Jia Comas (1676245)
Emma Juanico (1672960)
GrupLab-3

QUESTIONS TO ANSWER IN THE REPORT

1. **(1.5 points) Provide the order and size of the graphs g_B and g_D.**
    a. **Explain why, having explored the same number of nodes, the order of the two graphs (g_B and g_D) differs.**

Malgrat explorar el mateix nombre de nodes en ambdós grafs, l'ordre d'aquests difereix, és a dir, no tenen el mateix nombre de nodes. En el graf g_B hi ha **443** nodes i, en el g_D **610**.

La diferència és causada pel nombre de veïns dels nodes explorats (*crawled*), aquells nodes que únicament se'n coneix l'existència (*discovered*). Pel que fa en el graf g_B la probabilitat de que els veïns dels nous nodes a explorar siguin els mateixos dels nodes ja explorats, és a dir, ja estiguin en el graf, és més alta que en el graf g_D.

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

Tal com s'ha mencionat a l'apartat anterior, l'ordre dels dos grafs no és el mateix. Considerant que aquest fet és causat pel nombre de veïns coneguts (*discovered*) a partir dels explorats (*crawled*), en el graf generat mitjançant l'algorisme de DFS l'ordre serà major.

Aquest fet estaria causat per la probabilitat que en els veïns de l'algorisme BFS fossin els mateixos, com s'ha justificat anteriorment, a l'algorisme BFS els nodes explorats es troben cèntrics al node inicial, i per tant, la possibilitat que entre ells siguin veïns és gran. Aquest fet, evitaria que es generessin nous nodes, ja que aquests ja es trobarien en el graf i únicament s'afegiria una aresta. En canvis, a l'algorisme DFS, cada node explorat s'allunya més de l'original i disminueix la possibilitat que els veïns coincideixin. En altres paraules, augmenta la probabilitat que es generin nous nodes de veïns coneguts, nou node i nova aresta.

    c. **Explain what size the two graphs should have.**

A partir dels nodes explorats (max_nodes_to_crawl) i considerant que cada exploració retorna 20 nodes veïns, el nombre màxim d'arestes correspondrà a 100*20 = 2 000. Tot i això, considerant la possibilitat de repetició d'arestes, les quals no s'afegeixen, el nombre d'arestes serà menor.

I seguint amb la mateixa base que els apartats anteriors, el graf generat amb DFS tindrà una mida major que el generat amb BFS. Aquest fet es deu a que la possibilitat de repetir una aresta és menor en el darrer cas en comparació en el primer. Confirmant aquest argument, el graf g_B ha resultat una mida de **2000** i el g_D de **1901**.

Jia Comas (1676245)
Emma Juanico (1672960)
GrupLab-3

2. **(1 point) Indicate the minimum, maximum, and median of the in-degree and out- degree of the two graphs (g_B and g_D). Justify the obtained values.**

	grau d'entrada		grau de sortida	
	g_B	g_D	g_B	g_D
màxim	39	15	20	20
mínim	1	0	0	0
mitjana	4.51	3.12	4.51	3.12

CODI

```
p = dict(gx.in_degree())
```

```
p = dict(gx.out_degree())
```

```
g_max = max(p.values())
g_min = min(p.values())
g_mean = sum(p.values())/len(p)
```

3. **(0.5 points) Indicate the number of songs in the dataset *D* and the number of different artists and albums that appear in it.**

Nombre de cançons: 1516
Nombre d'artistes: 192
Nombre d'àlbums: 1057

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

El nombre d'artistes resulta **192**. Aquest està dins les diverses possibilitats d'explorar diferents nodes, ja que els dos grafs entrats tenen 100 nodes explorats i, per tant, el valor ha d'estar entre 100 -tots els nodes dels dos grafs son els mateixos- i 200 -cap dels nodes coincideixi-.

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

Amb dos grafs de 100 nodes explorats, és a dir, 100 artistes diferents, el valor total de cançons estarà dins l'interval de 1000 i 2000, ja que per cada artista s'agafen les 10 cançons més *top*.

Així doncs, s'ha de considerar la possibilitat que es repeteixin artistes entre els nodes, nodes que únicament es consideren una vegada, i que el node sigui l'artista principal de la cançó, sinó no es guardarà aquella cançó.

   c. **Justify why the number of retrieved albums is correct.**

El valor d'àlbums trobats hauria de variar entre el mateix nombre de cançons, és a dir, entre 1000 i 2000. Si tenim que el nombre de cançons és 1516, el nombre màxim d'àlbums hauria de ser 1516, però com que no totes les cançons es troben en àlbums diferents, hi ha repeticions, el nombre d'àlbums és inferior al de cançons, **1057**.