1.4 Questions to answer in the report

1. (1.5 points) Comment on the results obtained in Exercise 4 (include the results obtained in Exercise 4 in the report): (a) What are the degree distributions of the three obtained undirected graphs like? (b) Are the two selected artists similar based on their audio features? Comment on the comparison regarding the relationships between artists provided by Spotify (graphs gB and gD). (c) What can you infer from the similarity heatmap regarding the algorithm that selects related artists on Spotify?

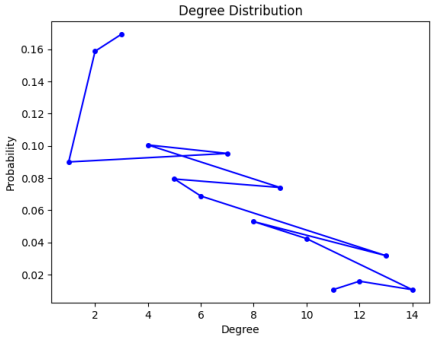
Answer:

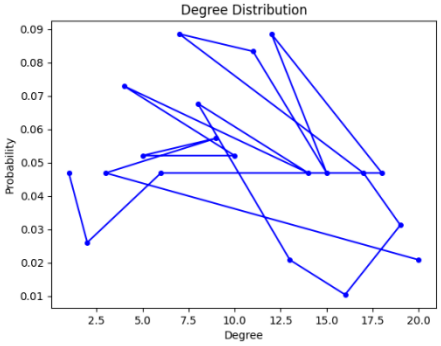
(a) The degree distributions of the three obtained undirected graphs, g'B, g'D, and gW, represent the frequency of occurrence of node degrees in each graph. To analyze the shape of the degree distributions, degree distribution plots were generated.

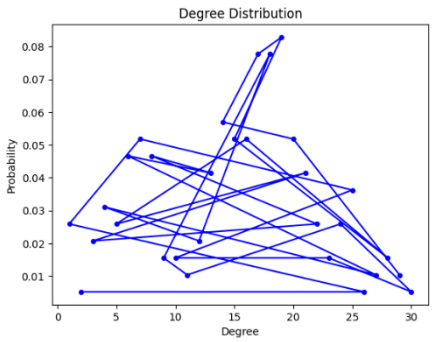
In the degree distribution plot of g'B, it can be observed that most nodes have a low degree (higher probability on lower degrees), indicating that there are many artists with few related artists based on audio features.

In the degree distribution plot of g'D, a more uniform distribution is observed, with some nodes having a high degree. This suggests that there are artists with a large number of related artists based on their audio features, since the range is also higher, but the higher probabilities remain between 7.5 and 12.5.

In the degree distribution plot of gW, a more chaotic distribution is observed telling us that there is a low probability of having degrees within the extreme and high probability of nodes with mid-degree, emphasising on degrees over the range 15-20, indicating that this graph shows nodes within those degree mostly. (GRAPHS ARE ORDERED: G'B, G'D, GwD)

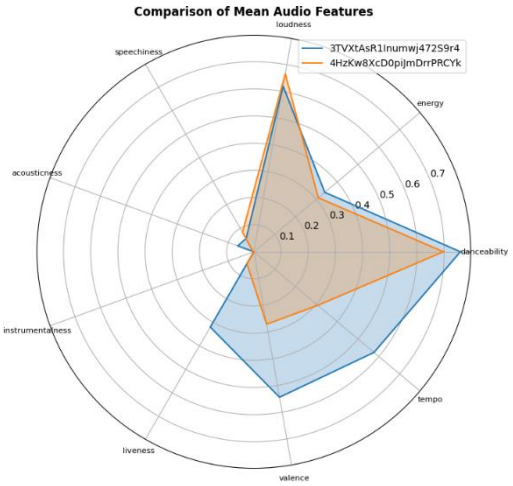


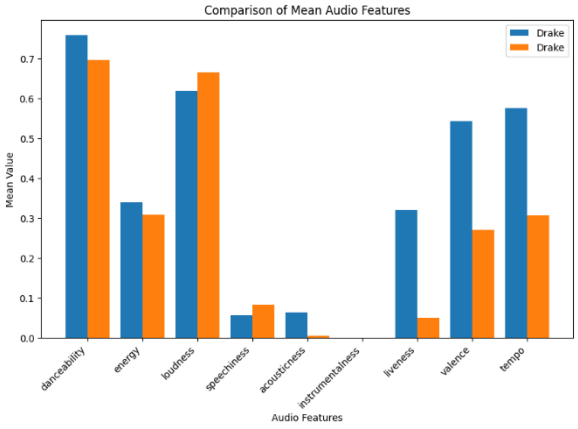




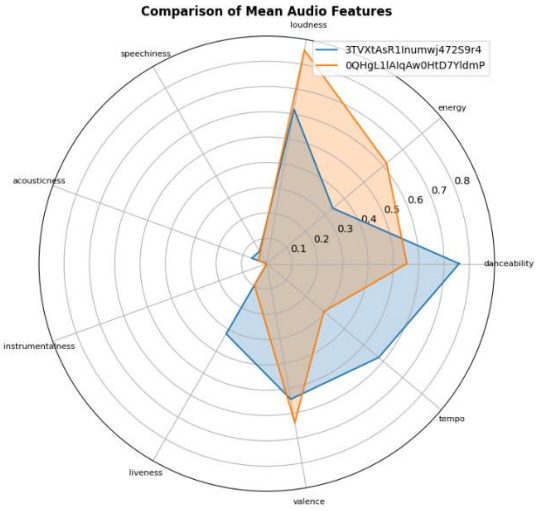
(b)

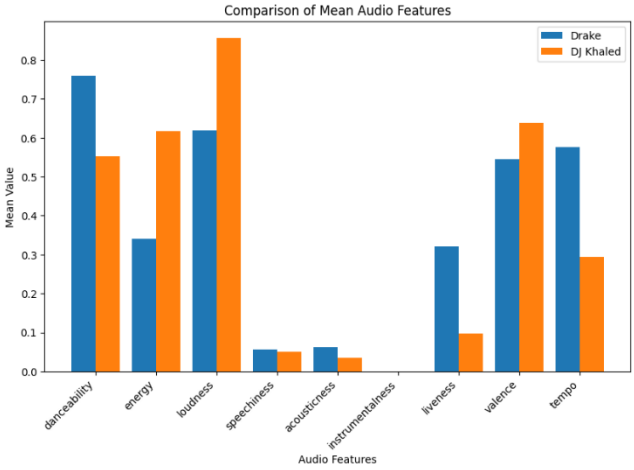
When the two selected artists are similar based on their audio features, it shows that the audio features shown in the plots are similar and share some similarities in terms of valleys, peaks, and general trends, the shapes of the figures shown are mostly equal regarding that the ranges are lightly smaller on the other graph.



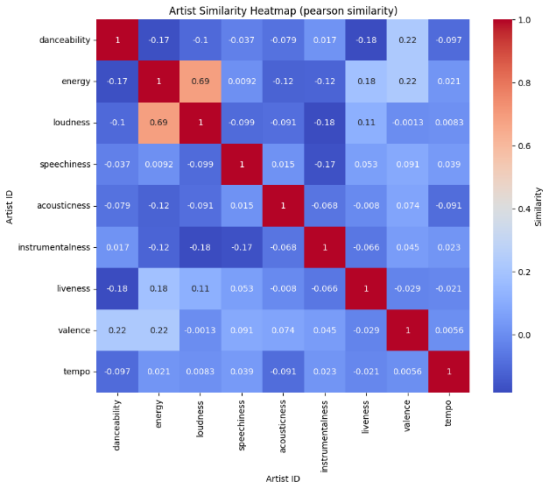


When If the graphs are not similar based, the plots show inconsistency in terms of shape and sharing values within the features, meaning that their songs tend to be different on the evaluated features.





(c) The similarity heatmap shows the similarity between the artists. This provides insights into how the similarities between the features of the audios are related within them.

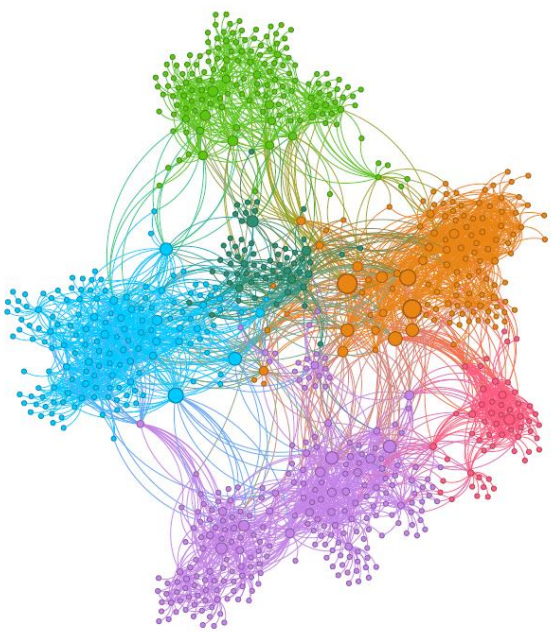


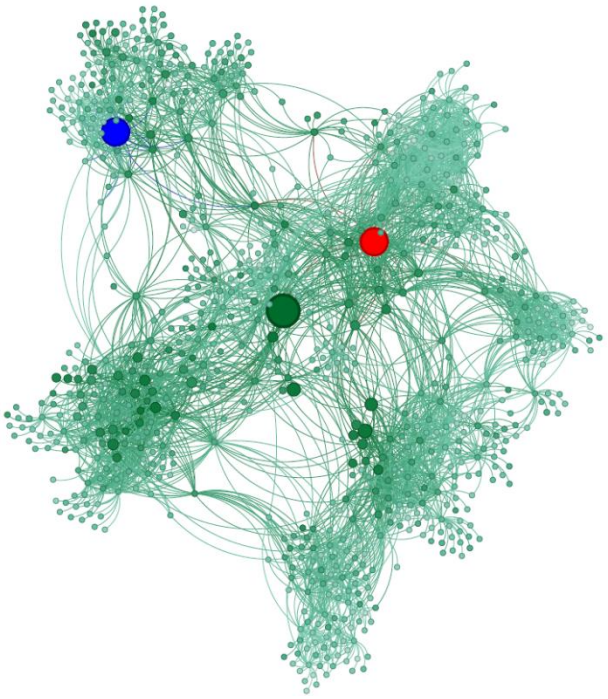
The similarity heatmap reveal patterns and clusters among artists based on the similarity of their audio features. When there are clear clusters and blocks of high similarity in the heatmap, it indicates high relationships among the audio features of the selected artists.

However, most of the heatmap shows dispersion and a lack of clear clustering, it indicates that there are not relationships or that there is variability in the similarity of audio features among the artists.

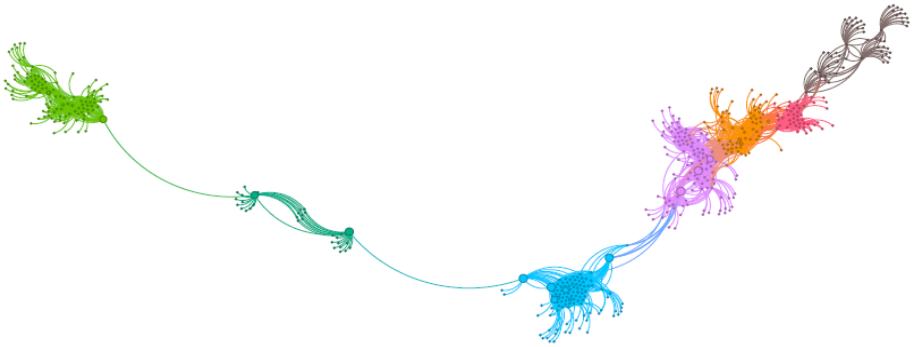
GEPHI

GRAPHS B a) and b)





GRAPHS D a) and b)



When comparing the two visualized graphs on Gephi, we can easily see that when applying a good distribution for each of the graphs is remarkable to say that the traversal mechanism used on crawling the given graph has to do with its distribution over the space, in other words, on the BFS graph, the nodes are distributed on a more dense way, more circular, with the most important node in the middle, and the nodes are expanding surrounding it.

On the other hand, when talking about the DFS graph, the principal node its on the extrema and the graph is being expanding in more depth within it, that is the reason why the DFS graph has a longer shape since when traversing the nodes in this form, it tends to acquire more depth and, hence inducing these tendencies.

Artists belonging to the same community its mostly produced by it similarity withing them when talking about features, genres, kind of music, popularity and centrality.