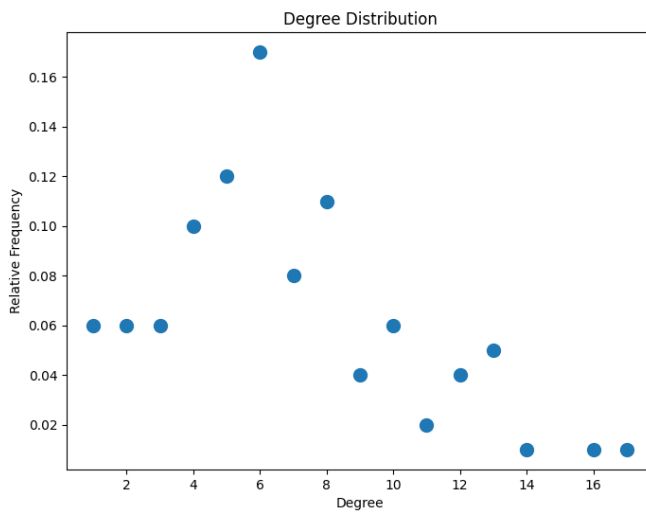


## Lab activity: analysis of music networks

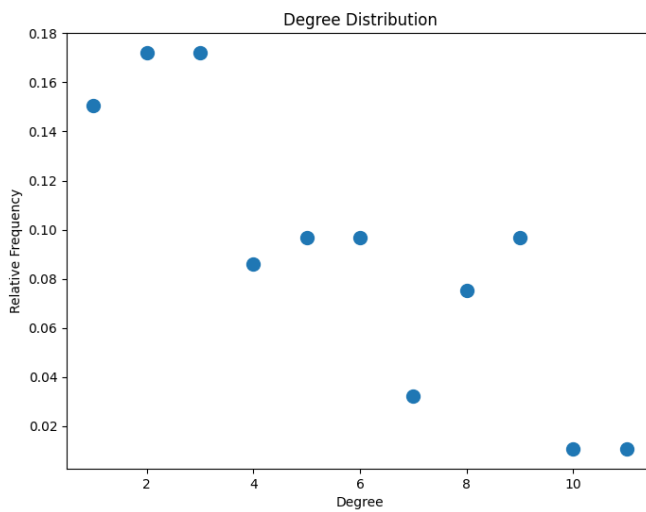
### P4

Plots of the **degree distribution** of the graphs  $gB'$ ,  $gD'$ , and  $gBw$ :

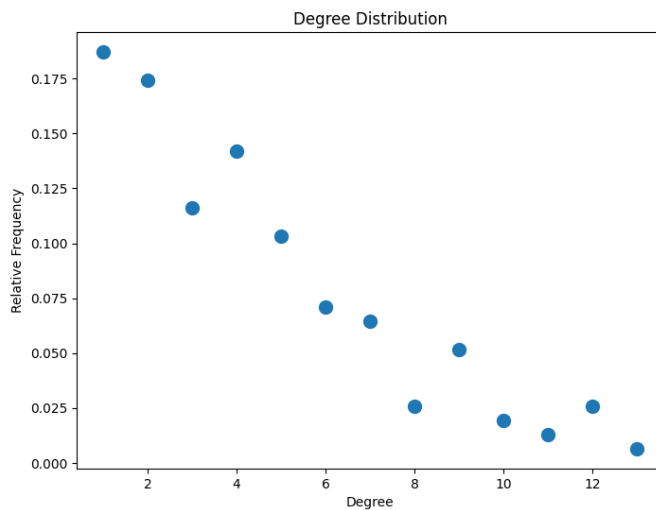
Graf  $gB'$ :



Graf  $gD'$ :



### Graf **gBw**:



a) What are the degree distributions of the three obtained undirected graphs like?

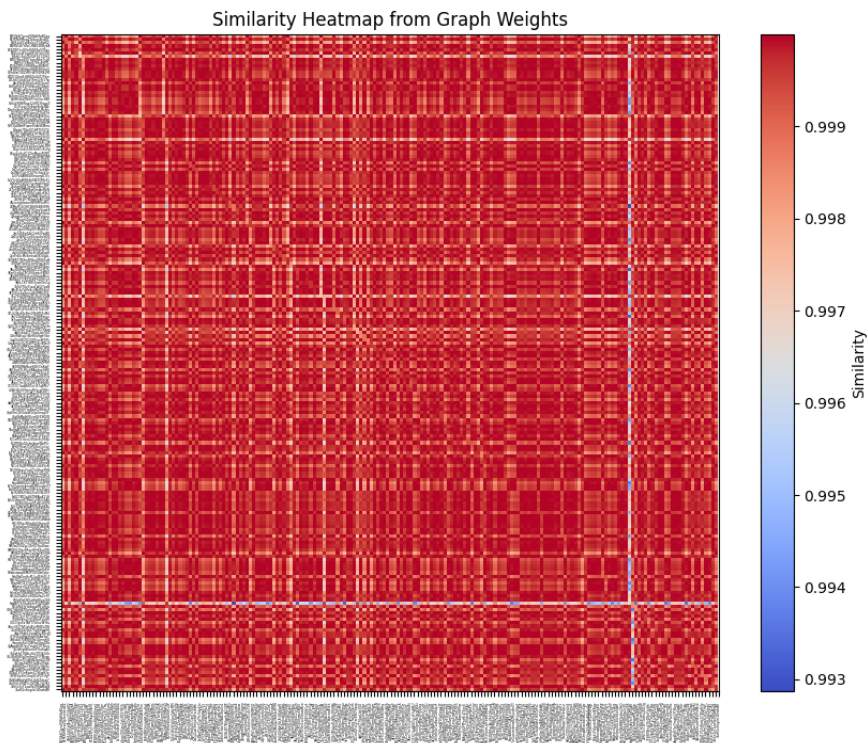
Podem veure com en els grafs **gB'** i **gD'**, existeix poca relació entre el grau dels nodes i freqüència amb la que apareixen en el graf.

En canvi, en el graf **gBw** s'observa una dependència inversament proporcional entre el grau i la freqüència relativa. És a dir, hi ha més nodes amb grau baix, que amb grau alt, i això evoluciona en forma de recta al gràfic com podem observar.

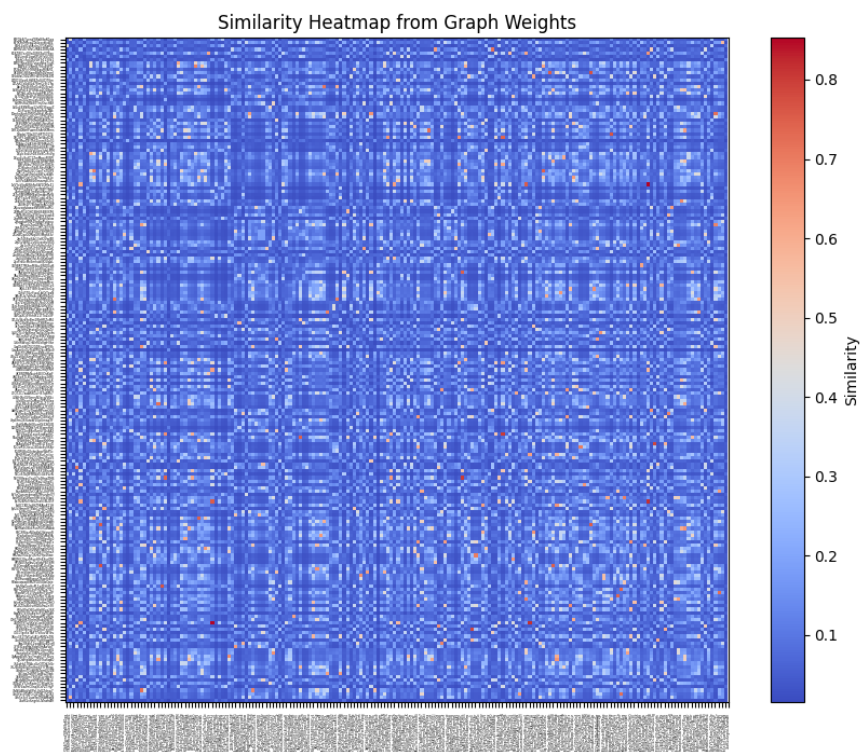
Podem comprovar que la funció `prune_low_weight_edges()`, que és la que retorna el graf **gBw** (i **gDw**),

**Heatmap** showing the **similarity** between all artists in the dataset using the **plot similarity heatmap** function:

**Cosinus similarity:**



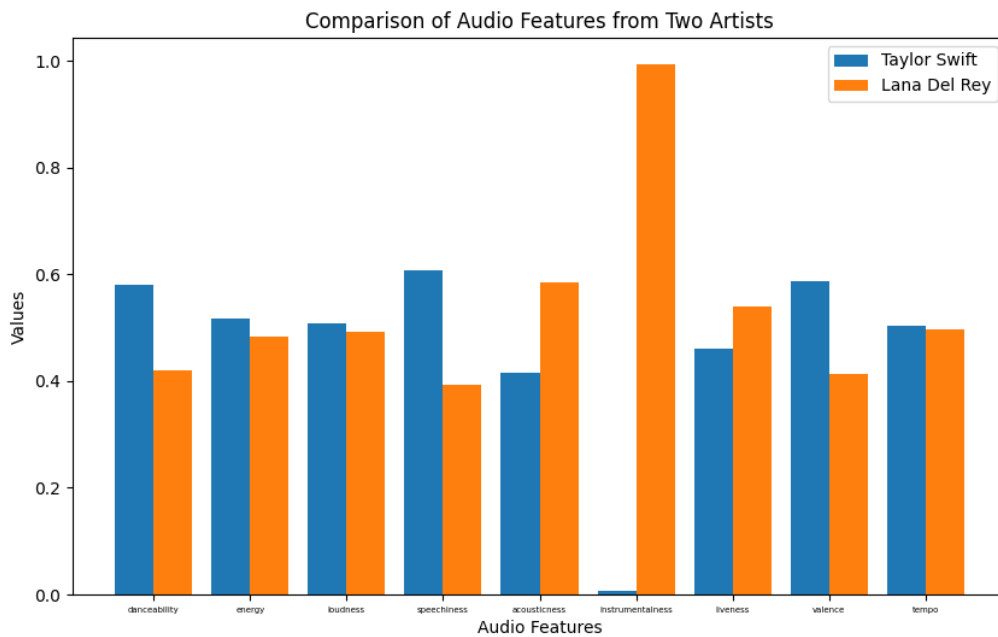
**Euclidean distance:**



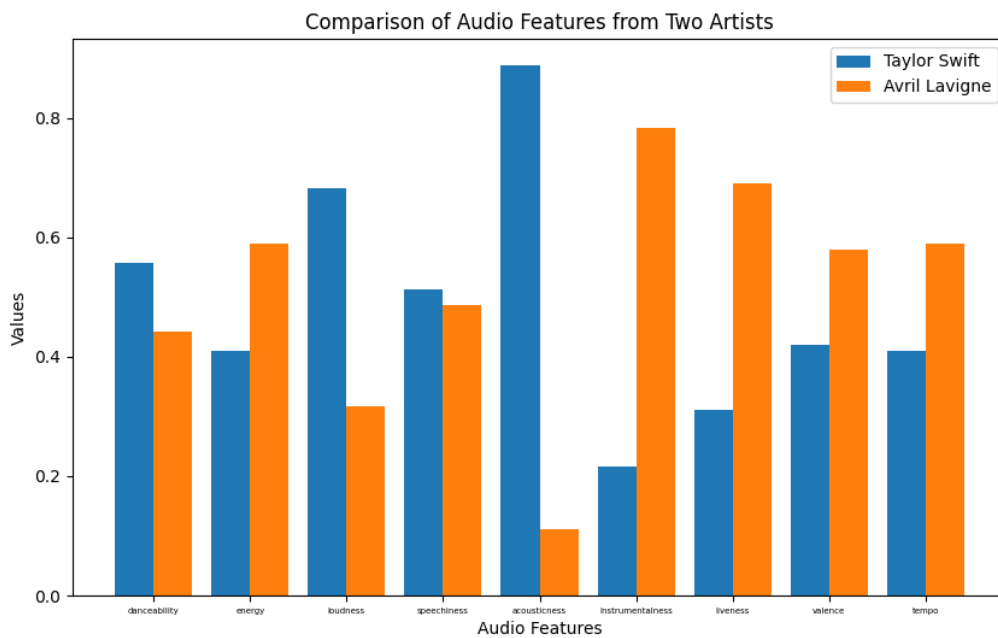
b) What can you infer from the similarity heatmap regarding the algorithm that selects related artists on Spotify?

Es que es pot observar (tant en la matriu de similaritats com la de distàncies), és que els artistes que en formen part, estan molt relacionats entre ells, cosa que no sorprèn quan sabem que l'origen d'aquestes dades d'artistes no ha sigut arbitrari, sinó que hem partit d'un artista (Taylor Swift) i hem explorat d'altres basant-nos en les seves col·laboracions (crawler). Observeu que hi ha valors que sí que destaquen per desviar-se d'aquestes similituds, però són una part molt reduïda i quasi inapreciable.

Comparison between the **most similar** artist to **Taylor Swift** from the graph gB using the plot **audio features** function:



Comparison between the **least similar** artist to **Taylor Swift** from the graph gB using the plot **audio features** function:



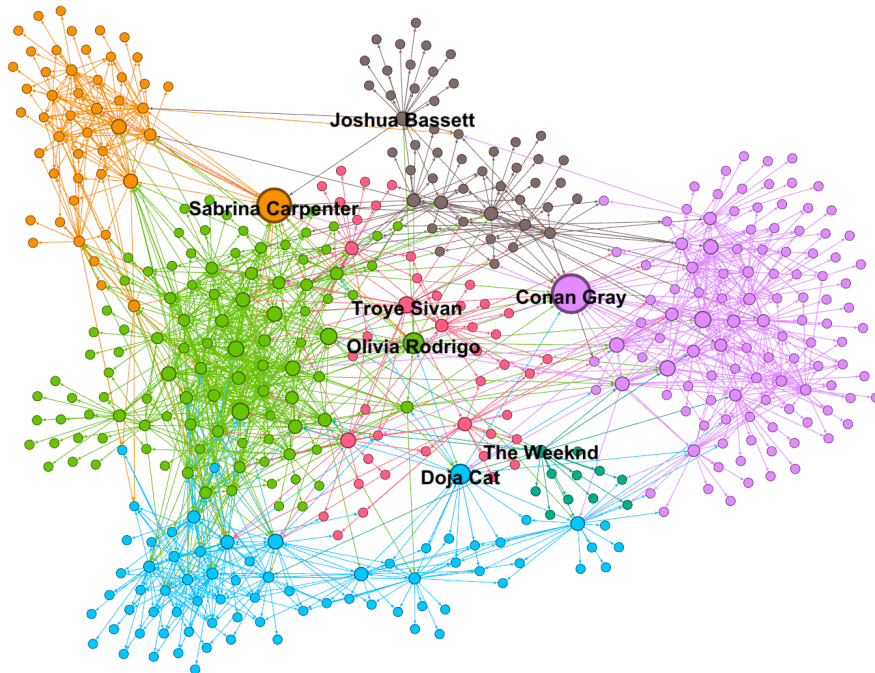
- c) Are the two selected artists similar based on their audio features?  
Comment on the comparison regarding the relationships between artists provided by Spotify (graphs gB and gD).

Com es pot observar, en el primer diagrama, excepte en la categoria *instrumentalness*, **Taylor Swift** i **Lana Del Rey** coincideixen gairebé en el mateix nivell de cada categoria, cosa que els dona una relació de similaritat molt alta. En canvi, en el segon diagrama s'observa com hi ha molt poques categories on el les categories coincideixin. Això fa que **Avril Lavigne** sigui l'artista més distant respecte a **Taylor Swift** (segons el criteri de les nostres funcions i el rang dels nostres grafs)

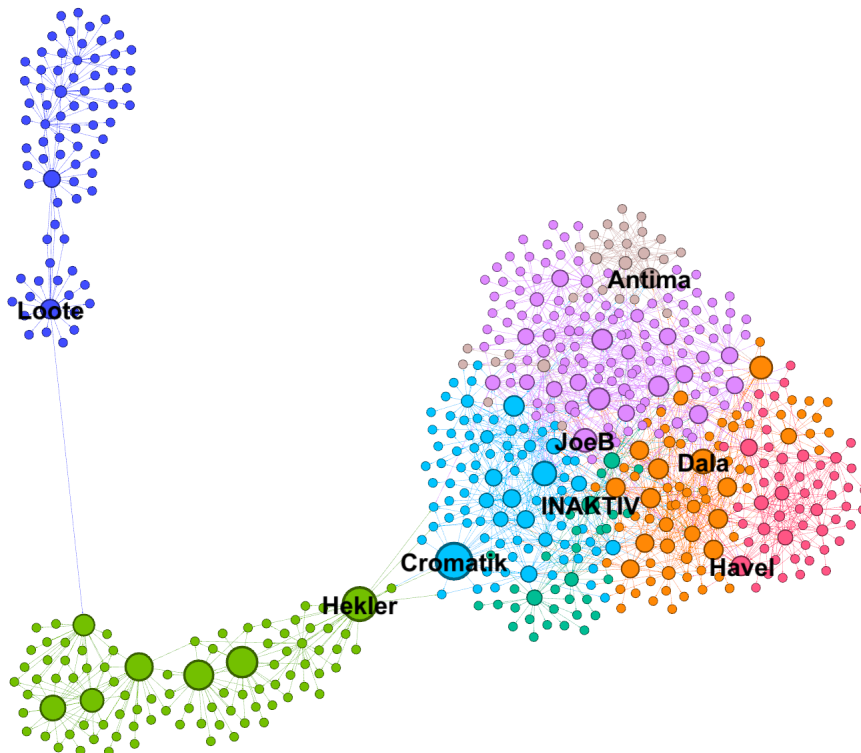
### Visualization of the graphs using Gephi:

- Assigns a color to the nodes based on the community they belong to.
- Sizes the nodes proportionally to their betweenness centrality.
- Shown the names of the most important artists (highest betweenness centrality) in each community.

#### Graf BFS:



#### Graf DFS:



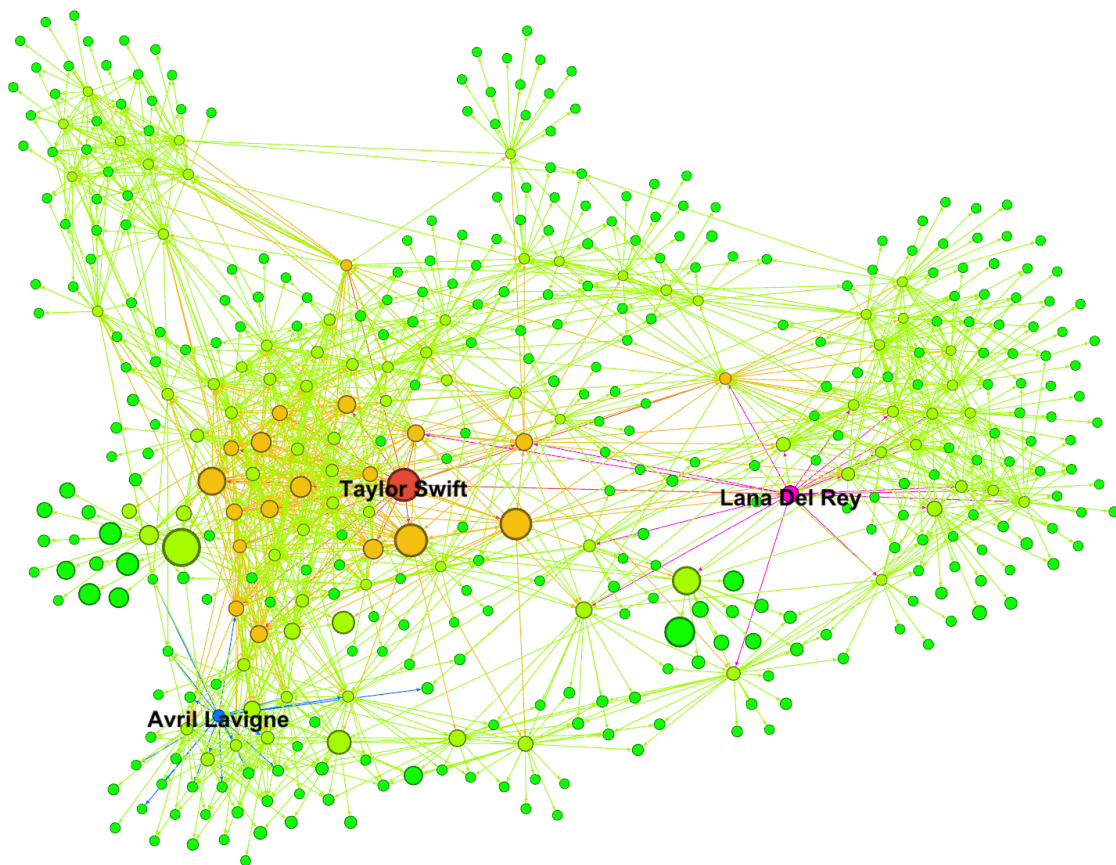
### Visualization of the graphs using Gephi:

- Same node positioning as the previous graphs
- Node size based on their number of followers
- Node color based on the distance of each node from the initial node of the crawler (the node representing the artist Taylor Swift).
- Highlight the two artists selected for the plot audio features comparison (the less and most similar artists to Taylor Swift).

En color **magenta** -> **Més** semblant

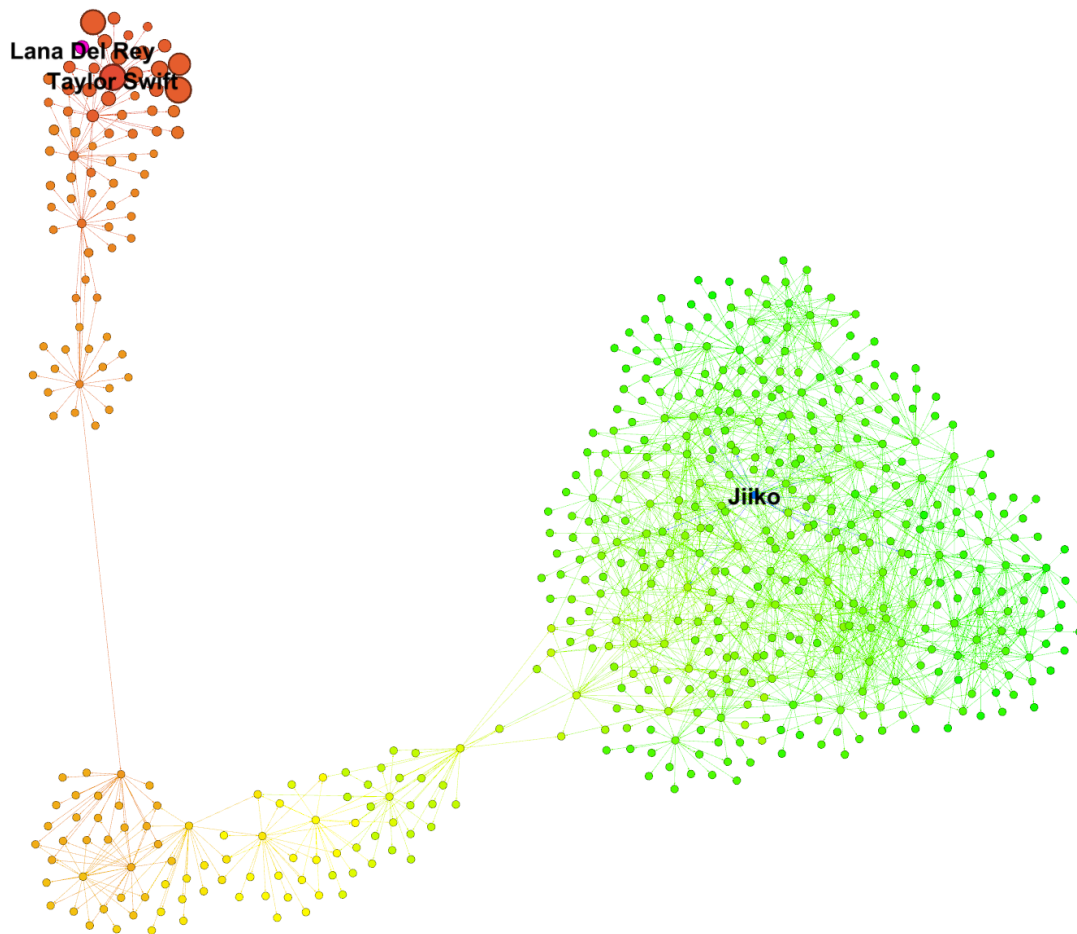
En color **blau** -> **Menys** semblant

### Graf BFS:





## Graf DFS:



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

El primer que es pot percebre entre els grafs és la **forma** que tenen:

- Els grafs provinents d'una cerca BFS formen una comunitat més densa a simple vista, i no s'observen grups distingits quan el mostrem per pantalla a Gephi.

En els grafs DFS en canvi, podem comprovar que han sigut explorats fent servir la profunditat, ja que observem comunitats separades per una aresta (una iteració del crawler).

També notem que el **node inicial** està, visualment, en llocs diferents a la representació:

- En el graf BFS, s'observa que el node inicial, Taylor Swift, es mostra al centre, i que tots els nodes han "crescut" arrel d'aquesta. Per tant, molts dels veïns d'una iteració del BFS coincideixen en els veïns de la següent.
- D'altra banda, en el DFS veiem que per cada aresta de l'iteració del crawler surten aproximadament 20 nodes (~20 related artists), fent que comparteixin un nombre més reduït de veïns, i que s'explorin artistes menys coneguts i que tenen menys a veure amb el node inicial

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

Suposem que l'algoritme fa servir el *clustering coefficient* (per exemple) per a agrupar nodes en comunitats, llavors així Gephi pot separar nodes per grups fent una estimació basada en paràmetres com aquests. I per a que un conjunt de nodes tingui un *clustering coefficient* elevat, han de tenir una sèrie de col·laboracions (arestes) en comú.

Així, predim que el que agrupa els artistes en una comunitat són les col·laboracions que fan entre ells, i les comunitats es podrien etiquetar en conjunts de condicions que faciliten poder tenir al menys una col·laboració, com, per exemple:

- Idiomes
- Països
- Gèneres musicals
- Edat dels artistes
- Gènere dels artistes