

Lab activity: analysis of music networks

P2

1. Provide the order and size of the four obtained undirected graphs (gB', gD', gBw, and gDw).

*Method used: cosine similarity

Graf	Ordre	Mida
gB'	100	330
gD'	93	202
gBw	155	329 (min. %: 98.3)
gDw	133	201 (min. weight: 0.99999805)

```
# 6 B)
compiled_songs = compute_mean_audio_features(songs)
grafic = create_similarity_graph(compiled_songs, "cosine", "Similarity.graphml")
#print(f"Similarity graph: {grafic}")

min_percentile=98.3
gWB = prune_low_weight_edges(grafic, out_filename="gBw.graphml",
min_percentile=min_percentile)
min_weight=0.99999805
gWD = prune_low_weight_edges(grafic, out_filename="gDw.graphml",
min_weight=min_weight)
```

2. Justify the strategy used to obtain gBw and gDw.

El primer pas consisteix en obtenir les mitjanes de les característiques de les cançons de cada artista. Una vegada es tenen les mitjanes s'executa una funció (create_similarity_graph) que agafa les característiques de cada artista i crea un graf amb pesos que representen la similitud de les característiques entre els artistes. En el nostre cas s'opta per utilitzar com a mètrica de similitud "cosine", tot i que la distància euclidiana està disponible.

L'últim pas es basa en eliminar les arestes que tenen un pes per sota del demanat. El criteri és que el graf resultant tingui una mida semblant a gB' i gD' respectivament.

Per tal de trobar el valor s'han realitzat diferents proves fins a aproximar-se al màxim a la mida dels grafs esmentats. Per introduir el valor existeixen dues opcions, a partir del percentil, o bé amb un valor mínim fixe de pes. Per crear el gwB s'ha utilitzat un percentil de 98.3 i per gwD un pes mínim de 0.99999805. Com es pot observar a la taula de l'exercici 1 les mides s'aproximen molt, es compleix el criteri.

- 3. Justify whether the directed graphs obtained from the initial exploration of the crawler (gB and gD) can have more than one weakly connected component and strongly connected component, and explain why. Indicate the relationship with the selection of a single seed.**

La possibilitat de que els grafs tinguin més d'una component feblement o fortament connectada depèn de si s'ha utilitzat una exploració en profunditat (DFS) o en amplada (BFS), així com de la selecció de la llavor (node inicial) en aquesta exploració.

Pel que fa a les components feblement connectades, amb DFS hi ha més probabilitat de que apareguin respecte BFS, ja que la funcionalitat del DFS consisteix en recorre una branca en profunditat (busca relacions més complicades).

En el cas de BFS la probabilitat disminueix perquè recorre el graf per nivells i, donat que és molt probable que els artistes estableixin relacions si es troben en un mateix nivell, les components feblement connexes perden probabilitat i les fortament connexes en guanyen.

Tanmateix, la selecció de la llavor inicial també resulta rellevant. Si la llavor està col·locada de tal manera que permet recorre el màxim d'un graf o estratègicament més centrada, és més probable que arribi a connectar molts components. Si es selecciona en una part específica (apartada), pot deixar components separats, cosa que resulta en connexions febles en les components.

- 4. Also justify the relationship between the previous results and the number of connected components in the undirected graphs (gB' and gD').**

Donat que gB' / gD' son les versions d'un graf no dirigit respecte els grafs dirigits gB / gD, el nombre de components connectades disminuirà. Per dur a terme la creació dels grafs gB' / gD' només s'agafaven aquelles arestes que fossin bidireccionals per tant el graf resultant té menys maneres de viatjar entre els nodes i, per tant, es creen menys components connectades.

- 5. Compute the size of the largest connected component from gB' and gD'. Which one is bigger? Justify the result.**

En el cas de gB' el màxim és 100, mentres que a gD' és de 87. En el cas de BFS s'exploren els nodes de manera més uniforme, així doncs, es tarda més en connectar nodes que estan a molta distància, no pas com a DFS. També, si l'exploració es per nivells hi ha més arestes bidireccionals, ja que hi ha més probabilitat de que els artistes es relacionin entre ells perquè estan més pròxims.

Com existeixen més arestes bidireccionals és més probable que es formin components connexes durant la transformació a grafs no dirigits.