Clustering of the dimensionality reduction results

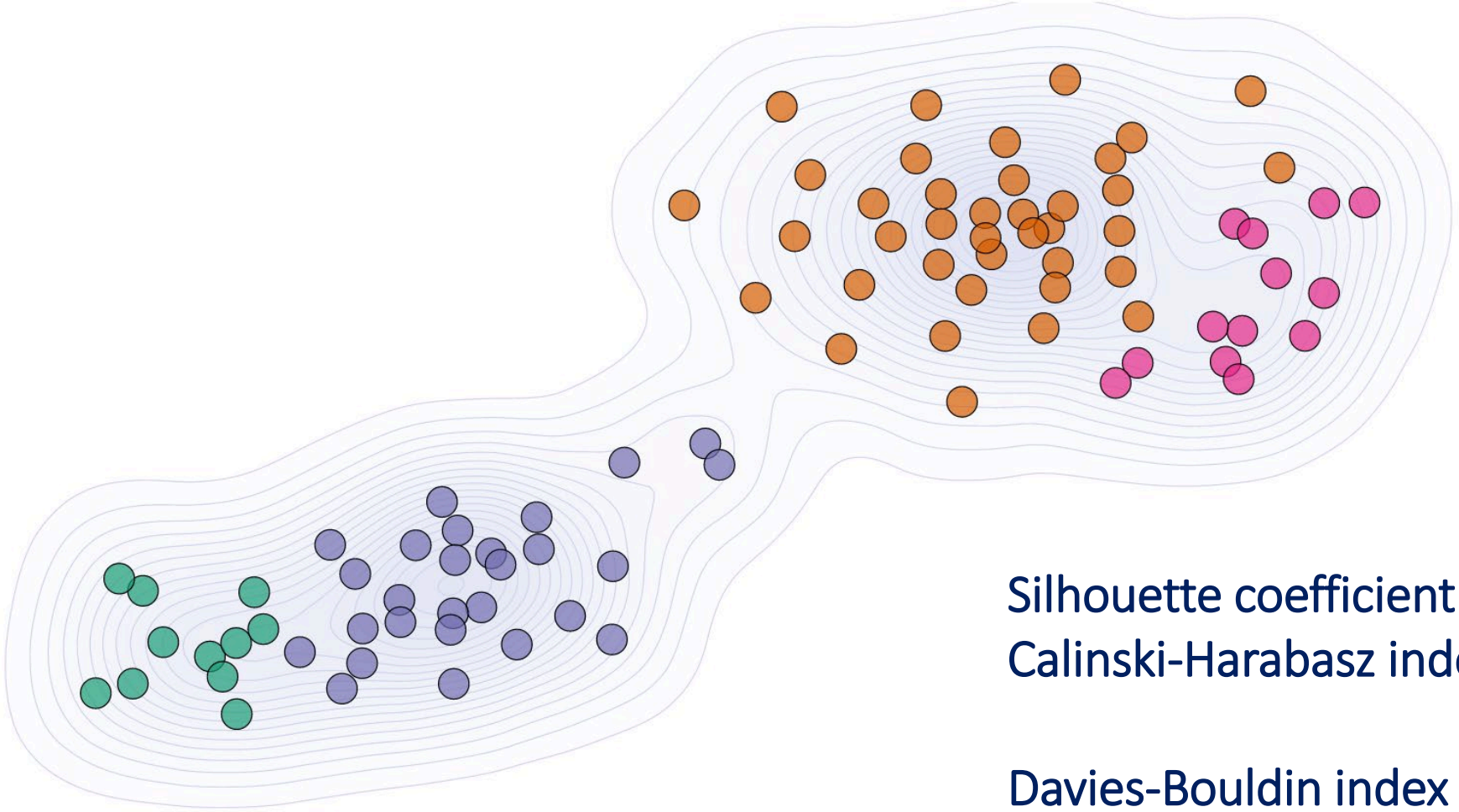
Summary

Findings on our assessed prostate cancer data

- t-SNE with k-means/mean-shift leads to two **visual** clusters
- PCA/FAMD lead to a **dense** region and two outliers
- UMAP/MDS have the poorest cluster separation scores

t-distributed Stochastic Neighbor Embedding (t-SNE)

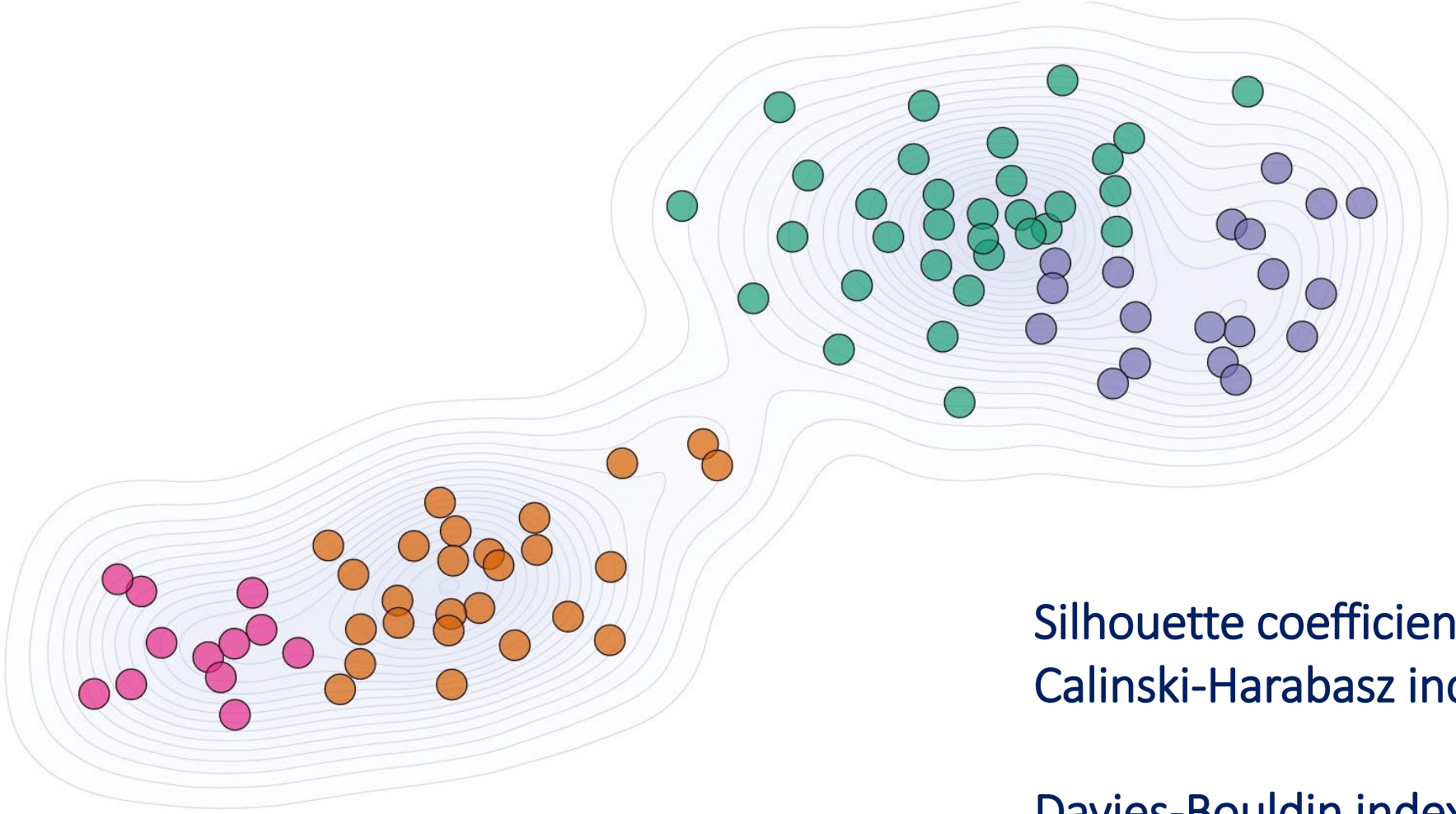
Gaussian mixture models (distribution-based)



Silhouette coefficient : 0.50
Calinski-Harabasz index : 239.68

Davies-Bouldin index : 0.88

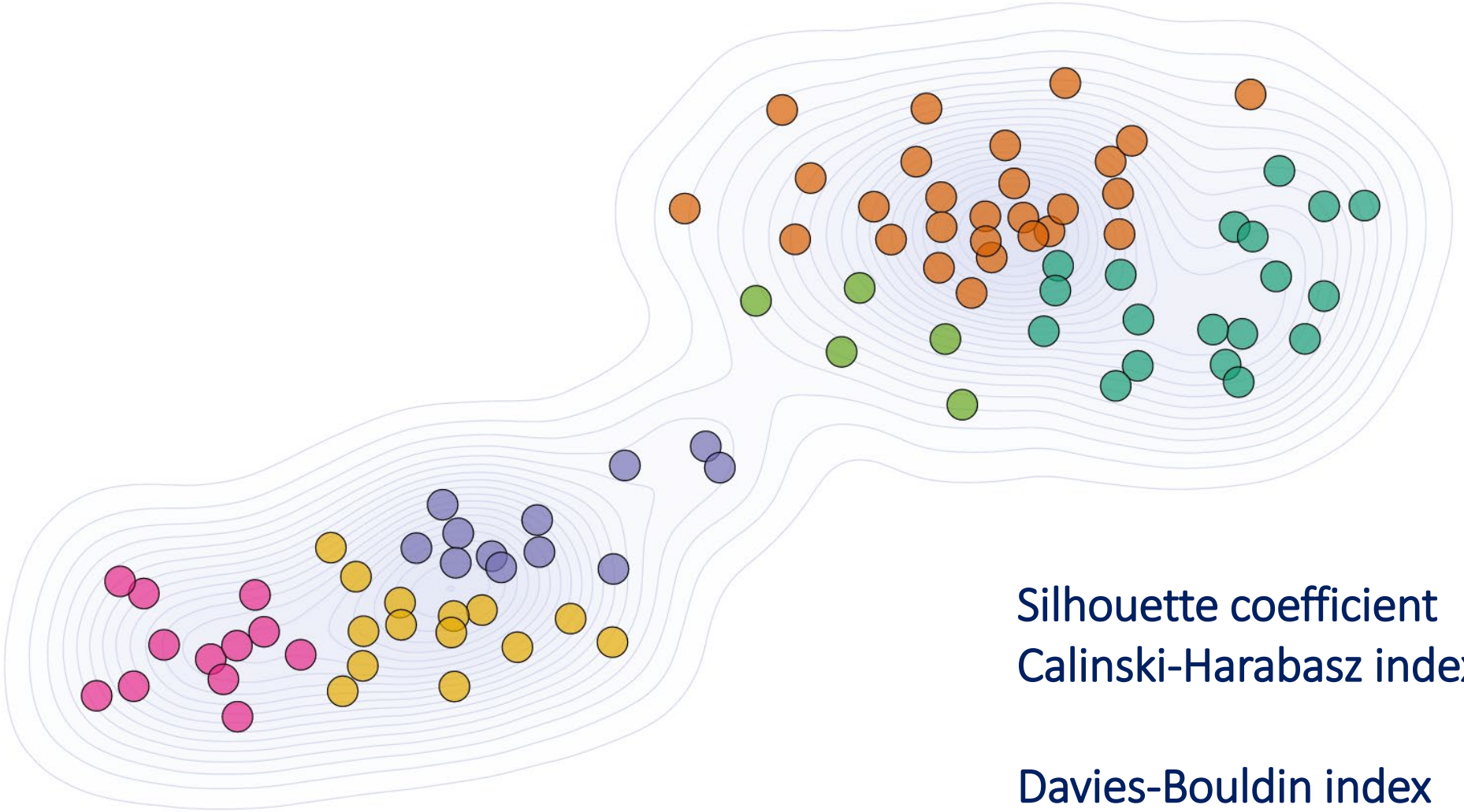
Hierarchical Clustering (4 Clusters)



Silhouette coefficient : 0.52
Calinski-Harabasz index : 255.10

Davies-Bouldin index : 0.92

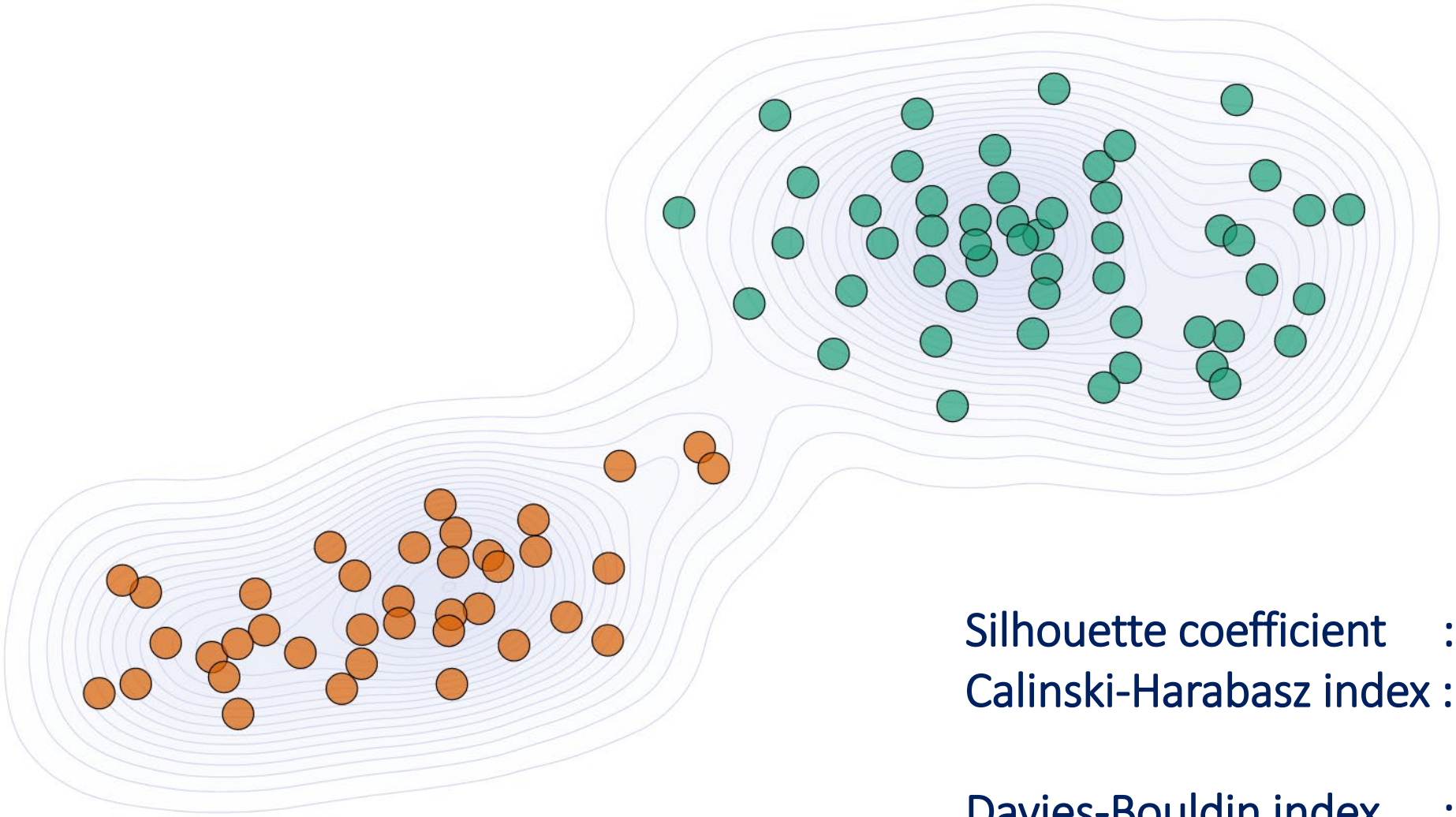
Hierarchical Clustering (6 Clusters)



Silhouette coefficient : 0.47
Calinski-Harabasz index : 219.86

Davies-Bouldin index : 0.89

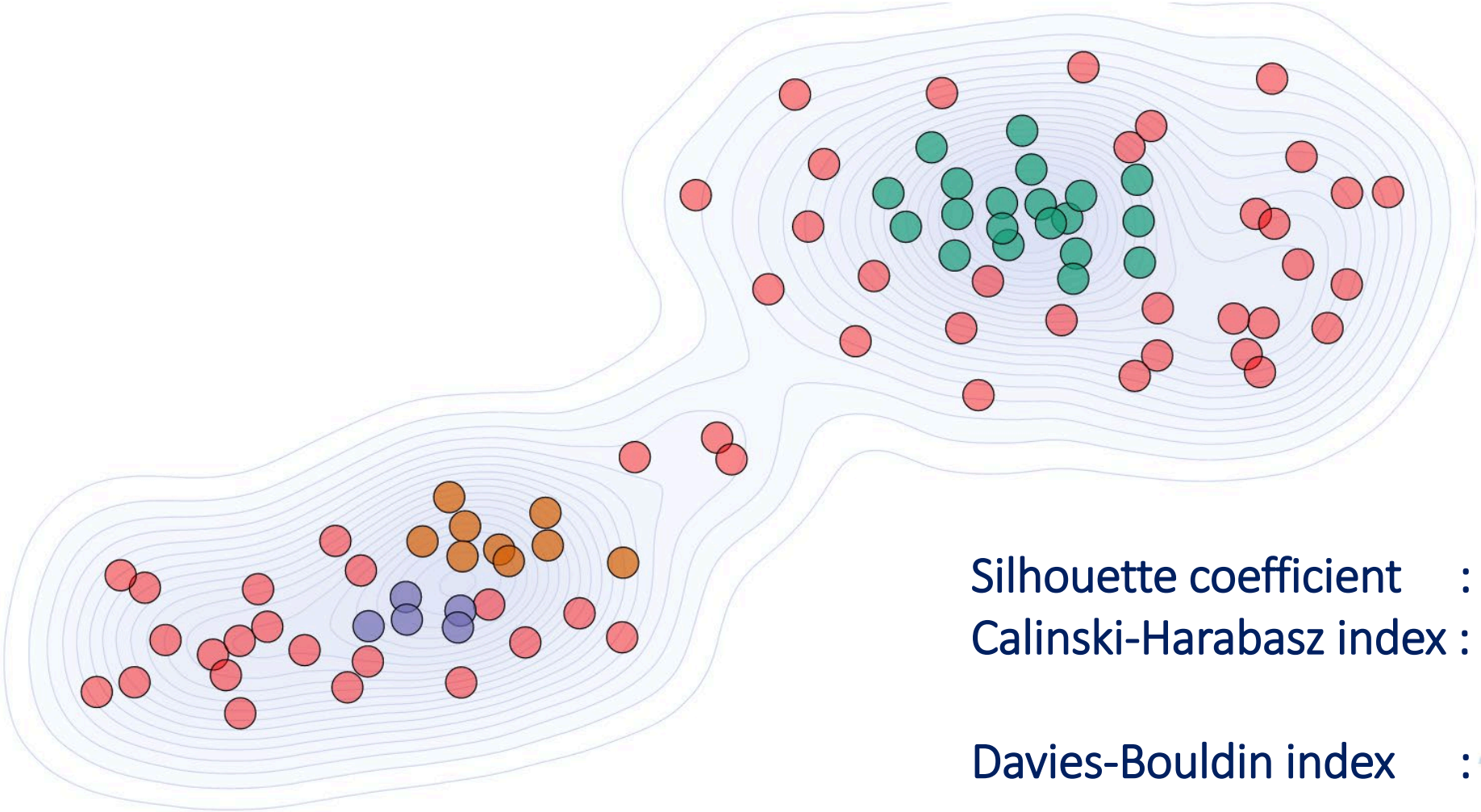
K-means/mean-shift (centroid-based)



Silhouette coefficient : 0.89
Calinski-Harabasz index : 414.15

Davies-Bouldin index : 0.40

Ordering points to identify the clustering structure (density-based)

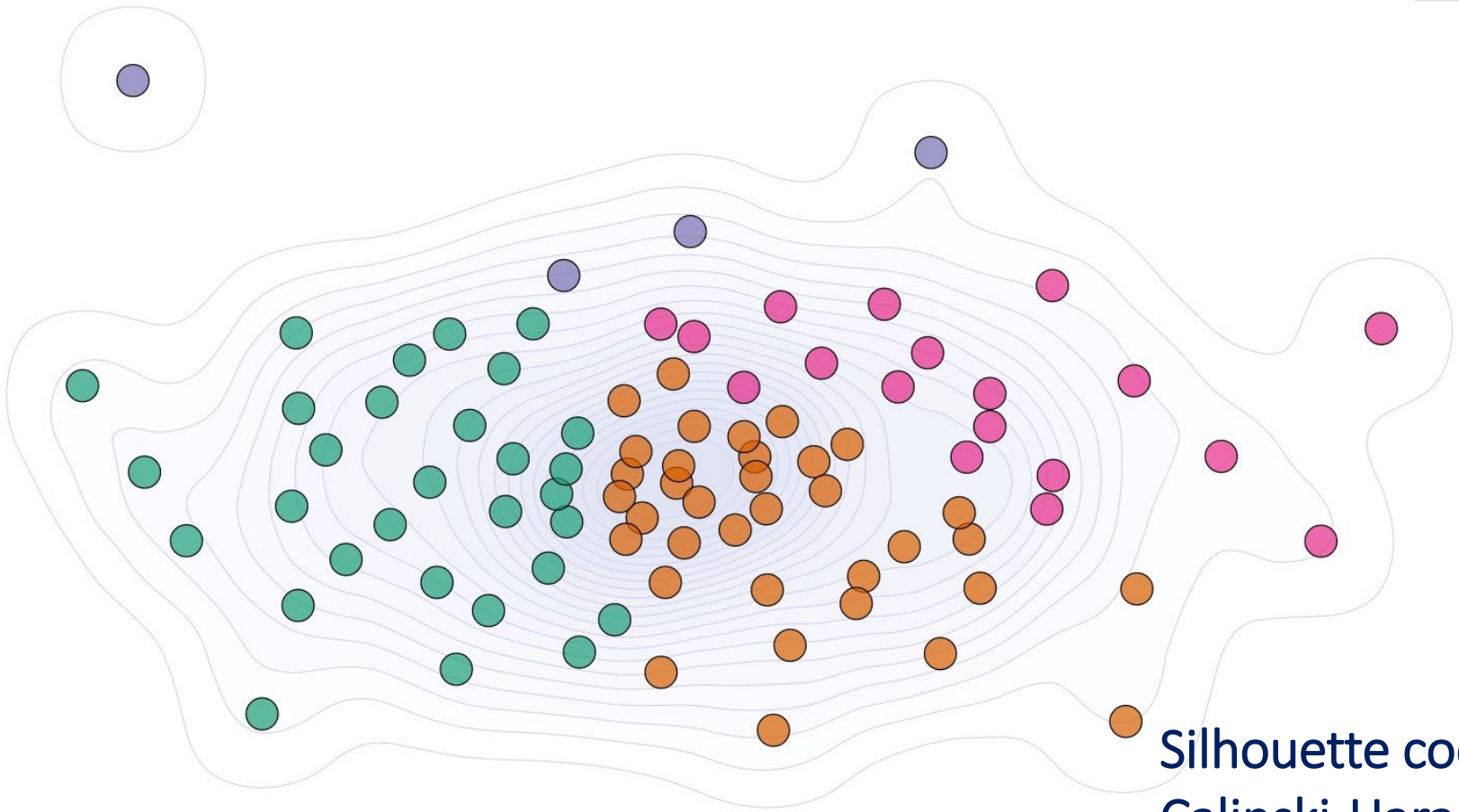


Silhouette coefficient : 0.21
Calinski-Harabasz index : 10.51

Davies-Bouldin index : 1.41

Multidimensional Scaling (MDS)

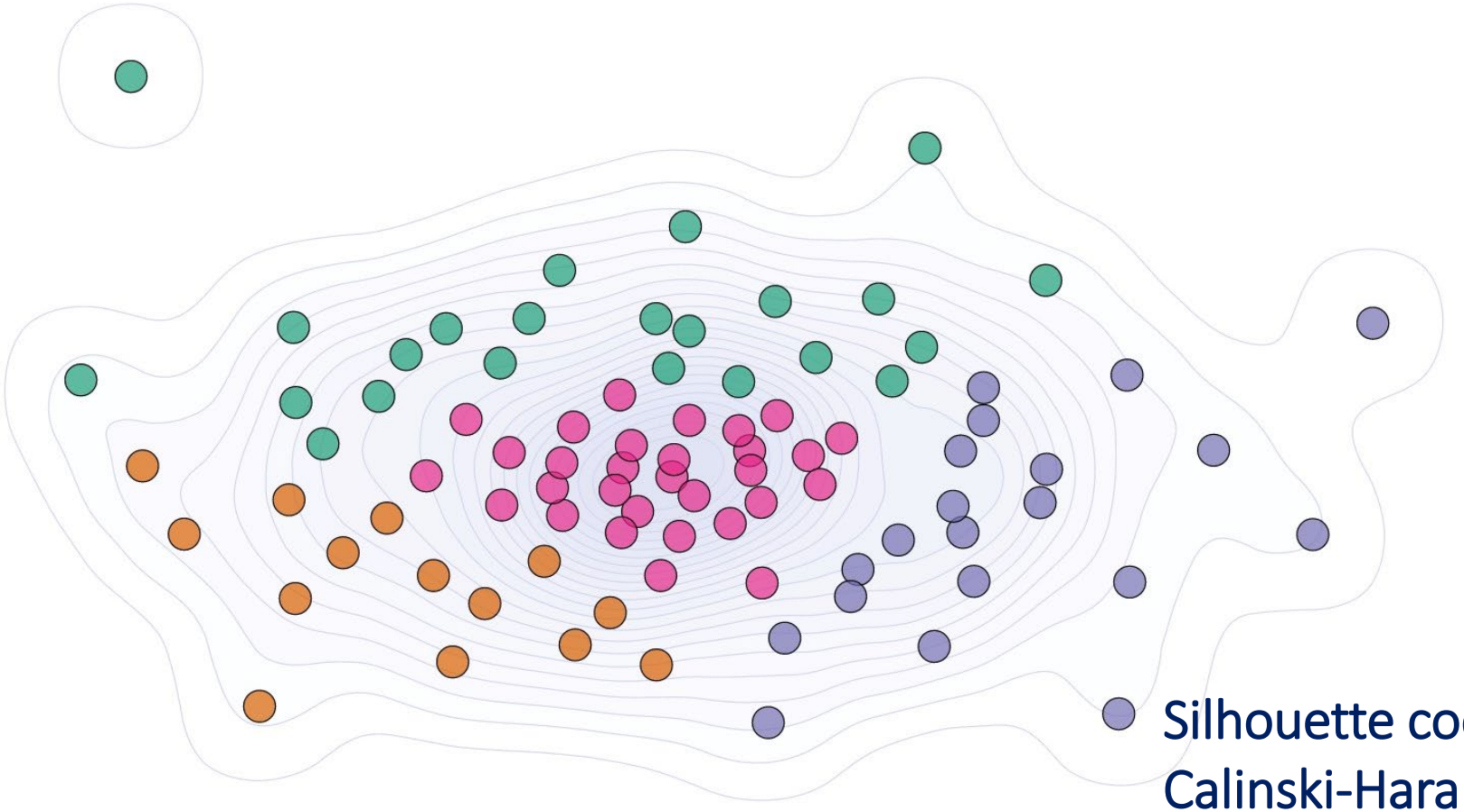
Gaussian mixture models (distribution-based)



Silhouette coefficient : 0.36
Calinski-Harabasz index : 33.19

Davies-Bouldin index : 1.13

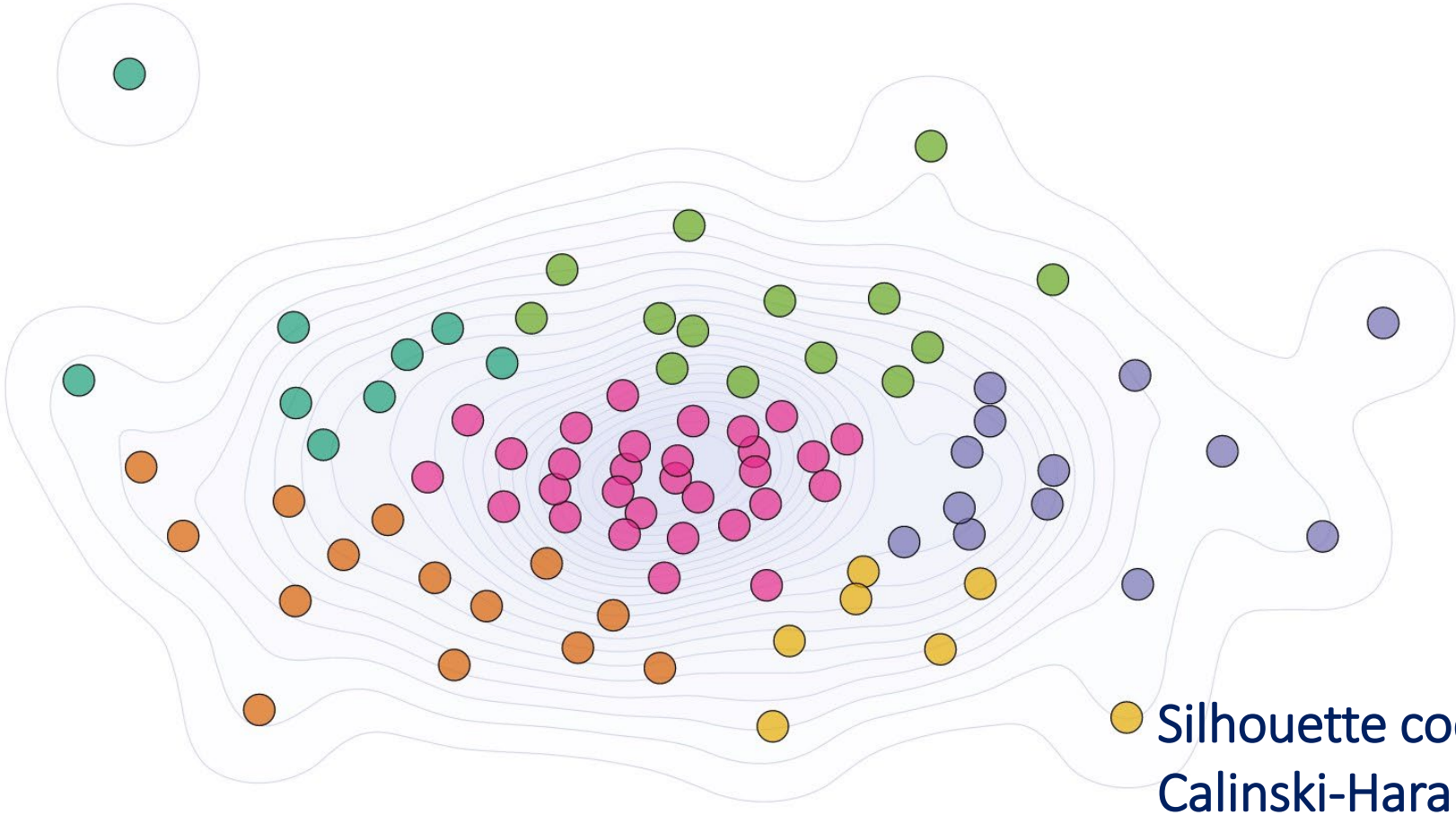
Hierarchical Clustering (4 Clusters)



Silhouette coefficient : 0.36
Calinski-Harabasz index : 38.67

Davies-Bouldin index : 1.06

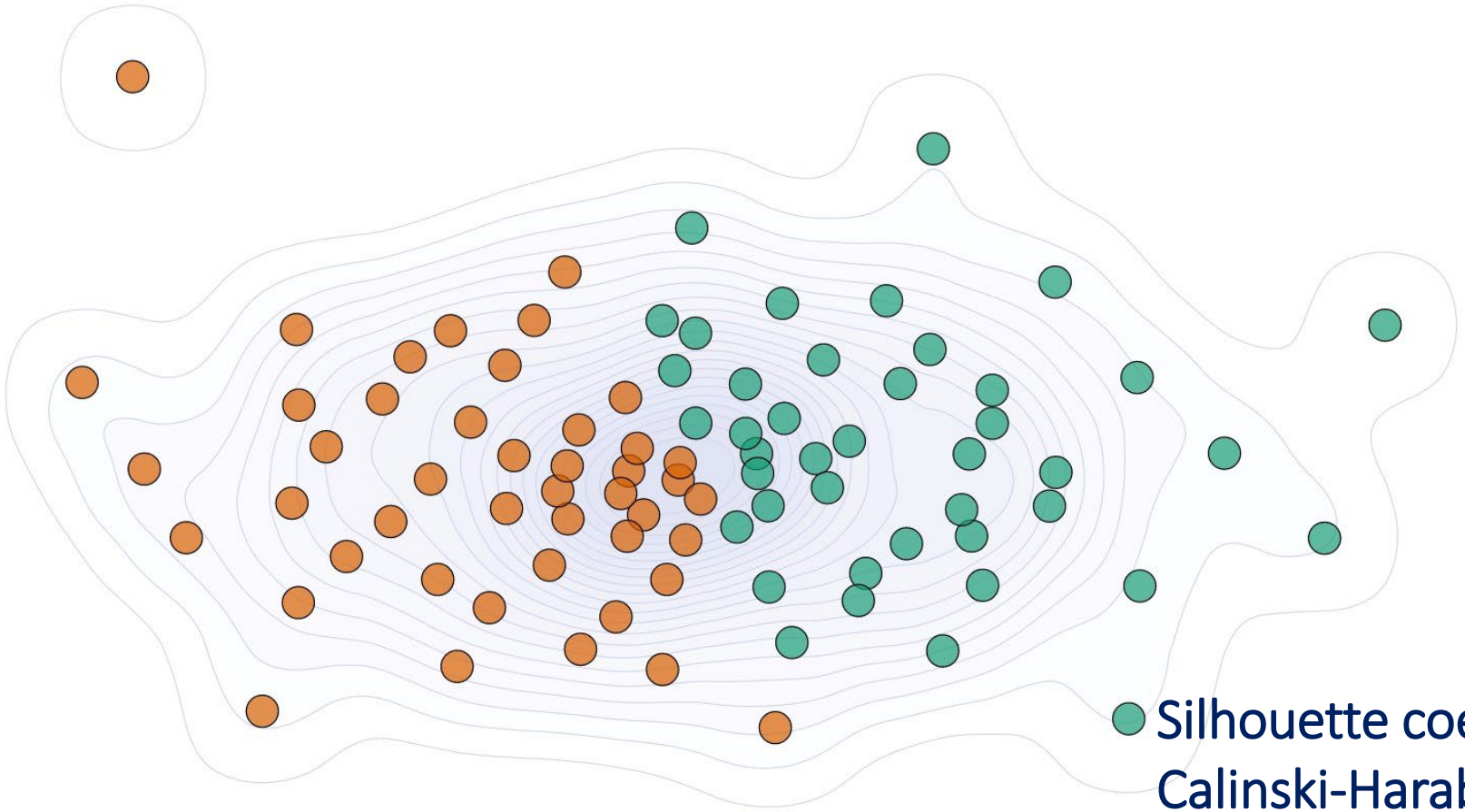
Hierarchical Clustering (6 Clusters)



Silhouette coefficient : 0.47
Calinski-Harabasz index : 45.97

Davies-Bouldin index : 0.86

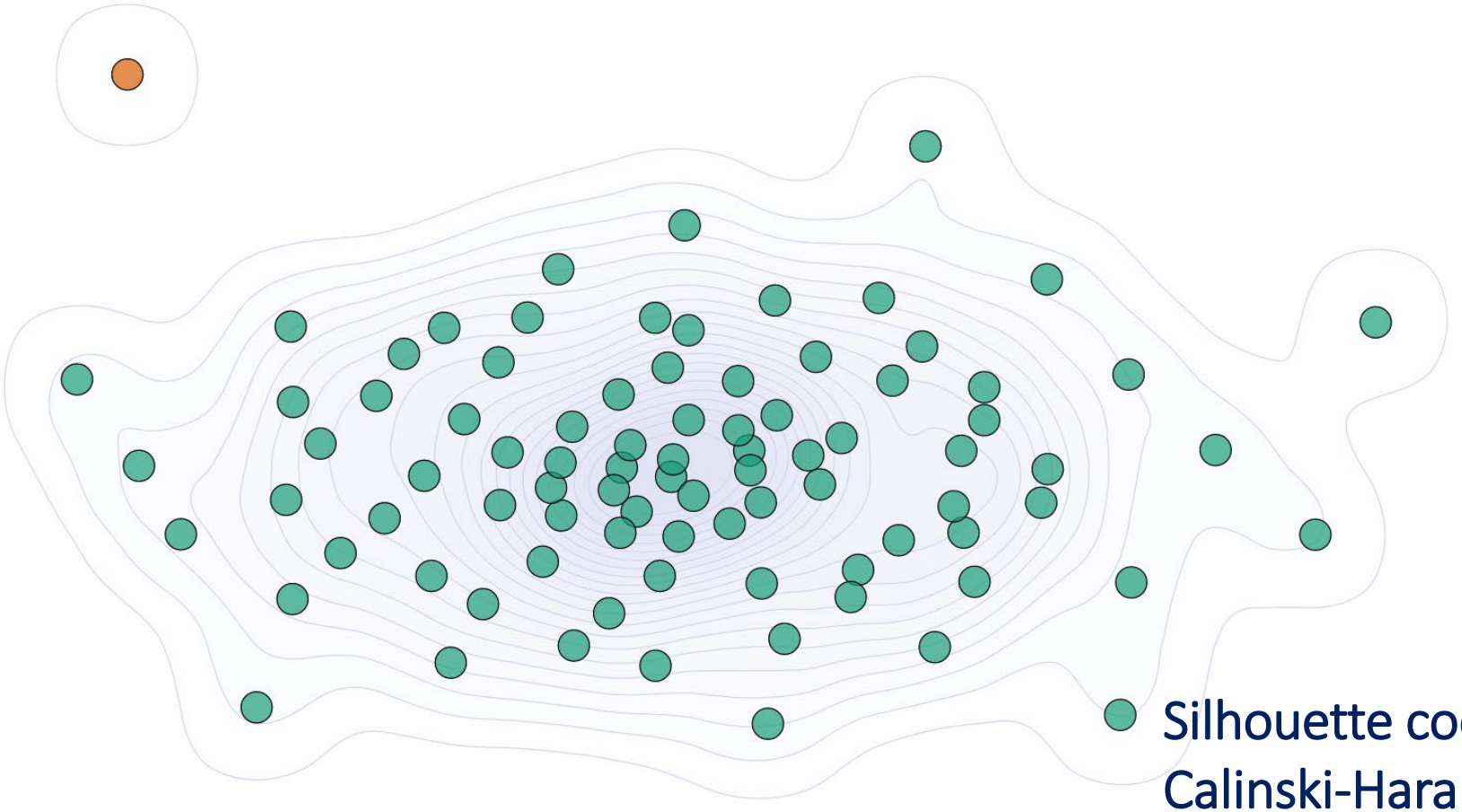
K-means (centroid-based)



Silhouette coefficient : 0.42
Calinski-Harabasz index : 37.70

Davies-Bouldin index : 1.30

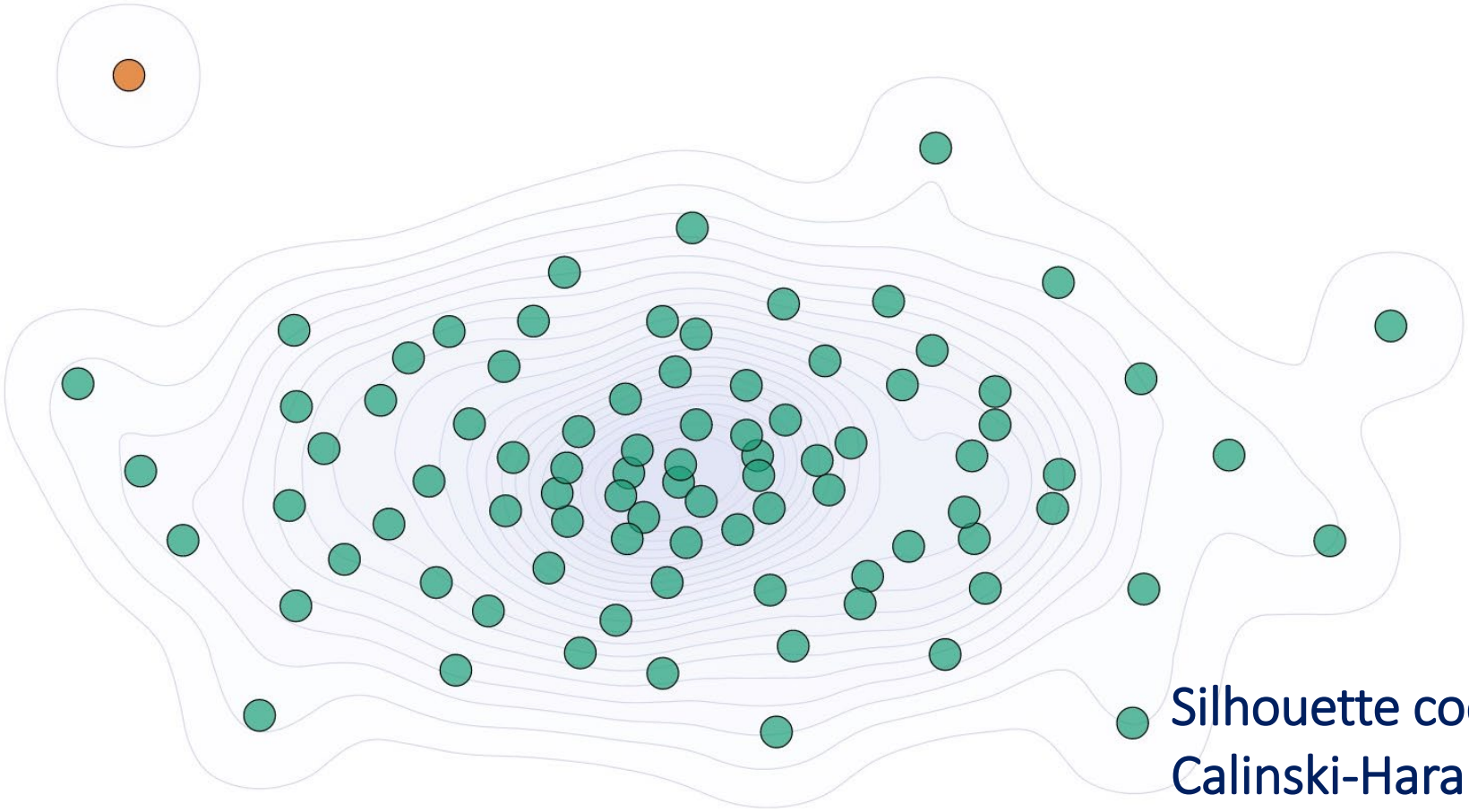
Mean-shift (centroid-based)



Silhouette coefficient : 0.89
Calinski-Harabasz index : 414.15

Davies-Bouldin index : 0.40

Ordering points to identify the clustering structure (density-based)

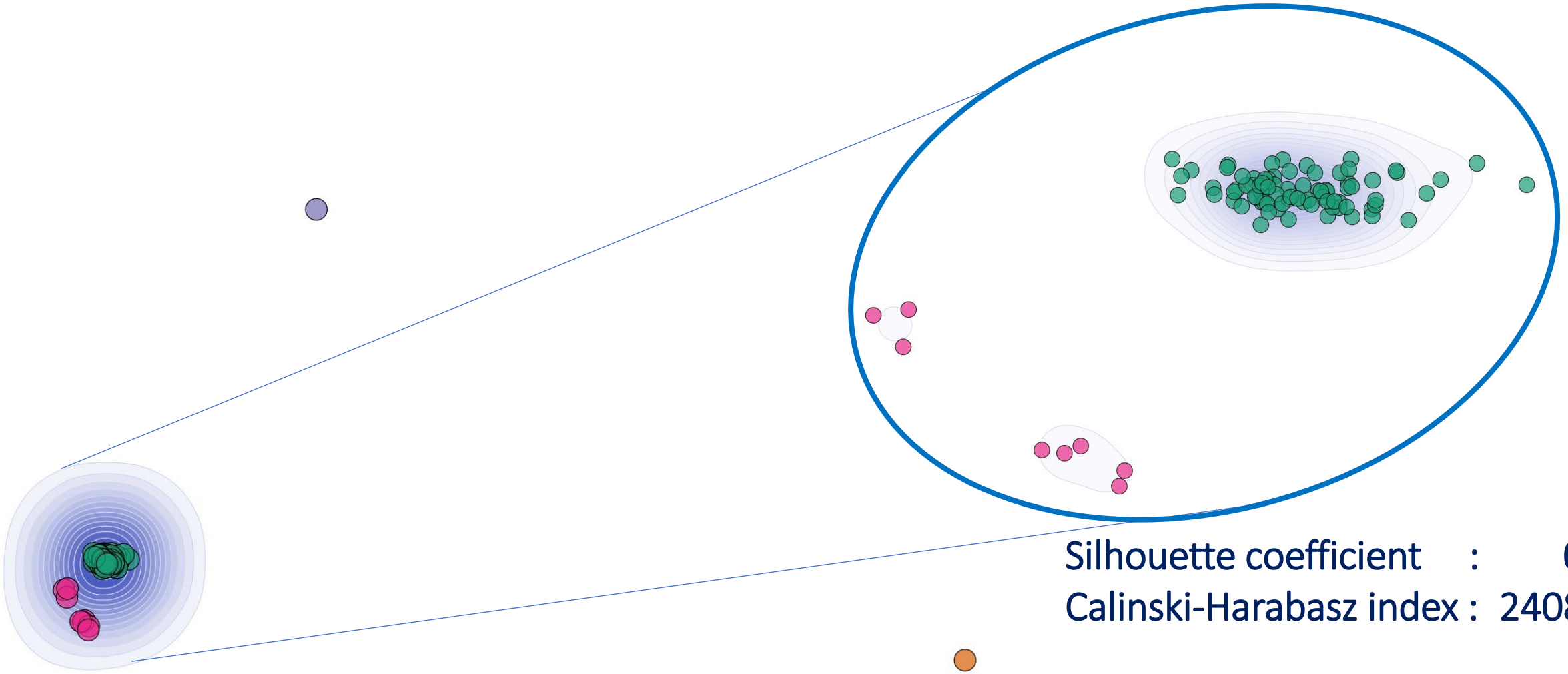


Silhouette coefficient : -0.39
Calinski-Harabasz index : 0.53

Davies-Bouldin index : 4.32

Factor Analysis of Mixed Data (FAMD)

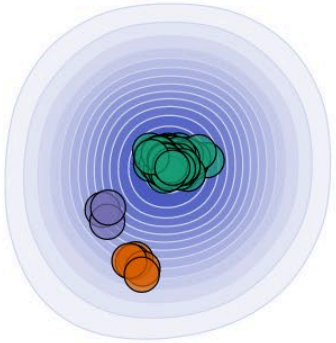
Gaussian mixture models (distribution-based)



Silhouette coefficient : 0.94
Calinski-Harabasz index : 2408.29

Davies-Bouldin index : 0.21

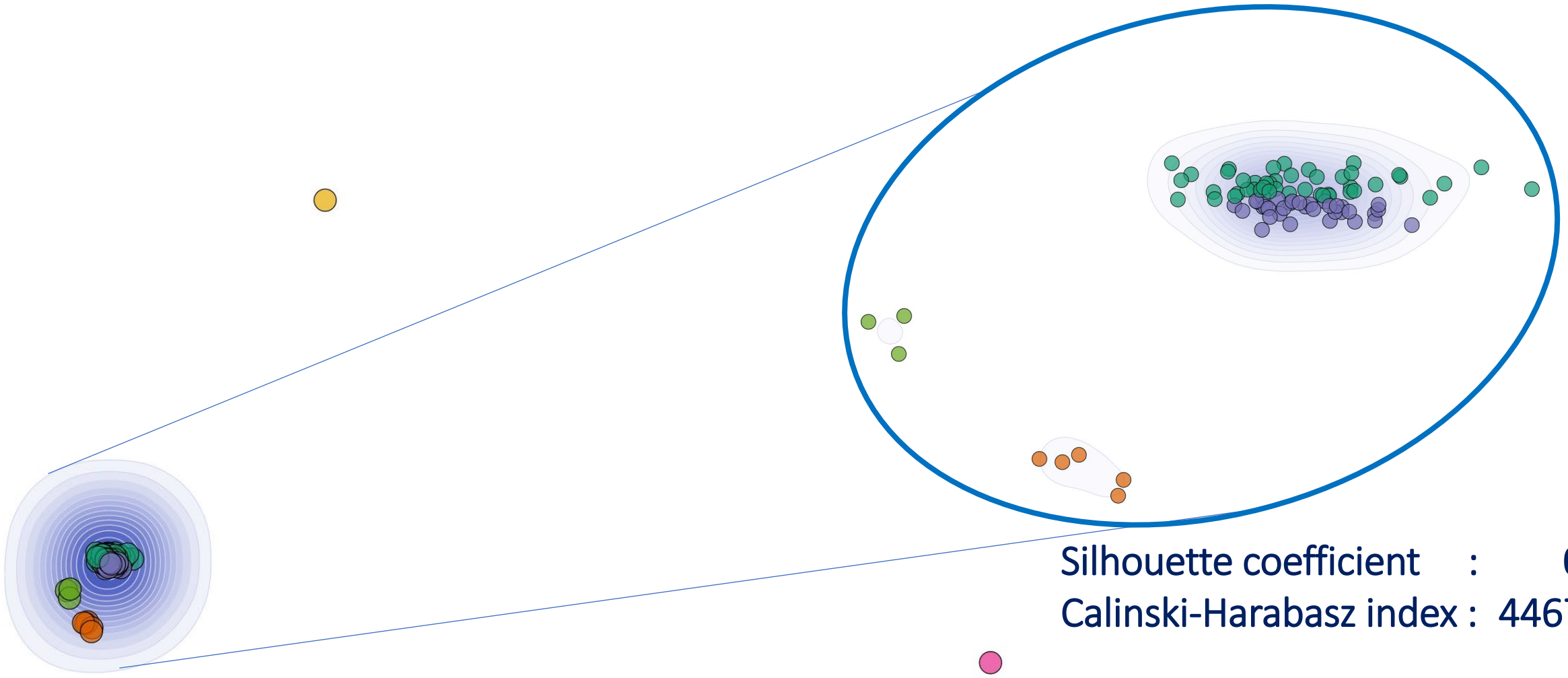
Mean-Shift



- Silhouette coefficient : 0.94
  Calinski-Harabasz index : 3671.70

  Davies-Bouldin index : 0.13

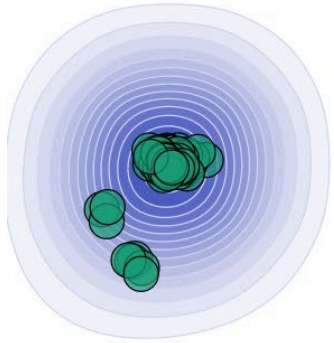
Hierarchical Clustering (6 Clusters)



Silhouette coefficient : 0.49
Calinski-Harabasz index : 4467.11

Davies-Bouldin index : 0.44

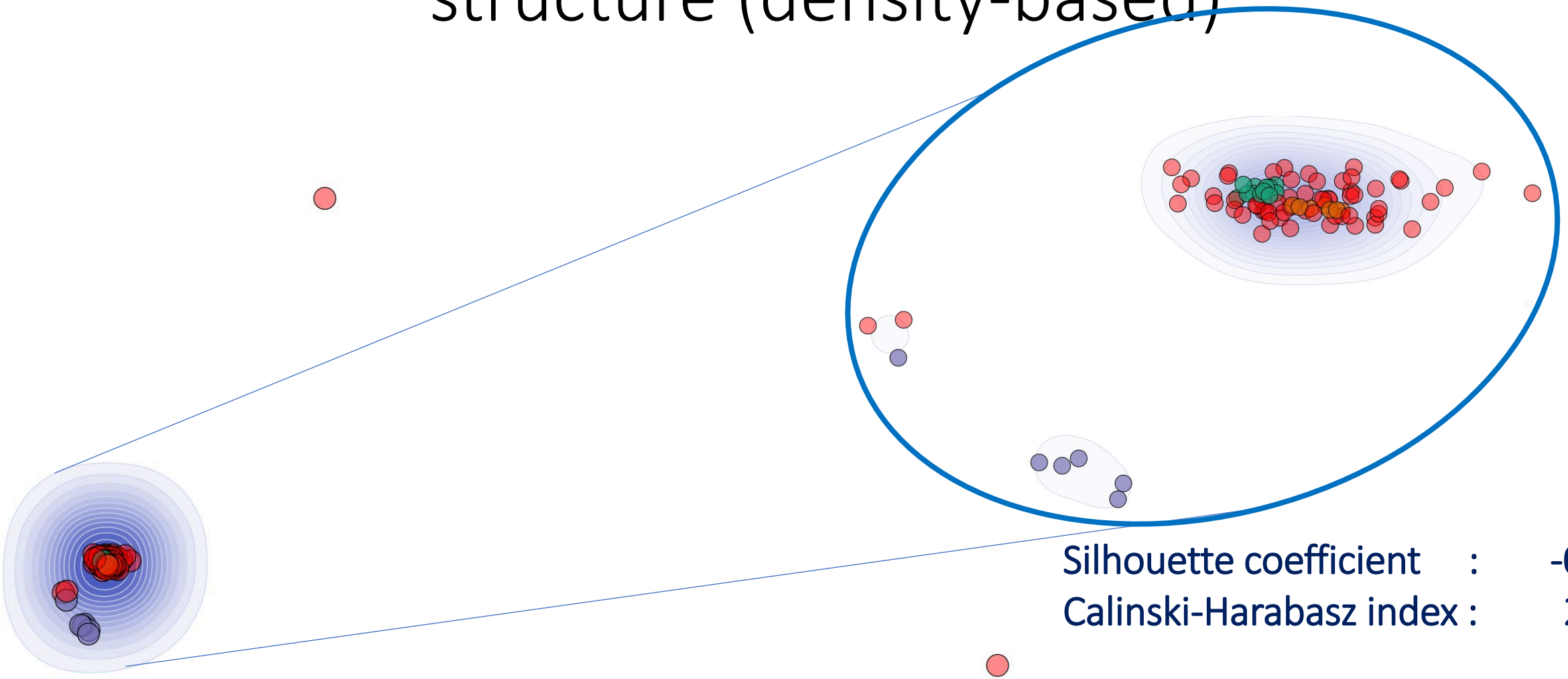
K-means (centroid-based)



- Silhouette coefficient : 0.97
  Calinski-Harabasz index : 104.65

  Davies-Bouldin index : 0.03

Ordering points to identify the clustering structure (density-based)

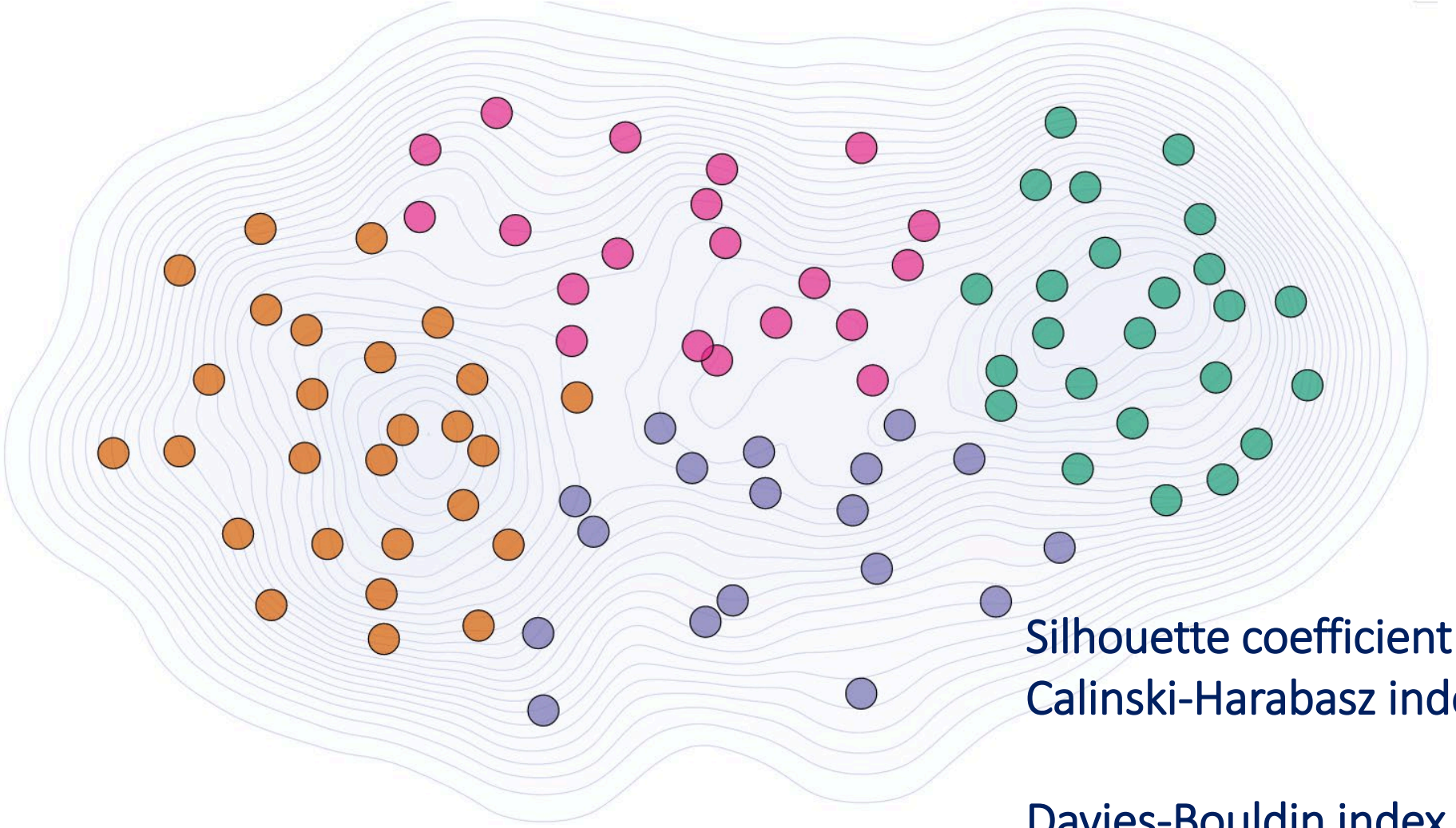


Silhouette coefficient : -0.44
Calinski-Harabasz index : 2.08

Davies-Bouldin index : 2.01

Uniform Manifold Approximation and Projection (UMAP)

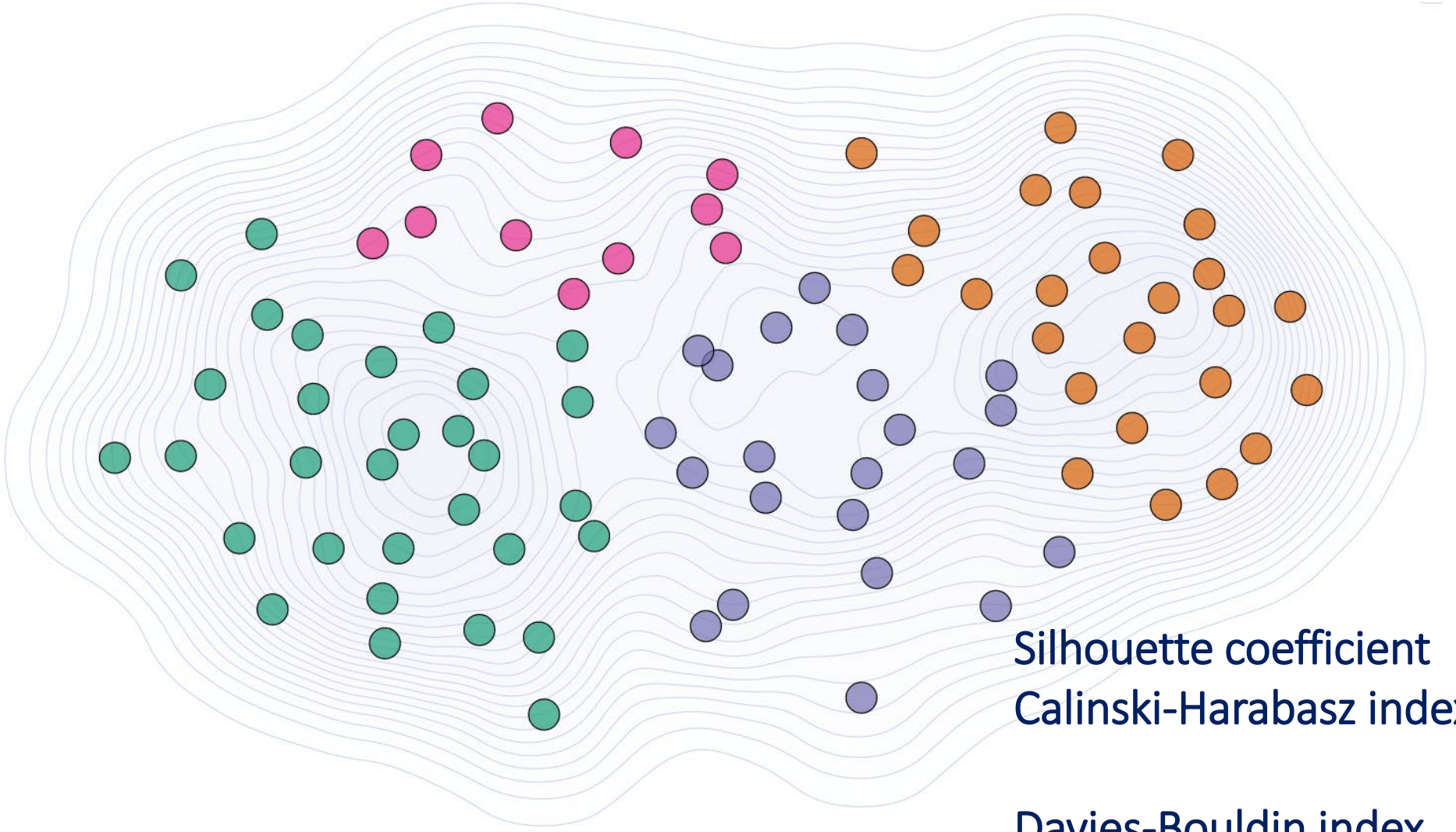
Gaussian mixture models (distribution-based)



Silhouette coefficient : 0.56
Calinski-Harabasz index : 90.58

Davies-Bouldin index : 0.82

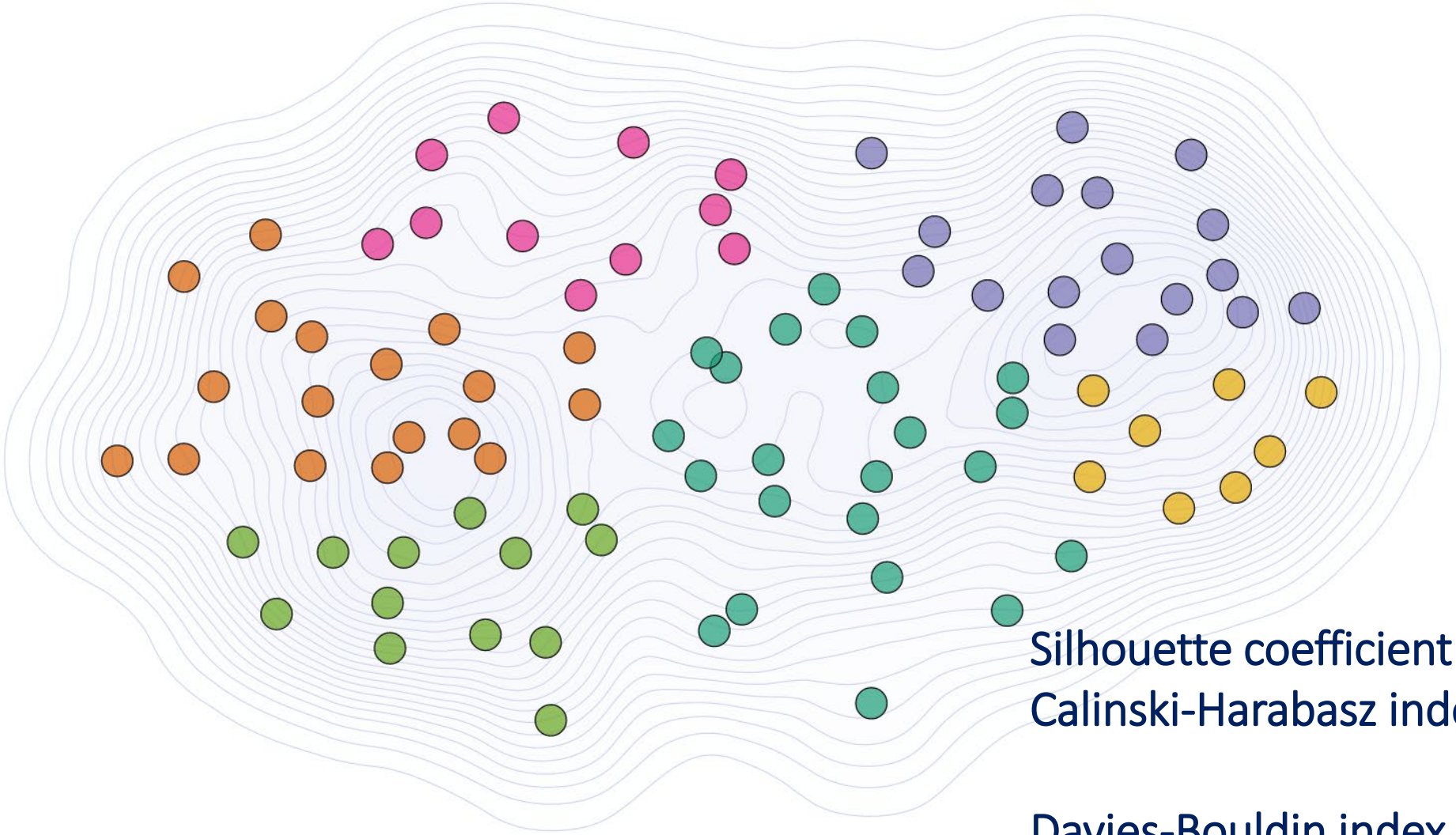
Hierarchical Clustering (4 Clusters)



Silhouette coefficient : 0.50
Calinski-Harabasz index : 86.18

Davies-Bouldin index : 0.89

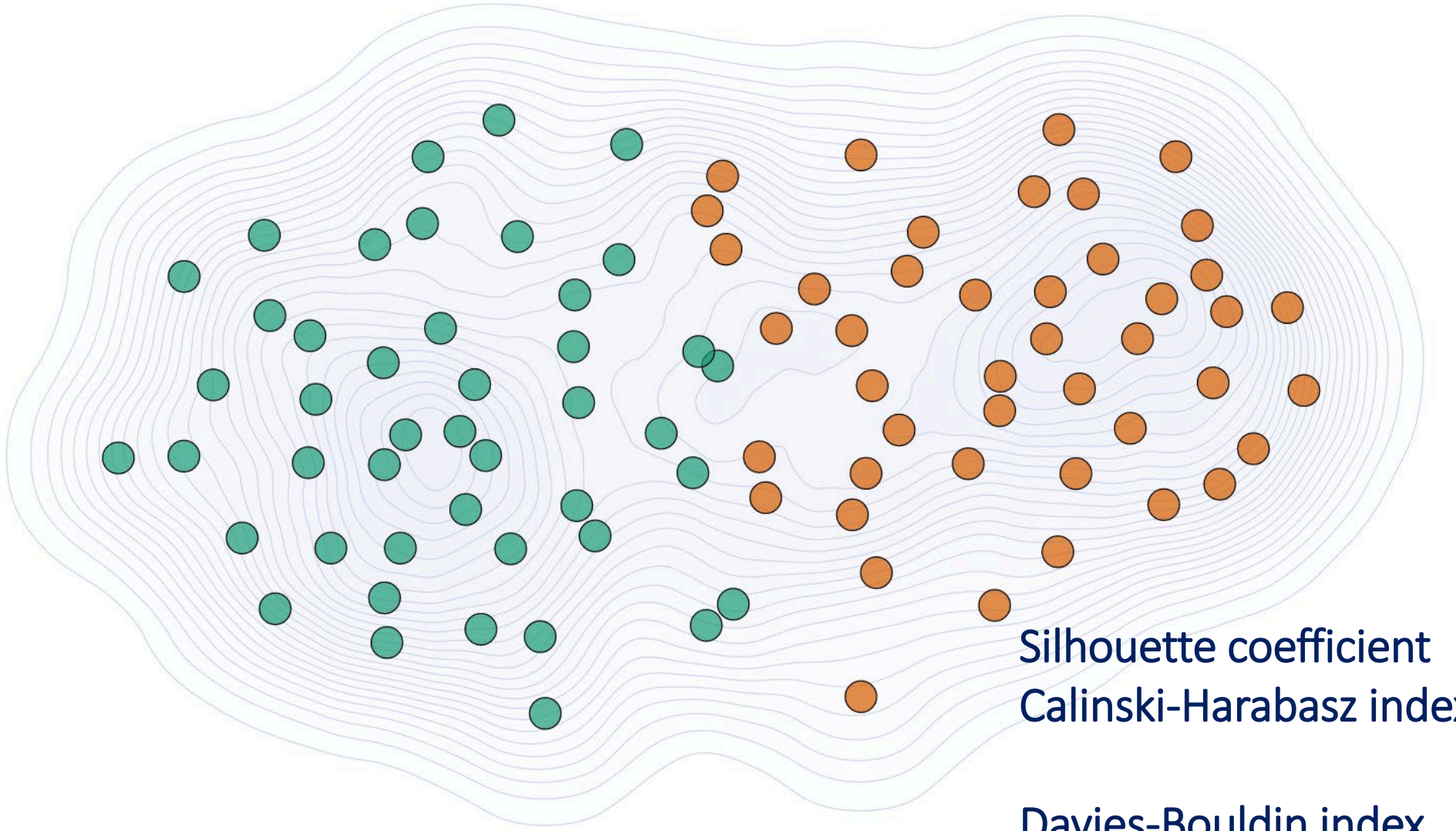
Hierarchical Clustering (6 Clusters)



Silhouette coefficient : 0.44
Calinski-Harabasz index : 82.32

Davies-Bouldin index : 0.88

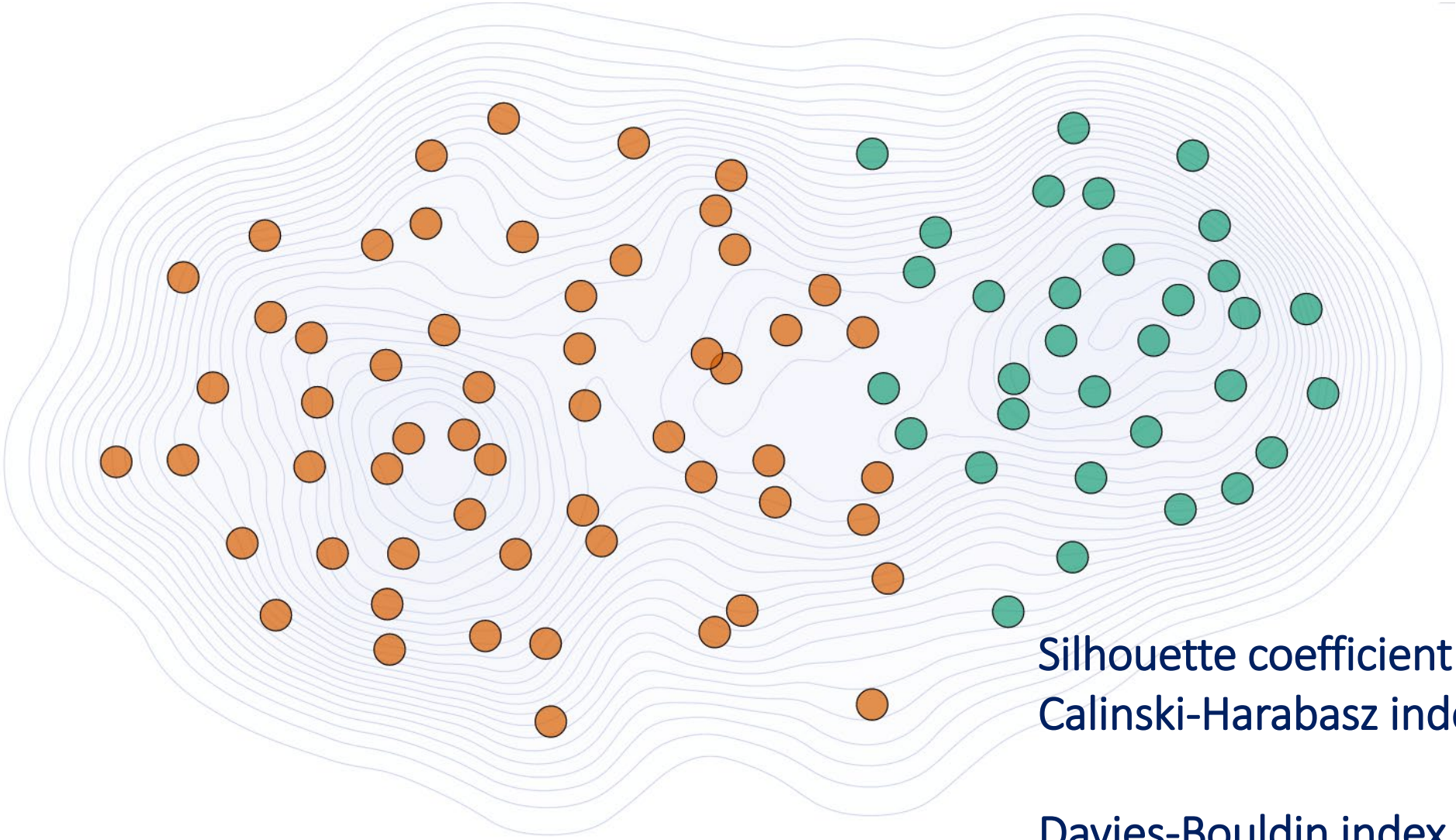
K-means (centroid-based)



Silhouette coefficient : 0.65
Calinski-Harabasz index : 111.35

Davies-Bouldin index : 0.81

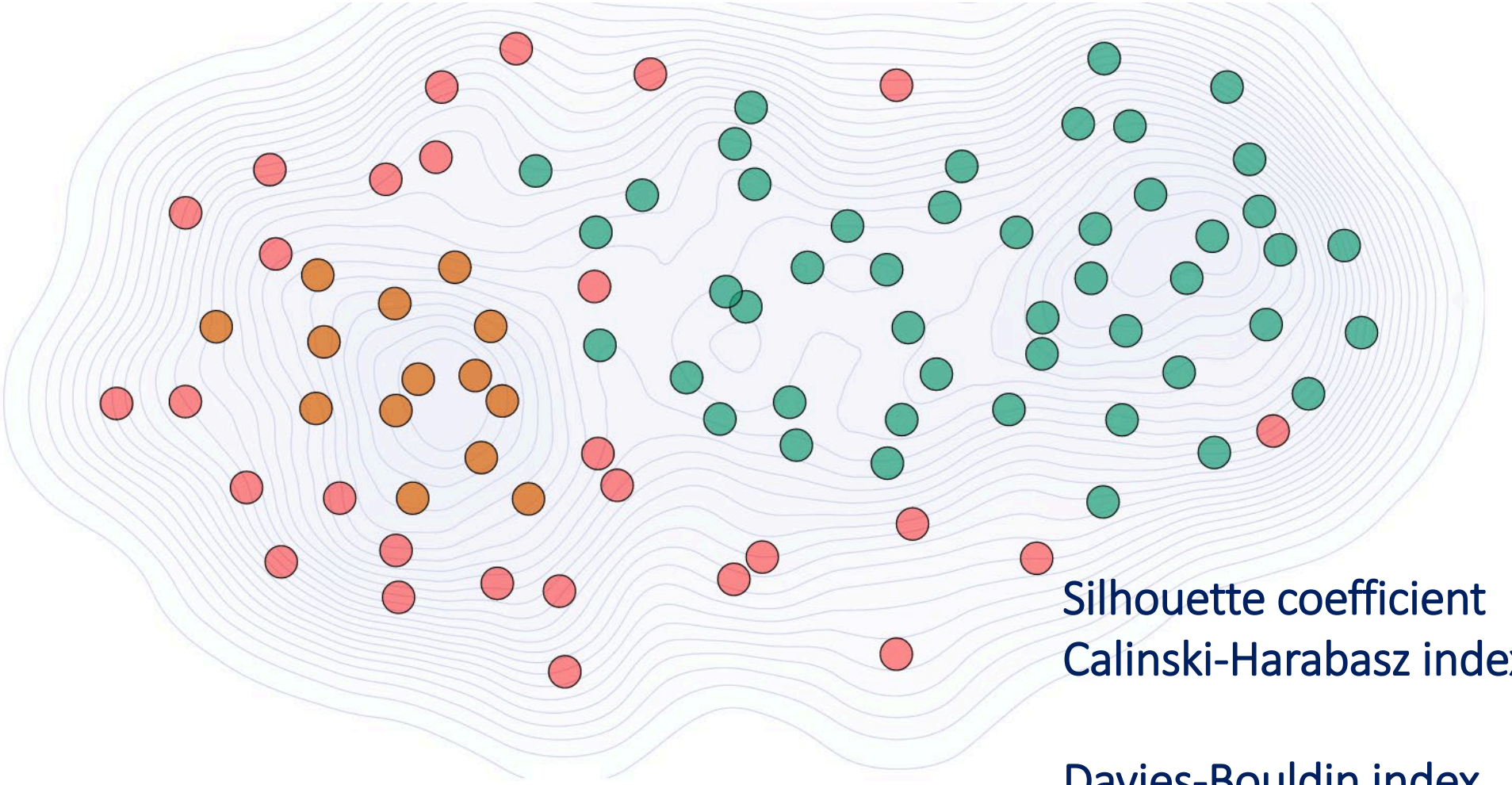
Mean-shift (centroid-based)



Silhouette coefficient : 0.60
Calinski-Harabasz index : 98.33

Davies-Bouldin index : 0.79

Ordering points to identify the clustering structure (density-based)

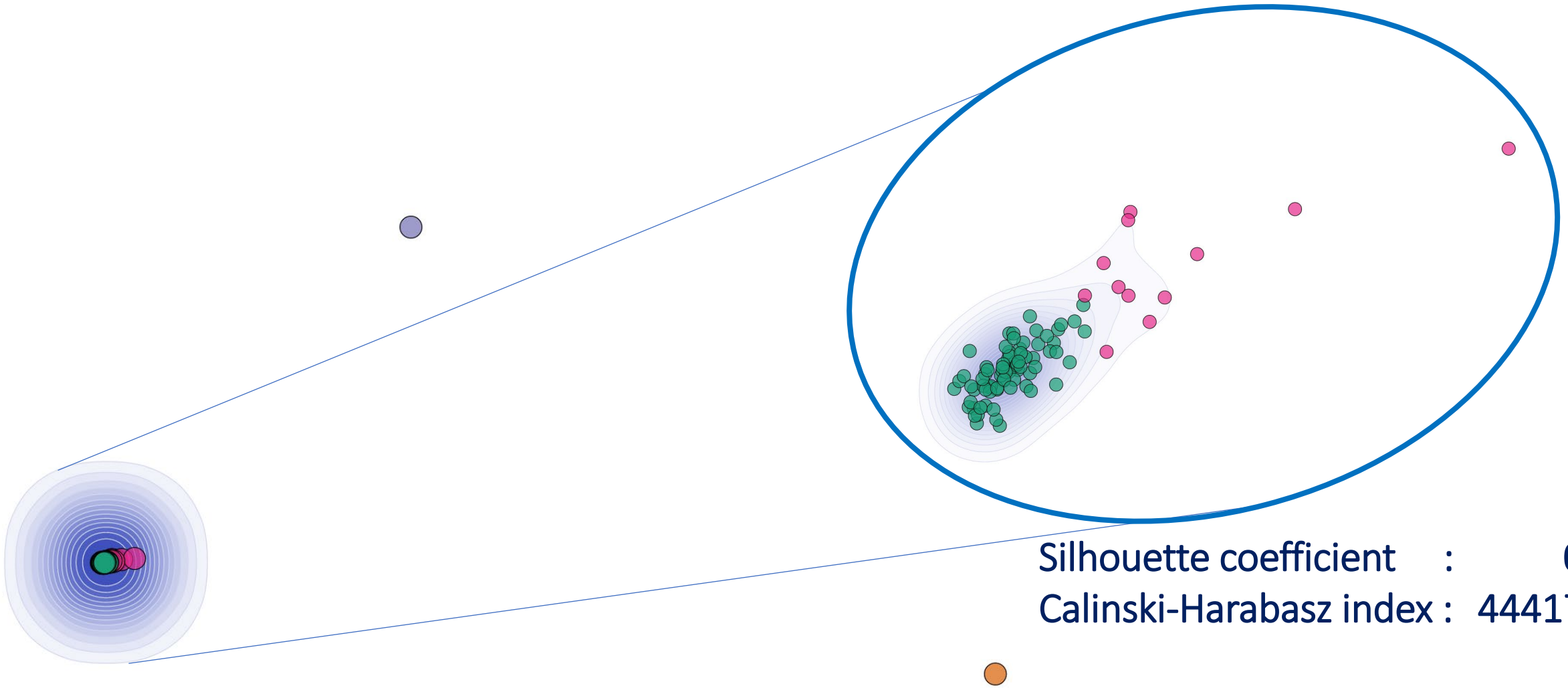


Silhouette coefficient : 0.21
Calinski-Harabasz index : 33.47

Davies-Bouldin index : 2.84

Principal Component Analysis (PCA)

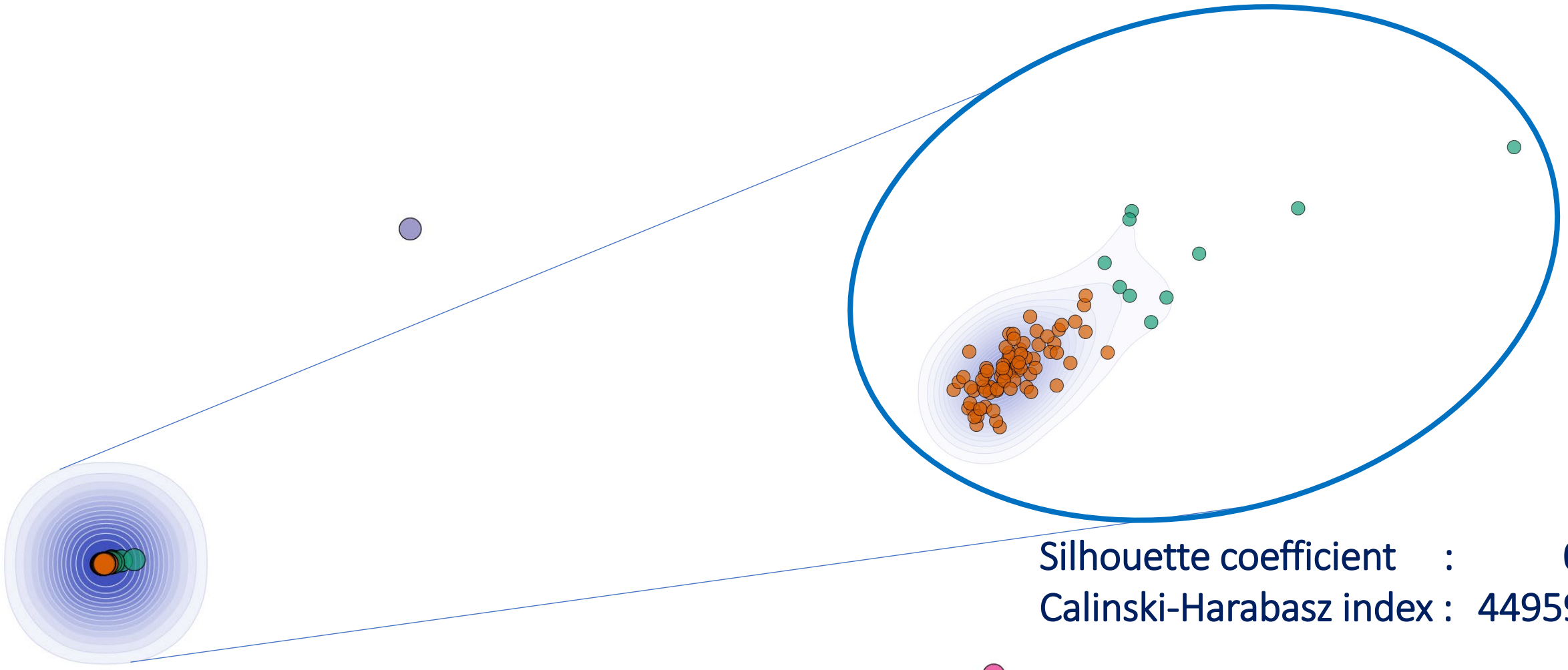
Gaussian mixture models (distribution-based)



Silhouette coefficient : 0.81
Calinski-Harabasz index : 44417.79

Davies-Bouldin index : 0.32

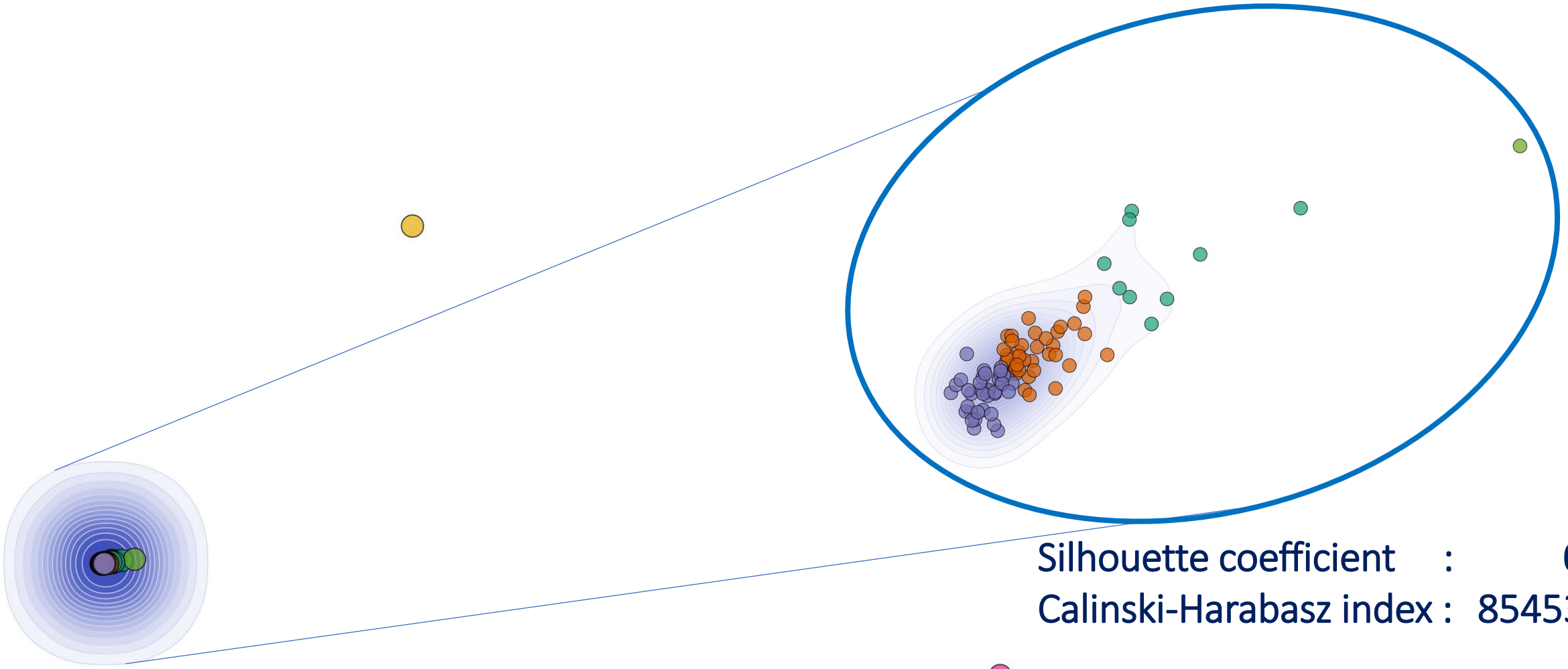
Hierarchical Clustering (4 Clusters)



Silhouette coefficient : 0.83
Calinski-Harabasz index : 44959.24

Davies-Bouldin index : 0.31

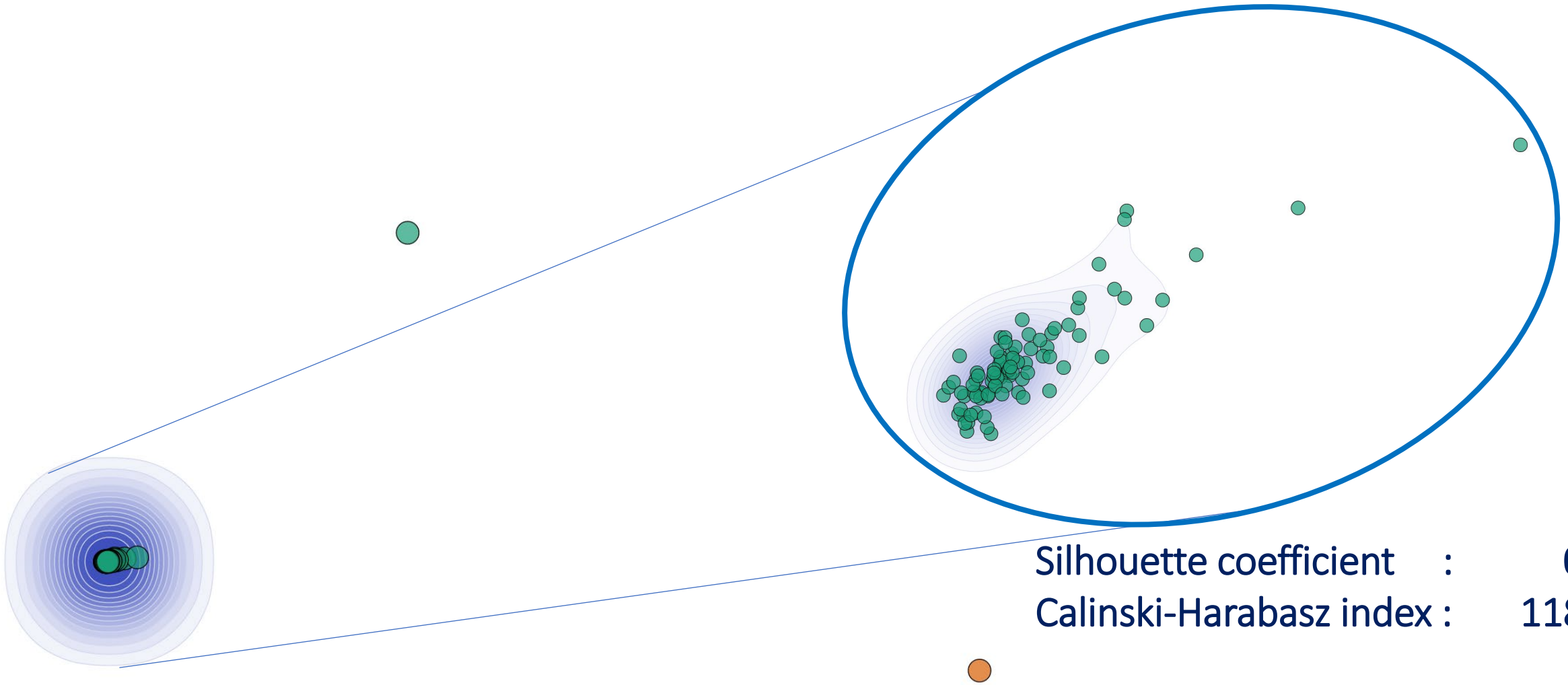
Hierarchical Clustering (6 Clusters)



Silhouette coefficient : 0.58
Calinski-Harabasz index : 85453.22

Davies-Bouldin index : 0.37

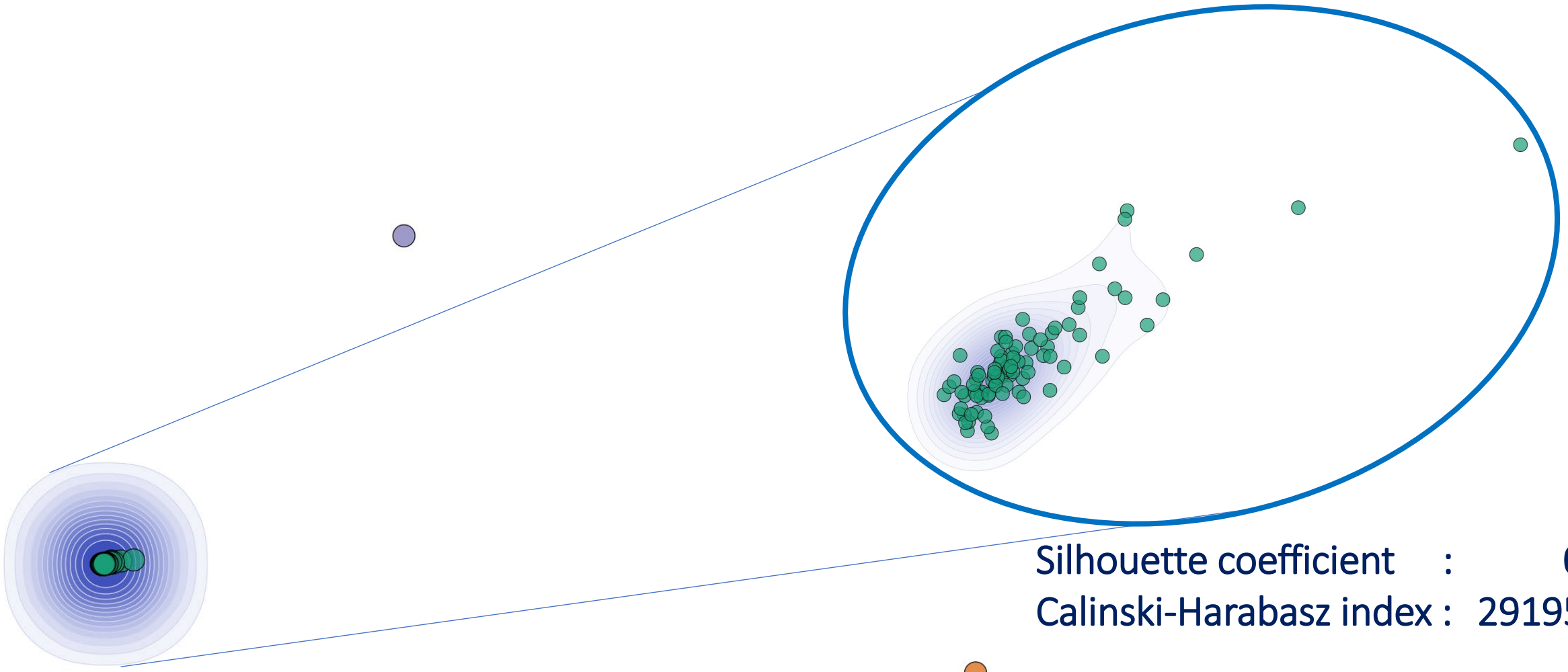
K-means (centroid-based)



Silhouette coefficient : 0.89
Calinski-Harabasz index : 118.47

Davies-Bouldin index : 0.02

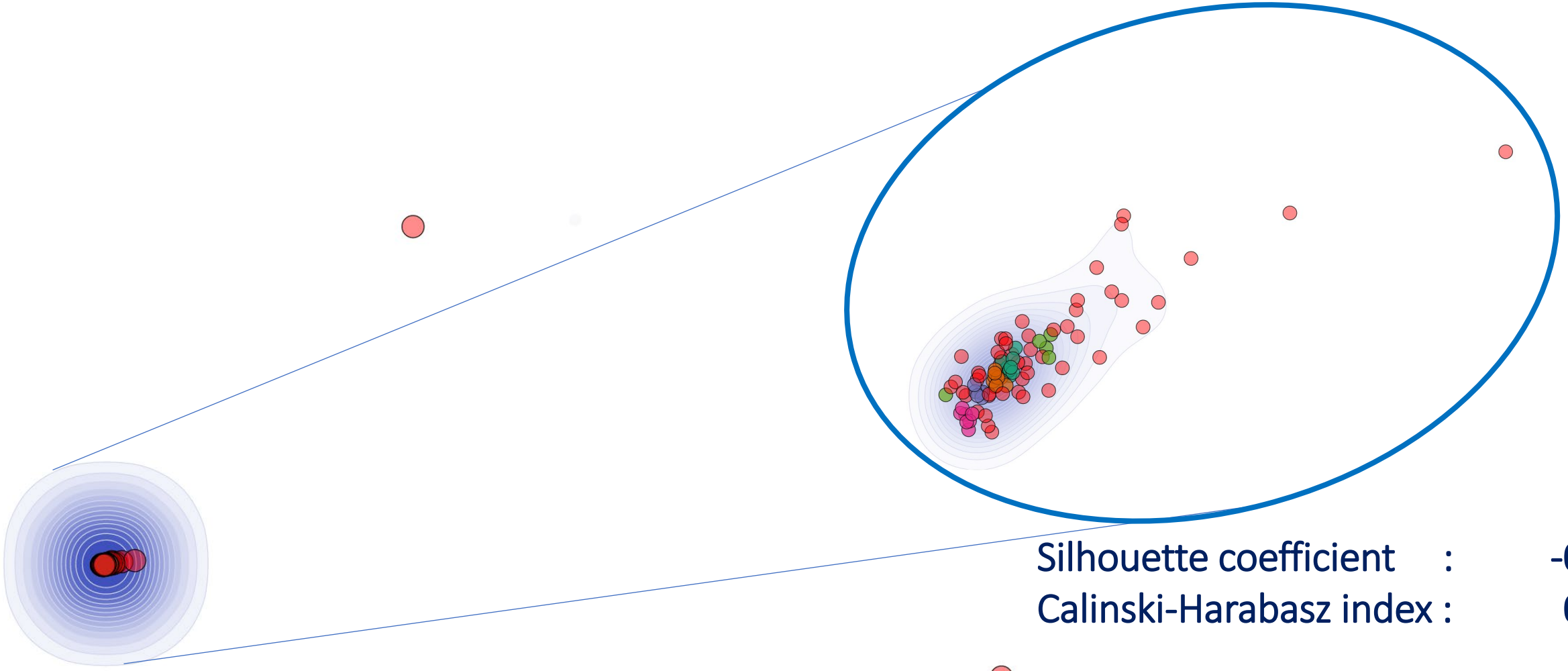
Mean-shift (centroid-based)



Silhouette coefficient : 0.98
Calinski-Harabasz index : 29195.39

Davies-Bouldin index : 0.0039

Ordering points to identify the clustering structure (density-based)



Silhouette coefficient : -0.31
Calinski-Harabasz index : 0.20

Davies-Bouldin index : 2.09