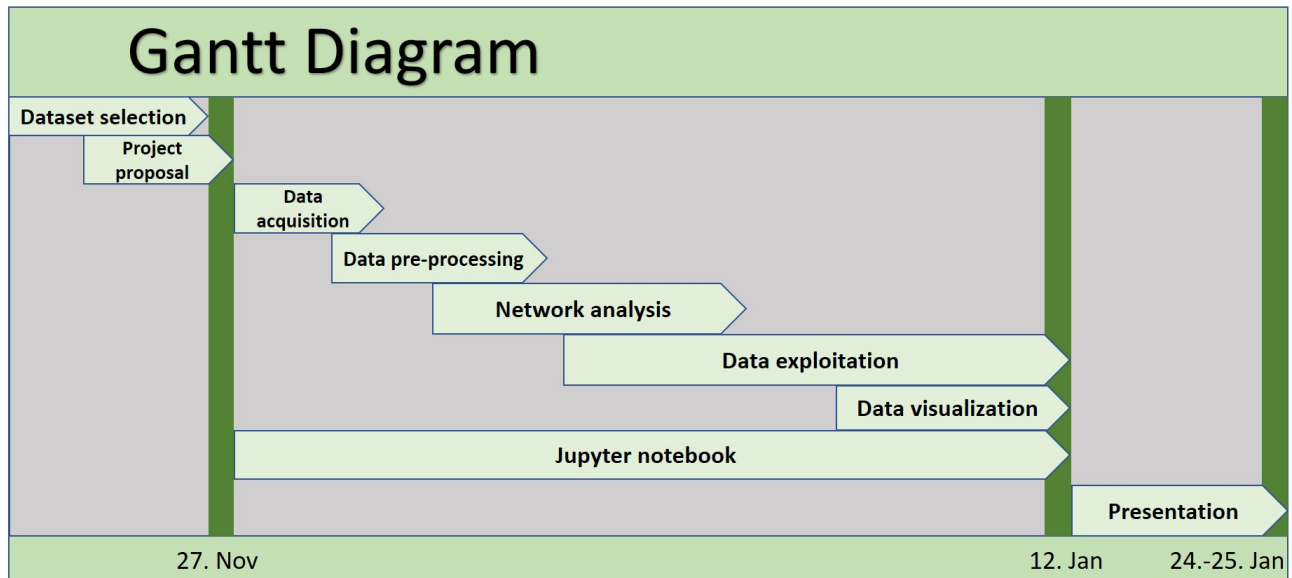**Conclusion**: The conclusive part of the project involves finishing the notebook, summarizing the research questions and preparing the presentation.

Figure 1: Gantt Diagram of the project plan