David Morillo Masagué (1666540)
Adrià Muro Gómez (1665191)

Lab activity: analysis of music networks.

1. Provide the order and size of the graphs gB and gD.

BFS:
- Ordre: **439**
- Mida: **2000**

DFS:
- Ordre: **569**
- Mida: **2000**

(a) Explain why, having explored the same number of nodes, the order of the two graphs (gB and gD) differs.

Si s'utilitza per recórrer un graf l'algoritme DFS hi ha menys probabilitat que els artistas relacionats del node en el temps *t* estiguin relacionats amb el node *t-1* i, per tant, es descobreixen més nodes a cada iteració. No obstant, el BFS recorre per nivells, així doncs les relacions entre els artistes relacionats tenen més probabilitat de connectar, i per tant, es visiten menys nodes nous.

(b) Justify which of the two graphs should have a higher order.

BFS: **439** nodes / **2000** arestes

DFS: **569** nodes / **2000** arestes

Com es pot observar, el graf recorregut en BFS té menys nodes que el DFS. Així doncs, es confirma l'esdeveniment explicat a l'apartat a). Donat que el DFS s'allunya tot lo possible de la seed les relacions entre els diferents nodes envers els artistes relacionats esdevindran menys probables d'ocòrrer.

(c) Explain what size the two graphs should have.

Teòricament hauria d'haver-hi 2000 arestes, ja que el valor màxim de nodes a explorar es 100 i a cada iteració s'executa la funció *artist_related_artists(act_artist)* que retorna els 20 artistes més relacionats. (20 * 100) = 2000.

2. Indicate the minimum, maximum, and median of the in-degree and outdegree of the two graphs (gB and gD). Justify the obtained values.

gB	in-degree	outdegree
minium	1	0
maximum	38	20
median	4.556	4.556

Graf BFS:

Pel que fa al mínim de l'in-degree és 1 i de l'outdegree és 0 que correspondria a nodes que reben una relació i no s'exploren perquè el màxim de nodes s'excedeix.
Tanmateix, en el màxim observem valors més grans. El valor in-degree màxim és 38 que correspondria a un node que està relacionat fortament amb molts artistes, l'outdegree màxim és 20 ja que, com s'ha explicat anteriorment, la funció d'artistes relacionats retorna els 20 artistes més relacionats amb el node actual com a màxim. S'observa una mitjana de 4.556 que coincideix tant al paràmetre in-degree com al outdegree perquè es compleix sum(in-degree) = sum(outdegree) en tots els grafs.

gD	in-degree	outdegree
minium	1	0
maximum	20	20
median	3.515	3.515

Graf DFS:

El paràmetre mínim no varia amb BFS per la mateixa raó explicada de que es tracta d'un node que rep una relació i després no es segueix iterant a causa del màxim de nodes a explorar. El valor més sorprenent de la taula és el màxim del in-degree que és 20 perquè teòricament en el model DFS un node té moltes menys probabilitats de que tingui un grau molt alt d'in-degree com s'ha explicat a l'apartat a) de l'exercici 1. No obstant, el valor 20 < 38 corresponent al BFS. Pel que fa al màxim de l'outdegree passa lo mateix que en el BFS, la funció retorna els 20 artistes més relacionats. Finalment, la mitjana en el DFS és inferior que el BFS perquè tenen arestes semblants però en el DFS l'ordre és bastant superior al BFS.

3. Indicate the number of songs in the dataset D and the number of different artists and albums that appear in it.

Número de cançons registrades: **1894**
Número d'artistes diferents: **197**
Número d'albums diferents: **1305**

(a) Explain why the number of artists is between 100 and 200, considering the input graphs.

Quan cridem a la funció, un dels paràmetres que li passem és una llista de grafs (en els quals en cadascun hem explorat 100 nodes). Per tant, les dades que analitzarem seran de màxim 200 artistes, suposant que no existeixen artistes repetits. Com que partim d'un mateix artista ("Taylor Swift"), és impossible que no existeixin dos nodes en comú en els dos grafs. En el nostre cas, hi ha 3 nodes exlorats en comú en els dos grafs, per tant, ens quedem en un total de 197.

(b) Justify why the number of songs you obtained is correct, considering the input graphs.

Com que sabem que la funció *sp.artist_top_tracks()* retorna com a màxim 10 cançons, i que en total tenim 197 artistes a analitzar, com a màxim obtindrem 1970 cançons. Degut a que hi ha artistes amb menys de 10 cançons a la seva discografia, segurament n'obtenim, com a màxim, 1970 (molt proper a aquest).
Efectivament, en la pràctica obtenim 1894 cançons, un valor inferior però molt apropat al predit.

(c) Justify why the number of retrieved albums is correct.

Hem obtingut 1305 àlbums diferents, dada que considerem que té sentit, ja que:
El cas mínim seria que totes les millors cançons de tots els artistes fossin del mateix álbum (gràcies a colaboracions), és a dir, un únic àlbum.
L'altre extrem seria que cada millor cançó de cada artista fos d'un àlbum (i que en aquest no hi haguéssin col·laboracions amb altres nodes del graf). (197*10 com a màxim, en la realitat veiem que serien 1894)
Clarament veiem que la segons opció és més probable que la primera (encara que les dues són escenaris quasi impossibles), així que considerem que el resultat d'aproparà més a 1894 que a 1. Comprovem que realment és 1305.