



Community Detection and Friendship Recommendation
in online Music Streaming Services

Julien CRABIÉ | Paul-Henri CASTETS | Oleg STRUYANSKIY | Chun-Han CHANG
Under the supervision of Fragiskos Malliaros

Motivation and Problem Definition

Some research has already been made on Music Streaming Services (MSS) networks, however, the research found has always been focusing mostly on the recommendation systems of those streaming platforms. Some elements have been omitted and replacing those is what this project will be trying to achieve.

From a user perspective, it is obvious that recommendation systems are relevant, but lack some precision in their suggestions, which can often be disappointing. As algorithms, the recommendation systems also lack the human side that a person-to-person recommendation would provide.

Human recommendation are not necessarily more accurate and don't necessarily lead to better results, but the level of trust is higher between two individuals and one will tend to follow another person's recommendation more often than an algorithm's recommendation. Further friendships could also emerge between 2 passionate music listeners, thanks to community belonging feelings and personality appreciations.

Therefore, from a business perspective, although human recommendation haven't been encouraged so far in online MSSs, it is probably a very relevant path to further improve the user experience, both from a musical standpoint, and from a human standpoint.

Thus, we will try to answer the following question:
Can we detect some communities of musical experts to provide more accurate and more human recommendations to other users of an online musical streaming service?

The goal will be to find those communities of experts and to link them to other users as friends, based on those users tastes, so that they can then provide further recommendations.

As the datasets studied contains data for different countries, we will also assess our model across countries and see if it is possible to link people from different countries together. We will also compare the experts and user tastes between countries and see if they vary in a similar pattern.

Dataset

The datasets used in this project represent the friendship networks of the users of the music streaming service Deezer. The data was collected in November 2017 for 3 European countries: Romania, Croatia and Hungary.

Country	#Nodes	#Edges
RO	41,773	125,826
HR	54,573	498,202
HU	47,538	222,887

In these datasets, the nodes represent the users and the edges are the mutual friendships. In particular:

- The csv files contain the edges: nodes are indexed from 0.
- The json files contain the genre preferences of users: each key is a user id and the genres loved are given as lists.

Genre notations are consistent across users. In each dataset, users could like 84 distinct genres. Liked genre lists were compiled based on the liked song lists.

The nodes were reindexed in order to achieve a certain level of anonymity.

Methodology

To address the problem, we propose to detect communities within the graph and then find the most influential users within every community who we think will be eligible for giving recommendations to other members. We will use eigenvector centrality for ranking user importance and try different algorithms for community detection to see which one will perform best:

- Girvan-Newman's Method
- Louvain Algorithm
- Spectral clustering

Evaluation

In order to understand how well we clustered our communities we have to evaluate how good our algorithm is doing. To do that, we need to test two different things which are the accuracy of our algorithm and its computational efficiency. We will be using different metrics for measuring the quality of clustering:

- modularity
- purity
- connectivity

Furthermore, we are going to use the second part of the dataset, which shows preferred genres of music for different users, to try and assess whether our results make sense or not. We also plan to visualize the data to get a better understanding of this network.

References

- B. Rozemberczki, R. Davies, R. Sarkar and C. Sutton. GEMSEC: Graph Embedding with Self Clustering. 2018.
- Guo, C., & Liu, X. (2015). Automatic Feature Generation on Heterogeneous Graph for Music Recommendation. *SIGIR*.
- Chen, C., Yang, C., Hsia, C., Chen, Y., & Tsai, M. (2016). Music Playlist Recommendation via Preference Embedding. *RecSys Posters*.
- Heitmann, B. (2012). An open framework for multi-source, cross-domain personalisation with semantic interest graphs. *RecSys*.
- Sangkeun Lee , Sang-il Song , Minsuk Kahng , Dongjoo Lee , Sang-goo Lee, Random walk based entity ranking on graph for multidimensional recommendation, Proceedings of the fifth ACM conference on Recommender systems, October 23-27, 2011, Chicago, Illinois, USA