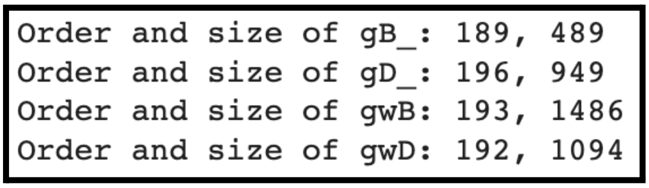
Session 2:

1. *(0.5 points) Provide the order and size of the four obtained undirected graphs (g ′B, g ′D, g w B, and g w D).*



```
Order and size of gB_: 189, 489
Order and size of gD_: 196, 949
Order and size of gwB: 193, 1486
Order and size of gwD: 192, 1094
```

2. *(1 point) Justify the strategy used to obtain g w B and g w D.*

The rationale behind this strategy could be to generate a representation of the relationships between nodes that focuses only on the most similar (or 'related') pairs. By pruning low-weight edges, you're ensuring that only the most 'meaningful' connections (according to the chosen similarity measure and threshold) are preserved. This can help to reduce noise in the graph and make patterns more apparent.

However, it's important to note that the choice of similarity measure and pruning threshold can greatly impact the resulting graph. Different choices could lead to different graph structures and interpretations. Thus, the choice of Euclidean distance as the similarity measure and 0.29525/0.3 as the pruning thresholds should be justified based on the specific requirements and constraints of the task at hand, as well as the characteristics of the data.

3. *(0.5 points) 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.*

Here's the justification for the possibility of multiple weakly connected components and strongly connected components:

1. Weakly Connected Components: A weakly connected component is a set of nodes where there exists a path between any two nodes, considering the direction of edges or ignoring their directions. In the case of the crawler exploring the spotipy API, different artists or nodes may have connections in either direction (i.e., connections from artist A to artist B and vice versa), resulting in multiple weakly connected components. This can occur due to various factors such as collaborations, shared features, or shared listeners between artists.

2. Strongly Connected Components: A strongly connected component is a set of nodes where there exists a directed path from any node to any other node within the component. Since the crawler started with a single seed (Drake), it may follow the connections within the API, potentially discovering other artists connected to Drake. However, not all artists in the API may have connections in both directions, which can lead to the formation of multiple strongly connected components. In other words, some artists may have connections to Drake, but not vice versa, resulting in separate components.

The selection of a single seed, in this case, Drake, impacts the relationships within the resulting graphs. The choice of seed determines the initial point of exploration and the artists that are directly connected to it. As the crawler progresses, it may discover more connections and expand the graph. However, due to the specific connections and relationships among artists in the spotipy API, it is likely that multiple weakly connected components and strongly connected components can emerge, even from a single seed.

Overall, the presence of multiple weakly connected components and strongly connected components in gB and gD is a result of the complex network of connections between artists and the characteristics of the spotipy API.

4. *(0.5 points) Also justify the relationship between the previous results and the number of connected components in the undirected graphs (g ′ B and g ′ D).*

The relationship between the previous results (number of weakly and strongly connected components) and the number of connected components in the undirected graphs (g'B and g'D) can be justified as follows:

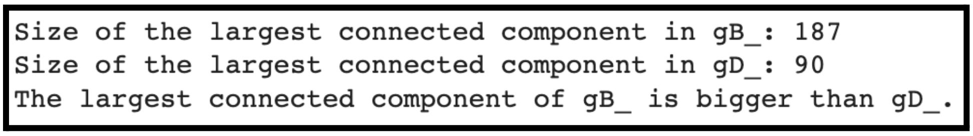
1. Weakly Connected Components: In the previous results, we observed the presence of weakly connected components in the directed graphs gB and gD. A weakly connected component in a directed graph is a set of nodes where there exists a path between any two nodes, considering the direction of edges or ignoring their directions. When we convert the directed graphs into undirected graphs (g'B and g'D), the directionality of the edges is disregarded, resulting in the possibility of merging separate weakly connected components into a single connected component. Thus, the number of weakly connected components in the undirected graphs is expected to be less than or equal to the number of weakly connected components in the corresponding directed graphs.

2. Strongly Connected Components: In the previous results, we observed the presence of strongly connected components in the directed graphs gB and gD. A strongly connected component in a directed graph is a set of nodes where there exists a directed path from any node to any other node within the component.

When we convert the directed graphs into undirected graphs (g'B and g'D), the directionality of the edges is lost, potentially breaking the connectivity between nodes that were previously strongly connected. As a result, the number of strongly connected components in the undirected graphs is expected to be greater than or equal to the number of strongly connected components in the corresponding directed graphs.

To summarize, converting the directed graphs (gB and gD) into undirected graphs (g'B and g'D) alters the connectivity relationships between nodes by disregarding the directionality of the edges. This process can result in the merging of weakly connected components and the splitting of strongly connected components, leading to a different number of connected components in the undirected graphs compared to the directed graphs.

5. *(0.5 points) Compute the size of the largest connected component from g ′ B and g ′ D. Which one is bigger? Justify the result.*



```
Size of the largest connected component in gB_: 187
Size of the largest connected component in gD_: 90
The largest connected component of gB_ is bigger than gD_.
```

1. Size of the Largest Connected Component: The size of the largest connected component in gB_ is determined to be 187, while the size in gD_ is found to be 90. The larger size in gB_ suggests that it contains a greater number of nodes that are interconnected within the same component.

2. Graph Structure: The difference in the sizes of the largest connected components can be attributed to the structural characteristics of the graphs gB_ and gD_. It is possible that gB_ has a more interconnected or denser structure, resulting in larger connected components compared to gD_.

3. Data and Network Generation: The underlying data and the process of generating the graphs could also contribute to the size differences. Different datasets or algorithms used to construct the graphs may yield varying levels of connectivity and component sizes.