

# Anomaly Detection & Density Clustering

Computational ClassNotes

[dara@lossogenerality.com](mailto:dara@lossogenerality.com)

## Multi-Modal Data

Mixed data e.g. numbers and text and images...

Out[*n*]=

CRIM	ZN	INDUS	CHAS
0.00632	18	2.31	tract does not bound Charles river
0.02731	0	7.07	tract does not bound Charles river
0.02729	0	7.07	tract does not bound Charles river
0.03237	0	2.18	tract does not bound Charles river
0.06905	0	2.18	tract does not bound Charles river
0.02985	0	2.18	tract does not bound Charles river
0.08829	12.5	7.87	tract does not bound Charles river
0.14455	12.5	7.87	tract does not bound Charles river
0.21124	12.5	7.87	tract does not bound Charles river
0.17004	12.5	7.87	tract does not bound Charles river
0.22489	12.5	7.87	tract does not bound Charles river
0.11747	12.5	7.87	tract does not bound Charles river
0.09378	12.5	7.87	tract does not bound Charles river
0.62976	0	8.14	tract does not bound Charles river
0.63796	0	8.14	tract does not bound Charles river
0.62739	0	8.14	tract does not bound Charles river
1.05393	0	8.14	tract does not bound Charles river
0.7842	0	8.14	tract does not bound Charles river
0.80271	0	8.14	tract does not bound Charles river
0.7258	0	8.14	tract does not bound Charles river

K < showing 1-20 of 506 > ↵

# AcceptanceThreshold

Set threshold and compute anomalies

Out[8]=

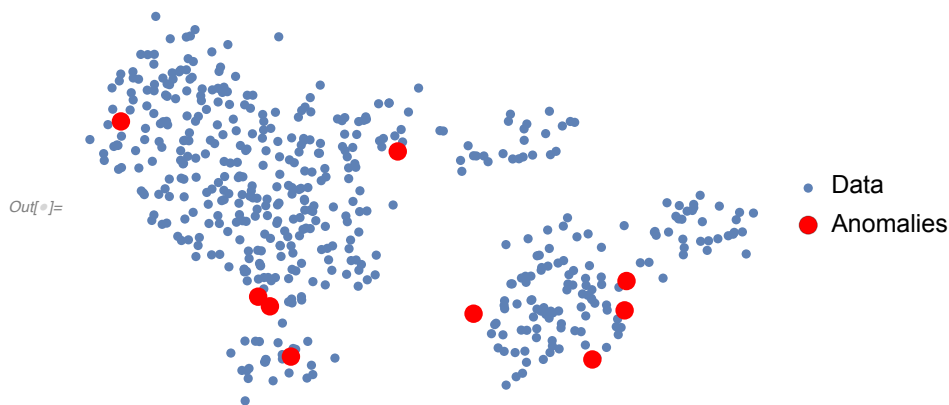
CRIM	ZN	INDUS	CHAS
0.21124	12.5	7.87	tract does not bound Charles river
0.09378	12.5	7.87	tract does not bound Charles river
0.80271	0	8.14	tract does not bound Charles river
1.25179	0	8.14	tract does not bound Charles river
1.13081	0	8.14	tract does not bound Charles river
1.38799	0	8.14	tract does not bound Charles river
1.61282	0	8.14	tract does not bound Charles river
0.25387	0	6.91	tract does not bound Charles river
0.0136	75	4	tract does not bound Charles river
0.01311	90	1.22	tract does not bound Charles river
0.01951	17.5	1.38	tract does not bound Charles river
0.02875	28	15.04	tract does not bound Charles river
0.08187	0	2.89	tract does not bound Charles river
0.22876	0	8.56	tract does not bound Charles river
0.97617	0	21.89	tract does not bound Charles river
1.62864	0	21.89	tract does not bound Charles river
2.37934	0	19.58	tract does not bound Charles river
1.12658	0	19.58	tract bounds Charles river
1.22358	0	19.58	tract does not bound Charles river
1.42502	0	19.58	tract does not bound Charles river

K < showing 1–20 of 81 > X

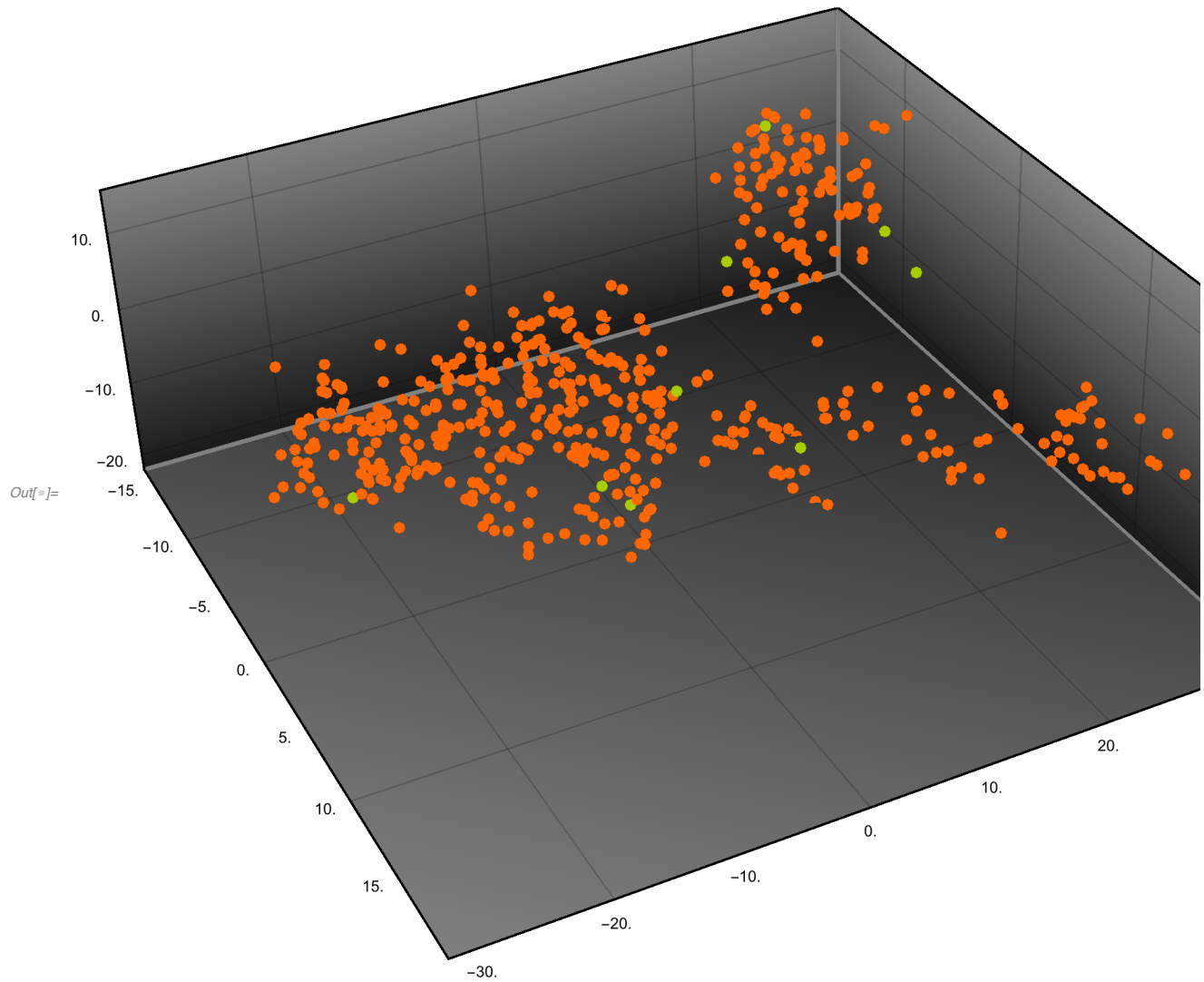
	CRIM	ZN	INDUS	CHAS
	1.38799	0	8.14	tract does not bound Charles river
	0.22876	0	8.56	tract does not bound Charles river
	1.62864	0	21.89	tract does not bound Charles river
	0.36894	22	5.86	tract does not bound Charles river
Out[ ]=	0.76162	20	3.97	tract does not bound Charles river
	3.47428	0	18.1	tract bounds Charles river
	4.55587	0	18.1	tract does not bound Charles river
	13.5222	0	18.1	tract does not bound Charles river
	6.53876	0	18.1	tract bounds Charles river

# Dimensional Reduction

2D reduction

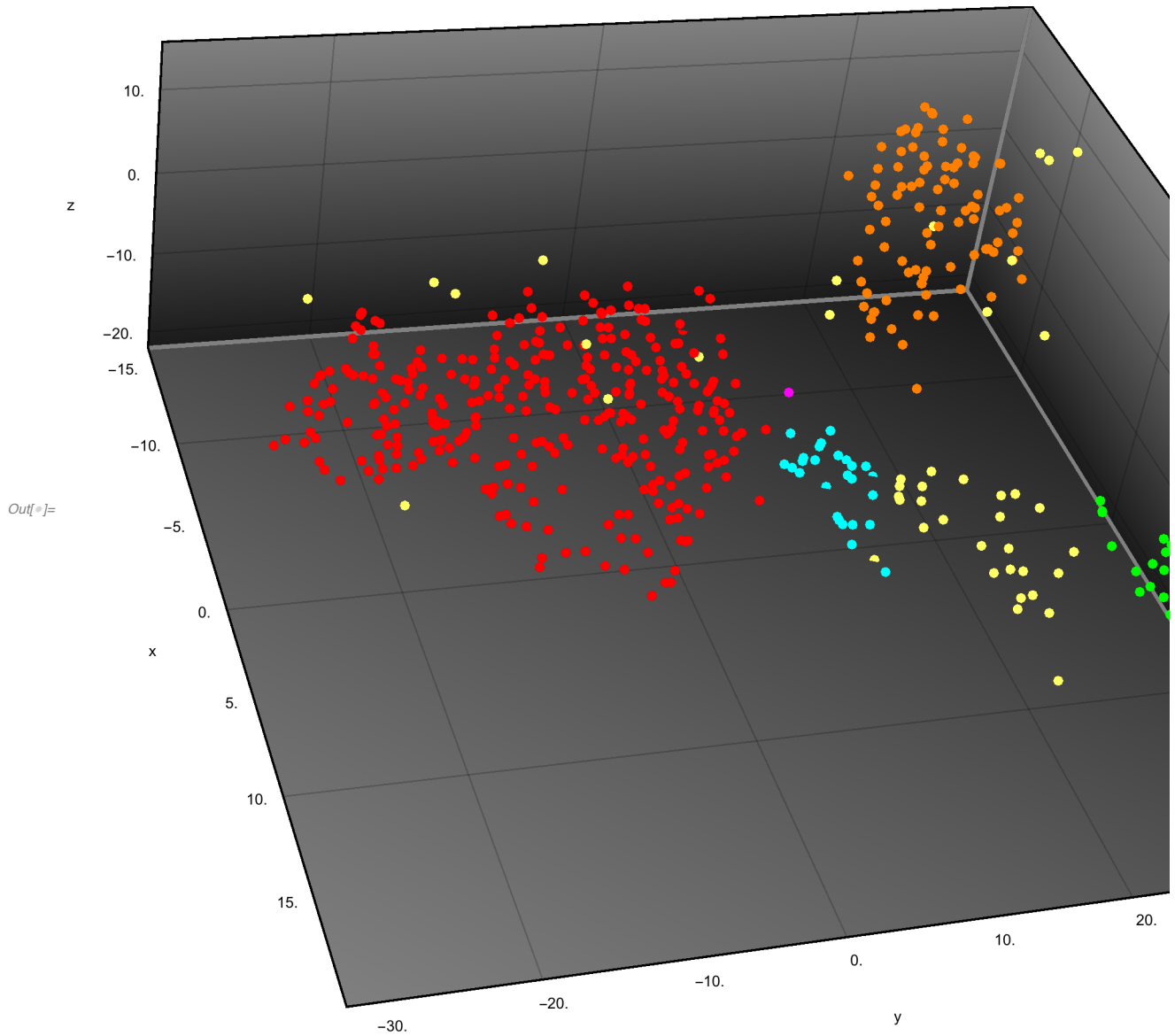


3D reduction



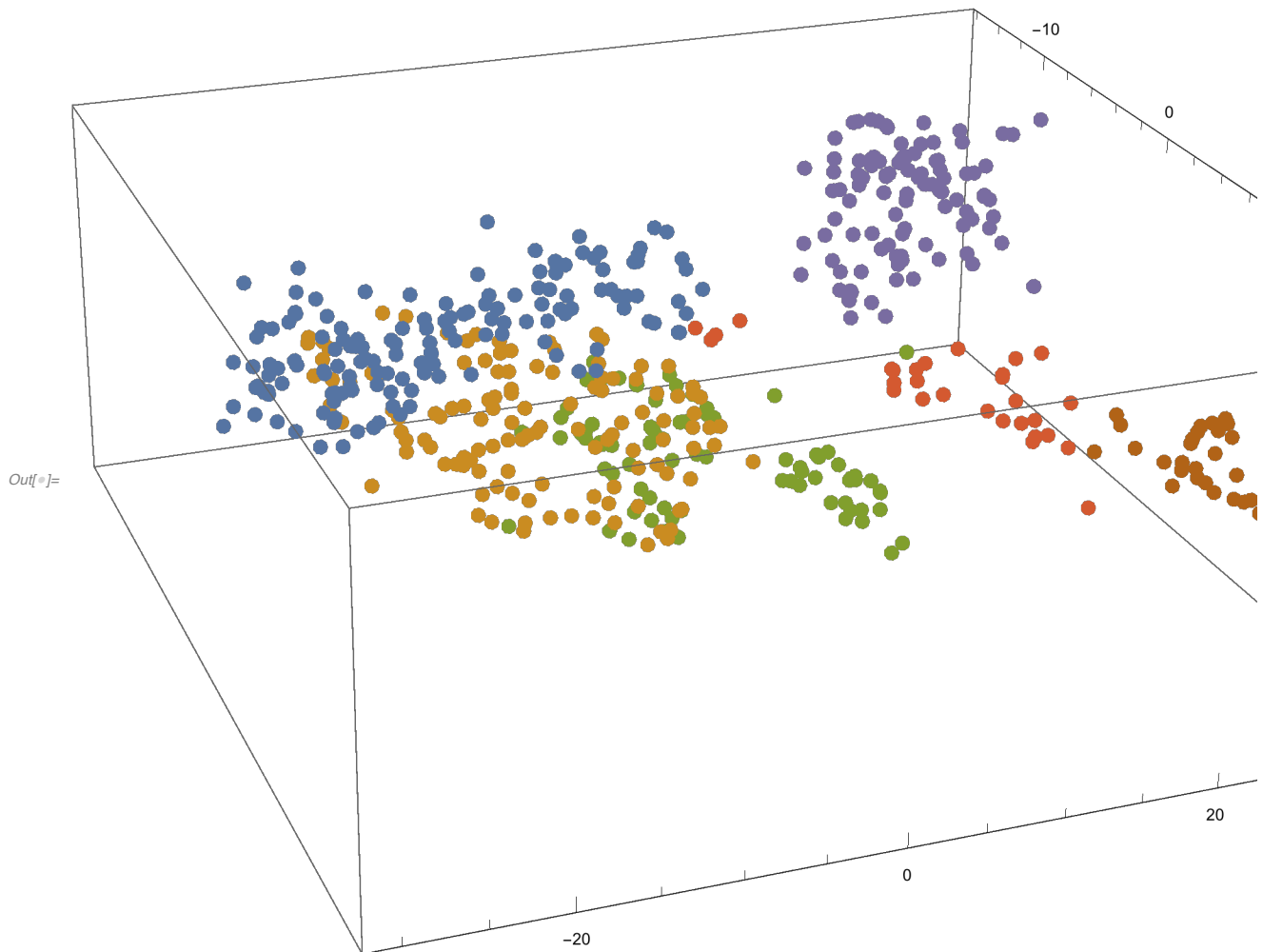
## Density Clusters

Data clustered according the density of local clusters, Red the highest Yellow the lowest or noise.



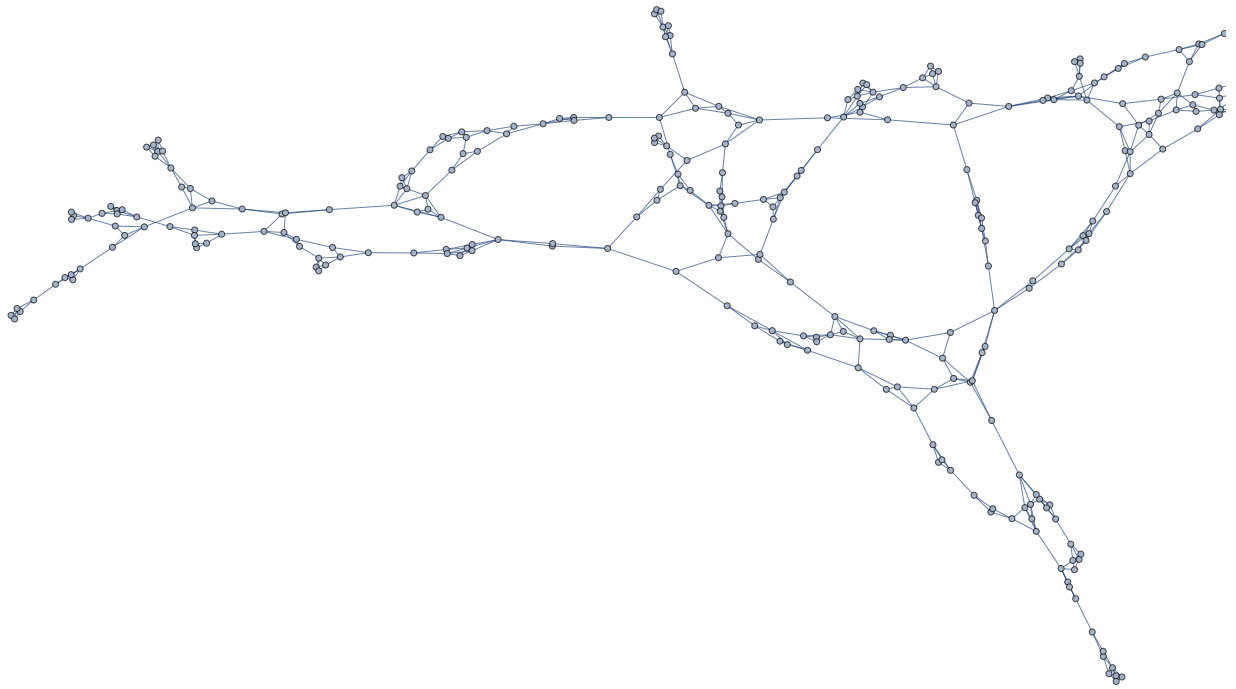
# Neighborhood Contraction Clustering

Near Neighbours are placed into the same cluster

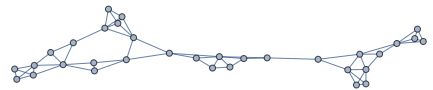
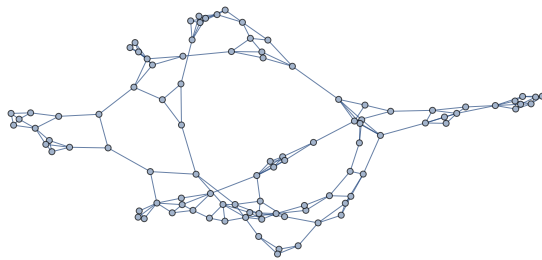


# Nearest Neighbor Graph

Compute the disjoint neighbourhoods



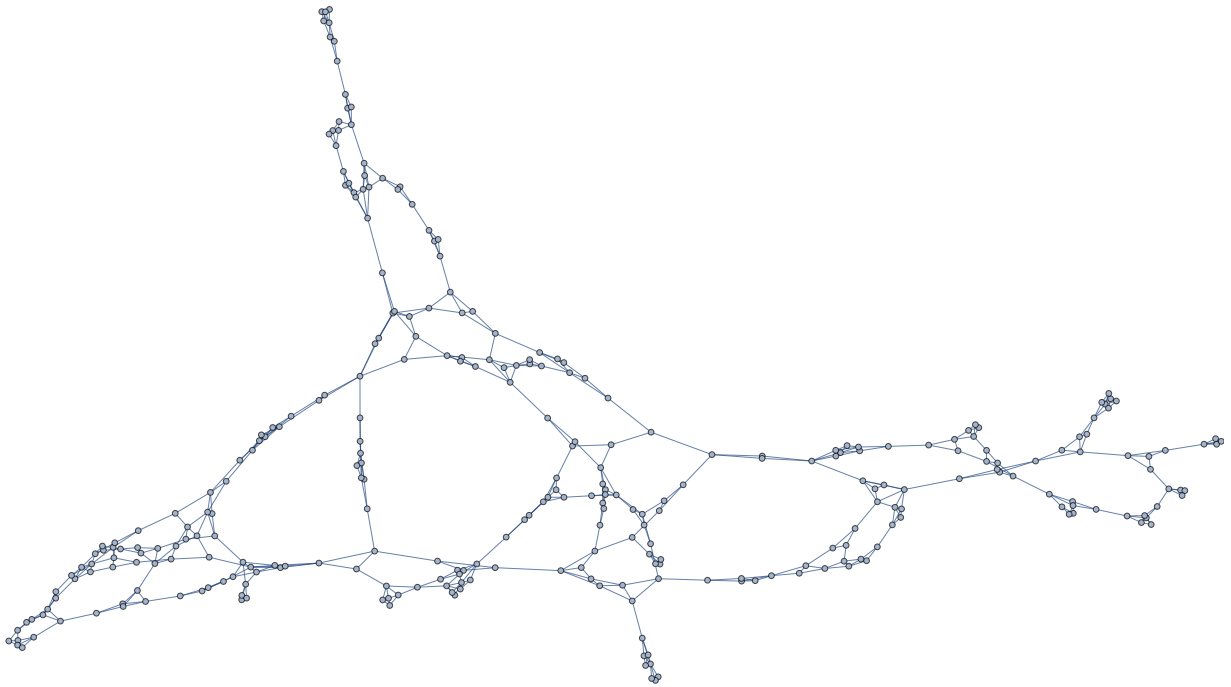
Out[ ]=-



Compute the largest neighbourhood

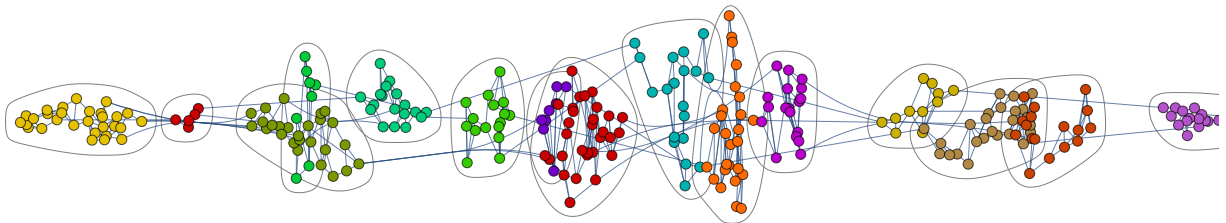


Out[ ]=

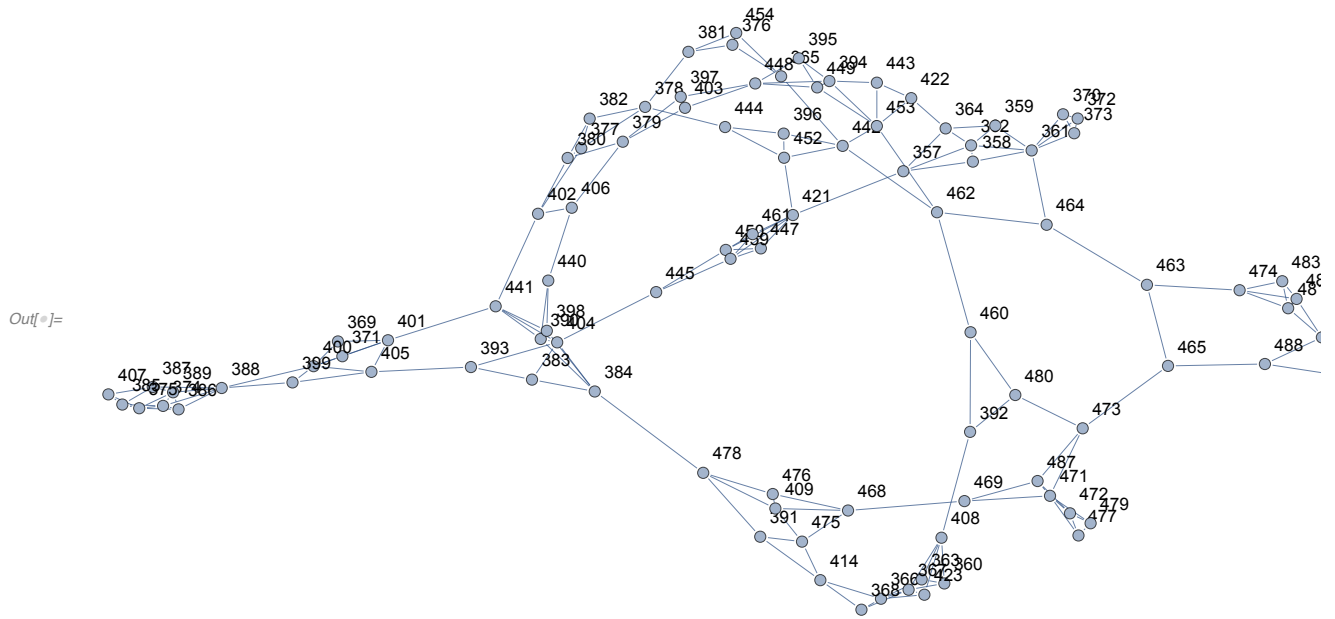


Within each cluster, there is yet another structure called Community which is denser more connected neighbours but connected to each other in fewer links (edges).

Out[ ]=



Compute the second largest neighbourhood



Numbers refer to the row number in the original excel file. The process of grabbing the original data could seamlessly be automated.

