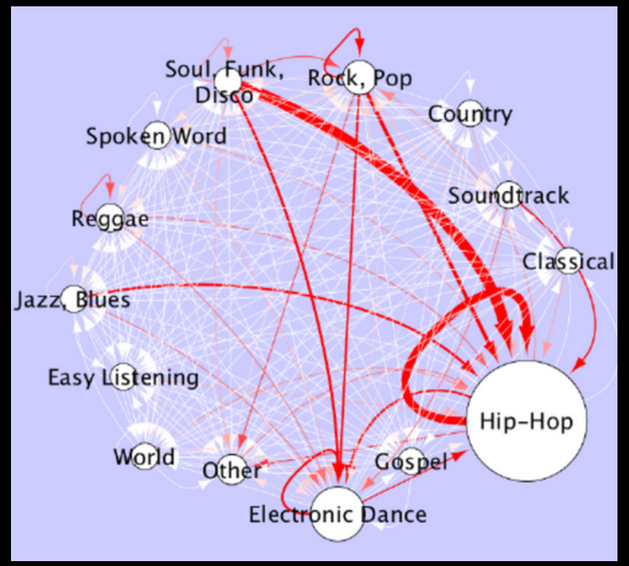
Analyzing Song Relationships with Network Science and Spotify Data

CS414 Spring 2023

Group Members:
Melike Soytürk, Burçak Kaymaz, Emir Kantül, Taha Dora Sonat, Deniz İncereis

Domain of Interest



- The analysis of music data, specifically using network science
- Understand the relationships between songs based on their musical properties
- Examine how these relationships have changed over time

Data Source

- We'll obtain the data for this project from Spotify's API, which includes a vast collection of songs and their corresponding musical properties, such as danceability, energy, key, and more.
- Also Kaggle provides mostly clean and easy to process data.
- Example Kaggle dataset: https://www.kaggle.com/datasets/vatsalmavani/spotify-dataset

kaggle

Main Hypotheses

- **The main hypotheses for this project are twofold:**

1. Songs that are similar in their musical properties are likely to be related to each other in a network.
2. The relationships between songs and their popularity have changed over time, and can be revealed through network analysis.

- Songs that are similar in their musical properties are likely to be related to each other in a network.

- This hypothesis suggests that songs that share similar properties, such as danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, and valence, are likely to be connected to each other in a network. In other words, the network would consist of clusters of songs that share similar properties, with connections between them based on their similarity.

- The relationships between songs and their popularity have changed over time, and can be revealed through network analysis.

- This hypothesis proposes that the relationships between songs and their popularity have changed over time and can be explored through network analysis. For example, songs that were popular in the past may not be as popular today, while other songs may have gained popularity over time. The network analysis can reveal how these changes have affected the relationships between songs in the network, and potentially uncover patterns or trends in the relationship between song popularity and other properties.

Current Literature

Musical Influence Network Analysis and Rank of Sample-Based Music

Nicholas J. Bryan and Ge Wang

- A novel approach to rank and sample music based on a network analysis of musical influence. The authors use data from Whosampled.com to construct a network of influence between music artists based on user listening data.

- The study uses several network analysis techniques to identify the most influential artists in the network and use this information to develop a ranking system for music.

Current Literature

Musical Influence Network Analysis and Rank of Sample-Based Music

Nicholas J. Bryan and Ge Wang

Some of these techniques include:

- **Node degree centrality:** This technique measures the number of connections each artist has in the network, and those with high degree centrality are considered more influential.
- **Betweenness centrality:** This technique measures how often an artist appears on the shortest path between other artists in the network, indicating their importance in connecting different parts of the network.
- **Eigenvector centrality:** This technique measures an artist's influence based on the influence of their connections, so an artist connected to many other influential artists would have a higher eigenvector centrality score.
- **Modularity clustering:** This technique identifies groups of highly connected artists (clusters) within the network based on the similarity of their connections.

Current Literature

Musical Influence Network Analysis and Rank of Sample-Based Music

Nicholas J. Bryan and Ge Wang

- Overall, the paper suggests that network analysis can be a useful tool for ranking and sampling music and generating personalized recommendations for users. The study provides insights into the social dynamics of the music industry and has important implications for music recommendation systems.
- The paper provides interesting insight to the sample network. We want to extend the aspect further to a broader range. Rather than sample data, we want to focus on the Spotify data which we believe will be more relevant compared to the data provided by Whosampled.com
- We also want to see the evolution of network formed by year.

Current Literature

Prediction of product success: explaining song popularity by audio features from Spotify data

Rutger Nijkamp

- The paper by Rutger Nijkamp aims to predict the **popularity of songs using audio features** from **Spotify data**. The author uses a dataset of over 160,000 songs to explore how various audio features, such as l**oudness, tempo, and key**, are related to a song's popularity.

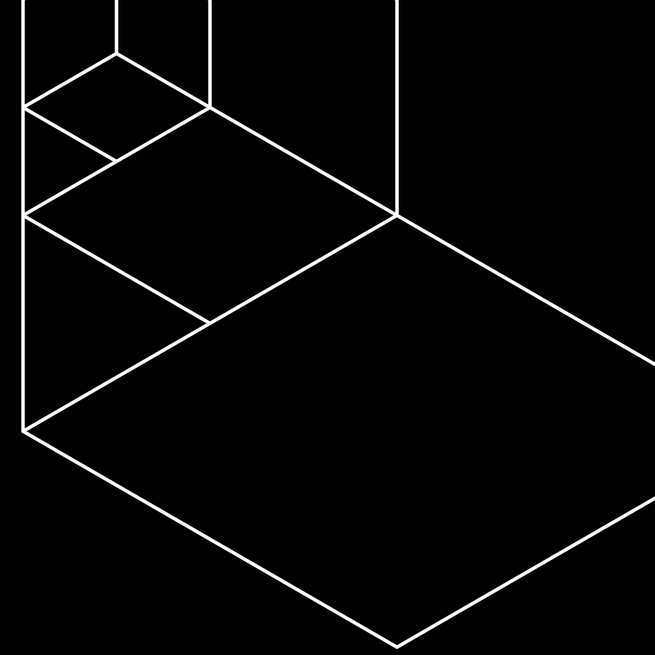
- The paper finds that certain audio features, such as danceability, energy, and loudness, **are strongly related to song popularity**, while others, such as acousticness and instrumentalness, have a weaker relationship. The paper also finds that **models based on audio features outperform models based on metadata**, such as artist and album information, in predicting song popularity.

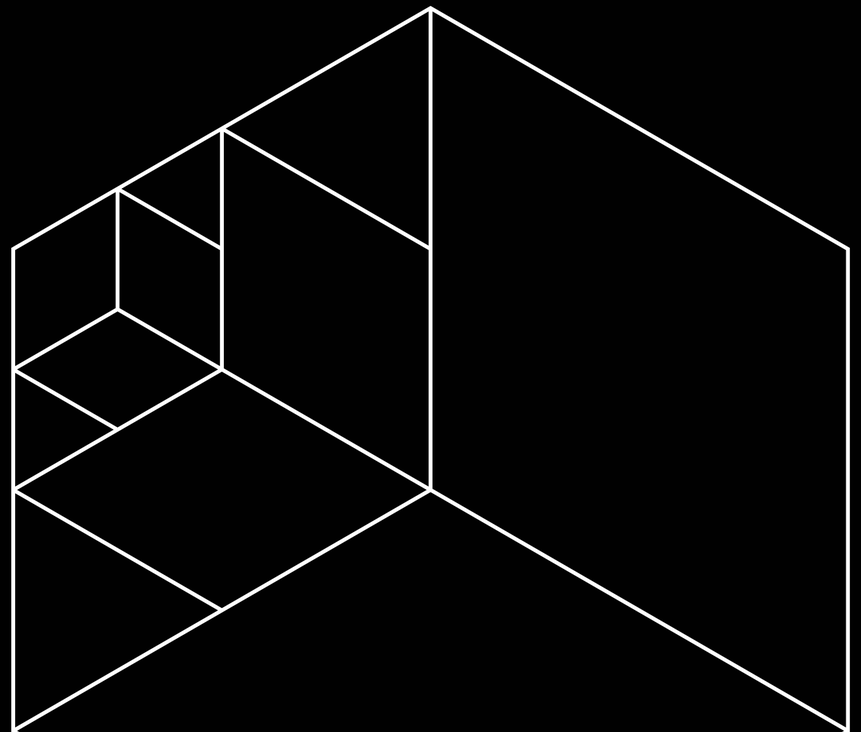
We also want to implement ML onto our project but we would like to see the time effect in terms of year and the formulation of sub groups within the network.

METHODOLOGIES

Python Programming

- NumPy
  - In order to handle large datasets of graph structures
  - Calculating various properties of networks
- Pandas
  - Analyzing large datasets of network properties
- NetworkX
  - Handling and visualizing graphs and networks
- Plotly
  - Visualizations for data analysis
- Similarity Coefficient
  - Measuring similarities between nodes and edges
- Gephi
  - Exploring and analyzing networks





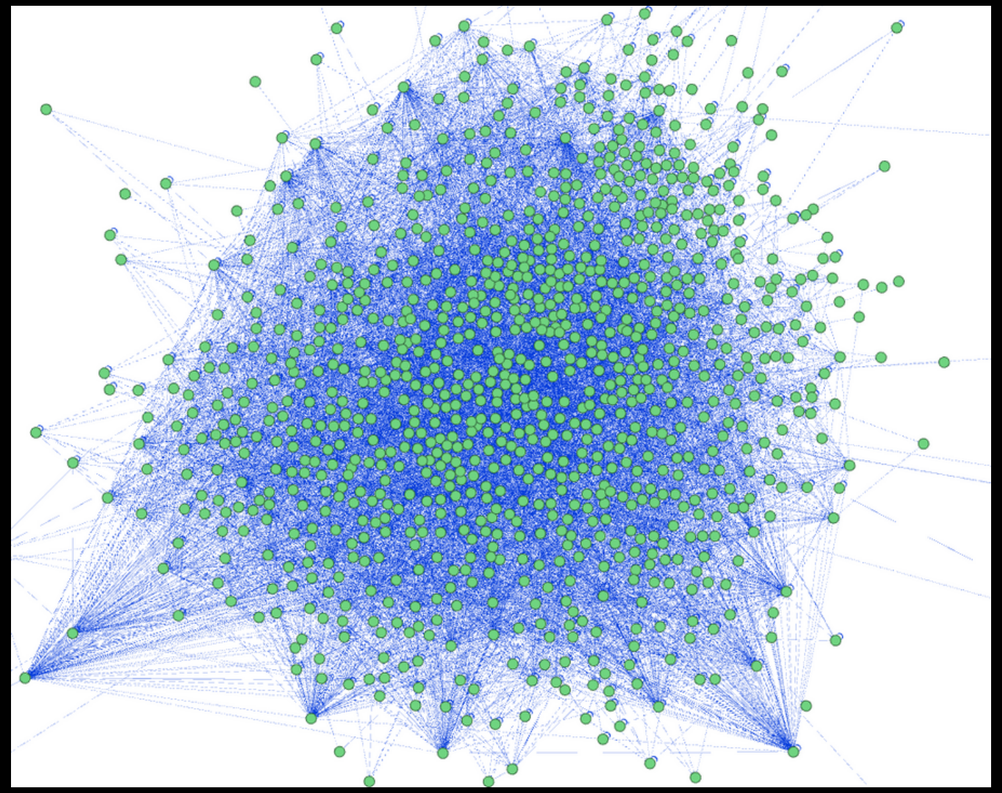
PLANNED ANALYSIS

- Creating a network of songs with their musical properties
- Using network analysis techniques to gain insights about the structure and properties about resulting network, year by year
- Using community detection algorithms to identify clusters of similar songs
- Observing how the network changes and grows year by year
- Examining how the relationships between song properties and popularity have evolved over time

RELAVANCE TO NETWORK SCIENCE

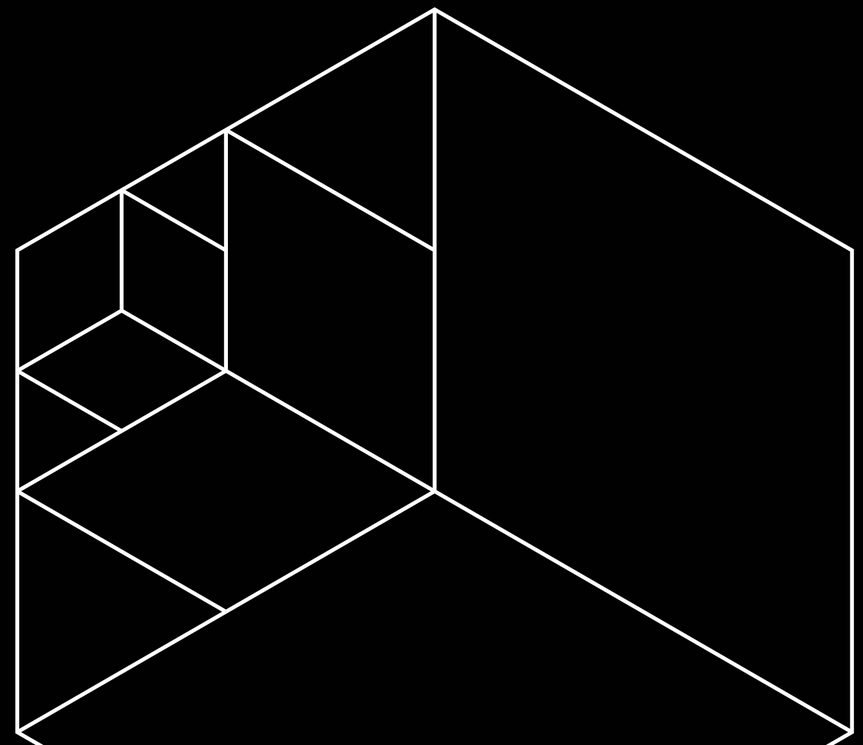
- This project involves the analysis of a network of songs based on their properties over time

- Techniques and methodologies that is going to be used



CONCLUSION

- With the help of network analysis techniques, we hope to gain insights into the structure and properties of the resulting network.
- Discover connections between songs and their popularity that might not be obvious through other methods



REFERENCES

- Prediction of product success: Explaining song popularity by audio ... (n.d.). Retrieved April 3, 2023, from https://essay.utwente.nl/75422/1/NIJKAMP_BA_IBA.pdf

- Musical influence network analysis and rank of sample-based music. (n.d.). Retrieved April 3, 2023, from https://www.researchgate.net/publication/220723768_Musical_Influence_Network_Analysis_and_Rank_of_Sample-Based_Music

- Mavani, V. (2021, December 17). Spotify dataset. Kaggle. Retrieved April 3, 2023, from https://www.kaggle.com/datasets/vatsalmavani/spotify-dataset