Session 3 Questions

1. (0.5 points) Indicate the number of nodes shared by the graphs gB and fB (seeds Drake and the last crawled artist from the DFS crawl, respectively); and gB and hB (seeds Drake and French Montana, respectively). Use the function num common nodes.

Common nodes gB and fB: 0
Common nodes gB and hB: 504

As we can see, between gB and fB, there are no common artists, which makes sense as fB starts from the last crawled node of the DFS, therefore, it is very far from Drake, and keeps going deeper and does not find any common nodes. For the gB and hB, there are lots of common nodes.

2. (0.5 points) Calculate the 25 most central nodes in the graph g′B using both degree centrality and betweenness centrality. How many nodes are there in common between the two sets? Explain what information this gives us about the analyzed graph.

We can see that there is a total of 5 common nodes between the 2 metrics, being them:

Vince Staples, A$AP Mob, Curren$y, Yo Gotti, Quality Control

These common nodes provide us valuable information about the structure and connectivity of the graph, meaning that these artists have a very important role in the graph. Them being very well-connected (degree centrality) but also act as intermediaries, being crucial in the information flow of the graph (betweenness centrality).

3. (0.5 points) Find cliques of size greater than or equal to min size clique in the graphs g′B and g′D . The value of the variable min size clique will depend on the graph. Choose the maximum value that generates at least 2 cliques. Indicate the value you chose for min size clique and the total number of cliques you found for each size. Calculate and indicate the total number of different nodes that are part of all these cliques and compare the results from the two graphs.

- g'B → minimum size to generate at least 2 cliques is 7, and there are 7 different cliques
- g'D → minimum size to generate at least 2 cliques is 10, and there are 10 different cliques

4. (0.5 points) Choose one of the cliques with the maximum size and analyze the artists that are part of it. Try to find some characteristic that defines these artists and explain it.
For the artist in the g'B cliques, we can see that all the artists are pretty famous, having the less famous 139487 followers and the minimum popularity being 40. But if we look at the crucial characteristic that define these artists is the genre, there are only 23 different genres, but 19 of them are a type of hip hop or rap, so the genres are very similar.

5. (0.5 points) Detect communities in the graph gD. Explain which algorithm and

parameters you used, and what is the modularity of the obtained partitioning. Do you consider the partitioning to be good?

Using the Louvain algorithm, which is computed much faster, the modularity obtained is 0.7, this value suggests that the community structure of the network is pronounced, having a big amount of densely connected nodes as it is very close to 1.

6. (1 point) Suppose that Spotify recommends artists based on the graphs obtained by the crawler (gB or gD). While a user is listening to a song by an artist, the player will randomly select a recommended artist (from the successors of the currently listened artist in the graph) and add a song by that artist to the playback queue.
(a) Suppose you want to launch an advertising campaign through Spotify. Spotify allows playing advertisements when listening to music by a specific artist. To do this, you have to pay 100 euros for each artist to which you want to add ads. What is the minimum cost you have to pay to ensure that a user who listens to music infinitely will hear your ad at some point? The user can start listening to music by any artist (belonging to the obtained graphs). Provide the costs for the graphs gB and gD, and justify your answer.

For the graph gB, as every artist has 20 successors and it has been created using BFS, we could select the 20 successors of an artist, and add the ad, so it would cost 20 * 100 = 2000€ and we would ensure that the ad is seen.

For the graph gD, as we are selecting 1 successor and going as deep as possible with it, we can't really ensure that one of the successors of the artist is selected, therefore we should add one ad to every artist in the graph.

(b) Suppose you only have 400 euros for advertising. Which selection of artists ensures a better spread of your ad? Indicate the selected artists and explain the reason for the selection for the graphs gB and gD.

Only having 400€ means that only 4 artists can be selected for each of the graphs. Therefore we would need to select the artist with a higher betweenness centrality, meaning that they are crucial in the information flow of the graph, and making them very likely to be traversed, making the user listen to the ad.

For the gB graph they would be:

Wale	0.03574463142334019
DJ Drama	0.03242108806018571
Yo Gotti	0.026792411958075336
A$AP Mob	0.0245935484690123

For the gD graph they would be:

Dat Boi T	0.08453551566537384
Fat Bastard	0.08336961592006914
Mr. Vic	0.08029570768128935
Thee Suspect	0.08010896838487057

7. (1 point) Consider a recommendation model similar to the previous one, in which the player shows the user a set of other artists (defined by the successors of the currently listened artist in the graph), and the user can choose which artist to listen to from that set. Assume that users are familiar with the recommendation graph, and in this case, the gB graph is always used.

(a) If you start by listening to the artist Young Dro and your favorite artist is Travis Porter, how many hops will you need at minimum to reach it? Give an example of the artists you would have to listen to in order to reach it.

In order to find the amount of hops and the artist we should visit we need to compute the shortest path going from Young Dro to Travis Porter, and the length of it is the minimum hops needed to reach it. So we would need 5 hops, and the nodes are:

Young Dro,Rocko, Yo Gotti,Young Money, Soulja Boy, Travis Porter