

Q4

```
Silhouette Coefficient, k=2: 0.0021
```

```
Silhouette Coefficient, k=3: 0.0018
```

```
Silhouette Coefficient, k=4: 0.0019
```

```
Silhouette Coefficient, k=5: 0.0018
```

```
Silhouette Coefficient, k=6: 0.0013
```

Discussion:

The intrinsic unsupervised cluster evaluation method I used was the Silhouette coefficient, which ranges from -1 to 1. With this evaluation metric 0 indicates clusters with a high degree of overlap, while 1 represents clusters with high separability and density among the data points. Since in the results I obtained, shown above, k=2 is the highest coefficient value, indicating the most cluster separability, I selected k=2 clusters as input for the Agglomerative hierarchical clustering algorithm. The results for this clustering algorithm with document titles as input are shown in the dendrogram below. I truncated the dendrogram to allow visibility of the clustered document titles. The truncation mode used is level with p=5, indicating 5 levels of clustering are shown prior to truncation.

Q5

