# Machine Learning
# Assignment 3 Report

The given dataset contains 150 points, where each point is a research paper, consisting of Title, Keywords, Topics, High-level Keywords, and Abstract, as shown below.

| | Title | Keywords | Topics | High-Level Keyword(s) | Abstract |
|---|---|---|---|---|---|
| 0 | The cascade auction – a mechanism for deterrin... | Mediators\nAuctions\nCollusion\nAd Exchanges | Auctions and Market-Based Systems\nE-Commerce\... | Multiagent Systems | We introduce a sealed bid auction of a single ... |
| 1 | Basis Adaptation for Sparse Nonlinear Reinforc... | Reinforcement learning\nSparsity\nMirror desce... | Dimension Reduction/Feature Selection\nOnline ... | Machine Learning | This paper presents a new approach to basis ad... |
| 2 | Optimal Coalition Structures in Cooperative Gr... | Cooperative Game Theory\nCoalition Structure G... | Coordination and Collaboration\nGame Theory | Multiagent Systems | Representation languages for coalitional game... |
| 3 | External Memory Best-First Search for Multiple... | External-Memory Search\nParallel Search\nMulti... | Heuristic Search\nEvaluation and Analysis (Sea... | Heuristic Search and Optimization | Multiple sequence alignment (MSA) is a central... |
| 4 | Posted Prices Exchange for Display Advertising... | Display Advertising\nDynamic Pricing\nMarket E... | Auctions and Market-Based Systems\nE-Commerce\... | Multiagent Systems | We propose a new market design for display adv... |

## Part - 1

1.) Using Bottom-up clustering with complete linkage:

Using this strategy, the clusters obtained are:

cluster 0 size = 9
[0, 4, 18, 76, 19, 146, 8, 79, 52]

cluster 1 size = 7
[1, 106, 137, 138, 117, 132, 133]

cluster 2 size = 10
[2, 96, 20, 87, 29, 95, 136, 67, 86, 123]

cluster 3 size = 5
[3, 30, 99, 71, 129]

cluster 4 size = 3
[5, 56, 33]

cluster 5 size = 2
[6, 101]

cluster 6 size = 4
[7, 35, 15, 59]

cluster 7 size = 3
[9, 114, 119]

cluster 8 size = 107
[10, 11, 93, 149, 12, 26, 85, 13, 48, 113, 134, 127, 14, 40, 16, 80, 97, 65, 17, 49, 135, 55, 31, 68, 21, 32, 42, 103, 54, 22, 110, 46, 60, 84, 104, 23, 118, 24, 37, 121, 25, 41, 89, 111, 63, 109, 94, 27, 36, 28, 74, 81, 131, 112, 34, 107, 38, 125, 43, 102, 98, 39, 50, 70, 116, 44, 140, 45, 128, 47, 83, 53, 51, 69, 72, 57, 78, 58, 82, 139, 105, 142, 120, 122, 61, 64, 147, 148, 115, 141, 62, 92, 66, 88, 73, 75, 77, 90, 145, 130, 91, 100, 144, 108, 124, 126, 143]

2.) Using Bottom-up clustering with single linkage

The clusters obtained here are:

cluster 0 size = 39
[0, 4, 18, 76, 2, 96, 8, 19, 146, 55, 29, 95, 136, 67, 86, 123, 20, 17, 49, 135, 31, 68, 70, 116, 24, 39, 50, 37, 1, 106, 117, 132, 133, 137, 138, 126, 79, 87, 121]

cluster 1 size = 100
[3, 30, 99, 5, 56, 33, 6, 101, 11, 14, 64, 90, 145, 130, 115, 141, 147, 148, 40, 93, 114, 119, 61, 57, 78, 32, 42, 103, 54, 16, 25, 41, 91, 100, 89, 80, 144, 111, 97, 36, 51, 63, 109, 65, 94, 72, 69, 22, 110, 46, 60, 84, 104, 143, 45, 21, 149, 140, 44, 27, 9, 7, 35, 15, 62, 92, 59, 12, 23, 118, 85, 26, 13, 48, 47, 83, 66, 88, 73, 113, 134, 127, 53, 34, 58, 82, 139, 108, 124, 38, 105, 142, 125, 120, 122, 43, 102, 98, 71, 129]

cluster 2 size = 1
[10]

cluster 3 size = 5
[28, 74, 81, 131, 112]

cluster 4 size = 1
[52]

cluster 5 size = 1
[75]

cluster 6 size = 1
[77]

cluster 7 size = 1
[107]

cluster 8 size = 1
[128]

# Part - 2

Clustering the given data points using Girvan Newmann Algorithm.

For this algorithm, we need to also decide upon a threshold value for initialising the cluster graph. For this we start with a small threshold of 0.1, and keep increasing it. The threshold values that I tried are: [0.1,0.11,0.12,0.13,0.14,0.15,0.16,0.17,0.18,0.19,0.2,0.25, 0.3]
The clusters obtained are:

----------Threshold = 0.1----------
Cluster 0 size = 40
[0, 2, 132, 4, 133, 135, 8, 136, 137, 138, 17, 18, 19, 20, 146, 24, 27, 29, 31, 39, 49, 50, 52, 55, 67, 68, 70, 76, 79, 86, 87, 95, 96, 106, 116, 117, 121, 123, 124, 126]
Cluster 1 size = 59
[128, 1, 130, 5, 9, 11, 12, 141, 14, 143, 145, 147, 148, 21, 149, 22, 25, 32, 33, 36, 40, 41, 42, 43, 44, 45, 46, 51, 54, 56, 57, 60, 61, 63, 64, 65, 69, 72, 75, 77, 80, 84, 89, 90, 91, 93, 94, 98, 99, 102, 103, 104, 109, 110, 111, 114, 115, 119, 125]
Cluster 2 size = 16
[129, 66, 3, 134, 71, 73, 13, 47, 48, 113, 83, 53, 88, 26, 30, 127]
Cluster 3 size = 6
[97, 101, 37, 6, 10, 140]
Cluster 4 size = 20
[34, 35, 38, 7, 122, 105, 107, 108, 139, 142, 15, 82, 85, 118, 23, 120, 58, 59, 92, 62]
Cluster 5 size = 1
[16]
Cluster 6 size = 5
[131, 74, 112, 81, 28]
Cluster 7 size = 1
[78]
Cluster 8 size = 2
[144, 100]

----------Threshold = 0.11----------
Cluster 0 size = 42
[0, 2, 132, 4, 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 1 size = 56
[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 2 size = 17
[129, 66, 3, 99, 134, 71, 73, 13, 47, 48, 113, 83, 53, 88, 26, 30, 127]
Cluster 3 size = 4
[16, 97, 101, 6]
Cluster 4 size = 21
[7, 139, 142, 15, 23, 34, 35, 38, 58, 59, 62, 77, 82, 85, 92, 105, 107, 108, 118, 120, 122]
Cluster 5 size = 5
[131, 74, 112, 81, 28]
Cluster 6 size = 2
[140, 44]

Cluster 7 size = 1
[78]
Cluster 8 size = 2
[144, 100]

----------Threshold = 0.12----------
Cluster 0 size = 40
[0, 2, 132, 4, 133, 135, 8, 136, 137, 138, 17, 18, 19, 20, 146, 24, 27, 29, 31, 39, 49, 50, 52, 55, 67, 68, 70, 76, 79, 86, 87, 95, 96, 106, 116, 117, 121, 123, 124, 126]
Cluster 1 size = 34
[128, 1, 9, 11, 12, 143, 16, 149, 21, 25, 32, 36, 40, 41, 42, 43, 44, 51, 54, 63, 65, 69, 72, 80, 89, 91, 94, 98, 102, 103, 109, 111, 114, 125]
Cluster 2 size = 18
[129, 66, 3, 99, 134, 71, 73, 13, 47, 48, 113, 83, 53, 88, 26, 92, 30, 127]
Cluster 3 size = 24
[130, 5, 141, 14, 145, 147, 148, 22, 33, 45, 46, 56, 57, 60, 61, 64, 75, 84, 90, 93, 104, 110, 115, 119]
Cluster 4 size = 6
[97, 101, 37, 6, 10, 140]
Cluster 5 size = 20
[34, 35, 38, 7, 122, 105, 107, 108, 139, 142, 15, 77, 82, 85, 118, 23, 120, 58, 59, 62]
Cluster 6 size = 5
[131, 74, 112, 81, 28]
Cluster 7 size = 1
[78]
Cluster 8 size = 2
[144, 100]

----------Threshold = 0.13----------
Cluster 0 size = 42
[0, 2, 132, 4, 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 1 size = 57
[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 2 size = 18
[129, 66, 3, 99, 134, 71, 73, 13, 47, 48, 113, 83, 53, 88, 26, 92, 30, 127]
Cluster 3 size = 3
[97, 101, 6]
Cluster 4 size = 20
[34, 35, 38, 7, 122, 105, 139, 108, 107, 142, 15, 77, 82, 85, 118, 23, 120, 58, 59, 62]
Cluster 5 size = 5
[131, 74, 112, 81, 28]
Cluster 6 size = 2
[140, 44]
Cluster 7 size = 1

[78]
Cluster 8 size = 2
[144, 100]

----------Threshold = 0.14----------
Cluster 0 size = 42
[0, 2, 132, 4, 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 1 size = 57
[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 2 size = 18
[129, 66, 3, 99, 134, 71, 73, 13, 47, 48, 113, 83, 53, 88, 26, 92, 30, 127]
Cluster 3 size = 3
[97, 101, 6]
Cluster 4 size = 20
[34, 35, 38, 7, 122, 105, 139, 108, 107, 142, 15, 77, 82, 85, 118, 23, 120, 58, 59, 62]
Cluster 5 size = 5
[131, 74, 112, 81, 28]
Cluster 6 size = 2
[140, 44]
Cluster 7 size = 1
[78]
Cluster 8 size = 2
[144, 100]

----------Threshold = 0.15----------
Cluster 0 size = 40
[0, 2, 132, 4, 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 1 size = 39
[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 2 size = 22
[129, 3, 131, 134, 13, 28, 30, 43, 47, 48, 53, 66, 71, 73, 74, 81, 83, 88, 99, 112, 113, 127]
Cluster 3 size = 19
[64, 33, 130, 5, 11, 75, 141, 145, 114, 147, 148, 149, 115, 119, 56, 57, 90, 93, 61]
Cluster 4 size = 7
[97, 101, 37, 6, 10, 140, 24]
Cluster 5 size = 11
[35, 7, 77, 15, 85, 118, 23, 26, 59, 92, 62]
Cluster 6 size = 1
[34]
Cluster 7 size = 10
[38, 122, 105, 139, 108, 107, 142, 82, 120, 58]

Cluster 8 size = 1
[78]

----------Threshold = 0.16----------
Cluster 0 size = 40
[0, 2, 132, 4, 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 1 size = 39
[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 2 size = 22
[129, 3, 131, 134, 13, 28, 30, 43, 47, 48, 53, 66, 71, 73, 74, 81, 83, 88, 99, 112, 113, 127]
Cluster 3 size = 19
[64, 33, 130, 5, 11, 75, 141, 145, 114, 147, 148, 149, 115, 119, 56, 57, 90, 93, 61]
Cluster 4 size = 7
[97, 101, 37, 6, 10, 140, 24]
Cluster 5 size = 11
[35, 7, 77, 15, 85, 118, 23, 26, 59, 92, 62]
Cluster 6 size = 1
[34]
Cluster 7 size = 10
[38, 122, 105, 139, 108, 107, 142, 82, 120, 58]
Cluster 8 size = 1
[78]

----------Threshold = 0.17----------
Cluster 0 size = 40
[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 1 size = 16
[129, 66, 3, 99, 134, 71, 73, 75, 13, 47, 113, 83, 53, 88, 30, 127]
Cluster 2 size = 33
[130, 5, 11, 141, 14, 145, 147, 148, 21, 149, 22, 40, 42, 45, 46, 51, 54, 56, 57, 60, 61, 64, 78, 84, 90, 93, 103, 104, 110, 114, 115, 119, 125]
Cluster 3 size = 9
[97, 33, 101, 37, 6, 10, 140, 24, 121]
Cluster 4 size = 12
[128, 35, 122, 7, 12, 15, 85, 118, 23, 26, 59, 62]
Cluster 5 size = 22
[9, 143, 16, 144, 25, 32, 36, 41, 43, 44, 63, 65, 69, 72, 80, 89, 91, 94, 100, 102, 109, 111]
Cluster 6 size = 7
[131, 74, 28, 48, 81, 112, 92]
Cluster 7 size = 10
[34, 38, 105, 107, 108, 139, 142, 82, 120, 58]
Cluster 8 size = 1
[77]

----------Threshold = 0.18----------
Cluster 0 size = 40
[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 1 size = 16
[129, 66, 3, 99, 134, 71, 73, 75, 13, 47, 113, 83, 53, 88, 30, 127]
Cluster 2 size = 33
[130, 5, 11, 141, 14, 145, 147, 148, 21, 149, 22, 40, 42, 45, 46, 51, 54, 56, 57, 60, 61, 64, 78, 84, 90, 93, 103, 104, 110, 114, 115, 119, 125]
Cluster 3 size = 9
[97, 33, 101, 37, 6, 10, 140, 24, 121]
Cluster 4 size = 12
[128, 35, 122, 7, 12, 15, 85, 118, 23, 26, 59, 62]
Cluster 5 size = 22
[9, 143, 16, 144, 25, 32, 36, 41, 43, 44, 63, 65, 69, 72, 80, 89, 91, 94, 100, 102, 109, 111]
Cluster 6 size = 7
[131, 74, 28, 48, 81, 112, 92]
Cluster 7 size = 10
[34, 38, 105, 107, 108, 139, 142, 82, 120, 58]
Cluster 8 size = 1
[77]

----------Threshold = 0.19----------
Cluster 0 size = 40
[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 1 size = 16
[129, 66, 3, 99, 134, 71, 73, 75, 13, 47, 113, 83, 53, 88, 30, 127]
Cluster 2 size = 33
[130, 5, 11, 141, 14, 145, 147, 148, 21, 149, 22, 40, 42, 45, 46, 51, 54, 56, 57, 60, 61, 64, 78, 84, 90, 93, 103, 104, 110, 114, 115, 119, 125]
Cluster 3 size = 9
[97, 33, 101, 37, 6, 10, 140, 24, 121]
Cluster 4 size = 12
[128, 35, 122, 7, 12, 15, 85, 118, 23, 26, 59, 62]
Cluster 5 size = 22
[9, 143, 16, 144, 25, 32, 36, 41, 43, 44, 63, 65, 69, 72, 80, 89, 91, 94, 100, 102, 109, 111]
Cluster 6 size = 7
[131, 74, 28, 48, 81, 112, 92]
Cluster 7 size = 10
[34, 38, 105, 107, 108, 139, 142, 82, 120, 58]
Cluster 8 size = 1
[77]

----------Threshold = 0.2----------

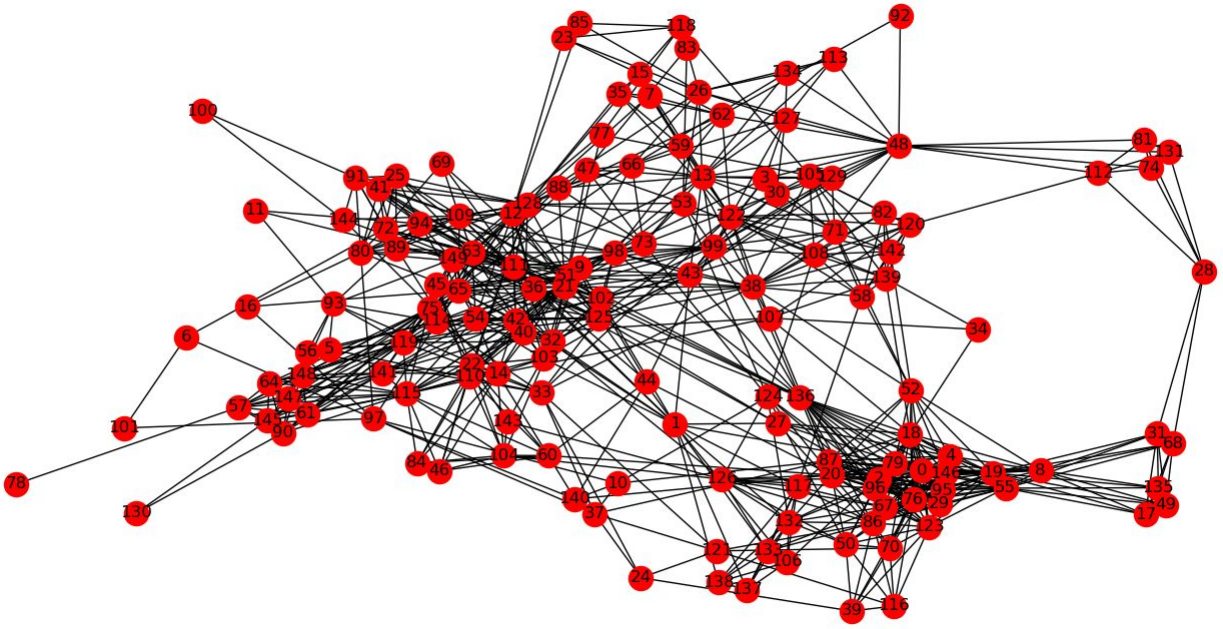
Cluster 0 size = 40
[0, 1, 2, 4, 132, 133, 135, 8, 136, 137, 138, 17, 18, 19, 20, 146, 27, 29, 31, 39, 49, 50, 52, 55, 67, 68, 70, 76, 79, 86, 87, 95, 96, 98, 106, 116, 117, 123, 124, 126]
Cluster 1 size = 16
[129, 66, 3, 99, 134, 71, 73, 75, 13, 47, 113, 83, 53, 88, 30, 127]
Cluster 2 size = 33
[130, 5, 11, 141, 14, 145, 147, 148, 21, 149, 22, 40, 42, 45, 46, 51, 54, 56, 57, 60, 61, 64, 78, 84, 90, 93, 103, 104, 110, 114, 115, 119, 125]
Cluster 3 size = 9
[97, 33, 101, 37, 6, 10, 140, 24, 121]
Cluster 4 size = 12
[128, 35, 122, 7, 12, 15, 85, 118, 23, 26, 59, 62]
Cluster 5 size = 22
[9, 143, 16, 144, 25, 32, 36, 41, 43, 44, 63, 65, 69, 72, 80, 89, 91, 94, 100, 102, 109, 111]
Cluster 6 size = 7
[131, 74, 28, 48, 81, 112, 92]
Cluster 7 size = 10
[34, 38, 105, 107, 108, 139, 142, 82, 120, 58]
Cluster 8 size = 1
[77]

----------Threshold = 0.25----------
Cluster 0 size = 46
[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 1 size = 17
[129, 34, 3, 99, 58, 38, 71, 105, 139, 108, 142, 82, 120, 122, 124, 125, 30]
Cluster 2 size = 60
[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 3 size = 22
[134, 7, 12, 13, 15, 23, 26, 35, 47, 48, 53, 59, 62, 66, 73, 83, 85, 88, 92, 113, 118, 127]
Cluster 4 size = 1
[52]
Cluster 5 size = 1
[75]
Cluster 6 size = 1
[77]
Cluster 7 size = 1
[107]
Cluster 8 size = 1
[128]

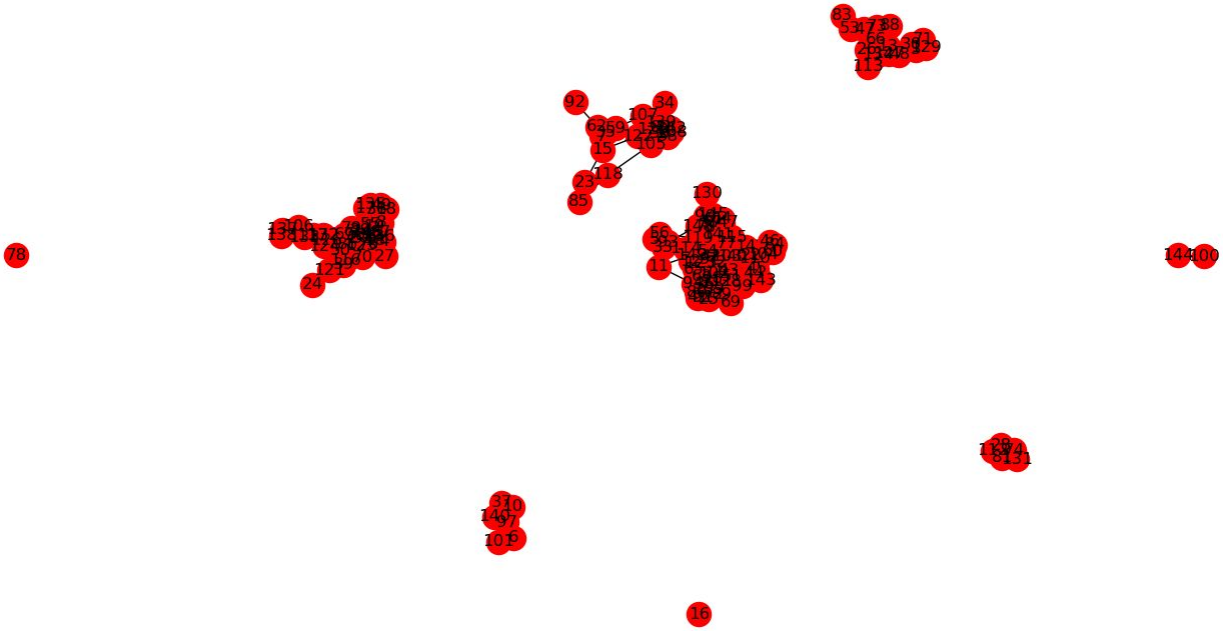
For threshold values of 0.3 and beyond, the initially obtained graph itself consists of more than 9 disconnected clusters. Hence obtaining nine clusters out of this graph is not possible.

The graphs obtained through the Girvan Newmann algorithm for some of the threshold values are:

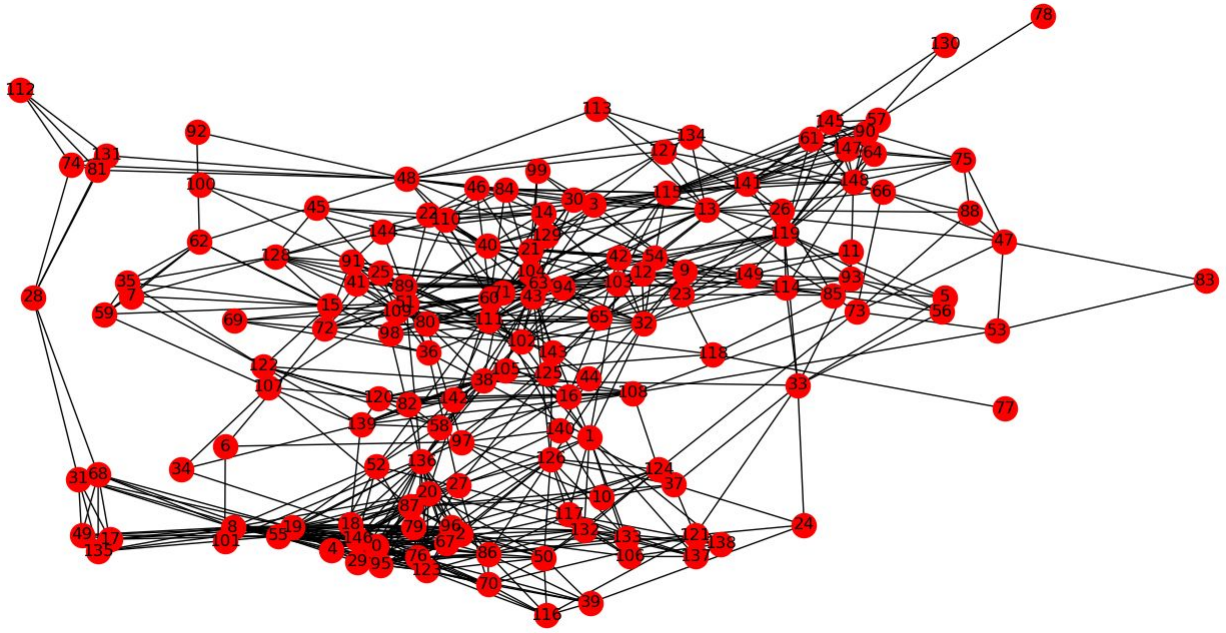
----------Threshold = 0.1---------- Initial



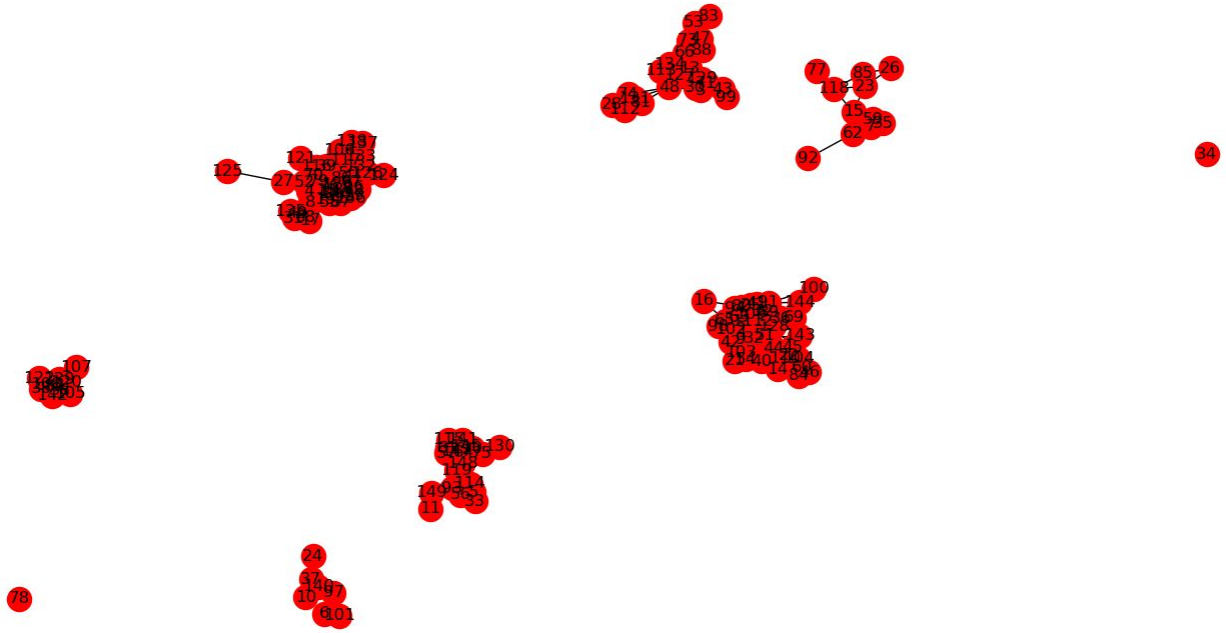
----------Threshold = 0.1---------- Final



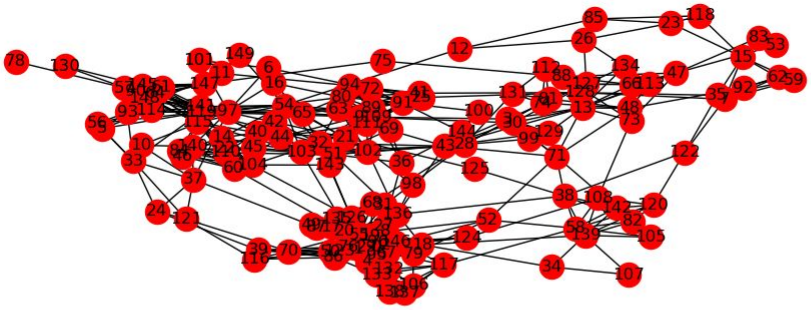
----------Threshold = 0.15---------- Initial



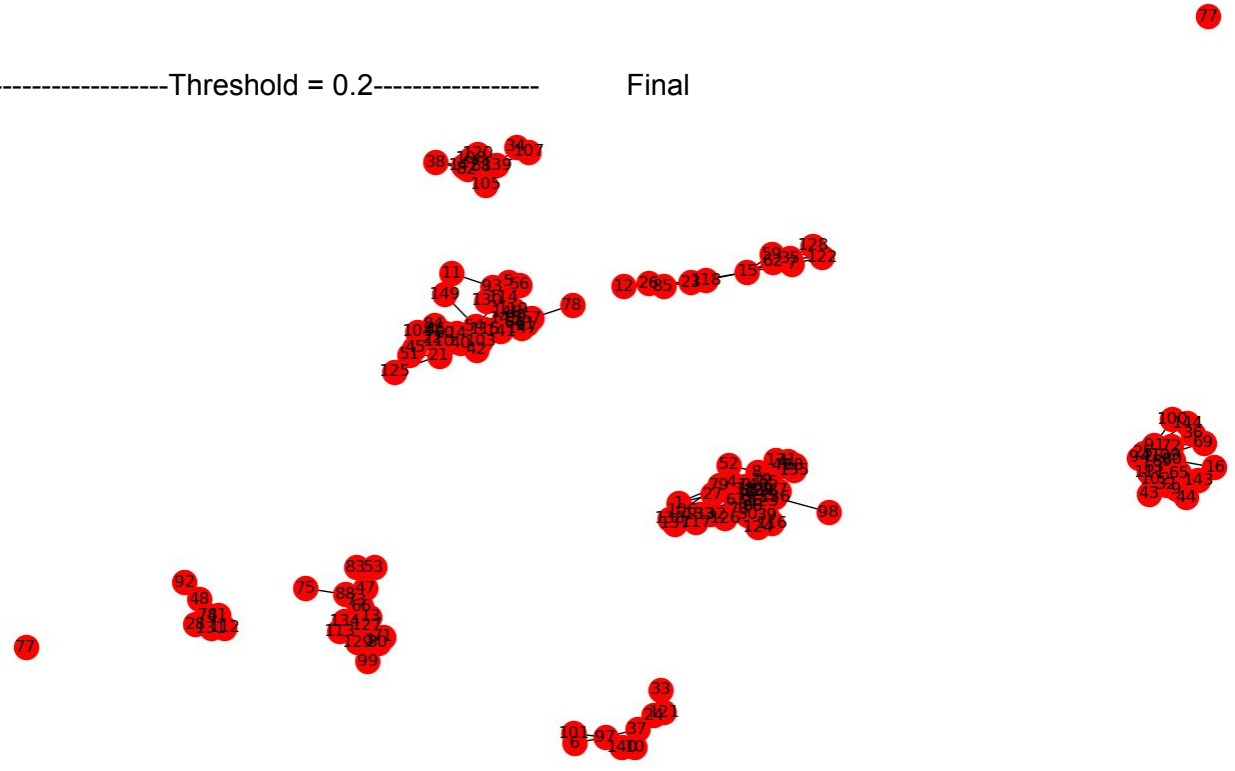
----------Threshold = 0.15----------  Final



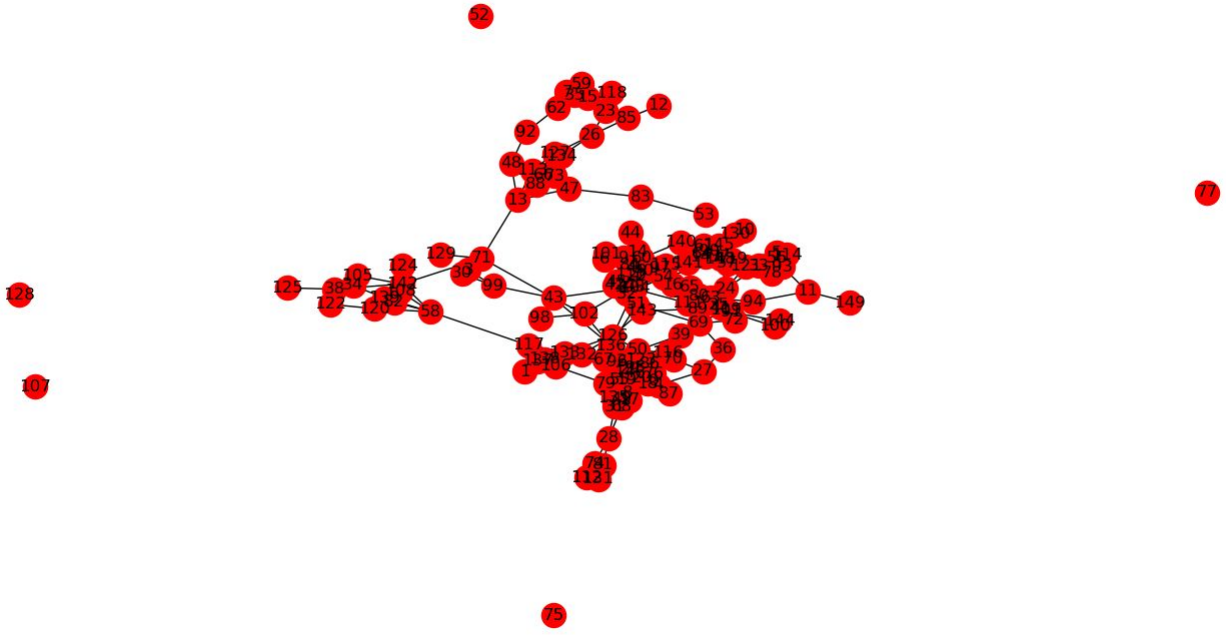
----------Threshold = 0.2----------  Initial



----------Threshold = 0.2----------  Final

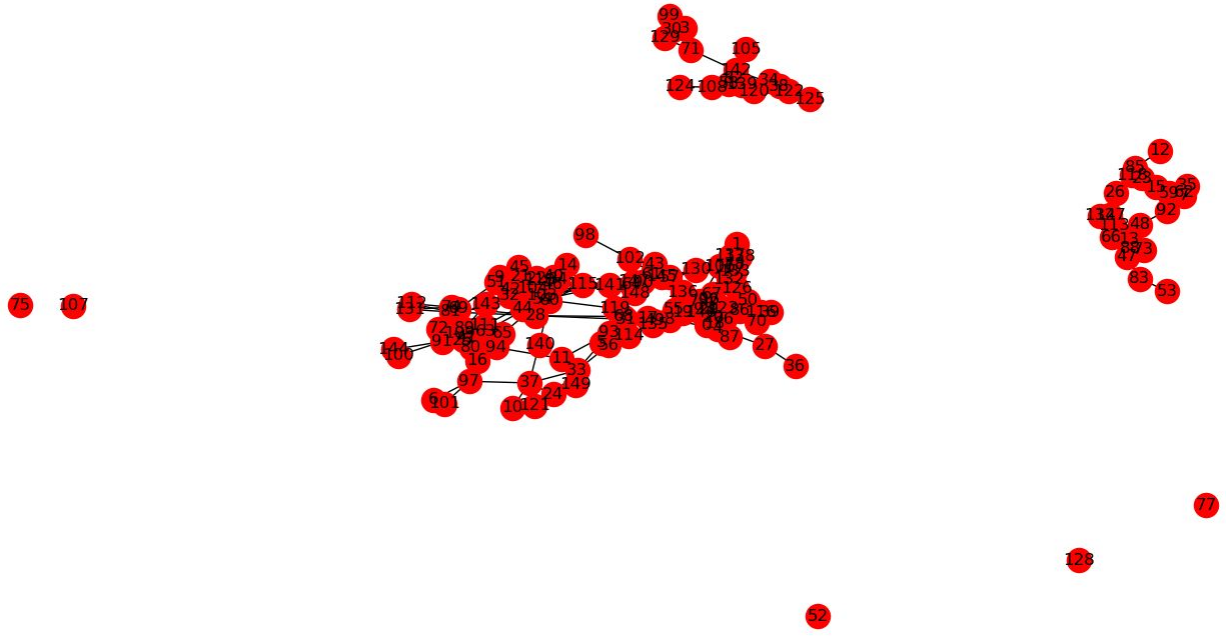


----------Threshold = 0.25----------  Initial



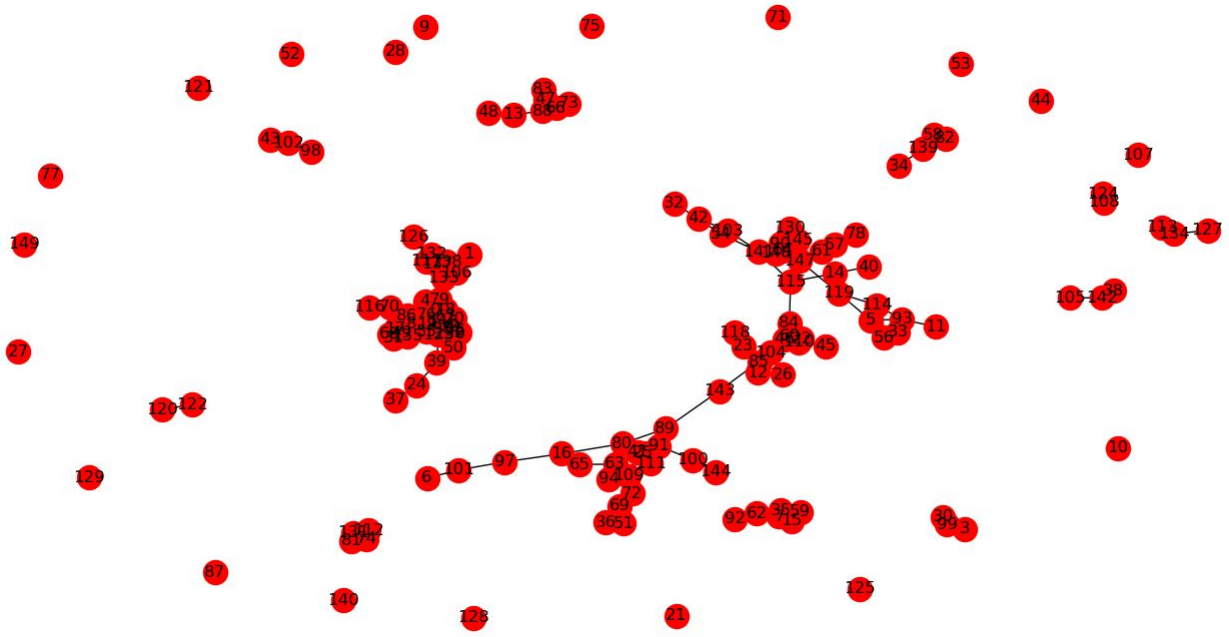
----------Threshold = 0.25----------

After



----------Threshold = 0.3----------

Initial ( as well as Final)



As we can see, the number of disjoint components for threshold 0.3 is more than 9, hence no clustering is possible here.

## Part 3

Using the Normalised Mutual Information metric to evaluate the clusters

The obtained values of the NMI scores of each cluster are:

| | |
|---|---|
| 1. Complete linkage clustering | 0.31934304860957213 |
| 2. Single linkage clustering | 0.3548502263990318 |
| 3. Girvan Newmann clustering with threshold 0.1 | 0.5308892970827201 |
| 4. Girvan Newmann clustering with threshold 0.11 | 0.5711823888460205 |
| 5. Girvan Newmann clustering with threshold 0.12 | 0.5599062753054901 |
| 6. Girvan Newmann clustering with threshold 0.13 | 0.5595718029698289 |
| 7. Girvan Newmann clustering with threshold 0.14 | 0.5595718029698289 |
| 8. Girvan Newmann clustering with threshold 0.15 | 0.6274579407960716 |
| 9. Girvan Newmann clustering with threshold 0.16 | 0.6274579407960716 |
| 10. Girvan Newmann clustering with threshold 0.17 | 0.5524100134489417 |
| 11. Girvan Newmann clustering with threshold 0.18 | 0.5524100134489417 |
| 12. Girvan Newmann clustering with threshold 0.19 | 0.5524100134489417 |
| 13. Girvan Newmann clustering with threshold 0.2 | 0.5524100134489417 |
| 14. Girvan Newmann clustering with threshold 0.25 | 0.5129800344956054 |

As we can see, the Girvan Newmann Algorithm gives better results as compared to the Bottom-up clustering algorithms. Also, based on the obtained NMI values, we can say that the optimum value of the threshold is around 0.15, which gives an NMI of 0.6274579407960716.