

Community Detection of Music Genres

Andrew Mokhtarzadeh | Brandon Tran | Dare Hunt | Junjie Li

Research Question

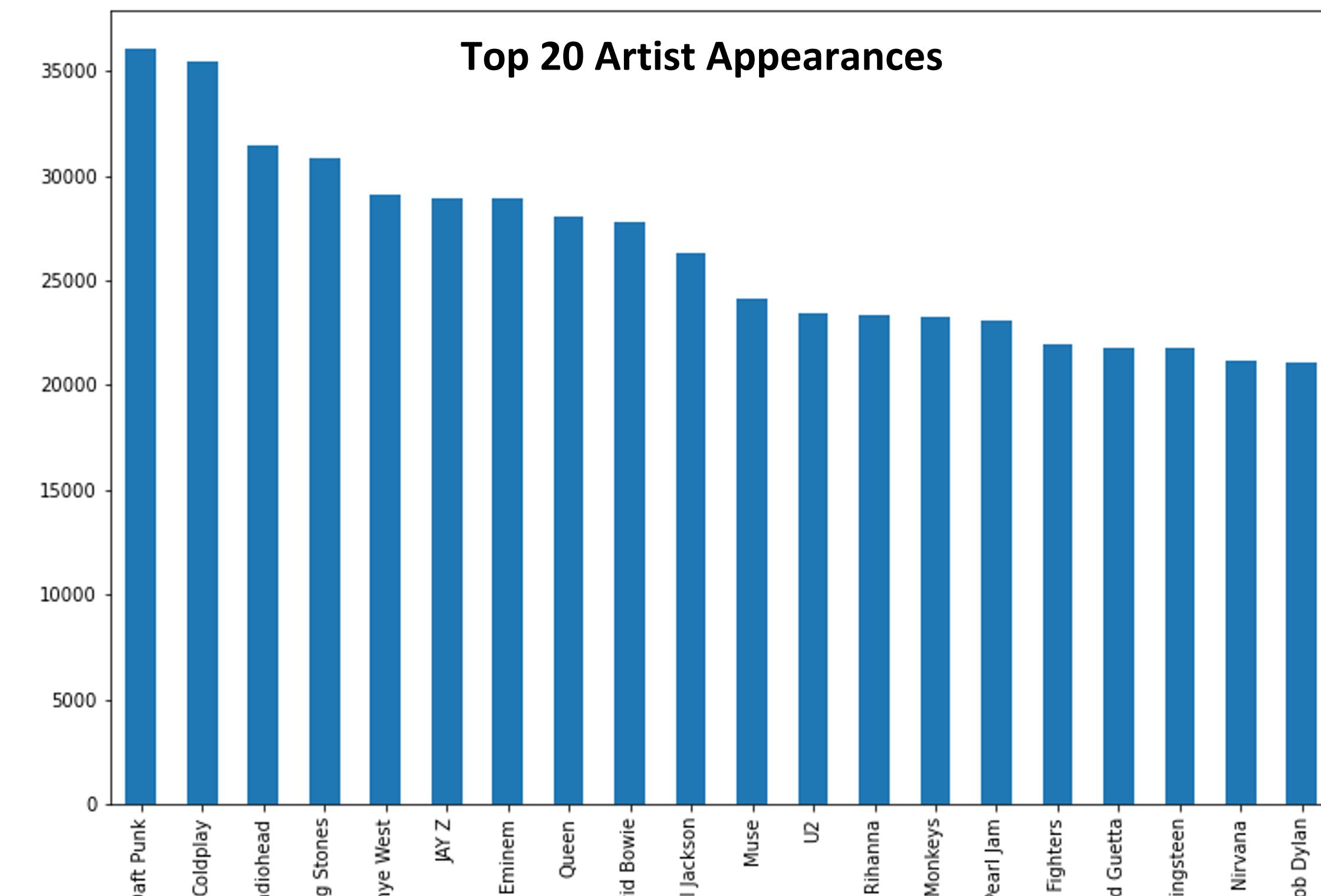
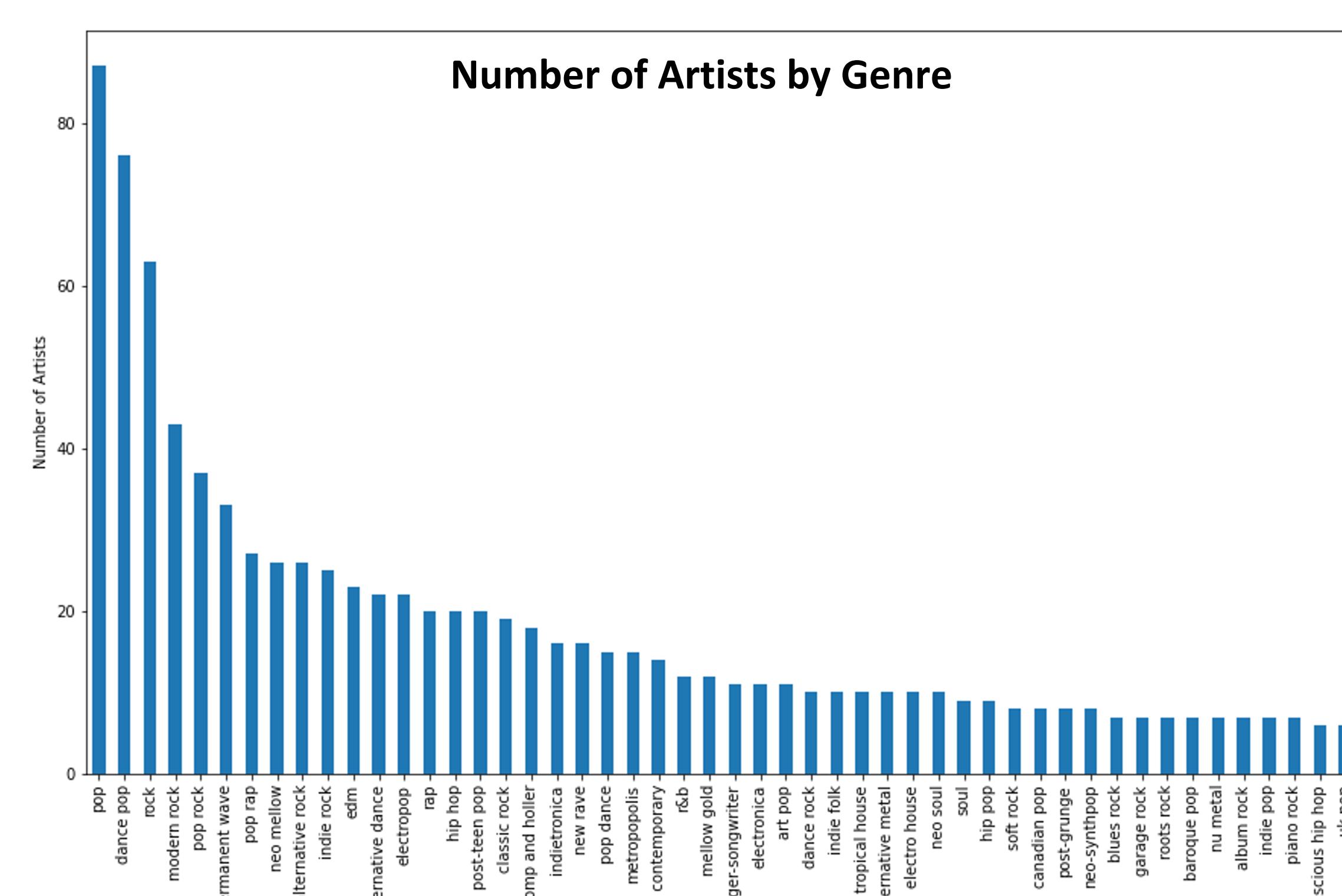
- Can we categorize artists by genres based on the playlists they are in?

Data Collection

- The Spotify Dataset is on Kaggle and artist genres were retrieved using Spotify's API

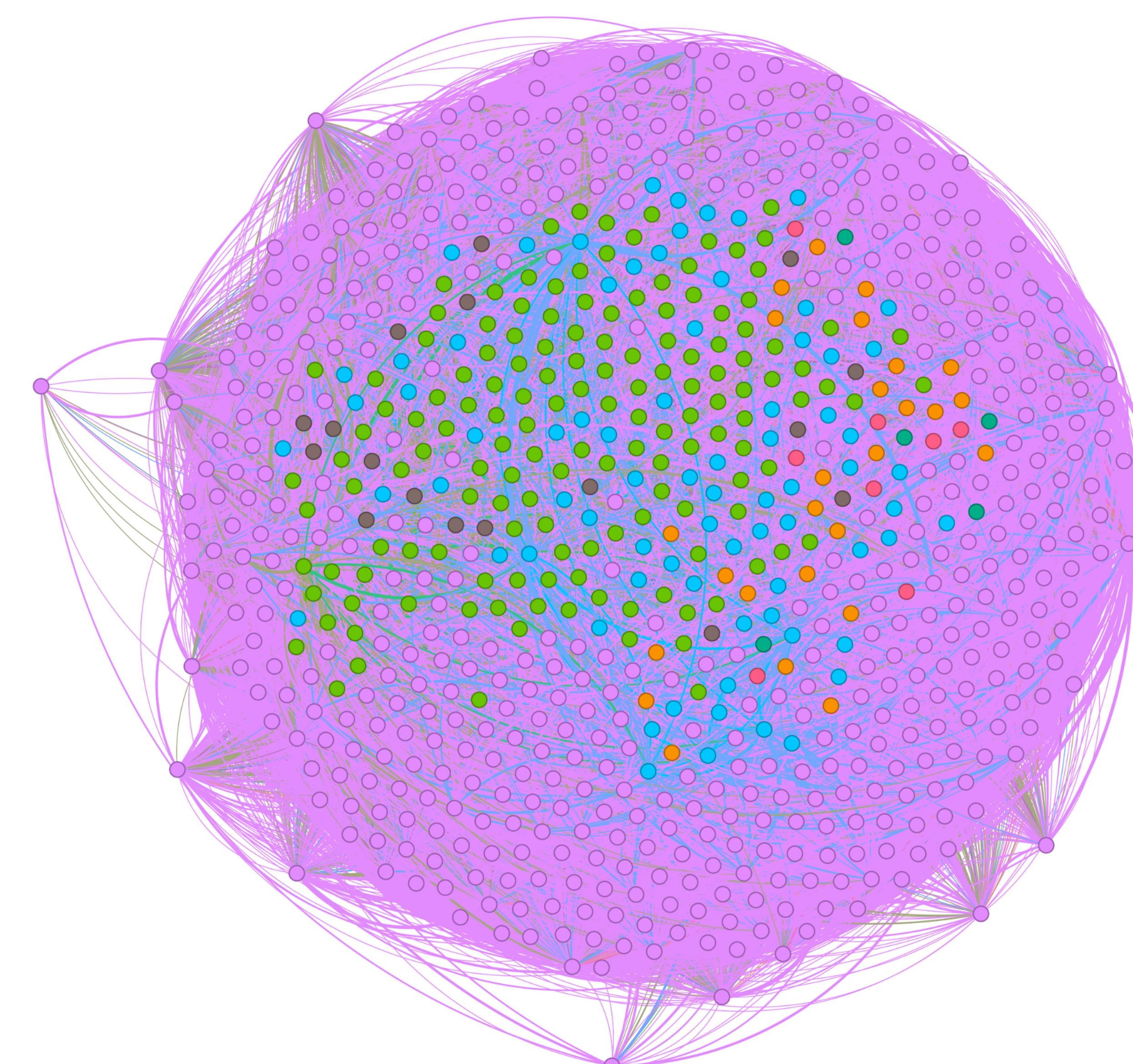
The Data

- Spotify Dataset of 161,530 playlists shared by various users
- We sample 2,000 playlists for our network to build in efficient time while producing significant results
- The sample includes:
 - 18,139 artists
 - 126,537 songs
 - 623 artists appearing on at least 10 playlists



Results

- When detecting 7 communities, the accuracy for the original genres associated with each artist is 53% and 82% when generalizing the genres (meshing indie-pop and dance-pop as pop)



How the accuracy is computed

- We take the top three Genres per community and check if the nodes have any of these three Genres

Example Communities from our Algorithm

frequency	artistname	frequency	artistname	frequency	artistname
232	John Mayer	251	Red Hot Chili Peppers	5033	Coldplay
149	Mariah Carey	115	Beck	1478	Arctic Monkeys
146	Maroon 5	100	Oasis	1441	The Smiths
139	Jack Johnson	73	Two Door Cinema Club	654	Kylie Minogue
135	Adele	71	Haim	605	Florence + The Machine

The Model

Communities from Edge Structure Attributes (CESNA)

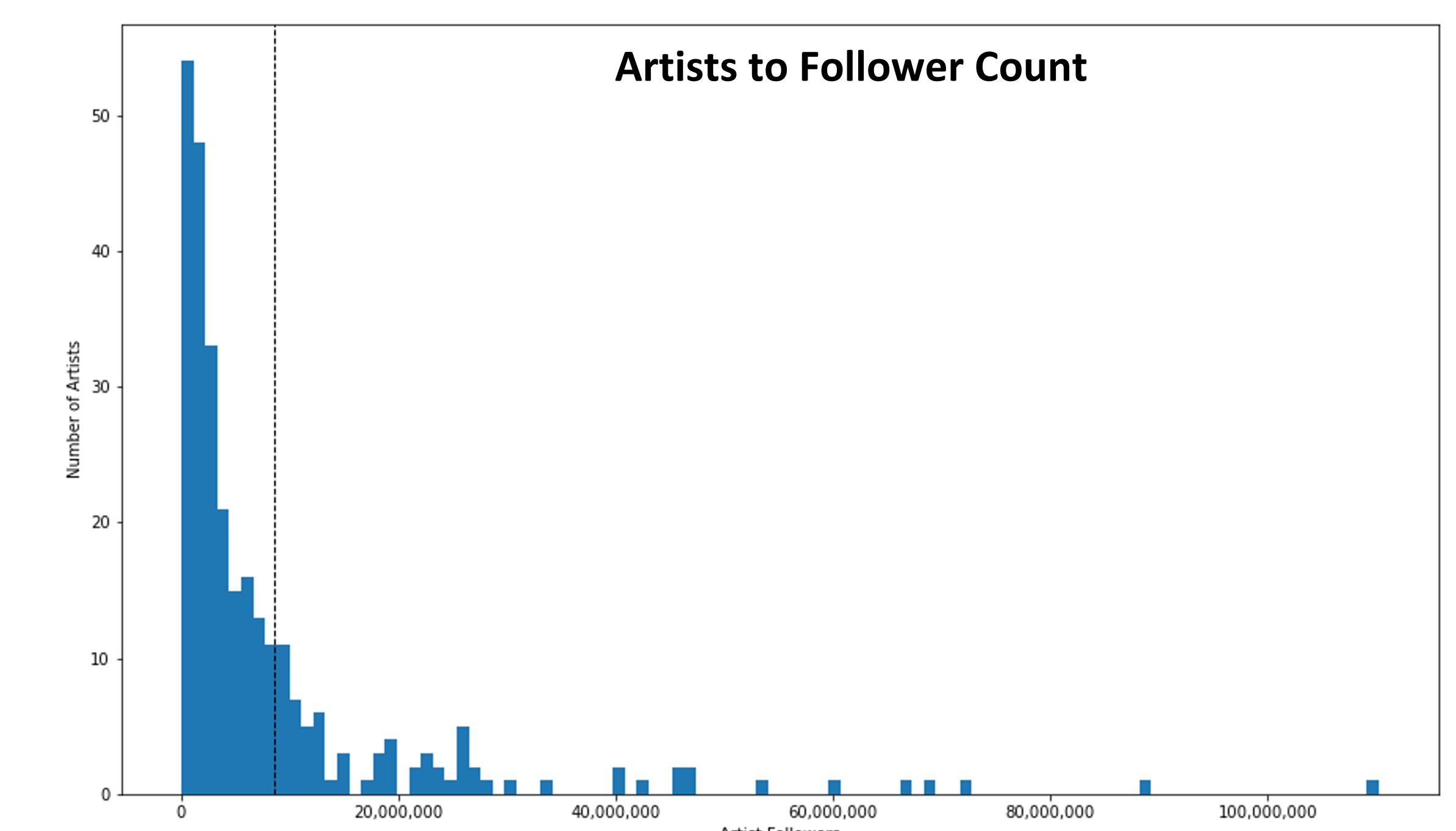
- This model uses a probability distribution and node attributes to predict the probability of a node belonging to each community before assignment
- Typical community detection just considers the graph structure and weights

Our implementation:

- Edges exist between two artists added to the same playlist
- Edges are weighted by the frequency that artists appear in the same playlist

Node Features:

- Artist popularity[0,100], Is top 25% of appearances [0, 1]



➤ Edges with a weight less than 2 is removed to improve model efficiency

➤ The algorithm involves maximizing the likelihood that a node belongs to a given community through gradient ascent. The Objective function is defined as $I(F) = LG + LX$ where $LG = \log P(G|F)$ and $LX = \log P(X|F, W)$ where G is the underlying Graph, F is the affiliation matrix, X is the attribute matrix, and W are the associated weight vectors of X .