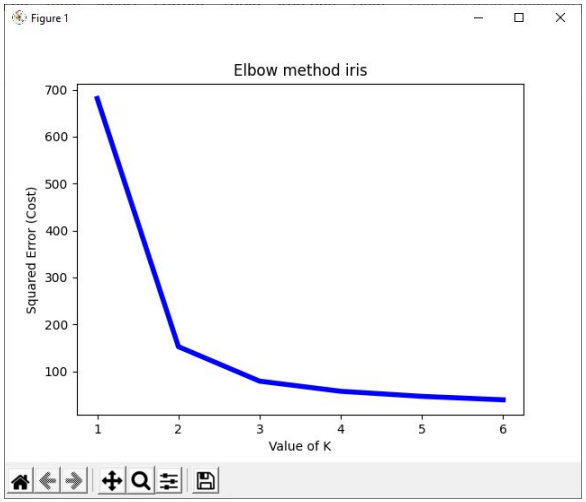
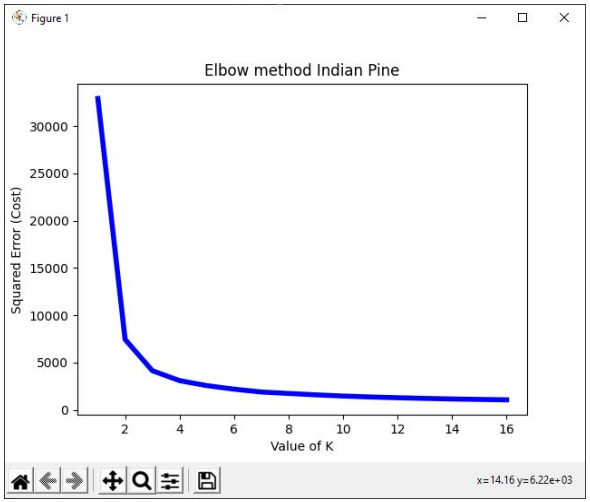
CS 488 Homework 4
Yingren Wang

1a.i) Elbow method plot iris



1a. ii) Elbow method plot Indian pines

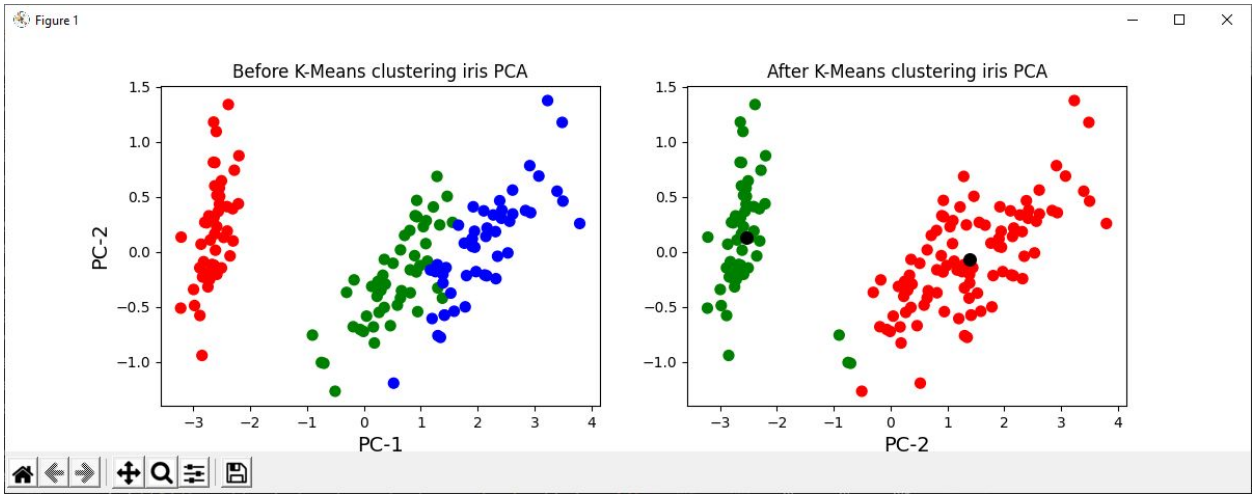


1b.i) I chose 2 as the k clusters for both Indian pines and Iris because that's where the elbow method plot shows the first rapid drop.

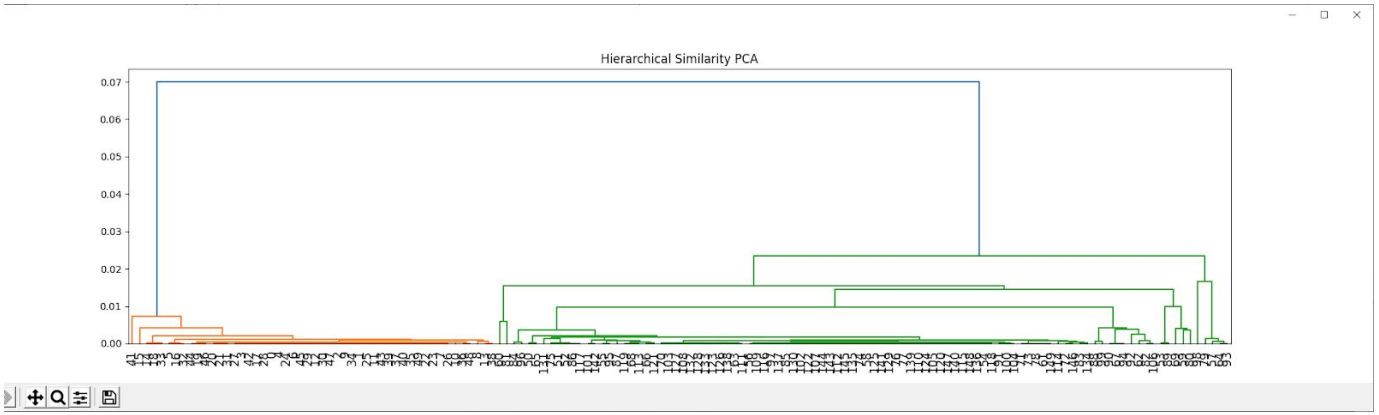
2.

Reduction Iris

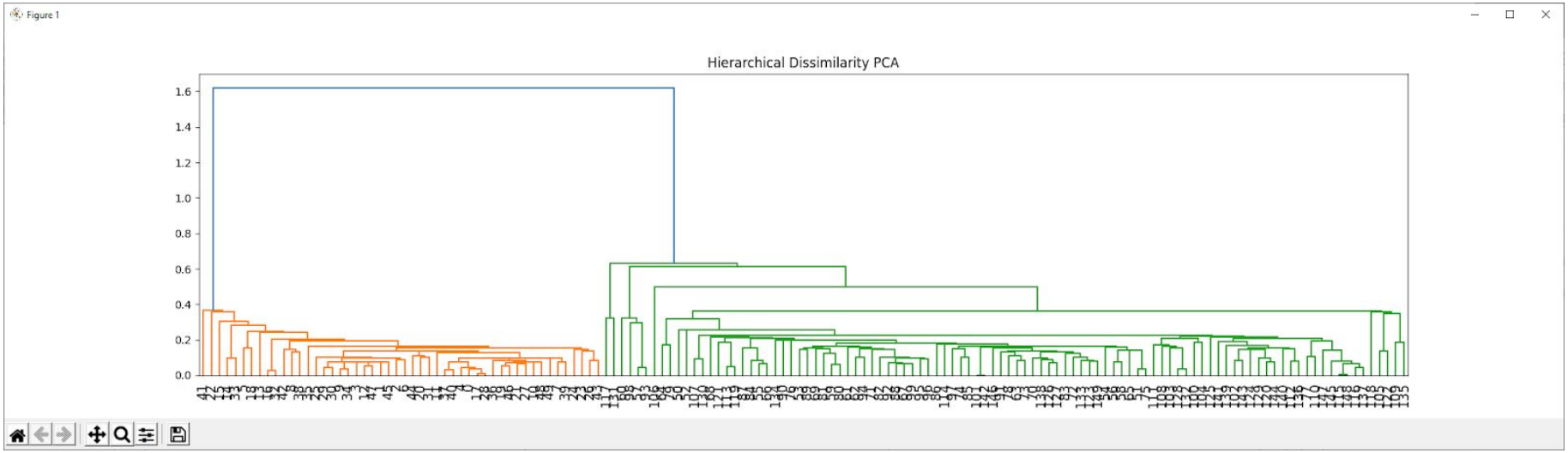
PCA K-Means Iris



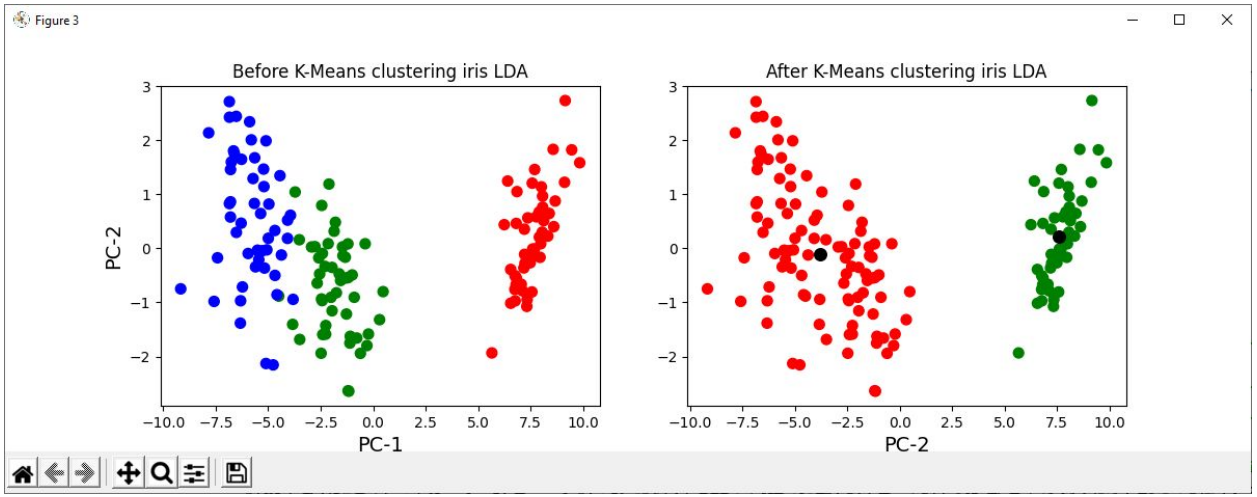
PCA Hierarchical similarity Iris



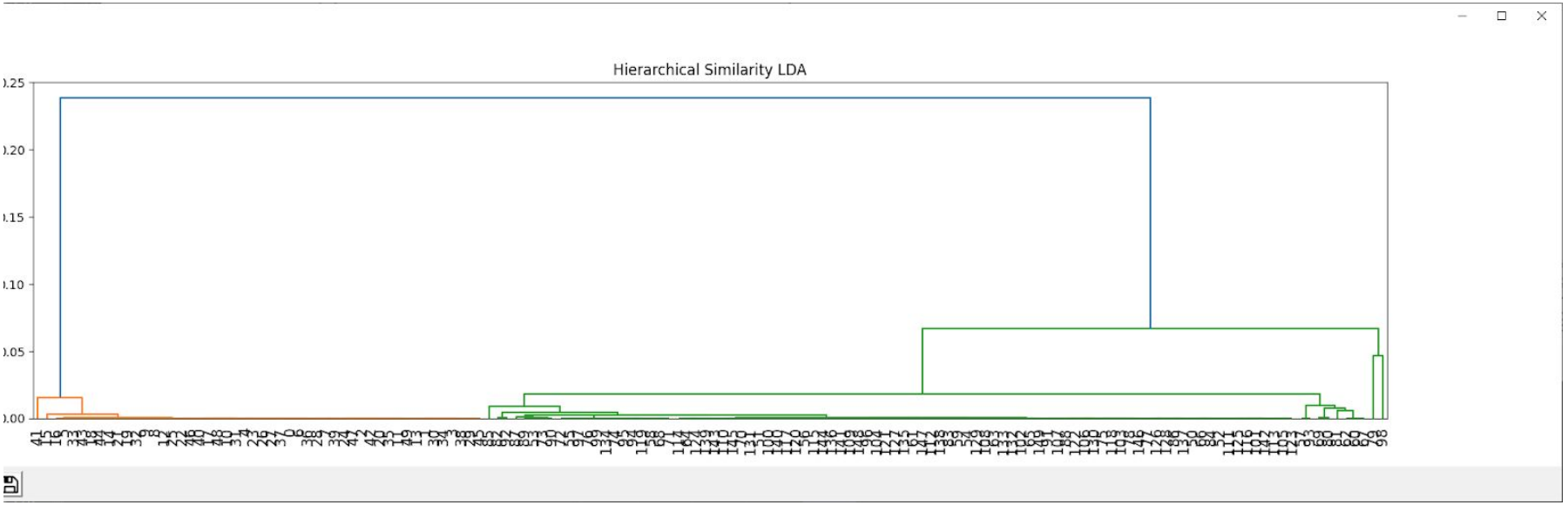
PCA Hierarchical dissimilarity Iris



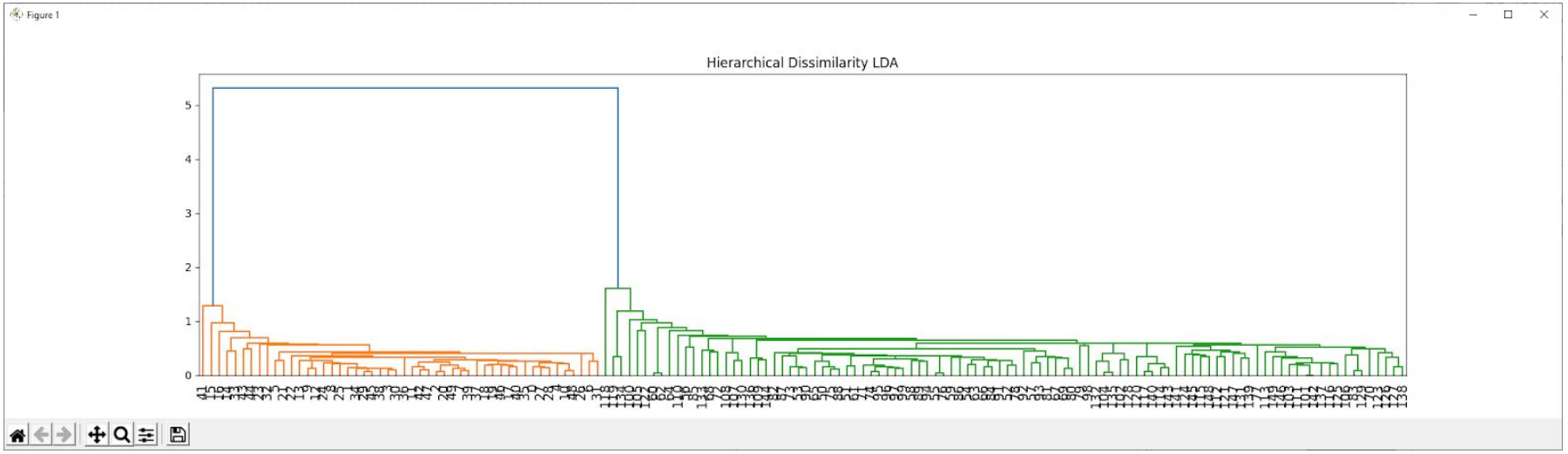
LDA K-Means Iris



LDA Hierarchical similarity Iris

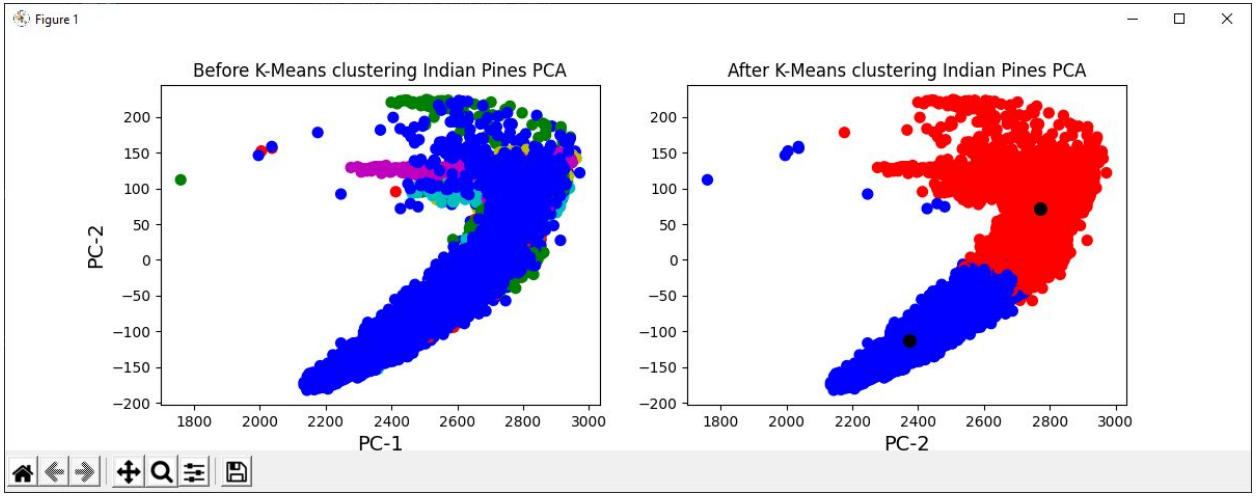


LDA Hierarchical dissimilarity Iris

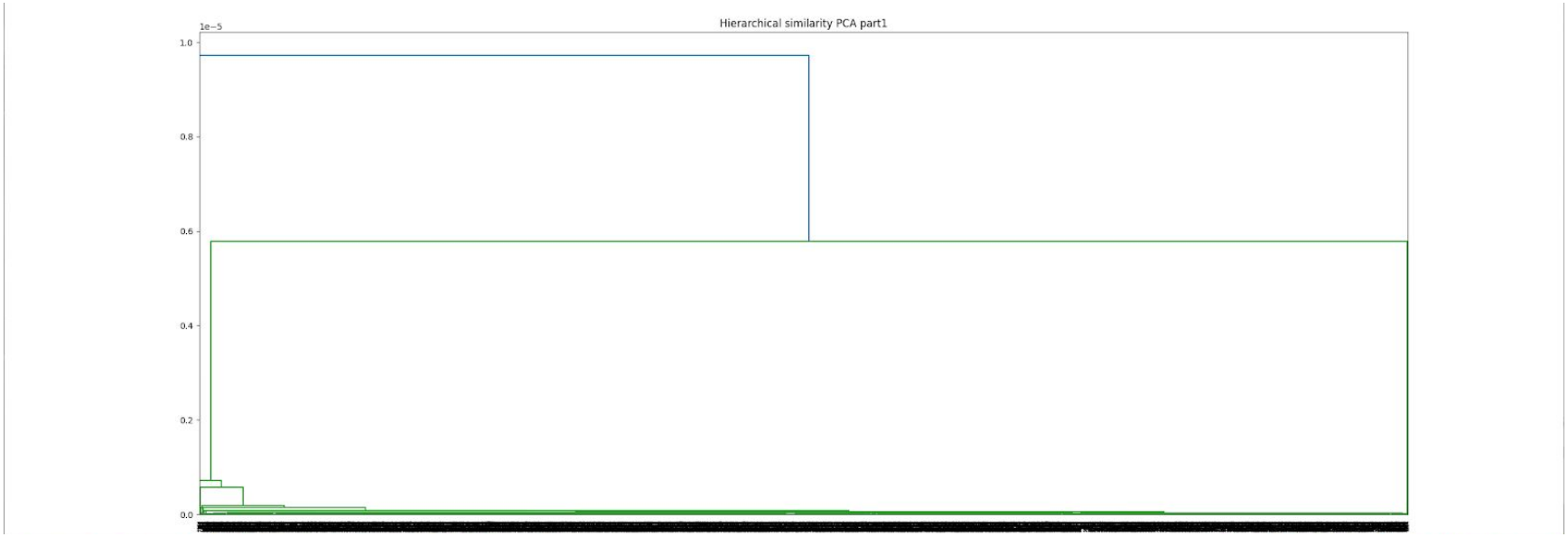


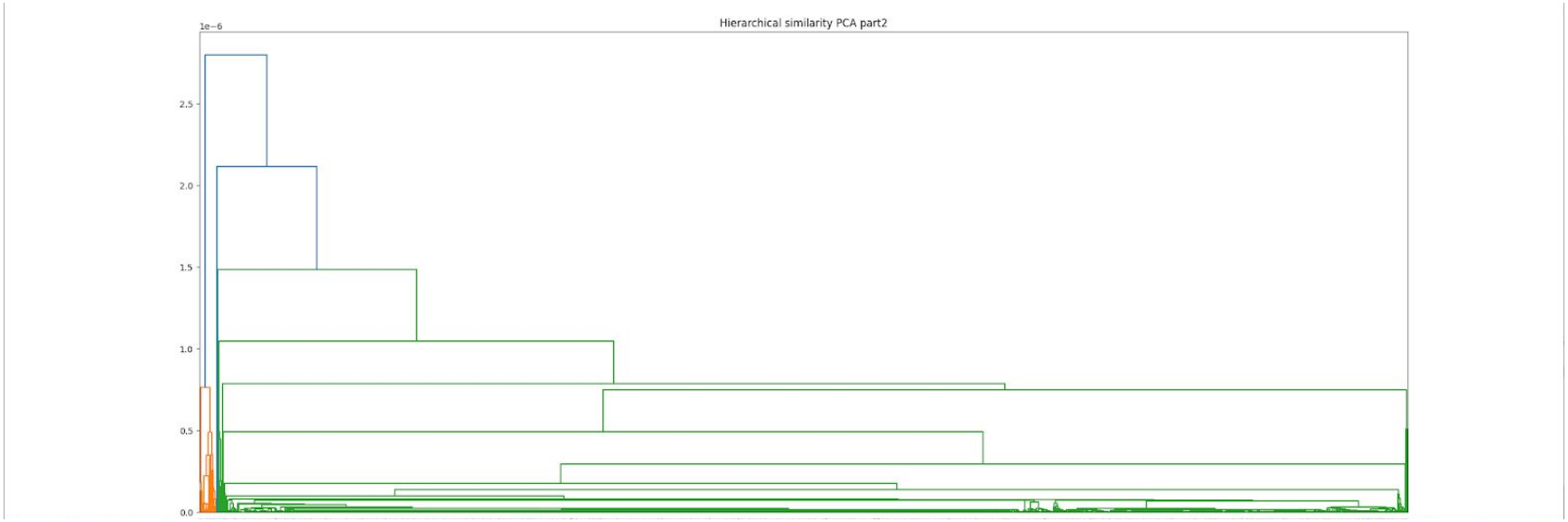
Reduction Indian Pines

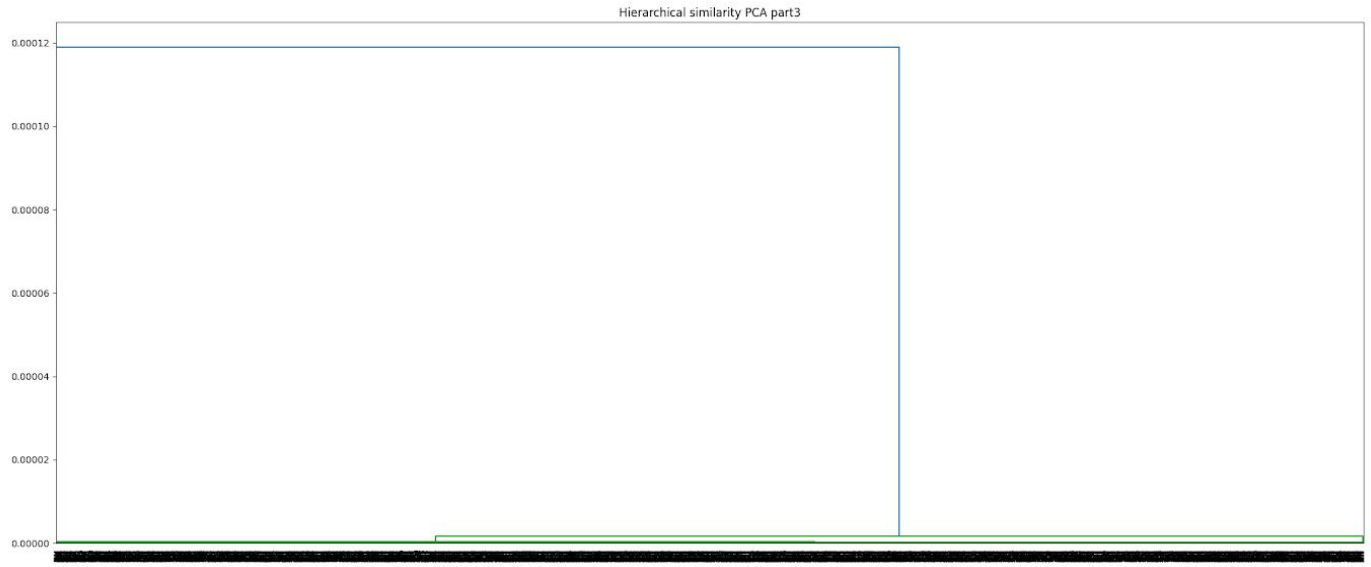
PCA K-Means Indian Pines

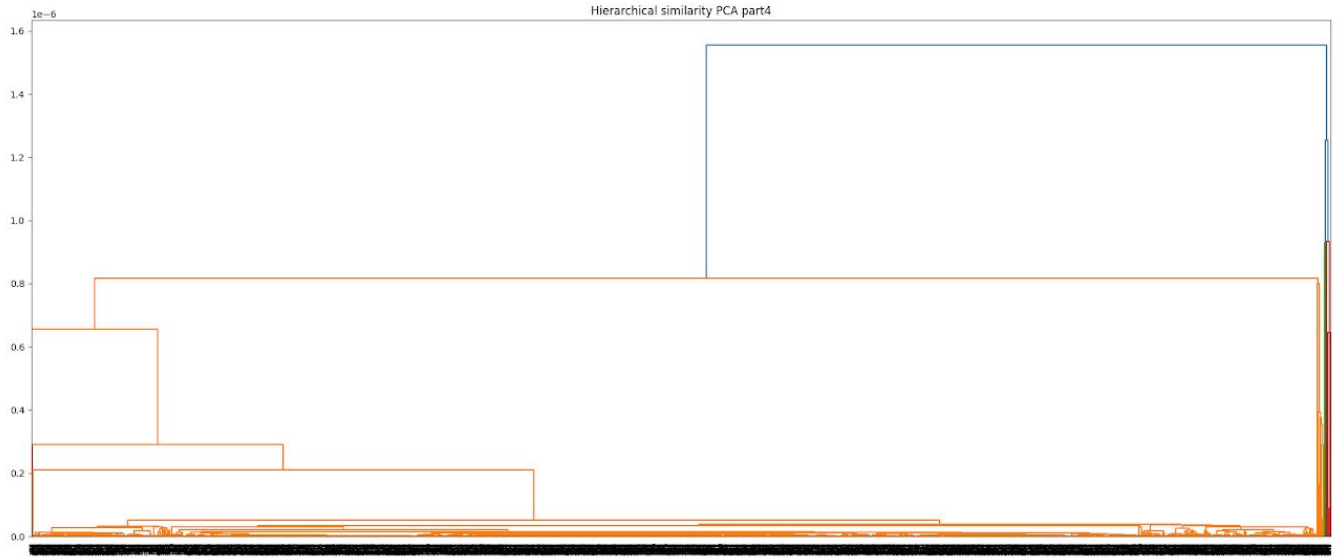


PCA Hierarchical similarity Indian Pines

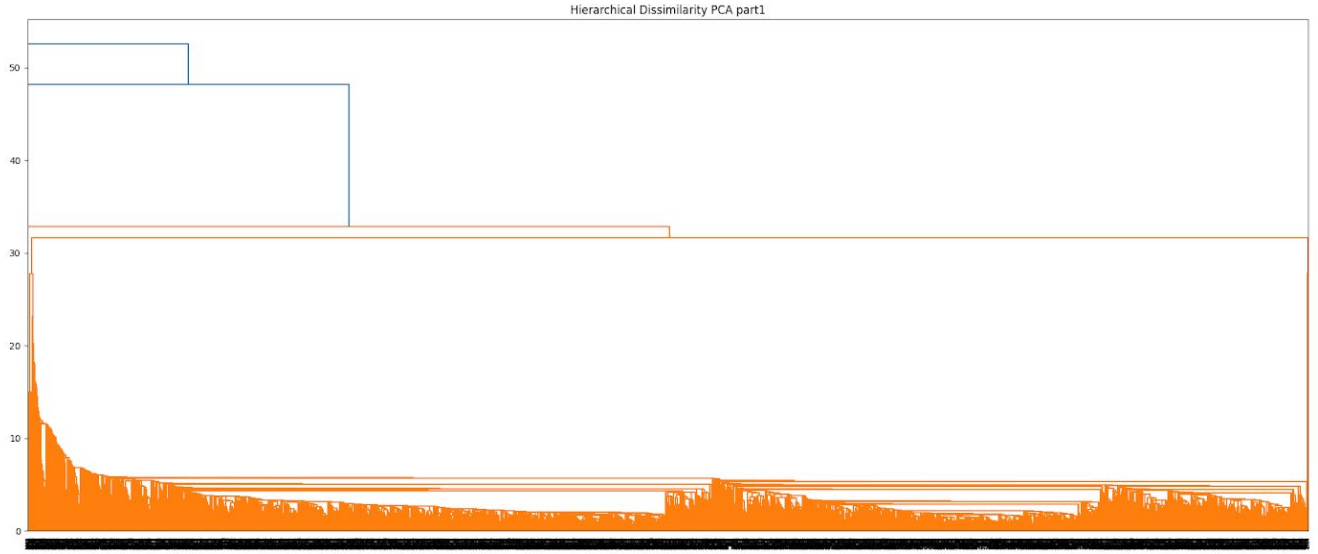


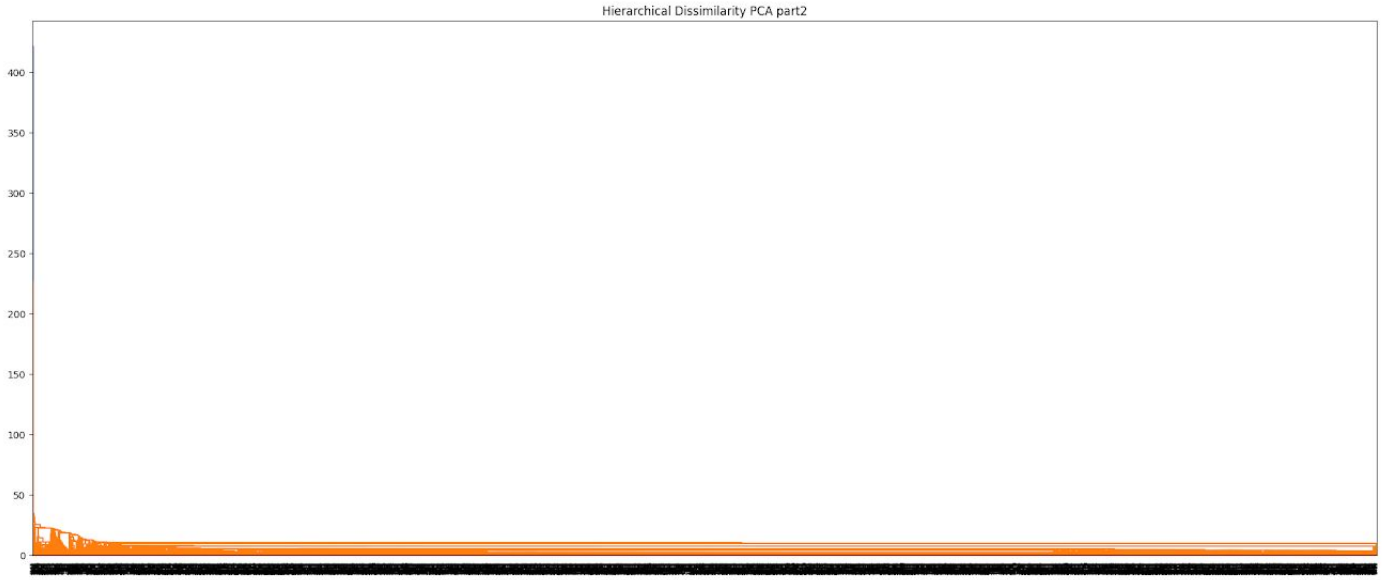


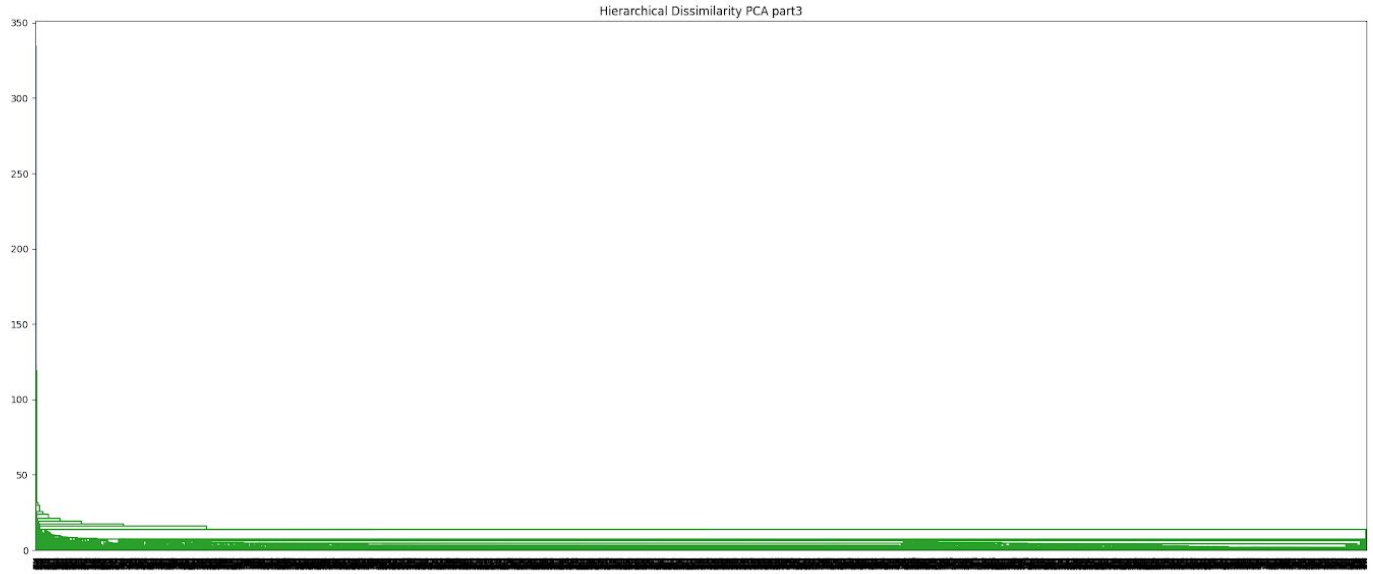

Hierarchical similarity PCA part3
0.00012
0.00010
0.00008
0.00006
0.00004
0.00002
0.00000

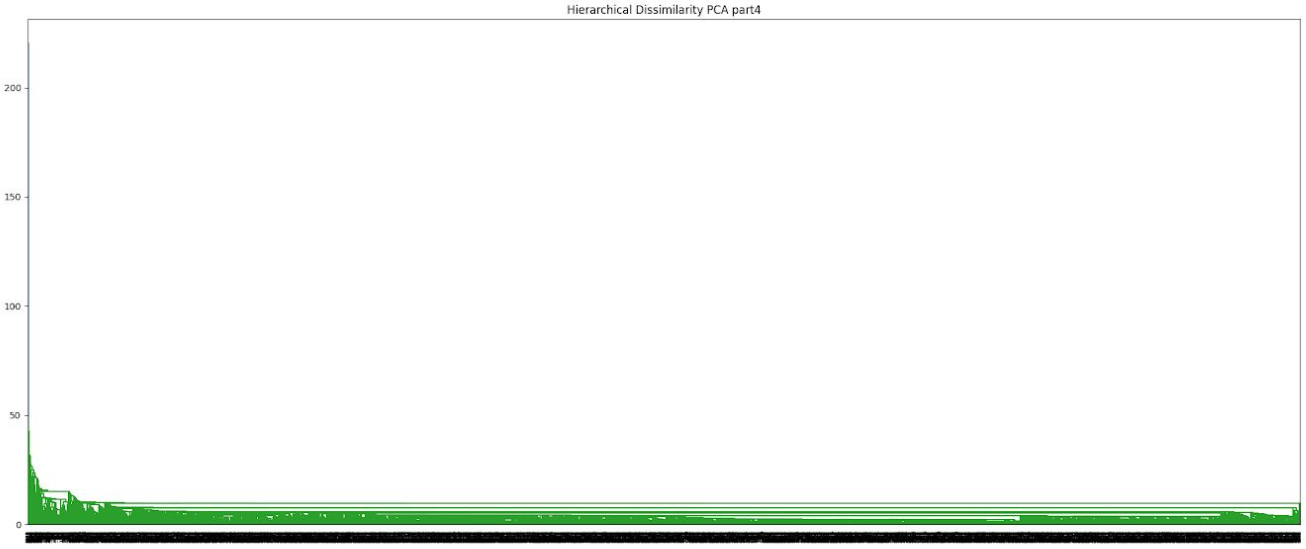

Hierarchical similarity PCA part4
1e−6
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0

PCA Hierarchical dissimilarity Indian Pines

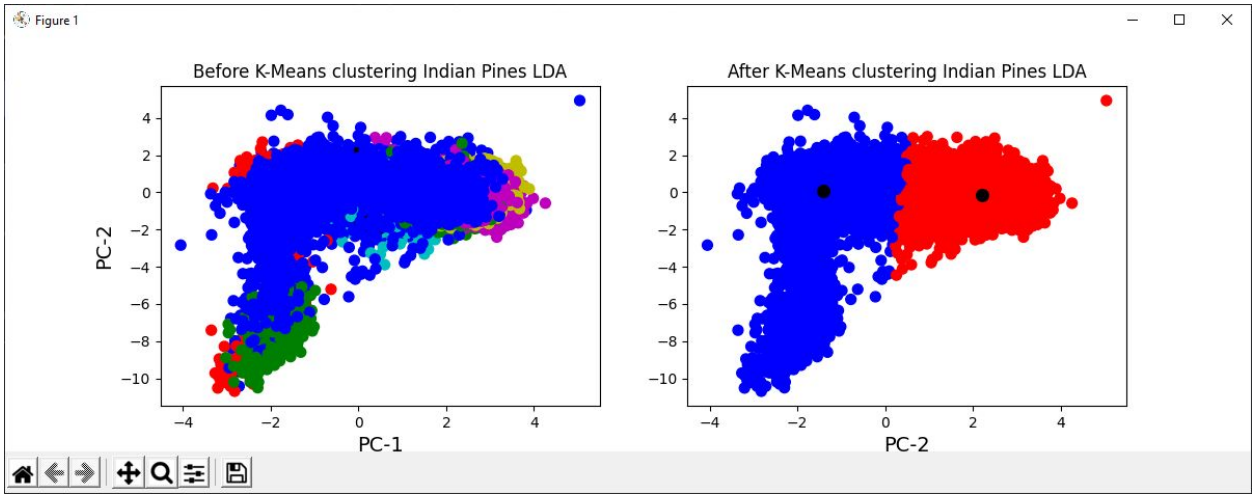




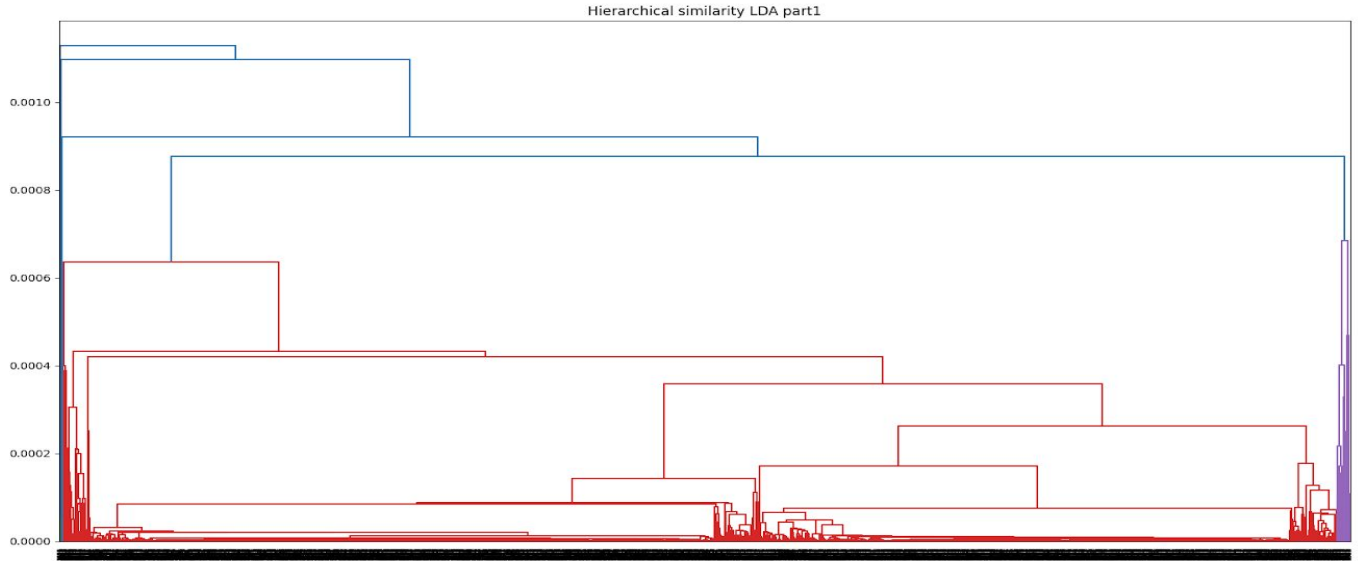

Hierarchical Dissimilarity PCA part3
350
300
250
200
150
100
50
0

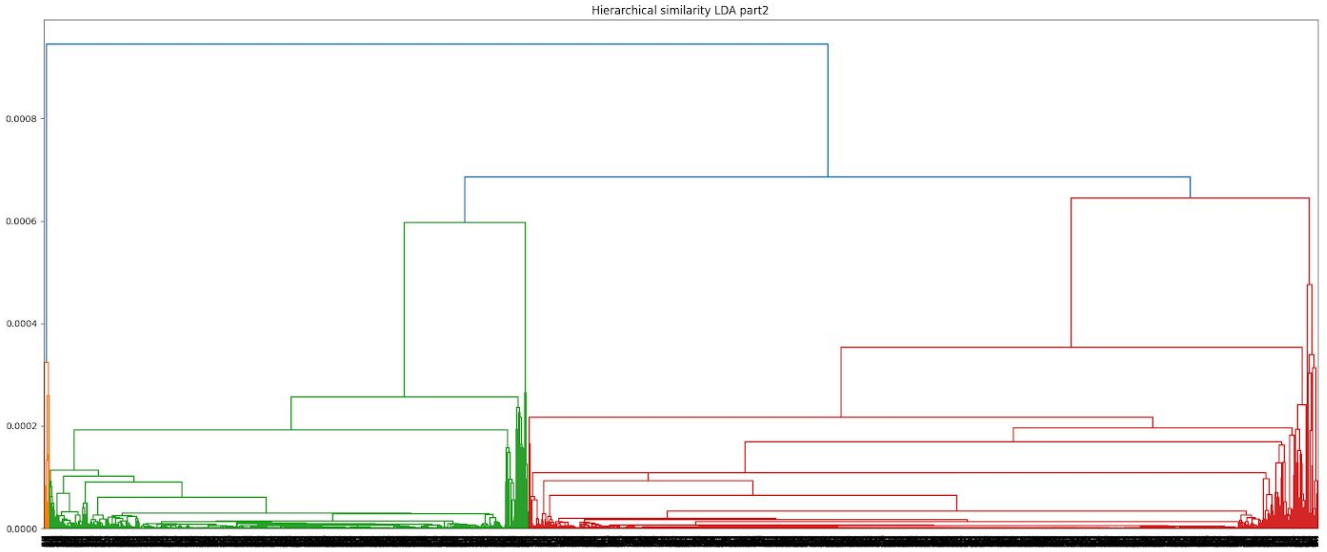

Hierarchical Dissimilarity PCA part4
200
150
100
50
0

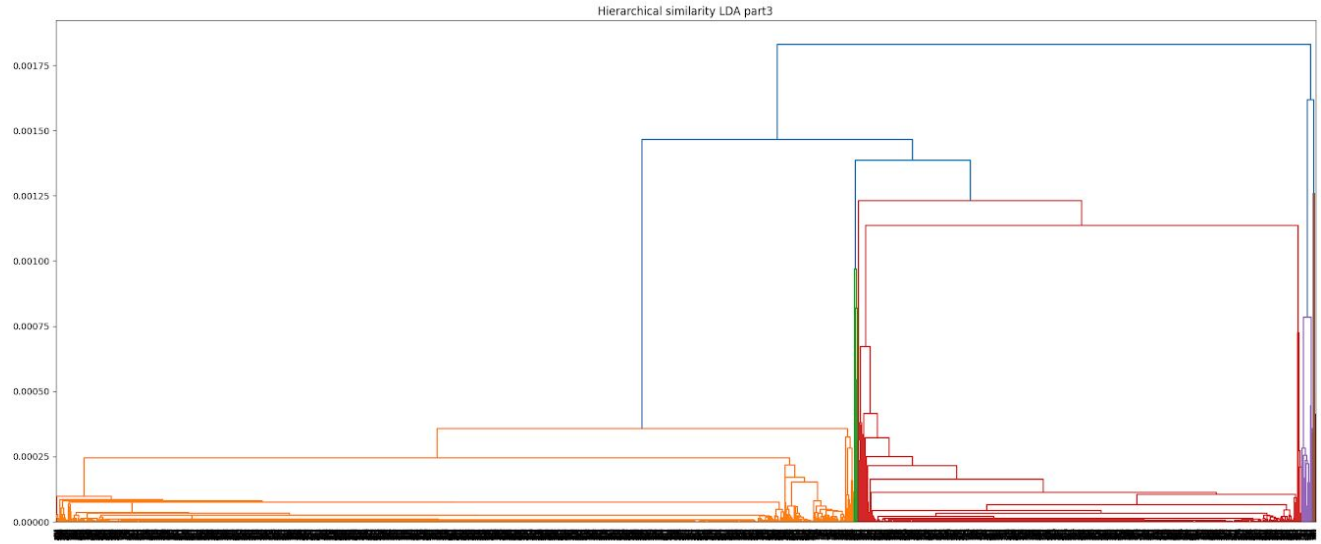
LDA K-Means Indian Pines

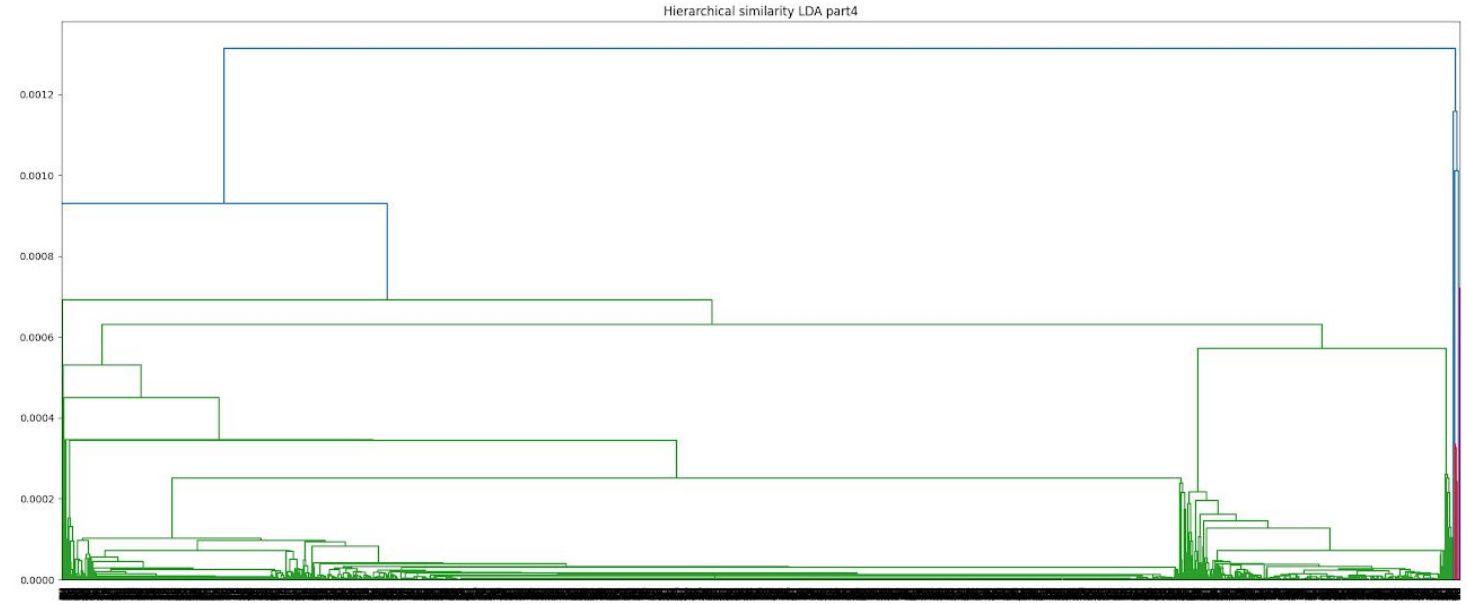


LDA Hierarchical similarity Indian Pines

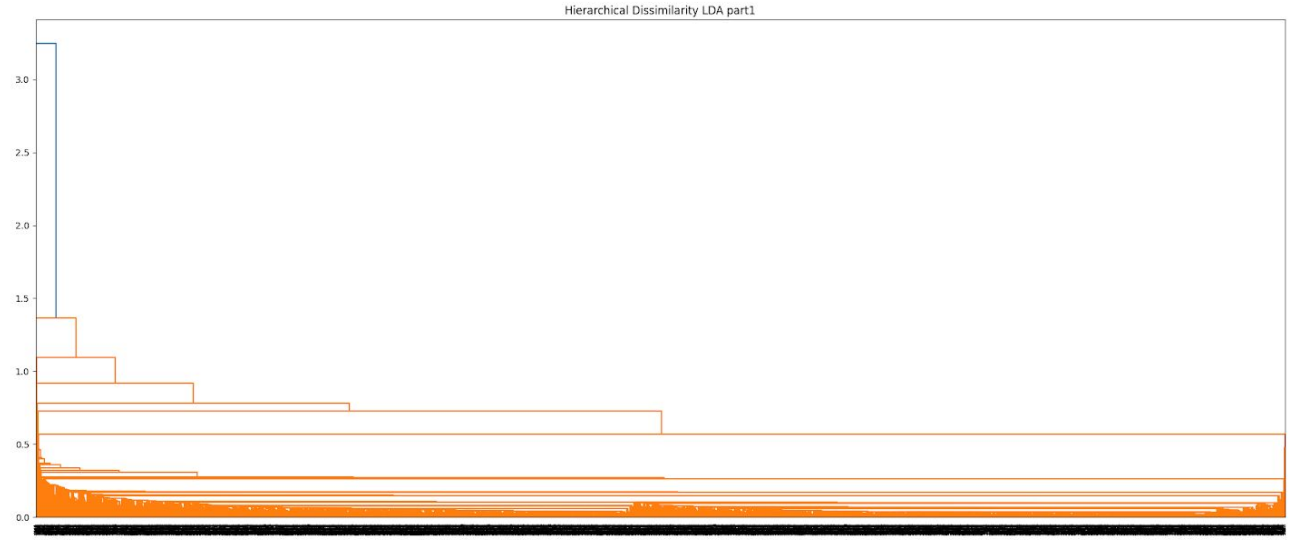


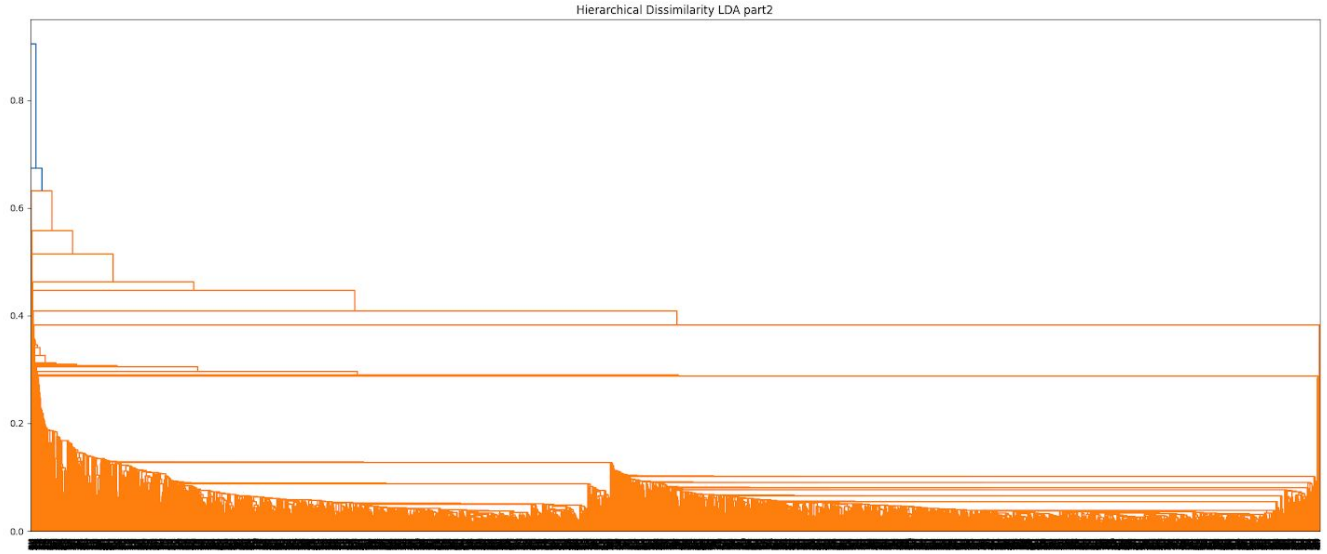


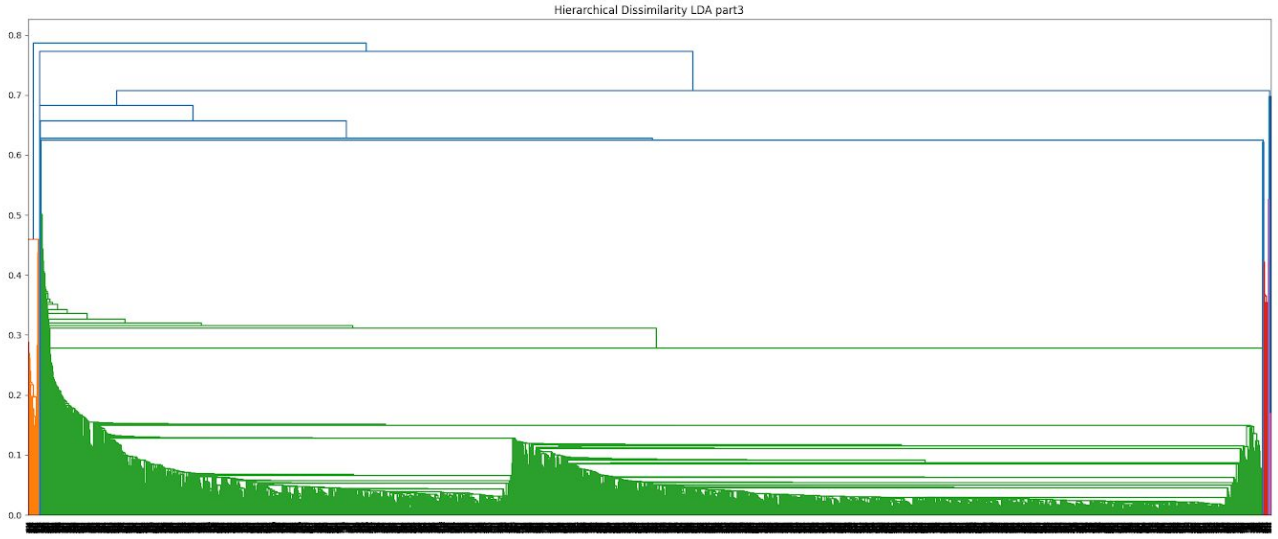

Hierarchical similarity LDA part3
0.00175
0.00150
0.00125
0.00100
0.00075
0.00050
0.00025
0.00000

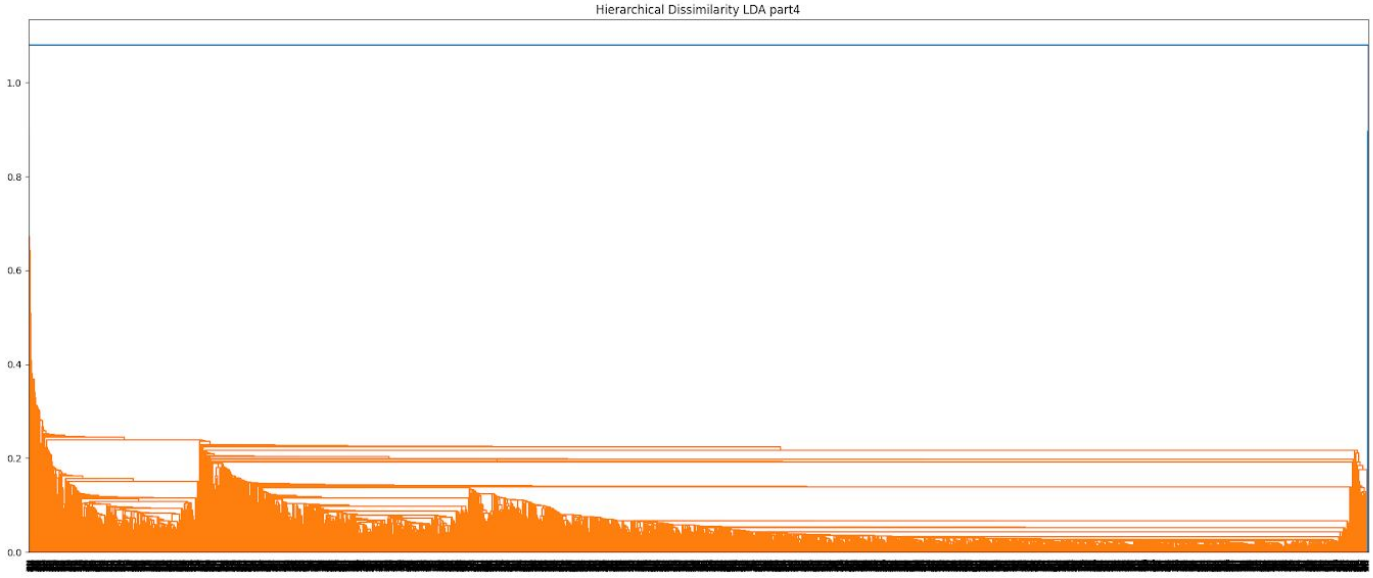

Hierarchical similarity LDA part4
0.0012
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000

LDA Hierarchical dissimilarity Indian Pines



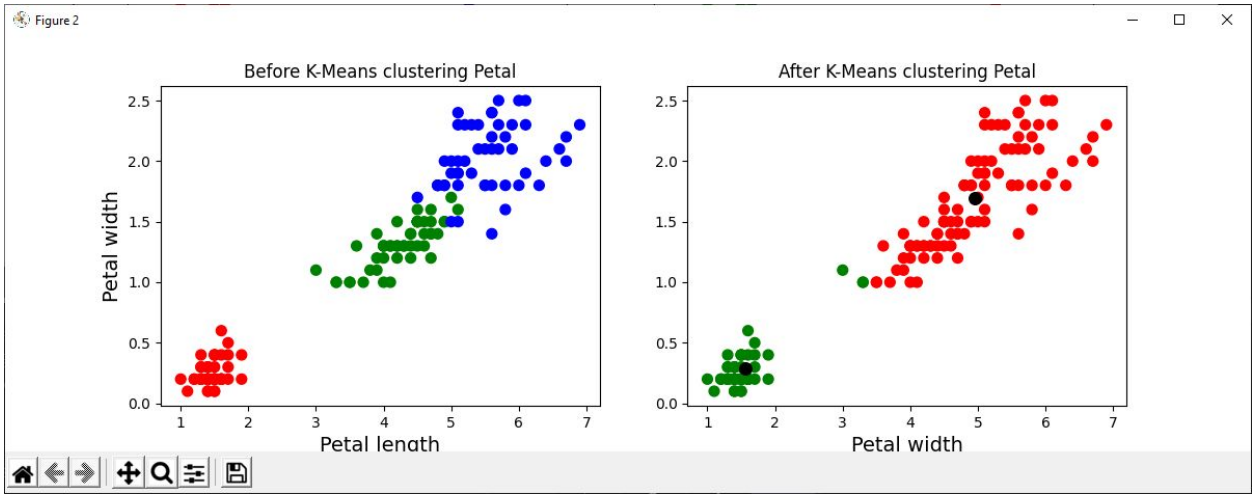


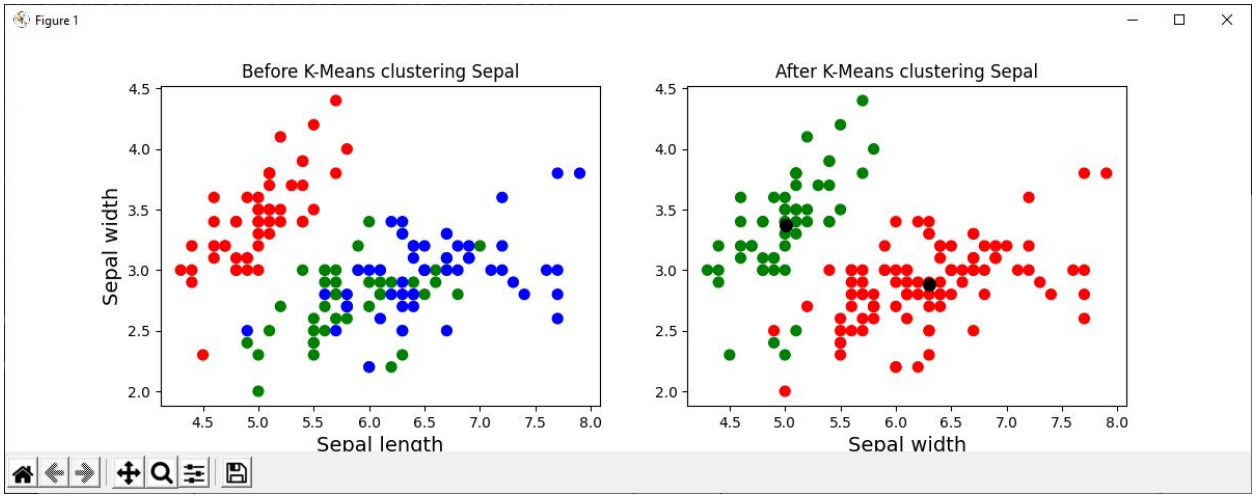

Hierarchical Dissimilarity LDA part3
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0


Hierarchical Dissimilarity LDA part4
1.0
0.8
0.6
0.4
0.2
0.0

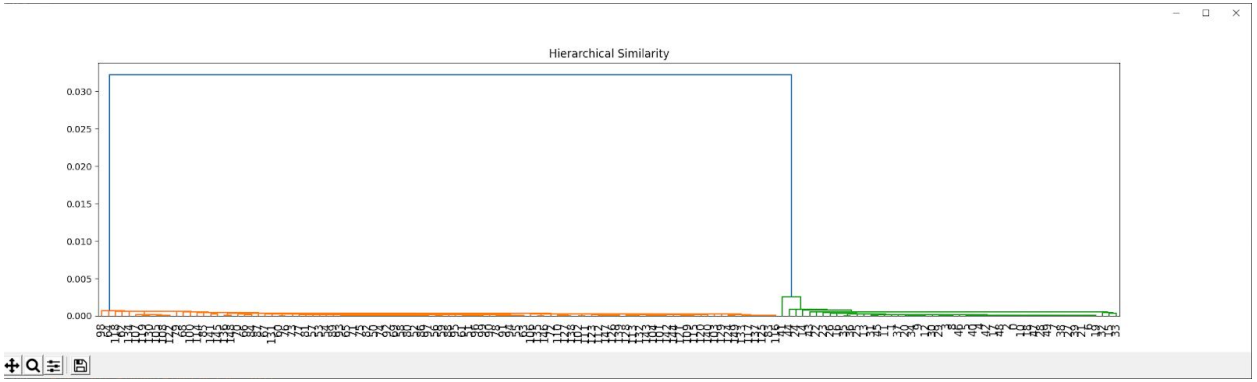
No reduction iris

K-Means no reduction Iris

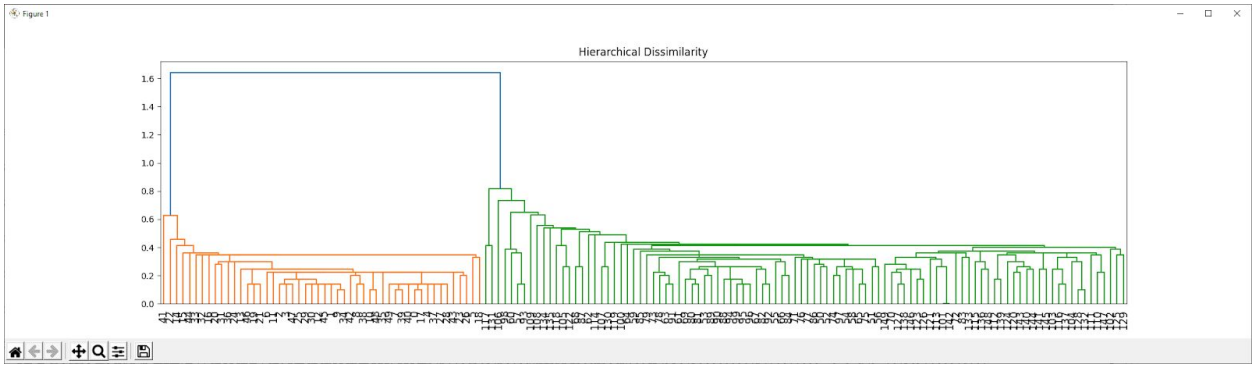




Hierarchical similarity no reduction iris

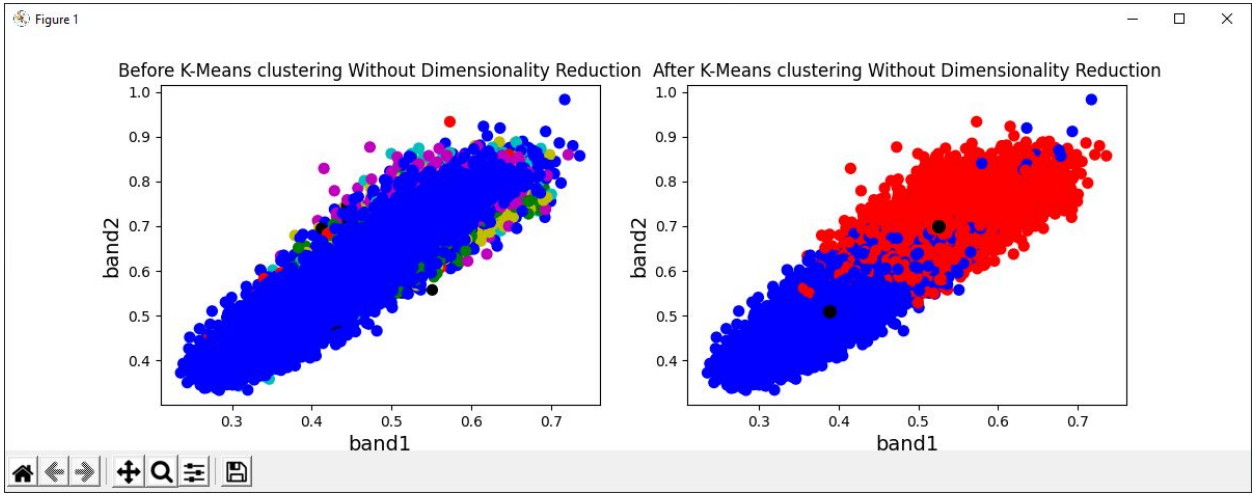


Hierarchical dissimilarity no reduction Iris

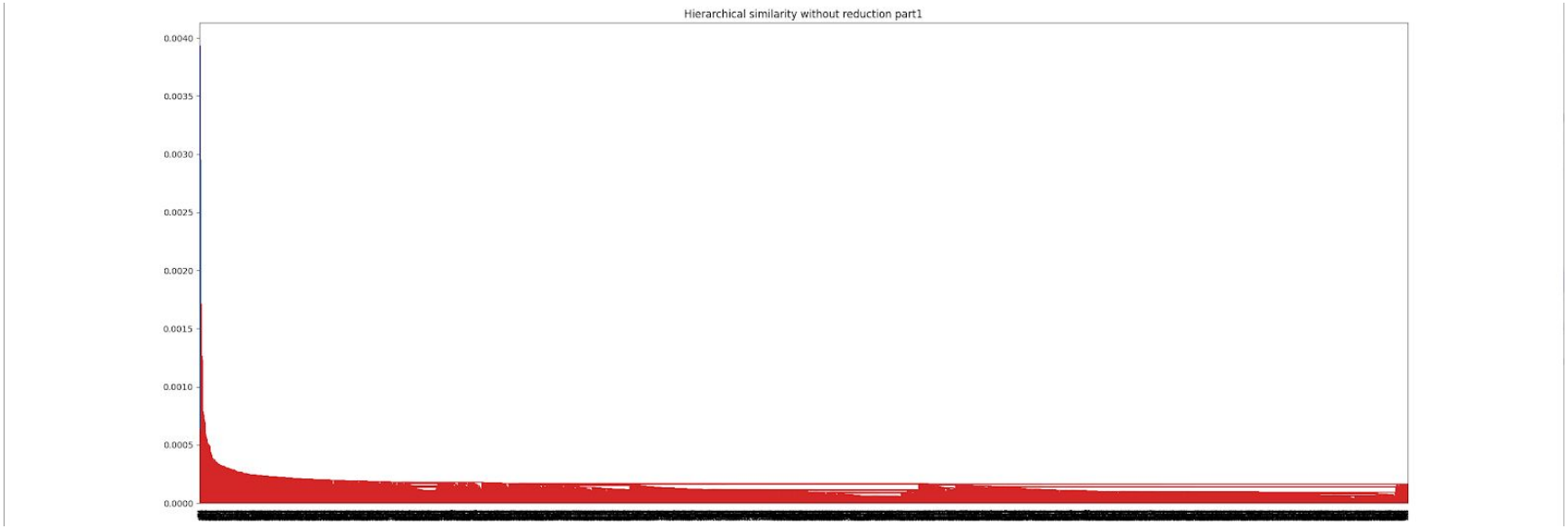


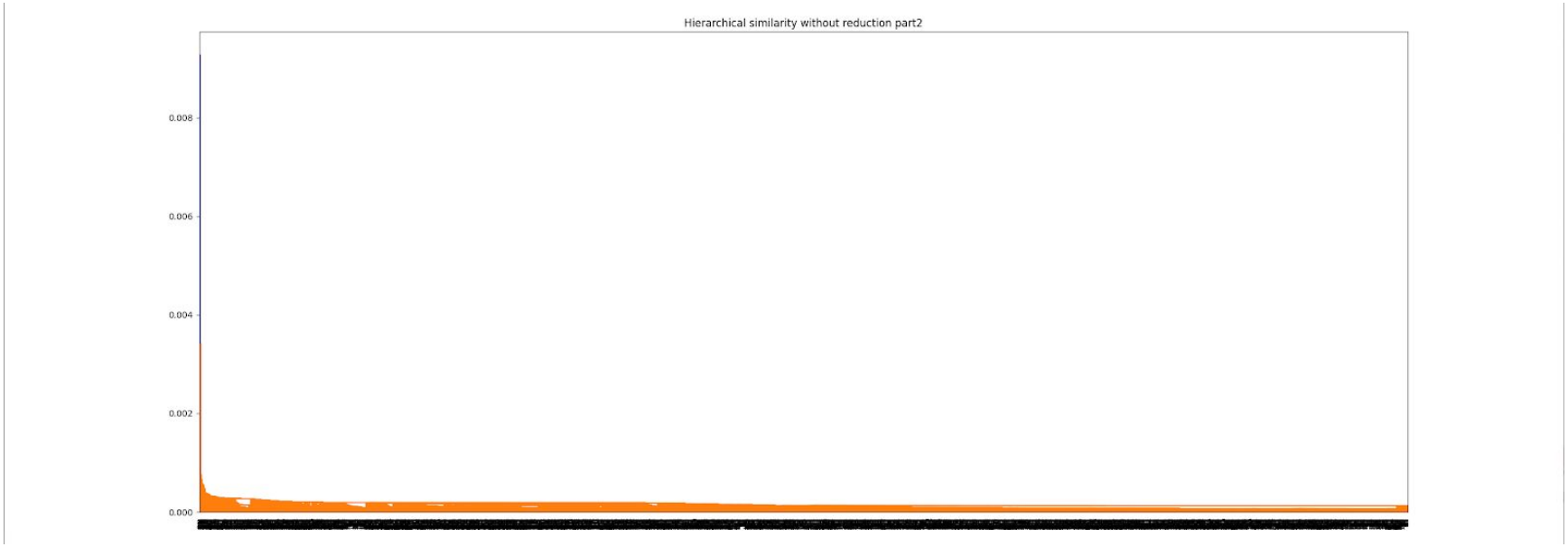
No reduction Indian Pines

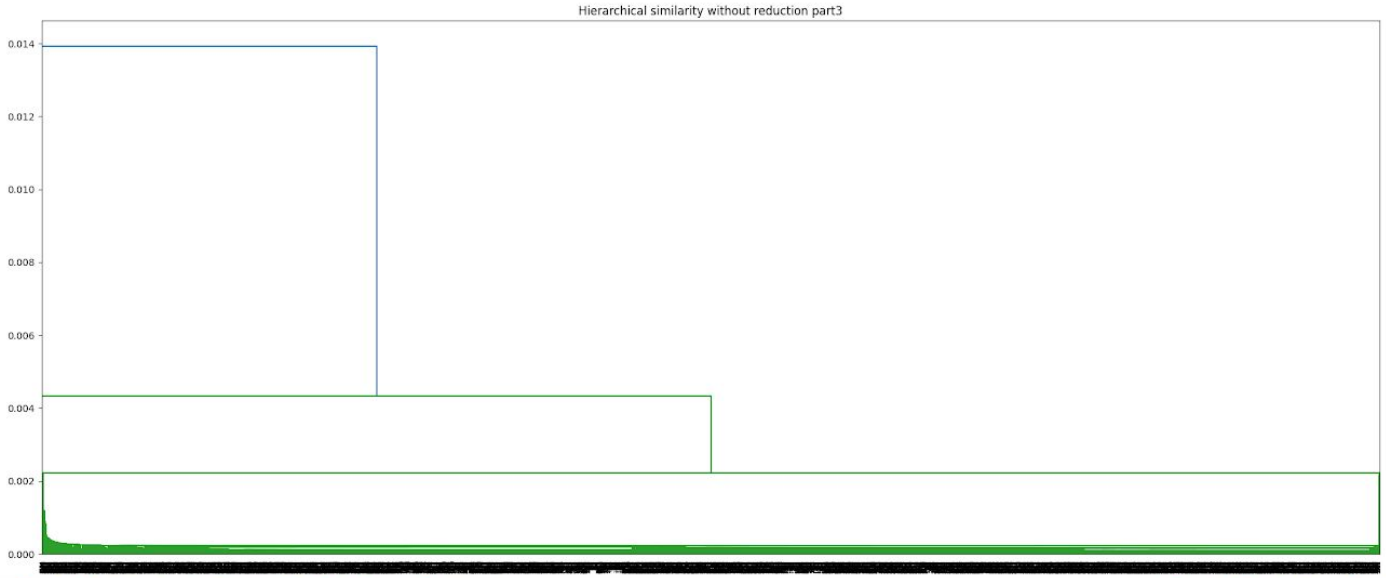
K-Means no reduction Indian Pines

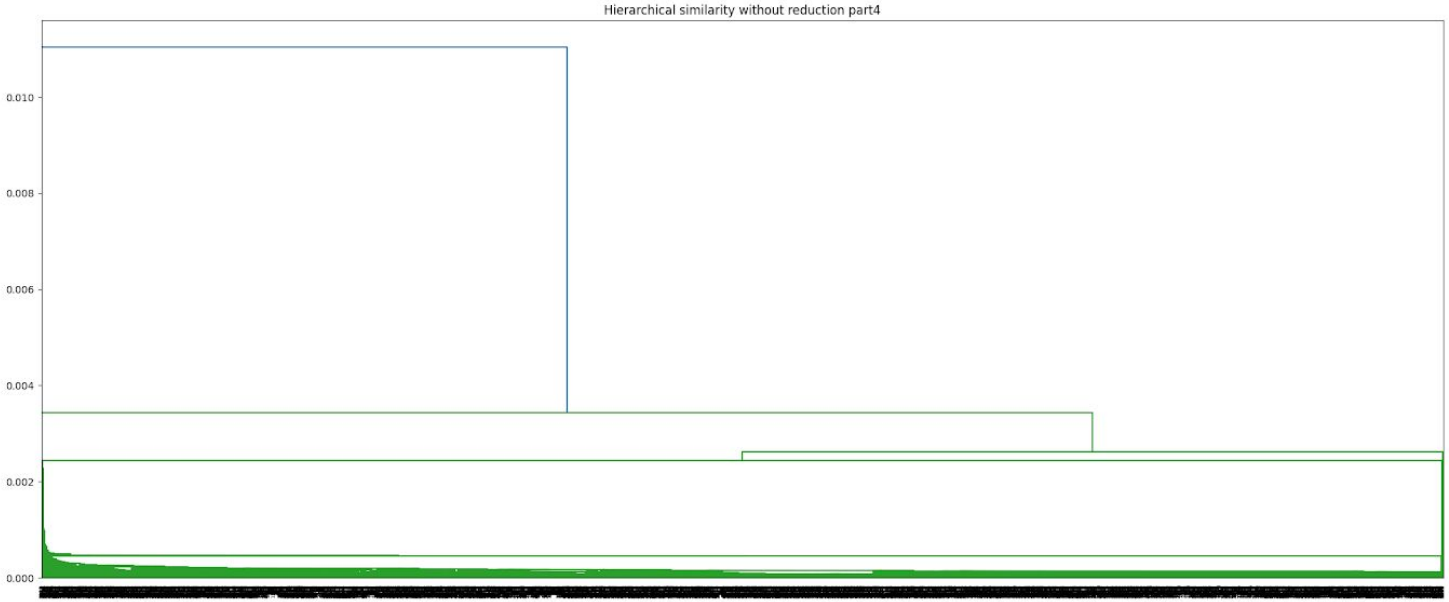


Hierarchical Similarity no reduction Indian Pines

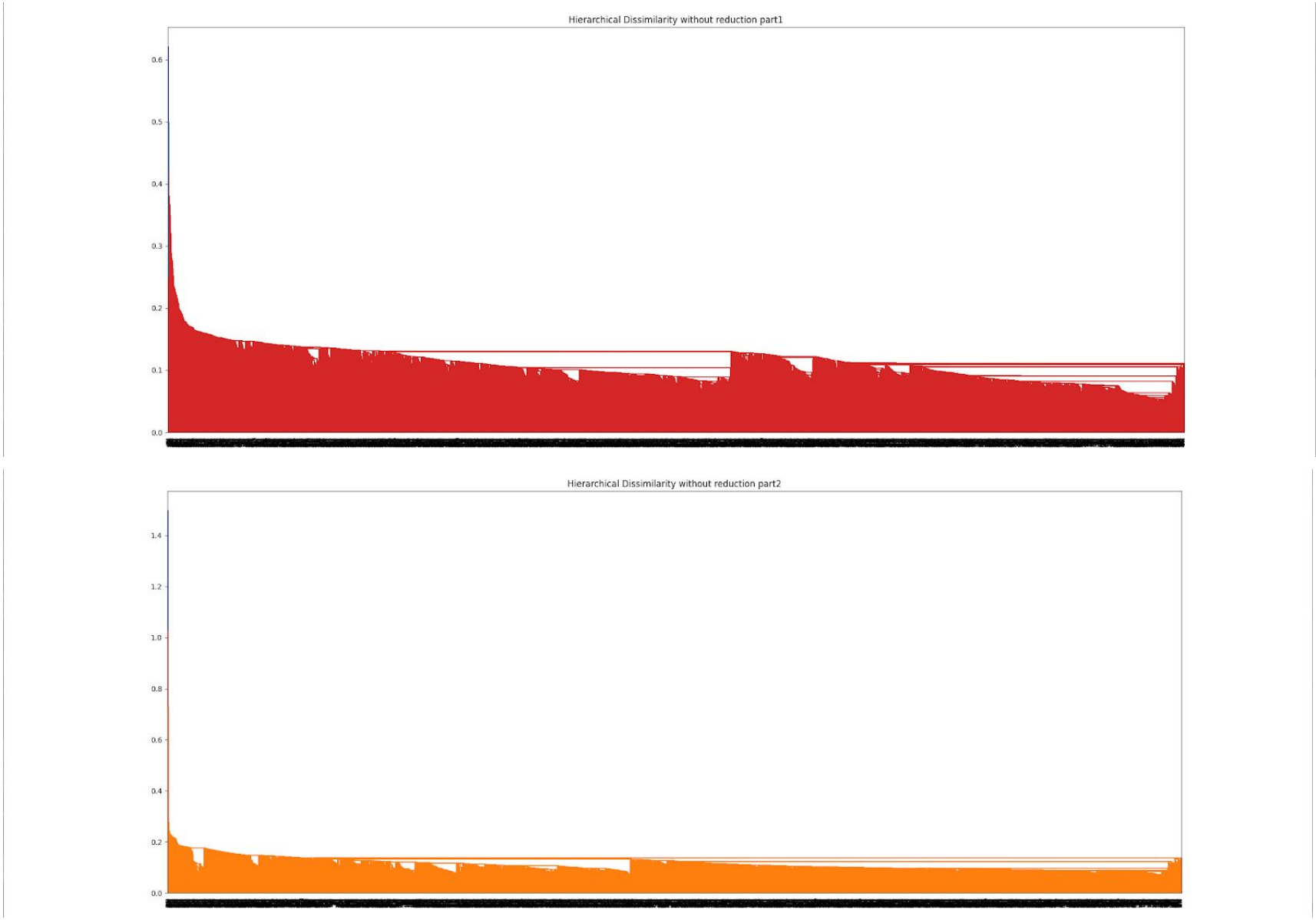


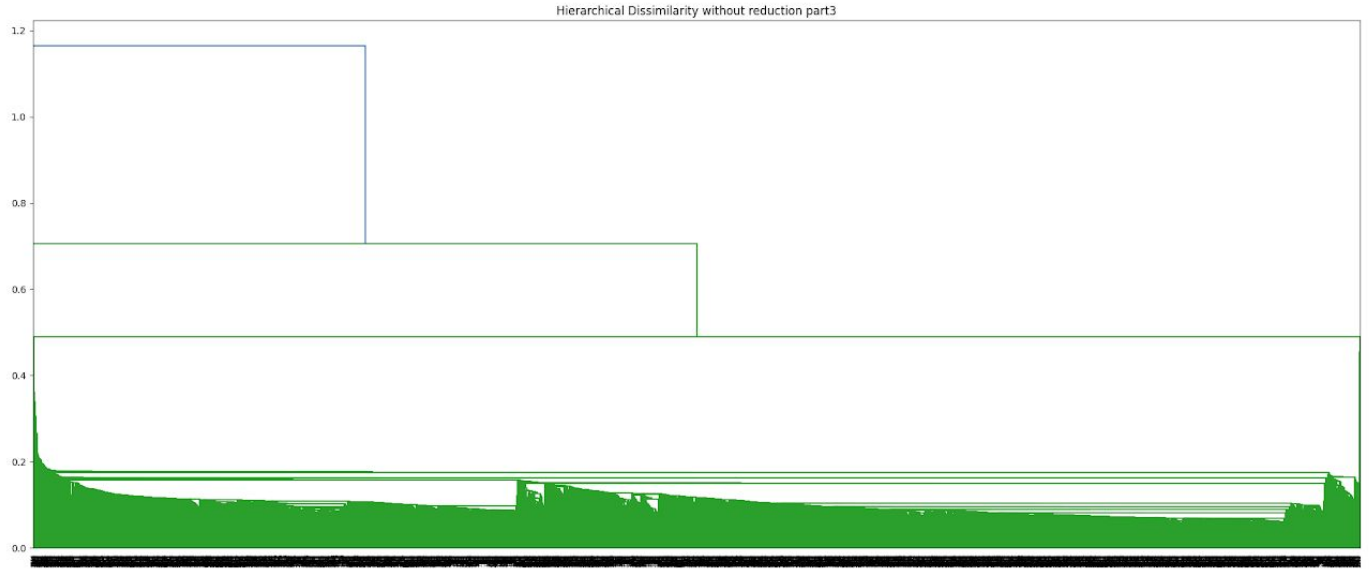


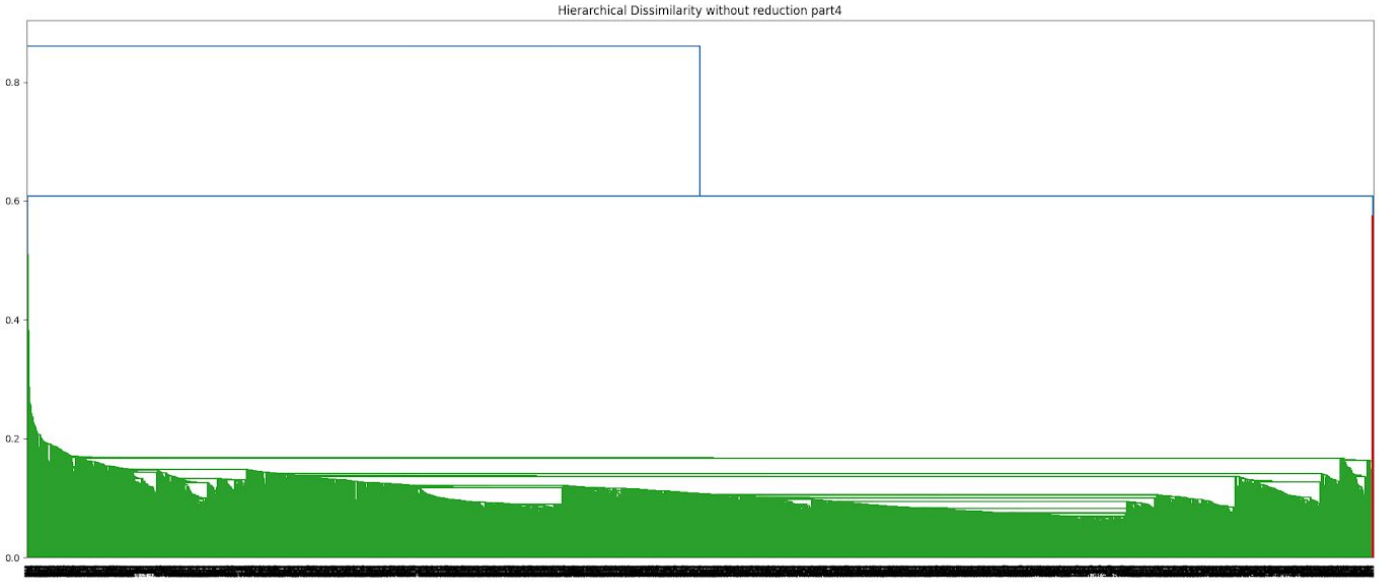

Hierarchical similarity without reduction part3
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000


Hierarchical similarity without reduction part4
0.010
0.008
0.006
0.004
0.002
0.000

Hierarchical Dissimilarity no reduction Indian Pines




Hierarchical Dissimilarity without reduction part3
1.2
1.0
0.8
0.6
0.4
0.2
0.0


Hierarchical Dissimilarity without reduction part4
0.8
0.6
0.4
0.2
0.0

3a.i) Cluster Validity Indices Iris

```
Davis Bouldin index for cluster = 2 is 0.4042928371730424
Silhouette index for clusters = 2 is 0.681046169211746
Davis Bouldin index for cluster = 3 is 0.6619715465007542
Silhouette index for clusters = 3 is 0.5528190123564091
Davis Bouldin index for cluster = 4 is 0.7803069838811072
Silhouette index for clusters = 4 is 0.4980505049972864
Davis Bouldin index for cluster = 5 is 0.8059652120182781
Silhouette index for clusters = 5 is 0.48874888709310454
Davis Bouldin index for cluster = 6 is 0.9141579726538072
Silhouette index for clusters = 6 is 0.36483400396700155
```

3a.ii) Cluster Validity Indices Indian Pines

```
Davis Bouldin index for cluster = 2 is 0.48721488000419705
Silhouette index for clusters = 2 is 0.6517245130695343
Davis Bouldin index for cluster = 3 is 0.6275585928875912
Silhouette index for clusters = 3 is 0.5703271112057529
Davis Bouldin index for cluster = 4 is 0.8060792032037776
Silhouette index for clusters = 4 is 0.4941480338116446
Davis Bouldin index for cluster = 5 is 0.8760410467193243
Silhouette index for clusters = 5 is 0.4579026802488645
Davis Bouldin index for cluster = 6 is 0.9132468400377141
Silhouette index for clusters = 6 is 0.40789548411053117
Davis Bouldin index for cluster = 7 is 0.9145617482009055
Silhouette index for clusters = 7 is 0.4120445033999344
Davis Bouldin index for cluster = 8 is 0.9616778040341105
Silhouette index for clusters = 8 is 0.3645883433380735
Davis Bouldin index for cluster = 9 is 0.9913064280520457
Silhouette index for clusters = 9 is 0.343071242952823
Davis Bouldin index for cluster = 10 is 1.028864996071575
Silhouette index for clusters = 10 is 0.3233867209060576
Davis Bouldin index for cluster = 11 is 1.0367750913259408
Silhouette index for clusters = 11 is 0.3169840985341388
Davis Bouldin index for cluster = 12 is 1.068410566673859
Silhouette index for clusters = 12 is 0.30407091232321415
Davis Bouldin index for cluster = 13 is 1.0667466248981639
Silhouette index for clusters = 13 is 0.3051155318062425
Davis Bouldin index for cluster = 14 is 1.065206883932703
Silhouette index for clusters = 14 is 0.30230418780677076
Davis Bouldin index for cluster = 15 is 1.0679585471701551
Silhouette index for clusters = 15 is 0.29263407744670267
Davis Bouldin index for cluster = 16 is 1.084103438159274
Silhouette index for clusters = 16 is 0.28492566303948746
```

3a.iii)

For the iris dataset, since the Davies Bouldin for 2 clusters is the smallest and Silhouette for 2 clusters is the biggest, 2 should be the k for the cluster value.

For the Indian Pines dataset, since the Davies Bouldin for 2 clusters is the smallest and Silhouette for 2 clusters is the biggest, 2 should be the k for the cluster value.

3b.i) Dimensionality reduction provides generally provides better separability, faster calculation, and more accurate clustering.

3b.ii)

For the iris dataset, LDA reduction worked better than no reduction at all because if you take a look at the after K-mean visualization graphs, LDA has more data separation than PCA and no reduction. There was no significant difference between the K-means and Hierarchical methods.

For the Indian pines dataset, K-means with LDA reduction worked better regarding the calculation speed and the data desperation.