Project Proposal: [Project Title]

Steven Canon-Almagro, Nathanael Huh

1. Nodes and Links   
   Nodes are song artists and Links are if they worked on a song together. A linked song can come from a feature, and/or co authoring an album. Scraping featured songs is the main way we will be determining the link.
2. Dataset  
   Drawing inspiration from the actor network. We will be using the Spotify API to get artist and feature data. Log in using personal Spotify account then make a ClientID account with Spotify for Developers. This account should then be approved relatively quick and supply team with a client ID and a Client Secret. We can then send the ClientID and Client Secret to get an access token. The access token then allows us access to the other end points in the API where we can then query the Spotify database and get data in return. Using the query function, I will be attempting to retrieve the top 500 artists, all their featured songs. Any featured artist on the songs will then be connected to the song artist on the network.

Artists like Kanye may pose a problem as they don’t put features in the track name, an array of artists is kept in the track data, but I don’t know how the data is stored.

The number of nodes is kept low because each artist has many songs to themselves so the data set will be very large.

My plan is to get the data and then create a quick program to save that to a csv file.

The csv file then will be used in python to create the network.

1. Expected size of the network  
   As of 2022 there are 11 million artists on Spotify. Our goal is to scrape the top 500 artists and map out the connection they have with their features. We are only scraping 500 artists from Spotify, but there will be much more information than that as the artist can work with multiple artists outside the top 500.
2. Questions you plan to ask and why we care  
   Is the music industry just as connected as the movie industry. What is the most well-connected artist on the Spotify platform.