Progress Update

1. We selected the right density o.12 which ensures connectivity as noted below:

A graph with a line

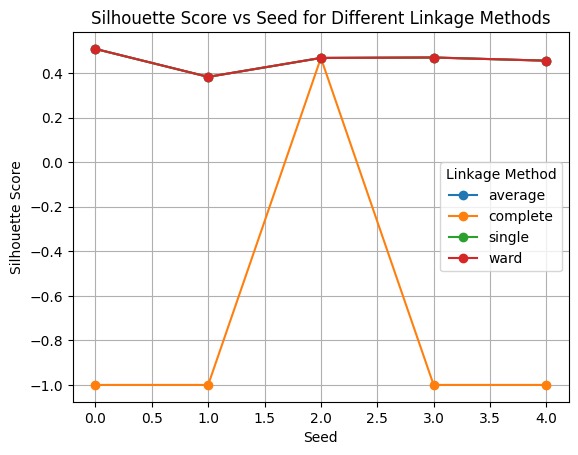
Description automatically generated

1. The overall Average Proportion of Preferred Color Nodes: 0.8750553303249897 for the entire dataset
2. A view of the path lengths

A diagram of a path length

Description automatically generated

1. We selected the ward method because it is more stable, scalable and outperforms the rest of the linkage methods.



1. Average accuracy of the ward method is as follows:

A graph with blue lines and points

Description automatically generated

1. The average accuracy vs silhouette score

A graph with a line and a line

Description automatically generated with medium confidence

1. Results of dominant color per cluster
2. Color Counts per Cluster:
3. Color Preference green red
4. Cluster
5. 1 75 0
6. 2 0 88
7. Dominant Color Results:
8. Cluster Dominant Color Dominant Color Percentage
9. 0 1 green 100.0
10. 1 2 red 100.0
11. Out[152]:

|  | **Cluster** | **Dominant Color** | **Dominant Color Percentage** |
| --- | --- | --- | --- |
| **0** | 1 | green | 100.0 |
| **1** | 2 | red | 100.0 |

1. The confusion matrix with 2 clusters show high accuracy rate

A blue squares with white text

Description automatically generated

8. Increasing to 4 colors, the clustering seems to be working well, we have to adjust the number of clusters though, and the confusion matrix doesn’t seem to be scaling. The biggest question is, how do you assign the correct ground truct?