Daniel Vidal 1634599
Neil de la Fuente 1630223
Jordi Longaron 1630483

Report on Session 2:
Data preprocessing

Introduction

In this session, we take the data loaded from session one and do the processing of the data using some functions to retrieve and compare information from graphs for easier management of the data.

Questions to answer in the report:

1. Provide the order and size of the four obtained undirected graphs (g'_B, g'_D, g^w_B, and g^w_D).

```
gB' size = 488   gB' order = 188

gD' size = 582   gD' order = 189

gwB' size = 506 gwB' order = 342  **revisar
```

2. Justify the strategy used to obtain g^w_B and g^w_D.

3. Justify whether the directed graphs obtained from the initial exploration of the crawler (g_B and g_D) can have more than one weakly connected component and one strongly connected component, and explain why. Indicate the relationship with the selection of a single seed.

It wouldn't make sense that there were multiple weakly connected components since it is all done by the crawler in connection to one node. On the other hand it makes sense that there would be many strongly connected components, thanks to loops in the tree/graf. If there were multiple seeds then it would be more likely there were more weakly connected components.

```
Breadth First Search:
Number of strongly connected components is 582
Number of strongly connected components with more than on node is
1
Number of weakly connected components is 1
```

```
Depth First search:
Number of strongly connected components is 807
Number of strongly connected components with more than one node is 4
```

4. Also justify the relationship between the previous results and the number of connected components in the undirected graphs (g'_B and g'_D).

5. Compute the size of the largest connected component from g'_B and g'_D. Which one is bigger? Justify the result.

The largest connected components for g'b and g'd are 186 and 94 respectively, making g'b bigger, which makes sense since it is much more likely for breadth first search to end up looping back to a predecessor than it is for depth first search.

```
The largest connected component of the DFS network has 186 nodes.
The largest connected component of the BFS network has 94 nodes.
```