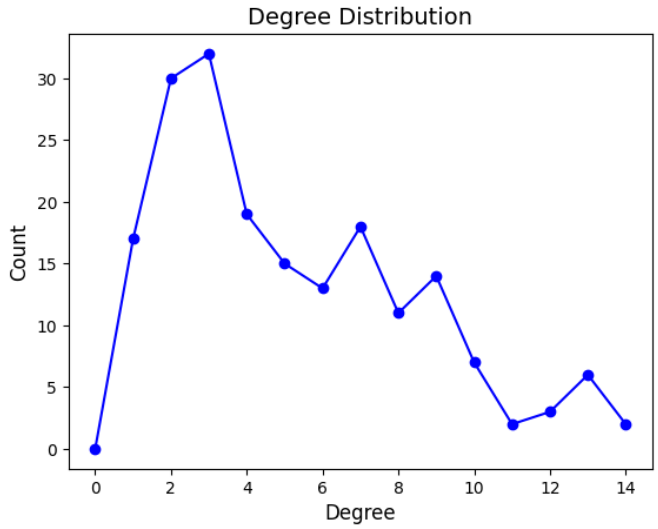
Session 4 Questions

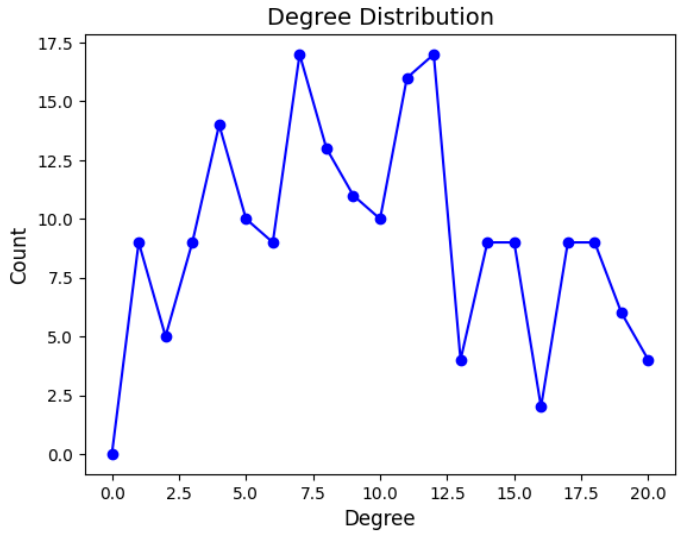
Questions to answer in the report
1. (1.5 points) Comment on the results obtained in Exercise 4 (include the results obtained in Exercise 4 in the report):

(a) What are the degree distributions of the three obtained undirected graphs like?
In all cases we have decided to use both parameters to false, as visually they are more appealing as well as more explainable, because it contain in the x-axis the degree, and in the y-axis the amount of nodes with that degree, instead of using a probability value or logarithmic values.

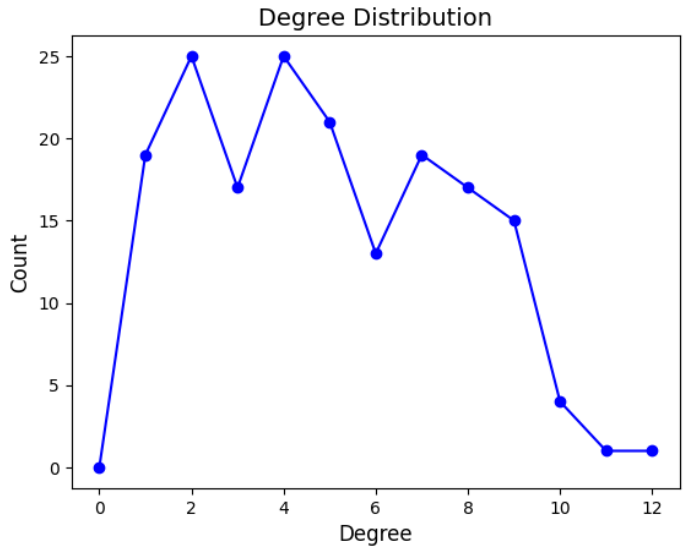
g'B → Normalized = False, loglog = False



g'D → Normalized = False , loglog = False



gwB → Normalized = False, loglog = False

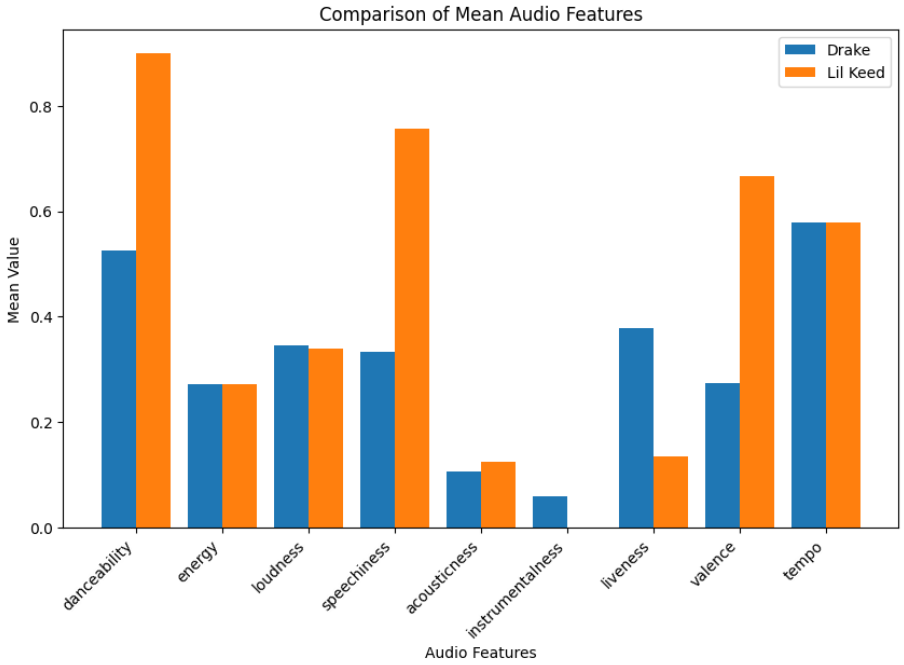


(b) Are the two selected artists similar based on their audio features? Comment on the comparison regarding the relationships between artists provided by Spotify (graphs gB and gD).

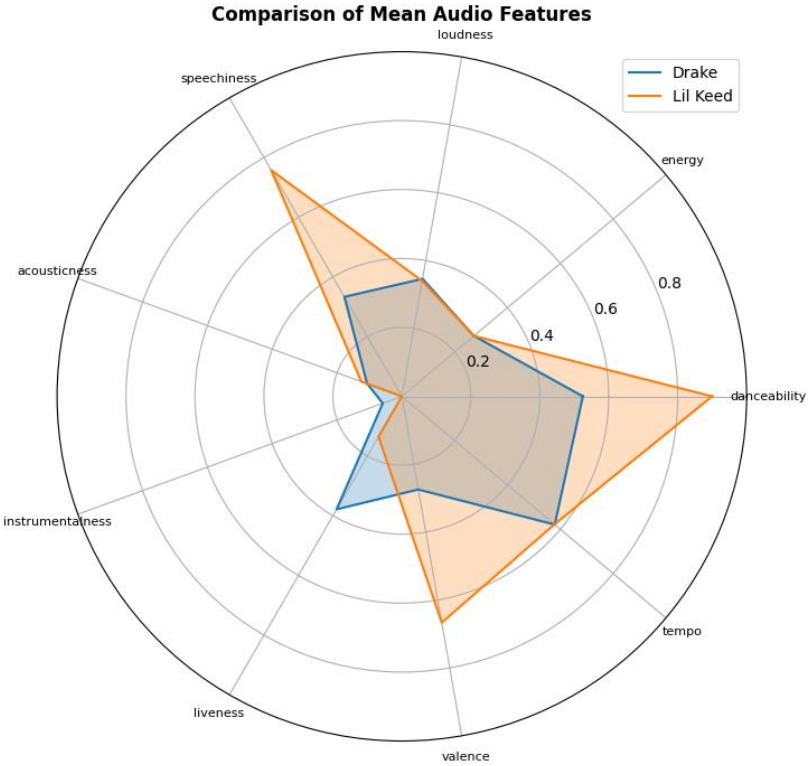
gB:

Comparison Drake and most similar (Lil Keed):

- Bar plot:

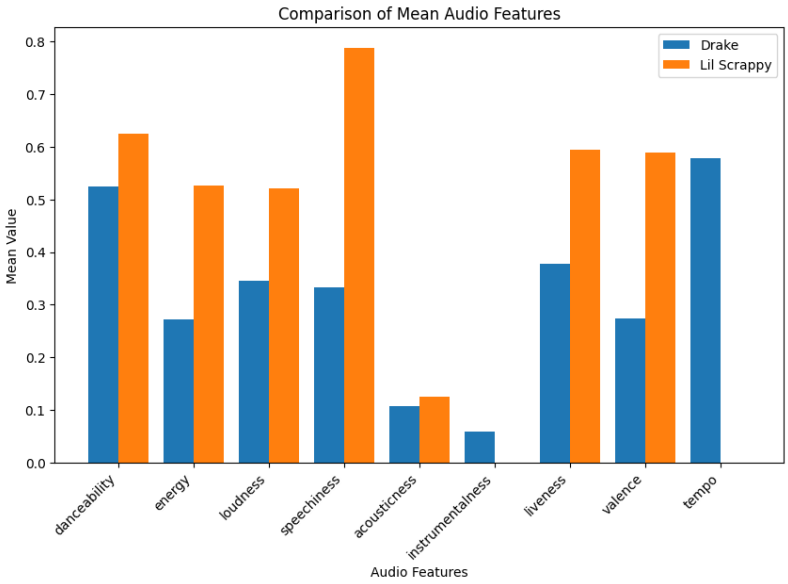


- Radar plot:

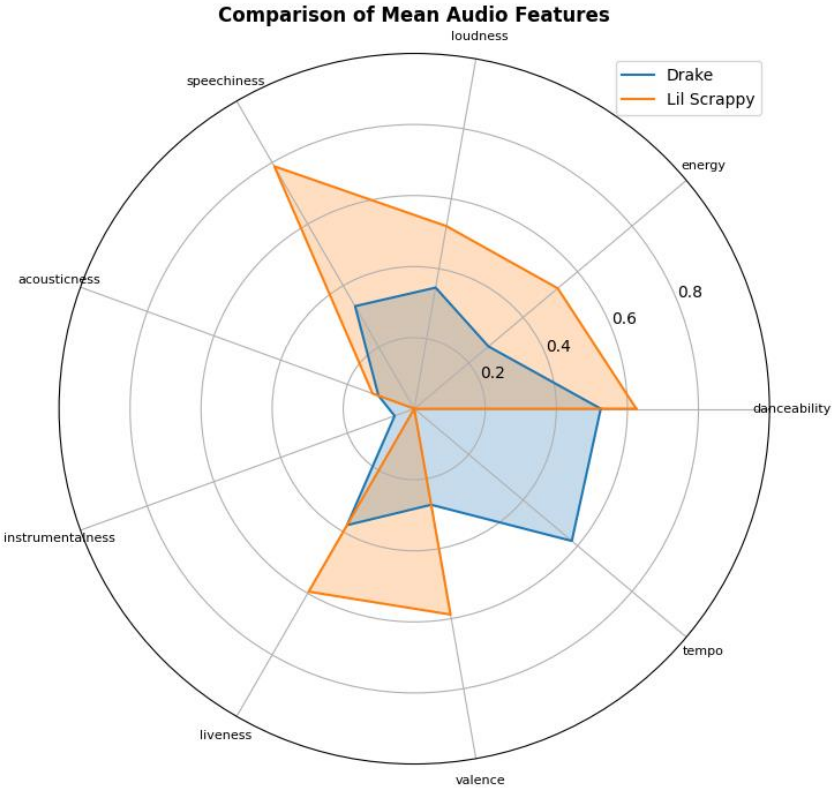


Comparison Drake and less similar (Lil Scrappy):

- Bar plot:



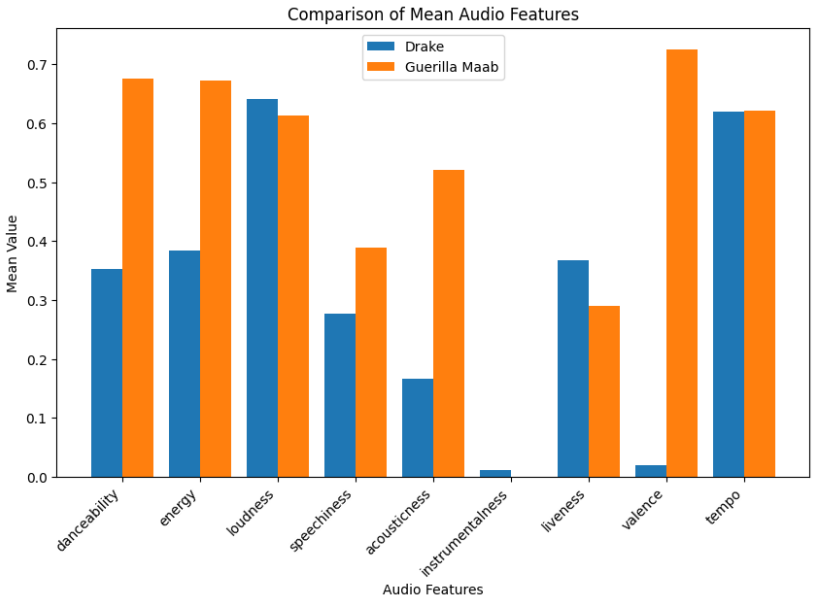
- Radar plot:



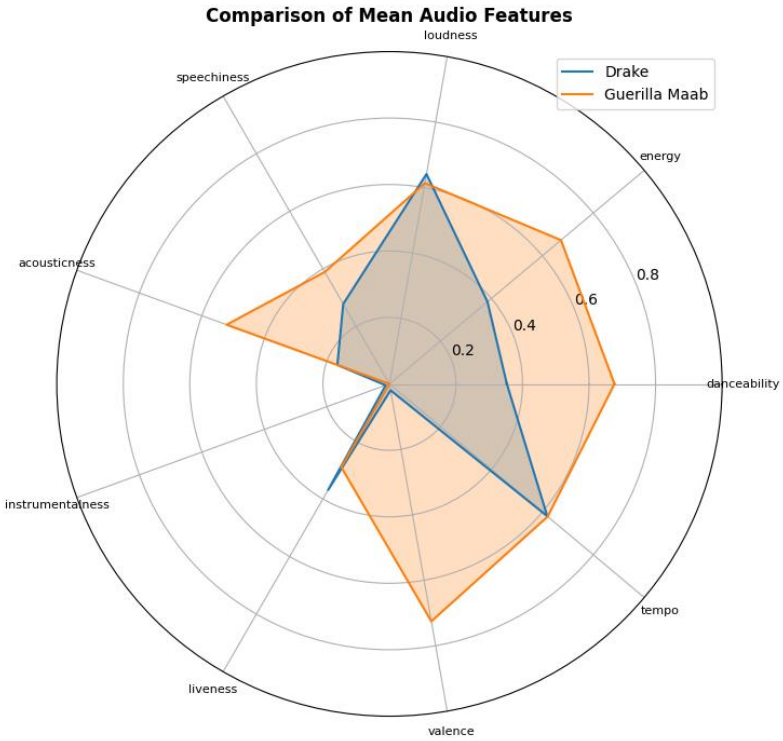
For the graph gB, we can see a notorious difference between the most and less similar artists to Drake, but still the results are not that far, as later in the similarity heat map we will see the reasons.

gD:
Comparison Drake and most similar (Guerilla Maab):

- Bar plot:

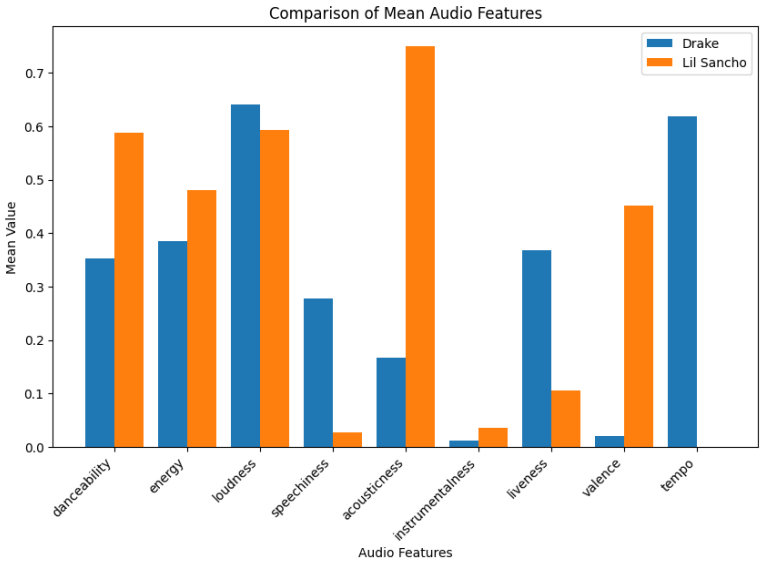


- Radar plot:

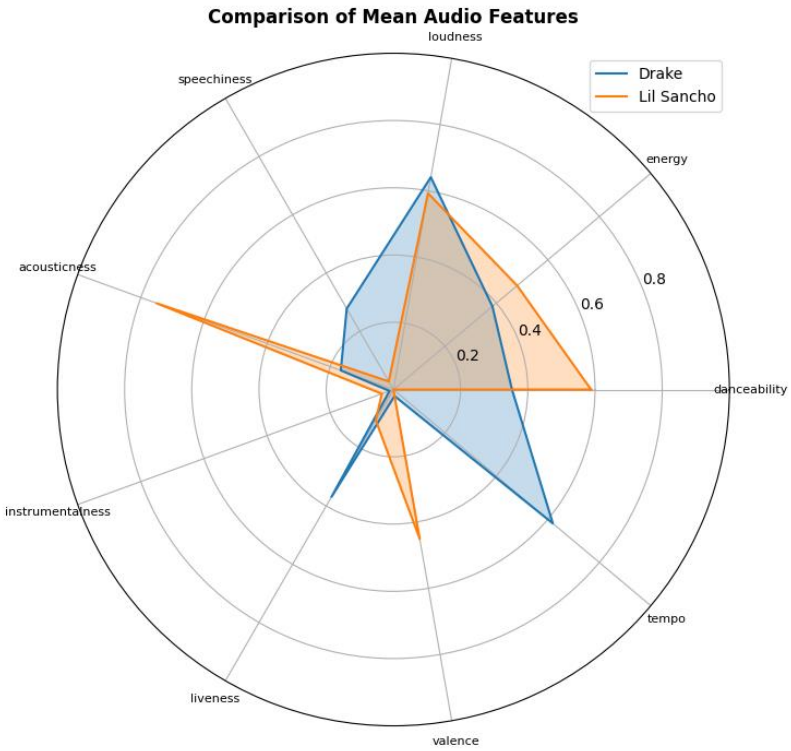


Comparison Drake and less similar (Lil Sancho):

- Bar plot:

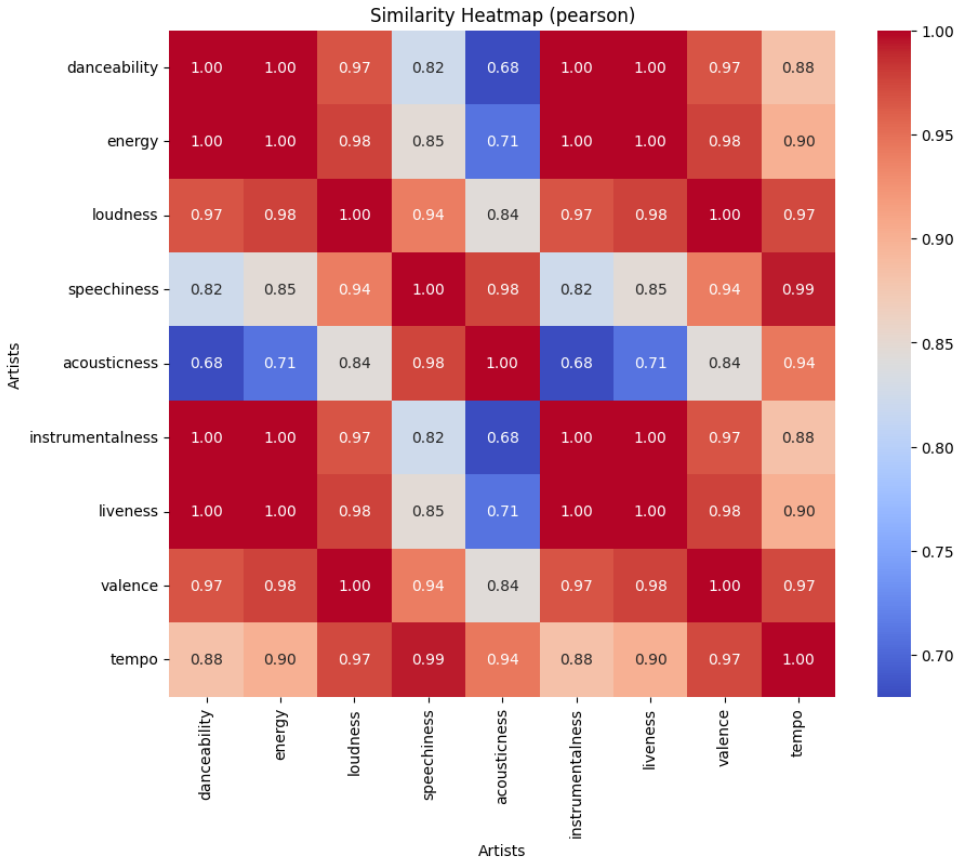


- Radar plot:



For the graph gD, despite a notorious difference between the most and less similar artists to Drake, we also can see that when comparing it with the ones of the gB, the results are more dissimilar as when using DFS we are going deeper, and obtaining less similar artist from the one of the top.

(c) What can you infer from the similarity heatmap regarding the algorithm that selects related artists on Spotify?



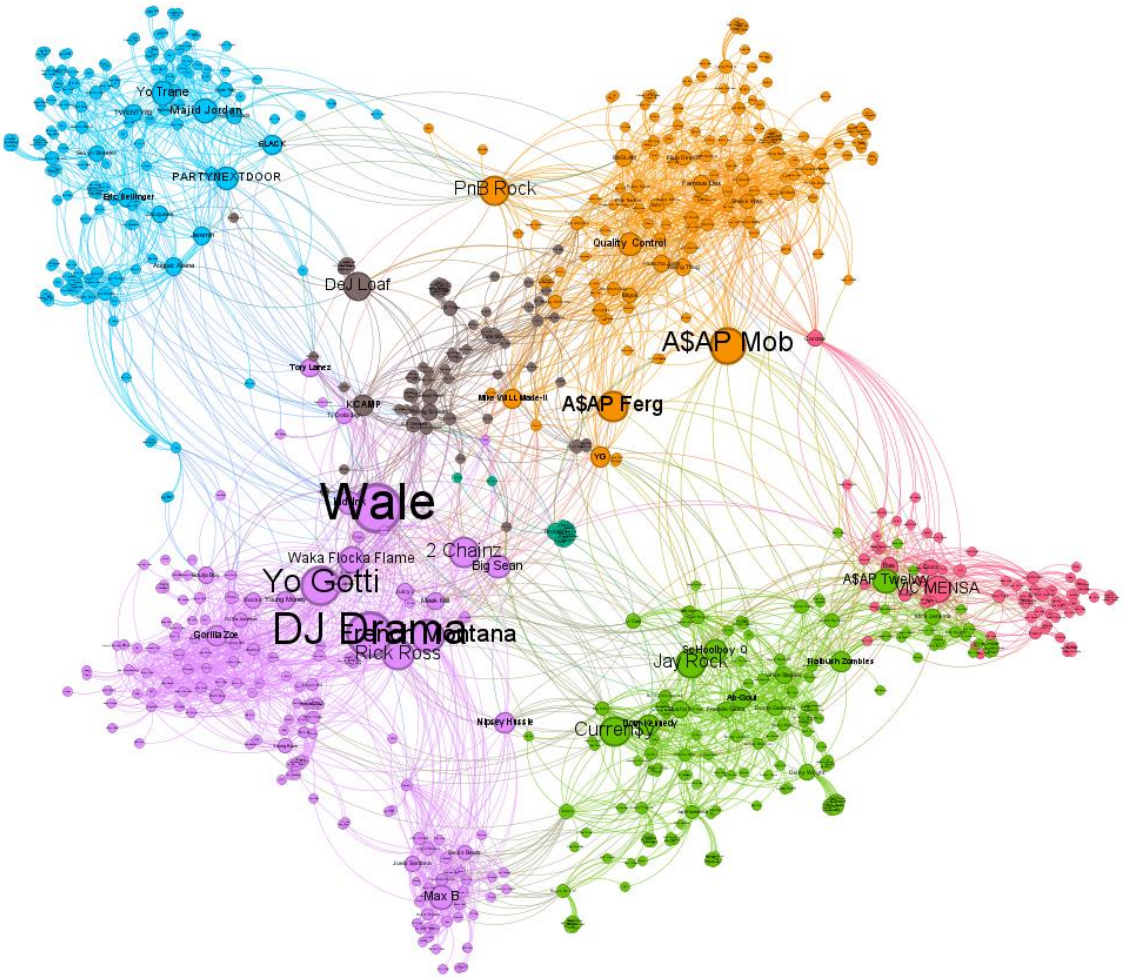
Being this image the similarity heat map, we can see that the minimum similarity value is 0.68, this high value indicates that there is a very big similarity among all the features compared for all the artists. Therefore we can assume that the related artists algorithm is based on the features of the songs, and that is why we obtain that very high similarity in all of them.

Data Visualization from Gephi

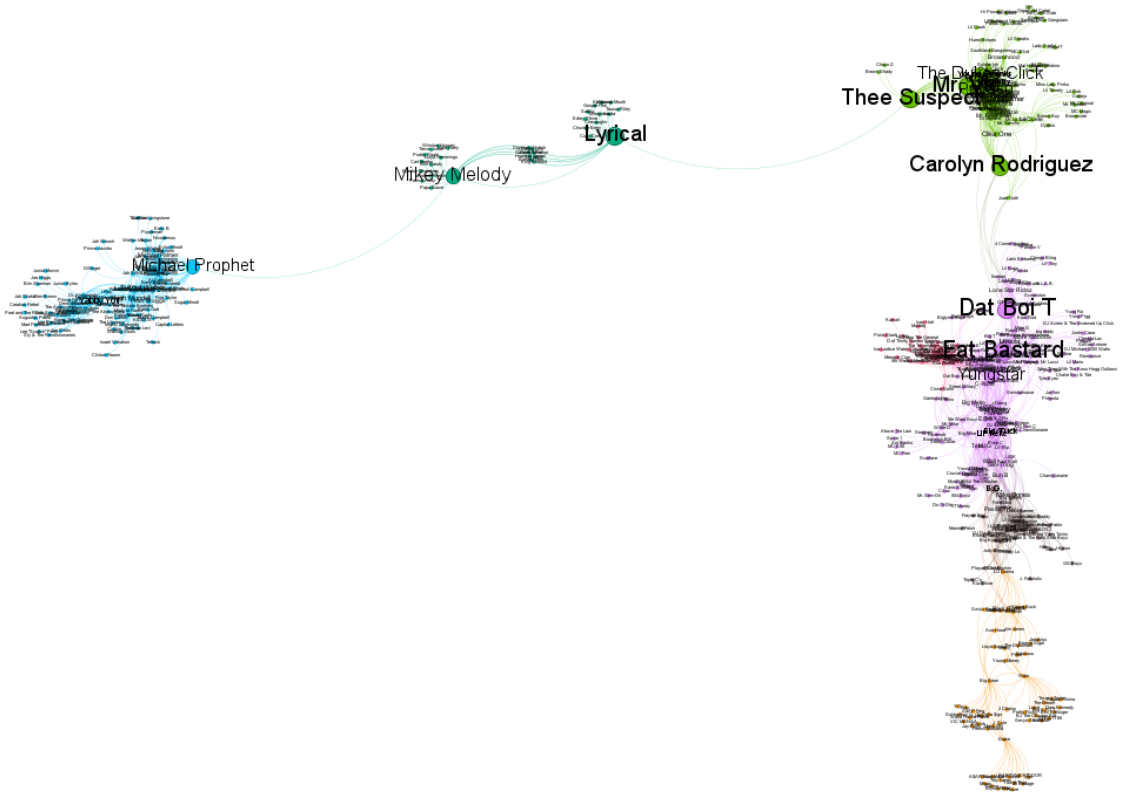
For each of the directed graphs gD and gB obtained in the first session of the practice:

1. (1.5 points) Generate a visualization of the graph using Gephi that assigns a color to the nodes based on the community they belong to and sizes the nodes proportionally to their betweenness centrality. Use a layout algorithm that allows for easy identification of the communities. Show the names of the most important artists (highest betweenness centrality) in each community.

gB:



gD:



2. (1.5 points) Comment on the visualizations generated with Gephi.

- **Compare graphs gB and gD. What can you say about their properties?**

We can clearly see how in the graph gB, despite being different communities formed, they are all very close to each other, while in the gD, the different communities are very distant. So the densities of both graphs are very different.

- **Can you identify common characteristics among artists belonging to the same community? Could you label the different communities?**

Despite not having accomplished the second gephi visualization, we will answer the question anyway. We can see that inside the different communities, many different genres are as well as song properties are found which are the main factors to determine the different communities. Moreover, we could assign a label to them, based on the genre found most times inside that community, so we would end up having several communities that would be easy to differentiate as they would be called for example: “rap”, “rock” or “pop” and many others.