Individual Coursework Submission Form

Specialist Masters Programme

Surname: Kiefer	**First Name: Marius**
MSc in: Business Analytics	**Student ID number: 240052822**
Module Code: SMM638	
Module Title: Network Analytics	
Lecturer: Dr Simone Santoni	**Submission Date: 11/12/2024**

Declaration:

By submitting this work, I declare that this work is entirely my own except those parts duly identified and referenced in my submission. It complies with any specified word limits and the requirements and regulations detailed in the coursework instructions and any other relevant programme and module documentation. In submitting this work, I acknowledge that I have read and understood the regulations and code regarding academic misconduct, including that relating to plagiarism, as specified in the Programme Handbook. I also acknowledge that this work will be subject to a variety of checks for academic misconduct.

We acknowledge that work submitted late without a granted extension will be subject to penalties, as outlined in the Programme Handbook. Penalties will be applied for a maximum of five days lateness, after which a mark of zero will be awarded.

Marker's Comments (if not being marked on-line):

Deduction for Late Submission:

Final Mark: **%**

Music Genres

Similarities, Groups and Social Phenomena
Final Course Project
SMM638 Network Analytics

Marius Kiefer

December 2024

Introduction

This analysis deals with analyzing Deezer music genres and user behavior in Croatia. The goal is to better understand similarities, groups and social phenomena within genres. This knowledge helps us to position new albums and musicians more efficiently.

Analysis Workflow

The data used for this analysis was scraped in 2017 and solely involves users in Croatia. Given information are friendship relationships between users and genres liked by each user based on their liked songs list. As a first step I create a matrix containing information which user (observation/row) liked which genre. Based on this information it is possible to create a second matrix containing the information, how many likes does a pair of users share. In other words, the count of music genres two users' share. Having this knowledge, allows us to create a third matrix. A matrix about how many users like a pair of genres. With this matrix we can assess similarities between genres. The same matrix also allows us to find communities/ groups within music genres. The algorithm used to detect groups is the Louvain Community Detection Algorithm (Blondel et al. 2008). The algorithm operates in two phases, optimizing the modularities of the groups. Modularity measures connection density within vs. between groups. Splitting genres into communities/ groups provides another layer of grounding positioning strategies.
Considering, that we are assessing similarity based on how many likes a pair of genre shares, we also want to know if friendships necessarily influences genres similarities. Based on these insights, Sonic is able to have a better understanding of positioning new albums and musicians. Concretely, a recommendation system for artist, albums, and songs would benefit from this knowledge. Another application could be the promotion of new songs or albums to the right audience.

Results

Starting with the similarity of music genres based on the number of users liking both genres, we can see in Table 1 that Dance and Pop are mutually liked by 24,604 users. It is visible that the first five pairs of genres all include Pop. Most of the pairs are consisting out of popular genres instead of more niche genres. To split genres into communities or subgroups we are using the Louvain Community Detection Algorithm. This algorithm is based on optimizing the modularity of the network's groups. The algorithm depends on the size and granularity for communities, defined by the resolution parameter. To determine this parameter, I compare each modularity for a range of resolution values.

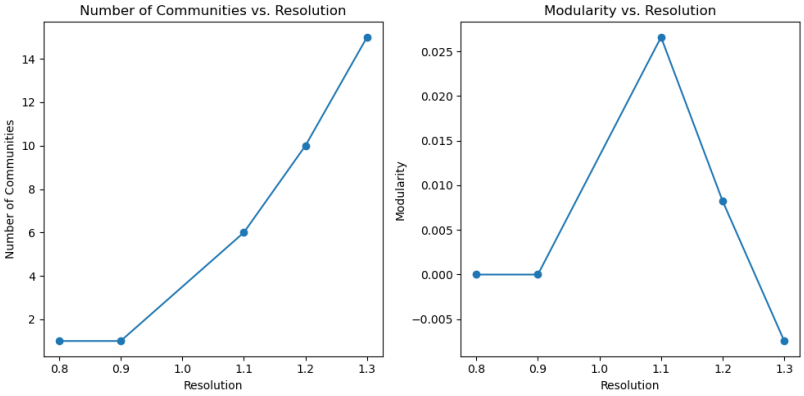

Figure 1: Modularity vs. Resolution

The highest level of modularity is achieved by a resolution of 1.1, resulting in 6 communities within genres. These 6 groups are visualized by the following graph:

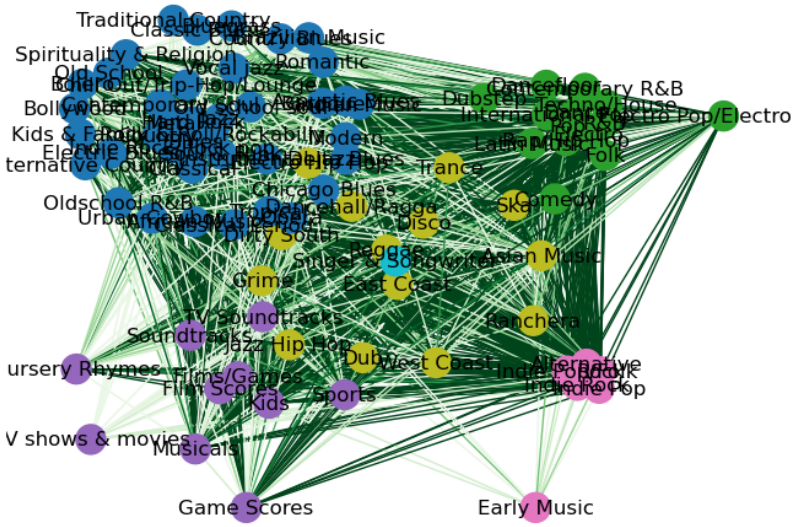

Figure 2: Communities of Genres

Four of the groups are consisting of a similar number of genres. One of the groups only consist out of one genre, Singer & Songwriter. Another group contains proportionally more genres than the other groups. The graph also shows that all clusters are strongly connected with each other through multiple ties.
We measure social ties based on a friendship network. To calculate the proportion of friends for a pair of genres, I first determine the total number of possible pairs among their mutual likers using the combination formula $\binom{n}{2}$. Then, I divide the number of actual friendships within this group by the total possible pairs, ensuring the proportion is zero if no pairs exist. Comparing Dance and Pop (24,604 mutual likers) with Country and Dirty South (75 mutual likers), we observe that the friendship proportion of Country and Dirty South is even higher than the friendship proportion of Dance and Pop.

Pair	**Likes**
Dance, Pop	24604
Pop, Rock	24079
Alternative, Pop	19653
Pop, Rap/Hip Hop	19442
Electro, Pop	18019
Dance, Electro	17917
Alternative, Rock	15917
Dance, Rap/Hip Hop	15791
Dance, Rock	15769
Pop, R&B	14804

Table 1: Similarity of genres based on mutual likes of users

Analysis Interpretation

As an example, to address the underlying business problem, I choose the genre Blues. According to the similarity of music genres, Blues is most similar to Pop, Indie Rock, Rock, Alternative and Hard Rock. Concluding, if a new album, song or artist is introduced to the platform, it is most efficient, if the product gets positioned close to the genres Pop, Indie Rock, Rock, Alternative and Hard Rock.

Pair	**Likes**
Blues, Rock	2961
Blues, Pop	2907
Alternative, Blues	2502
Blues, Indie Rock	2287
Blues, Hard Rock	2113
Blues, Indie Pop/Folk	1930
Blues, Dance	1862
Blues, R&B	1836
Blues, Indie Pop	1690
Blues, Electro	1672

Table 2: Number of mutual likes for genre pairs involving Blues.

Just from a basic understanding of music, these genres are not too similar sound wise but apparently users liking these genres also like Blues. If we inspect the group that Blues is allocated to, we notice that some of the genres are way closer to Blues than the ones from the similarity table. We have Chicago Blues, Classical Blues, Soul & Funk as well as Vocal Jazz. The group also contains a few genres that are not related to each other sound wise. This observation might be caused by the low modularity of the groups (0.025). We have a modularity close to zero, indicating a very weak community structure.

While Dance and Pop have a significantly larger number of mutual likers (24,604) compared to Country and Dirty South (75), the latter pair exhibits a higher proportion of friendships among their mutual likers. This suggests that the smaller, more niche audience of Country and Dirty South may have a tighter-knit social network, with mutual likers more likely to be friends. In contrast, the larger audience of Dance and Pop might reflect a broader, less interconnected network. This emphasizes that genre similarity doesn't solely depend on mutual likers but also on their social ties.

Taking these three layers into account, Sonic has a stronger foundation for determining how to position new content effectively.

References

Blondel, Vincent D. et al. (2008). "Fast unfolding of communities in large networks". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.10, P10008. DOI: 10.1088/1742-5468/2008/10/P10008.