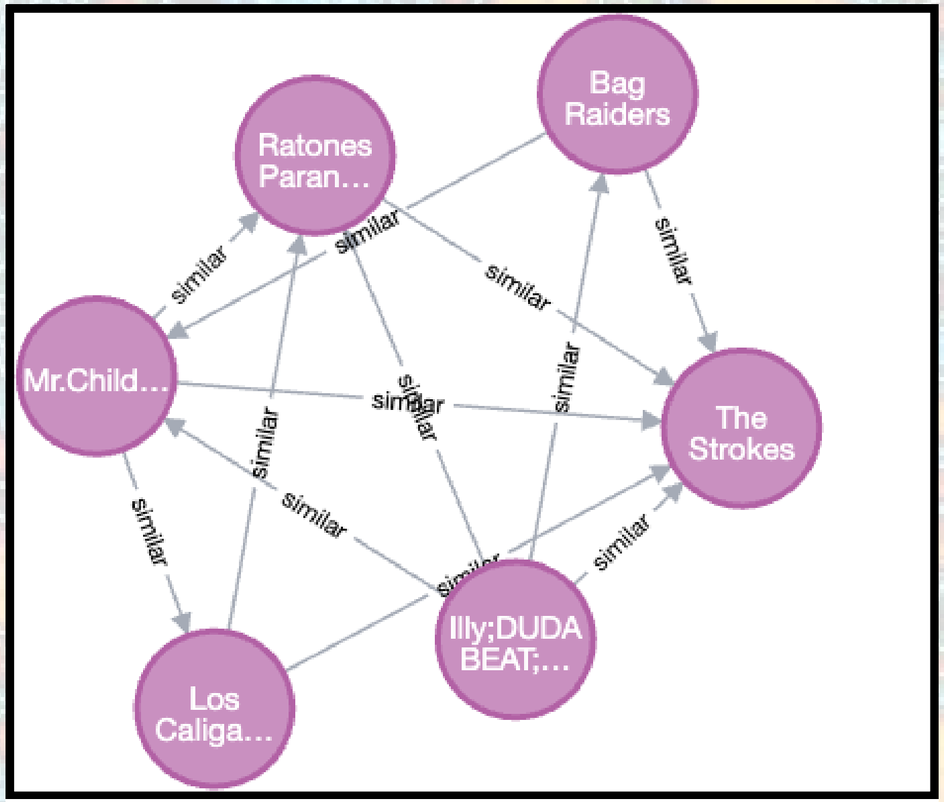
DS4300: Song Recommender Engine | Sarah Bernardo & Kate Lanman

Nodes: 5,040
Edges: 124,432

Our algorithm uses the 12 features from the Spotify dataset to calculate song similarities. We assign each feature a weight based on results from a simple linear regression using track genres as labels. Genres are how we naturally cluster songs, so using these labels to train the data, we can gain an idea of the importance of each feature in song similarity. We then compute a weighted Euclidean distance between each song in our sample and use this as our metric for similarity (a smaller distance implies greater similarity). Songs are considered “similar” if they are in the bottom 1% of all scores computed (smallest relative distances).

1. **One Headlight** by The Wallflowers
2. **That's Life** by Still Woozy
3. **En vän med en bil (Live)** by Håkan Hellström
4. **Di-Sk8 Eu Vou** by Charlie Brown Jr.
5. **Watch Me Bleed** by Tears For Fears



NOTE: since we randomly sample from the full dataset, the songs in the image are not the same as our recommended song