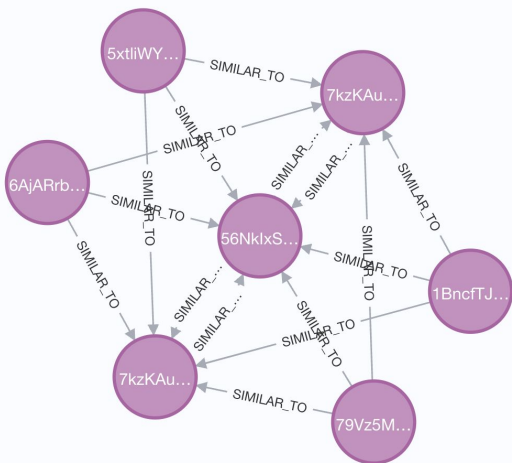


# Spotify Recommendation With Neo4j



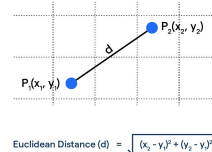
112 nodes 1330 relationships

Speed of our device severely limited  
our recommendation abilities:

OS: Mac  
CPU: Apple M2, Speed: 3.49 GHz  
Memory: 8 GB  
Cores: 8

Our sampling process relied on simplicity and randomness: we loaded the given spotify.csv data into Python, and used Python methods to put the csv rows into a Pandas dataframe, then dropping duplicates for data cleaning purposes, selecting a sample of 100 which include all songs from the given album. We found computer performance issues with using higher sample sizes [INSERT COMPUTER SPECS]. This can be easily adjusted by changing the album\_name global variable for replication with other albums. In the full graph data model, our nodes are songs which have a variety of features including [LIST], and each song has a relationship to another song as defined by its similarity score, which is calculated in python using the euclidean function using the SciPy library stats and zscore. The Euclidean distance metric is very useful for our purposes to measure song similarity, as compared to other distance metrics such as cosine similarity and manhattan distance. Manhattan distance would apply better to a binary dataset with high-dimensionality, but this is not true for our data or features. Cosine similarity improves Euclidean issues with high-dimensional data, but our data is not especially high-dimensional, and cosine distance is unable to capture differences in magnitude across dataset variables as well as a normalized Euclidean dataset. These are just a few of the reasons we chose Euclidean over another way to measure. Normalization is performed to allow for accessibility of the resulting similarity scores, in our cases, producing a scale from 0 to 100 where 0 is the least similar and 100 is most similar. This allows us to then in Neo4j, once the data has been loaded in with the nodes (songs) and edges (similarity score), order resulting songs where the similarity score is the highest and limit to the top 5, when the given comparison album is that given by the individual seeking recommendations (in our case, "Is This It" by the Strokes). We also limit results to popularity higher than 50.

## Euclidean Distance



Euclidean distance determines song similarity

## Our Song Recs:

**Song:** BLUE Artist: Tiesto; Stevie Appleton Album: BLUE  
**Song:** Experience Artist: Ludovico; Daniel Hope; I Virtuosi Italiani Album: In a Time Lapse  
**Song:** Choke Artist: OneRepublic Album: Oh My My  
**Song:** Friendship Years Artist: Jordan Chan, Elkin Change Album: Golden Job  
**Song:** Experience Artist: Ludovico Einaudi; Daniel Hope; I Virtuosi Italiani Album: In a Time Lapse