CS60050 - MACHINE LEARNING ASSIGNMENT 3
CLUSTERING AND EVALUATION

NAME: **Himanshu Mundhra**
ROLL No. : **16CS10057**

In this folder, we aim to build a Clustering out of the **AAAI Submitted Papers Dataset**, by measuring their similarities with each other using the **Jaccard Coefficient of Similarity** on the **TOPICS** of the Papers.

We use Single-Linkage and Complete-Linkage Hierarchical Agglomerative Algorithms in **PART1,** and then we use Girvan Newman Graph-Based Hierarchical Divisive Algorithm with different thresholds in **PART2.**

In Part3, we evaluate the quality of our clusters by adopting as 'gold-standard' the **HIGH-LEVEL KEYWORDS** of each paper, and use **NORMALISED MUTUAL INFORMATION** as the Metric of Evaluating the Clusterings.

The implementation of **PART1** has been done completely from scratch without the use of any specialized libraries for evaluation of proximities and for merging the clusters.
At each place where we need to choose a Maximum Value from the *Proximity Matrix* or the *Closest Matrix,* we use the **numpy.nanargmax()** function which returns the first occurrence of the maximum value.
To prevent the clusters from becoming too skewed towards the Smaller Indexed Clusters, when we merge two clusters, we assign it the larger of the two indices.
This tie-breaking is not specified anywhere in particular and is according to the whims of the programmer and the practical need of the Clustering. Results differ a lot, especially in such more or less discrete datasets, depending on the way these ties are broken. We prefer this to randomly choosing one of the max indices to get a deterministic answer from each run of the Algorithm on the same dataset, my whim nothing imperative.

9 Clusters formed using **Single Linkage** on the AAAI Submitted Papers Dataset using Jaccard Coefficient on "Topics" as the Parameter for measuring the Similarity of Two Papers

CLUSTER 1 INDEXED BY 52 HAS 1 OBJECTS = {52}

CLUSTER 2 INDEXED BY 75 HAS 1 OBJECTS = {75}

CLUSTER 3 INDEXED BY 77 HAS 1 OBJECTS = {77}

CLUSTER 4 INDEXED BY 107 HAS 1 OBJECTS = {107}

CLUSTER 5 INDEXED BY 128 HAS 1 OBJECTS = {128}

CLUSTER 6 INDEXED BY 142 HAS 42 OBJECTS = {129, 3, 134, 7, 139, 12, 13, 142, 15, 23, 26, 30, 34, 35, 38, 43, 47, 48, 53, 58, 59, 62, 66, 71, 73, 82, 83, 85, 88, 92, 98, 99, 102, 105, 108, 113, 118, 120, 122, 124, 125, 127}

CLUSTER 7 INDEXED BY 146 HAS 47 OBJECTS = {0, 1, 2, 131, 4, 132, 133, 135, 8, 136, 137, 138, 10, 140, 17, 146, 19, 18, 20, 24, 27, 28, 29, 31, 37, 39, 49, 50, 55, 67, 68, 70, 74, 76, 79, 81, 86, 87, 95, 96, 106, 112, 116, 117, 121, 123, 126}

CLUSTER 8 INDEXED BY 148 HAS 55 OBJECTS = {130, 5, 6, 9, 11, 141, 14, 143, 144, 145, 16, 147, 148, 21, 22, 25, 32, 33, 36, 40, 41, 42, 44, 45, 46, 51, 54, 56, 57, 60, 61, 63, 64, 65, 69, 72, 78, 80, 84, 89, 90, 91, 93, 94, 97, 100, 101, 103, 104, 109, 110, 111, 114, 115, 119}

CLUSTER 9 INDEXED BY 149 HAS 1 OBJECTS = {149}

Normalized Mutual Information Value = 0.508863

Dendogram stored in SingleLinkage.txt

9 Clusters formed using **Complete Linkage** on the AAAI Submitted Papers Dataset using Jaccard Coefficient on "Topics" as the Parameter for measuring the Similarity of Two Papers

CLUSTER 1 INDEXED BY 138 HAS 99 OBJECTS = {0, 1, 2, 3, 4, 5, 7, 9, 12, 13, 15, 17, 18, 20, 21, 22, 23, 26, 27, 28, 29, 30, 31, 32, 33, 35, 36, 42, 43, 44, 45, 46, 48, 49, 50, 51, 54, 56, 57, 58, 59, 60, 62, 63, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 81, 82, 84, 85, 86, 87, 88, 92, 94, 95, 96, 98, 99, 102, 103, 104, 106, 108, 109, 110, 112, 113, 116, 117, 118, 120, 122, 123, 124, 126, 127, 128, 129, 131, 132, 133, 134, 135, 136, 137, 138}

CLUSTER 2 INDEXED BY 139 HAS 3 OBJECTS = {107, 34, 139}

CLUSTER 3 INDEXED BY 140 HAS 9 OBJECTS = {37, 39, 10, 140, 47, 83, 53, 24, 121}

CLUSTER 4 INDEXED BY 142 HAS 4 OBJECTS = {38, 105, 125, 142}

CLUSTER 5 INDEXED BY 144 HAS 13 OBJECTS = {97, 100, 101, 6, 41, 143, 144, 111, 80, 16, 89, 91, 25}

CLUSTER 6 INDEXED BY 145 HAS 3 OBJECTS = {145, 90, 130}

CLUSTER 7 INDEXED BY 146 HAS 6 OBJECTS = {146, 19, 52, 55, 8, 79}

CLUSTER 8 INDEXED BY 148 HAS 8 OBJECTS = {64, 147, 148, 115, 141, 40, 61, 14}

CLUSTER 9 INDEXED BY 149 HAS 5 OBJECTS = {114, 149, 119, 11, 93}

Normalized Mutual Information Value = 0.343572

Dendogram stored in CompleteLinkage.txt

9 Clusters formed using **Graph Clustering** on the AAAI Submitted Papers Dataset using Jaccard Coefficient on "Topics" as the Parameter for measuring the Similarity of Two Papers with THRESHOLD for Edge (Similarity) between two Nodes (Papers) = **0.10**

CLUSTER 1 WITH 42 OBJECTS = {0, 2, 4, 132, 133, 135, 8, 136, 10, 137, 138, 17, 18, 19, 20, 146, 24, 27, 29, 31, 37, 39, 49, 50, 52, 55, 67, 68, 70, 76, 79, 86, 87, 95, 96, 106, 116, 117, 121, 123, 124, 126}

CLUSTER 2 WITH 56 OBJECTS = {128, 1, 130, 5, 9, 11, 12, 141, 14, 143, 145, 147, 148, 21, 149, 22, 25, 32, 33, 36, 40, 41, 42, 43, 45, 46, 51, 54, 56, 57, 60, 61, 63, 64, 65, 69, 72, 75, 80, 84, 89, 90, 91, 93, 94, 98, 102, 103, 104, 109, 110, 111, 114, 115, 119, 125}

CLUSTER 3 WITH 17 OBJECTS = {129, 66, 99, 3, 134, 71, 73, 13, 47, 48, 113, 83, 53, 88, 26, 30,127}

CLUSTER 4 WITH 4 OBJECTS = {16, 97, 101, 6}

CLUSTER 5 WITH 21 OBJECTS = {7, 139, 142, 15, 23, 34, 35, 38, 58, 59, 62, 77, 82, 85, 92, 105, 107, 108, 118, 120, 122}

CLUSTER 6 WITH 5 OBJECTS = {131, 74, 112, 81, 28}

CLUSTER 7 WITH 2 OBJECTS = {140, 44}

CLUSTER 8 WITH 1 OBJECTS = {78}

CLUSTER 9 WITH 2 OBJECTS = {144, 100}

Initial Number of Clusters = 1

Normalized Mutual Information Value = 0.571182

9 Clusters formed using **Graph Clustering** on the AAAI Submitted Papers Dataset using Jaccard Coefficient on "Topics" as the Parameter for measuring the Similarity of Two Papers with THRESHOLD for Edge (Similarity) between two Nodes (Papers) = **0.13**

CLUSTER 1 WITH 42 OBJECTS = {0, 2, 4, 132, 133, 135, 8, 136, 10, 137, 138, 17, 18, 19, 20, 146, 24, 27, 29, 31, 37, 39, 49, 50, 52, 55, 67, 68, 70, 76, 79, 86, 87, 95, 96, 106, 116, 117, 121, 123, 124, 126}

CLUSTER 2 WITH 57 OBJECTS = {128, 1, 130, 5, 9, 11, 12, 141, 14, 143, 16, 145, 147, 148, 149, 21, 22, 25, 32, 33, 36, 40, 41, 42, 43, 45, 46, 51, 54, 56, 57, 60, 61, 63, 64, 65, 69, 72, 75, 80, 84, 89, 90, 91, 93, 94, 98, 102, 103, 104, 109, 110, 111, 114, 115, 119, 125}

CLUSTER 3 WITH 18 OBJECTS = {129, 66, 99, 3, 134, 71, 73, 13, 47, 48, 113, 83, 53, 88, 26, 92, 30,127}

CLUSTER 4 WITH 3 OBJECTS = {97, 101, 6}

CLUSTER 5 WITH 20 OBJECTS = {7, 139, 142, 15, 23, 34, 35, 38, 58, 59, 62, 77, 82, 85, 105, 107, 108, 118, 120, 122}

CLUSTER 6 WITH 5 OBJECTS = {131, 74, 112, 81, 28}

CLUSTER 7 WITH 2 OBJECTS = {140, 44}

CLUSTER 8 WITH 1 OBJECTS = {78}

CLUSTER 9 WITH 2 OBJECTS = {144, 100}

Initial Number of Clusters = 1

Normalized Mutual Information Value = 0.559572

9 Clusters formed using **Graph Clustering** on the AAAI Submitted Papers Dataset using Jaccard Coefficient on "Topics" as the Parameter for measuring the Similarity of Two Papers with THRESHOLD for Edge (Similarity) between two Nodes (Papers) = **0.15**

CLUSTER 1 WITH 40 OBJECTS = {0, 2, 4, 132, 133, 135, 8, 136, 137, 138, 17, 18, 19, 20, 146, 27, 29, 31, 39, 49, 50, 52, 55, 67, 68, 70, 76, 79, 86, 87, 95, 96, 106, 116, 117, 121, 123, 124, 125, 126}

CLUSTER 2 WITH 39 OBJECTS = {128, 1, 9, 12, 14, 143, 144, 16, 21, 22, 25, 32, 36, 40, 41, 42, 44, 45, 46, 51, 54, 60, 63, 65, 69, 72, 80, 84, 89, 91, 94, 98, 100, 102, 103, 104, 109, 110, 111}

CLUSTER 3 WITH 22 OBJECTS = {129, 3, 131, 134, 13, 28, 30, 43, 47, 48, 53, 66, 71, 73, 74, 81, 83, 88, 99, 112, 113, 127}

CLUSTER 4 WITH 19 OBJECTS = {130, 5, 11, 141, 145, 147, 148, 149, 33, 56, 57, 61, 64, 75, 90, 93, 114, 115, 119}

CLUSTER 5 WITH 7 OBJECTS = {97, 37, 101, 6, 10, 140, 24}

CLUSTER 6 WITH 11 OBJECTS = {35, 7, 77, 15, 85, 118, 23, 26, 59, 92, 62}

CLUSTER 7 WITH 1 OBJECTS = {34}

CLUSTER 8 WITH 10 OBJECTS = {58, 38, 105, 139, 108, 107, 142, 82, 120, 122}

CLUSTER 9 WITH 1 OBJECTS = {78}

Initial Number of Clusters = 1

Normalized Mutual Information Value = 0.627458

9 Clusters formed using **Graph Clustering** on the AAAI Submitted Papers Dataset using Jaccard Coefficient on "Topics" as the Parameter for measuring the Similarity of Two Papers with THRESHOLD for Edge (Similarity) between two Nodes (Papers) = **0.16**

CLUSTER 1 WITH 40 OBJECTS = {0, 2, 4, 132, 133, 135, 8, 136, 137, 138, 17, 18, 19, 20, 146, 27, 29, 31, 39, 49, 50, 52, 55, 67, 68, 70, 76, 79, 86, 87, 95, 96, 106, 116, 117, 121, 123, 124, 125, 126}

CLUSTER 2 WITH 39 OBJECTS = {128, 1, 9, 12, 14, 143, 144, 16, 21, 22, 25, 32, 36, 40, 41, 42, 44, 45, 46, 51, 54, 60, 63, 65, 69, 72, 80, 84, 89, 91, 94, 98, 100, 102, 103, 104, 109, 110, 111}

CLUSTER 3 WITH 22 OBJECTS = {129, 3, 131, 134, 13, 28, 30, 43, 47, 48, 53, 66, 71, 73, 74, 81, 83, 88, 99, 112, 113, 127}

CLUSTER 4 WITH 19 OBJECTS = {130, 5, 11, 141, 145, 147, 148, 149, 33, 56, 57, 61, 64, 75, 90, 93, 114, 115, 119}

CLUSTER 5 WITH 7 OBJECTS = {97, 37, 101, 6, 10, 140, 24}

CLUSTER 6 WITH 11 OBJECTS = {35, 7, 77, 15, 85, 118, 23, 26, 59, 92, 62}

CLUSTER 7 WITH 1 OBJECTS = {34}

CLUSTER 8 WITH 10 OBJECTS = {58, 38, 105, 139, 108, 107, 142, 82, 120, 122}

CLUSTER 9 WITH 1 OBJECTS = {78}

Initial Number of Clusters = 1

Normalized Mutual Information Value = 0.627458

9 Clusters formed using **Graph Clustering** on the AAAI Submitted Papers Dataset using Jaccard Coefficient on "Topics" as the Parameter for measuring the Similarity of Two Papers with THRESHOLD for Edge (Similarity) between two Nodes (Papers) = **0.18**

CLUSTER 1 WITH 40 OBJECTS = {0, 1, 2, 4, 132, 133, 135, 8, 136, 137, 138, 17, 18, 19, 20, 146, 27, 29, 31, 39, 49, 50, 52, 55, 67, 68, 70, 76, 79, 86, 87, 95, 96, 98, 106, 116, 117, 123, 124, 126}

CLUSTER 2 WITH 16 OBJECTS = {129, 66, 99, 3, 134, 71, 73, 75, 13, 47, 113, 83, 53, 88, 30, 127}

CLUSTER 3 WITH 33 OBJECTS = {130, 5, 11, 141, 14, 145, 147, 148, 149, 22, 21, 40, 42, 45, 46, 51, 54, 56, 57, 60, 61, 64, 78, 84, 90, 93, 103, 104, 110, 114, 115, 119, 125}

CLUSTER 4 WITH 9 OBJECTS = {97, 33, 37, 101, 6, 10, 140, 24, 121}

CLUSTER 5 WITH 12 OBJECTS = {128, 35, 26, 7, 12, 15, 85, 118, 23, 122, 59, 62}

CLUSTER 6 WITH 22 OBJECTS = {9, 143, 16, 144, 25, 32, 36, 41, 43, 44, 63, 65, 69, 72, 80, 89, 91, 94, 100, 102, 109, 111}

CLUSTER 7 WITH 7 OBJECTS = {131, 74, 92, 48, 81, 112, 28}

CLUSTER 8 WITH 10 OBJECTS = {34, 38, 105, 107, 108, 139, 142, 82, 120, 58}

CLUSTER 9 WITH 1 OBJECTS = {77}

Initial Number of Clusters = 2

Normalized Mutual Information Value = 0.552410

9 Clusters formed using **Graph Clustering** on the AAAI Submitted Papers Dataset using Jaccard Coefficient on "Topics" as the Parameter for measuring the Similarity of Two Papers with THRESHOLD for Edge (Similarity) between two Nodes (Papers) = **0.20**

CLUSTER 1 WITH 46 OBJECTS = {0, 1, 2, 131, 4, 132, 133, 135, 8, 136, 137, 138, 17, 18, 19, 20, 146, 27, 28, 29, 31, 36, 39, 43, 49, 50, 55, 67, 68, 70, 74, 76, 79, 81, 86, 87, 95, 96, 98, 102, 106, 112, 116, 117, 123, 126}

CLUSTER 2 WITH 17 OBJECTS = {129, 34, 99, 3, 38, 71, 122, 105, 139, 108, 142, 82, 120, 58, 124, 125, 30}

CLUSTER 3 WITH 60 OBJECTS = {130, 5, 6, 9, 10, 11, 140, 141, 14, 143, 16, 145, 144, 147, 148, 149, 21, 22, 24, 25, 32, 33, 37, 40, 41, 42, 44, 45, 46, 51, 54, 56, 57, 60, 61, 63, 64, 65, 69, 72, 78, 80, 84, 89, 90, 91, 93, 94, 97, 100, 101, 103, 104, 109, 110, 111, 114, 115, 119, 121}

CLUSTER 4 WITH 22 OBJECTS = {134, 7, 12, 13, 15, 23, 26, 35, 47, 48, 53, 59, 62, 66, 73, 83, 85, 88, 92, 113, 118, 127}

CLUSTER 5 WITH 1 OBJECTS = {52}

CLUSTER 6 WITH 1 OBJECTS = {75}

CLUSTER 7 WITH 1 OBJECTS = {77}

CLUSTER 8 WITH 1 OBJECTS = {107}

CLUSTER 9 WITH 1 OBJECTS = {128}

Initial Number of Clusters = 6

Normalized Mutual Information Value = 0.512980

9 Clusters formed using **Graph Clustering** on the AAAI Submitted Papers Dataset using Jaccard Coefficient on "Topics" as the Parameter for measuring the Similarity of Two Papers with THRESHOLD for Edge (Similarity) between two Nodes (Papers) = **0.25**

CLUSTER 1 WITH 38 OBJECTS = {0, 1, 2, 4, 133, 132, 135, 8, 136, 137, 138, 17, 18, 19, 20, 146, 24, 29, 31, 37, 39, 49, 50, 55, 67, 68, 70, 76, 79, 86, 87, 95, 96, 106, 116, 117, 123, 126}

CLUSTER 2 WITH 3 OBJECTS = {99, 3, 30}

CLUSTER 3 WITH 53 OBJECTS = {130, 5, 6, 11, 141, 14, 143, 16, 145, 144, 147, 148, 21, 22, 25, 32, 33, 36, 40, 41, 42, 45, 46, 51, 54, 56, 57, 60, 61, 63, 64, 65, 69, 72, 78, 80, 84, 89, 90, 91, 93, 94, 97, 100, 101, 103, 104, 109, 110, 111, 114, 115, 119}

CLUSTER 4 WITH 6 OBJECTS = {35, 7, 15, 59, 92, 62}

CLUSTER 5 WITH 1 OBJECTS = {9}

CLUSTER 6 WITH 1 OBJECTS = {10}

CLUSTER 7 WITH 5 OBJECTS = {12, 85, 118, 23, 26}

CLUSTER 8 WITH 7 OBJECTS = {66, 73, 13, 47, 48, 83, 88}

CLUSTER 9 WITH 1 OBJECTS = {27}

CLUSTER 10 WITH 1 OBJECTS = {28}

CLUSTER 11 WITH 4 OBJECTS = {34, 139, 58, 82}

CLUSTER 12 WITH 6 OBJECTS = {38, 105, 142, 120, 122, 125}

CLUSTER 13 WITH 3 OBJECTS = {98, 43, 102}

CLUSTER 14 WITH 1 OBJECTS = {44}

CLUSTER 15 WITH 1 OBJECTS = {52}

CLUSTER 16 WITH 1 OBJECTS = {53}

CLUSTER 17 WITH 1 OBJECTS = {71}

CLUSTER 18 WITH 4 OBJECTS = {112, 81, 74, 131}

CLUSTER 19 WITH 1 OBJECTS = {75}

CLUSTER 20 WITH 1 OBJECTS = {77}

CLUSTER 21 WITH 1 OBJECTS = {107}

CLUSTER 22 WITH 2 OBJECTS = {124, 108}

CLUSTER 23 WITH 3 OBJECTS = {113, 134, 127}

CLUSTER 24 WITH 1 OBJECTS = {121}

CLUSTER 25 WITH 1 OBJECTS = {128}

CLUSTER 26 WITH 1 OBJECTS = {129}

CLUSTER 27 WITH 1 OBJECTS = {140}

CLUSTER 28 WITH 1 OBJECTS = {149}

Initial Number of Clusters = 28

Normalized Mutual Information Value = 0.619977

9 Clusters formed using **Graph Clustering** on the AAAI Submitted Papers Dataset using Jaccard Coefficient on "Topics" as the Parameter for measuring the Similarity of Two Papers with THRESHOLD for Edge (Similarity) between two Nodes (Papers) = **0.30**

CLUSTER 1 WITH 37 OBJECTS = {0, 1, 2, 4, 133, 132, 135, 8, 136, 137, 138, 17, 18, 19, 20, 146, 24, 29, 31, 37, 39, 49, 50, 55, 67, 68, 70, 76, 79, 86, 95, 96, 106, 116, 117, 123,126}

CLUSTER 2 WITH 3 OBJECTS = {99, 3, 30}

CLUSTER 3 WITH 52 OBJECTS = {130, 5, 6, 11, 141, 14, 143, 16, 145, 144, 147, 148, 22, 25, 32, 33, 36, 40, 41, 42, 45, 46, 51, 54, 56, 57, 60, 61, 63, 64, 65, 69, 72, 78, 80, 84, 89, 90, 91, 93, 94, 97, 100, 101, 103, 104, 109, 110, 111, 114, 115, 119}

CLUSTER 4 WITH 6 OBJECTS = {35, 7, 15, 59, 92, 62}

CLUSTER 5 WITH 1 OBJECTS = {9}

CLUSTER 6 WITH 1 OBJECTS = {10}

CLUSTER 7 WITH 5 OBJECTS = {12, 85, 118, 23, 26}

CLUSTER 8 WITH 7 OBJECTS = {66, 73, 13, 47, 48, 83, 88}

CLUSTER 9 WITH 1 OBJECTS = {21}

CLUSTER 10 WITH 1 OBJECTS = {27}

CLUSTER 11 WITH 1 OBJECTS = {28}

CLUSTER 12 WITH 4 OBJECTS = {34, 139, 58, 82}

CLUSTER 13 WITH 3 OBJECTS = {105, 142, 38}

CLUSTER 14 WITH 3 OBJECTS = {98, 43, 102}

CLUSTER 15 WITH 1 OBJECTS = {44}

CLUSTER 16 WITH 1 OBJECTS = {52}

CLUSTER 17 WITH 1 OBJECTS = {53}

CLUSTER 18 WITH 1 OBJECTS = {71}

CLUSTER 19 WITH 4 OBJECTS = {112, 81, 74, 131}

CLUSTER 20 WITH 1 OBJECTS = {75}

CLUSTER 21 WITH 1 OBJECTS = {77}

CLUSTER 22 WITH 1 OBJECTS = {87}

CLUSTER 23 WITH 1 OBJECTS = {107}

CLUSTER 24 WITH 2 OBJECTS = {124, 108}

CLUSTER 25 WITH 3 OBJECTS = {113, 134, 127}

CLUSTER 26 WITH 2 OBJECTS = {120, 122}

CLUSTER 27 WITH 1 OBJECTS = {121}

CLUSTER 28 WITH 1 OBJECTS = {125}

CLUSTER 29 WITH 1 OBJECTS = {128}

CLUSTER 30 WITH 1 OBJECTS = {129}

CLUSTER 31 WITH 1 OBJECTS = {140}

CLUSTER 32 WITH 1 OBJECTS = {149}

Initial Number of Clusters = 32

Normalized Mutual Information Value = 0.613114

Discussion on the Optimal Choice of the Value of Threshold

The threshold in Graph-Based Clustering Girvan Newman type of clustering determines which two nodes are similar enough to have a direct edge between them when we are initially making the graph.

The choice of the value of threshold is, like the value of **K** in **K-Means Clustering**, highly dependent on the need of the clustering, the practical purpose for which the clustering is being done. Here, since we need **9 Clusters** at the end of the clustering, we need to choose the threshold value in such a way that we do have exactly 9 clusters at the end of the process and the mutual information of the clustering is high as well.

Now intuitively, *as the value of threshold increases, the number of connected components (aka clusters) in the graph increases*. We need to limit our threshold to such a value that the initial graph itself should not have more than **9 clusters.**
Also, from the trend we observe from the table below, *the Mutual Information Value increases with an Increase in Threshold Value* as this ensures that only those Nodes (Papers) with a bare minimum similarity are connected to each other by a direct edge.

Threshold Value	NMI Value	#CC_Initial
0.100	0.571182	1
0.130	0.559572	1
0.140	0.559572	1
0.150	**0.627458**	**1**
0.160	**0.627458**	**1**
0.170	0.552410	2
0.180	0.552410	2
0.200	0.512980	6
0.250	**0.619977**	**28**
0.300	**0.613114**	**32**

We restrain ourselves from choosing the RED Values, as even though their NMI is high enough, the Initial Number of Connected Components is very high, way high then what we intend our clustering to have.

Hence, the **optimal choice** of the **Threshold Value** is around **0.15-0.16,** as this is the perfect trade-off between **High Normalized Mutual Information** and **Required Number of Clusters (**Connected Components).