**Question1:**

Dimensions of the TF-IDF matrix: (4732, 16080)

**Question2:**

Confusion matrix:

[[ 72 2271]

[2305 84]]

By inspecting the contingency table, we found that the clustering result was actually pretty good, most of the points are clustered correctly.

5 measures for the K-means:

Homogeneity Score: 0.791

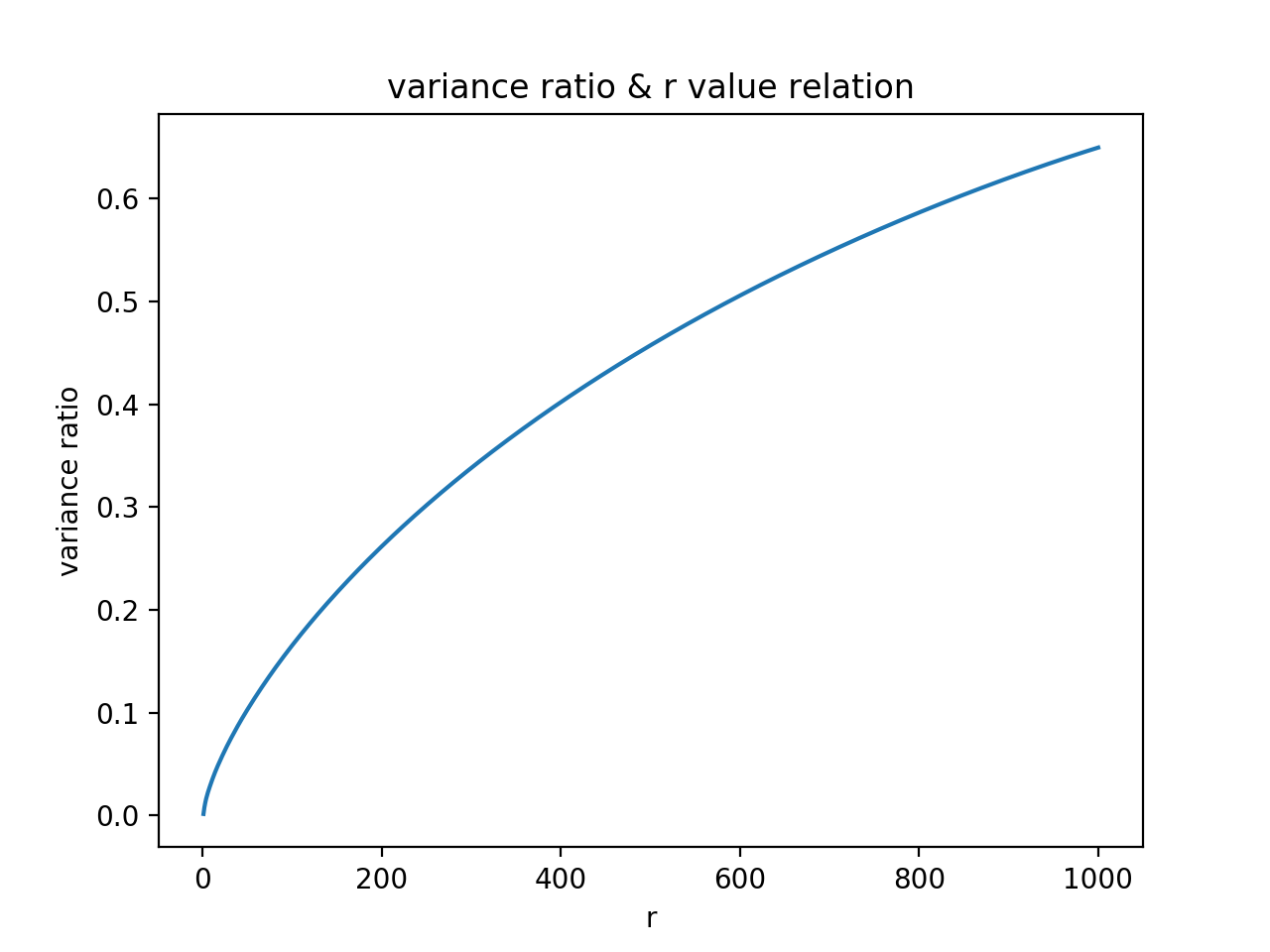
Completeness Score: 0.791

V-measure: 0.791

Adjusted Rand Score: 0.872

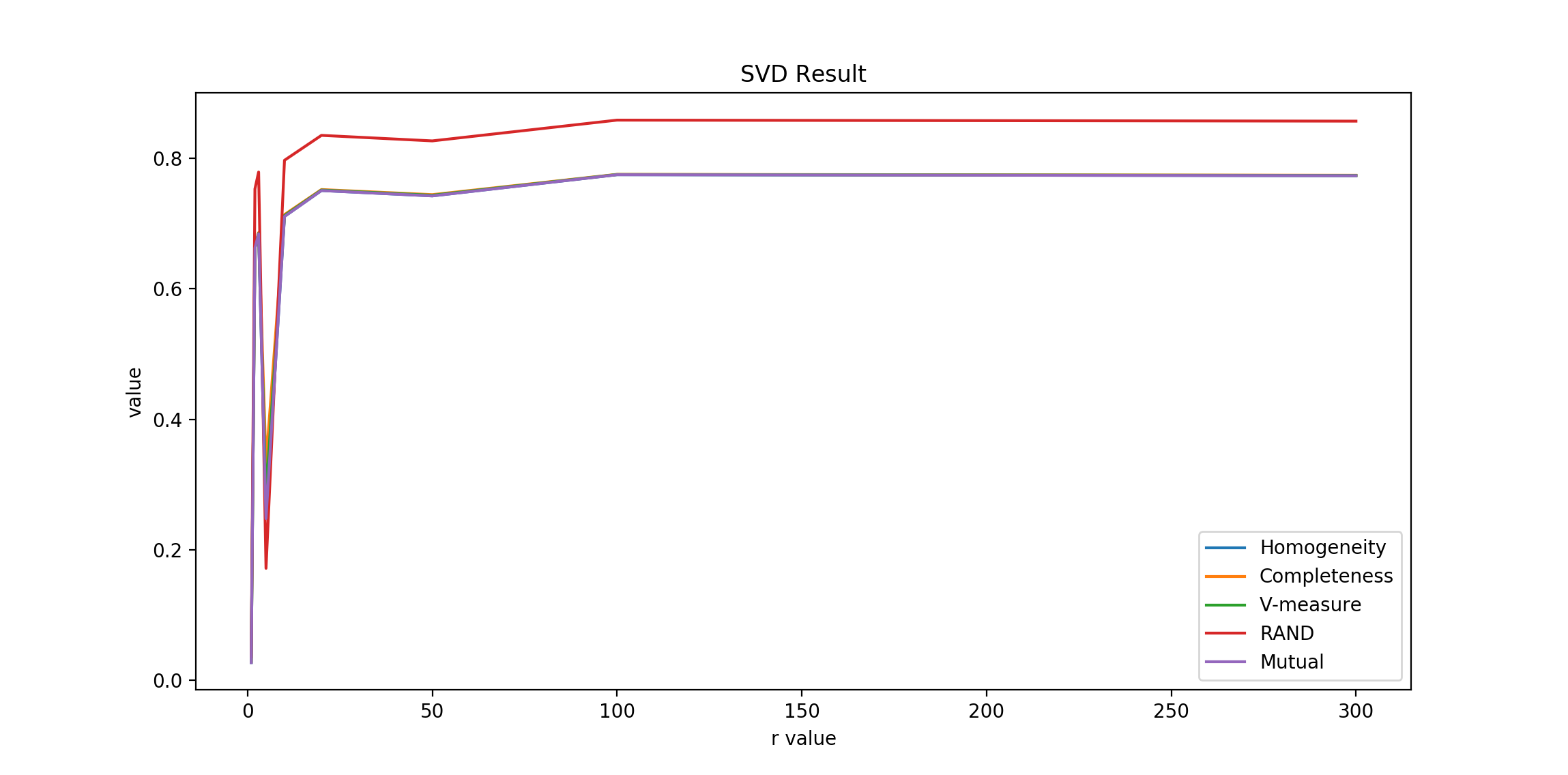
Adjusted Mutual Info Score: 0.791

**Question3:**

****

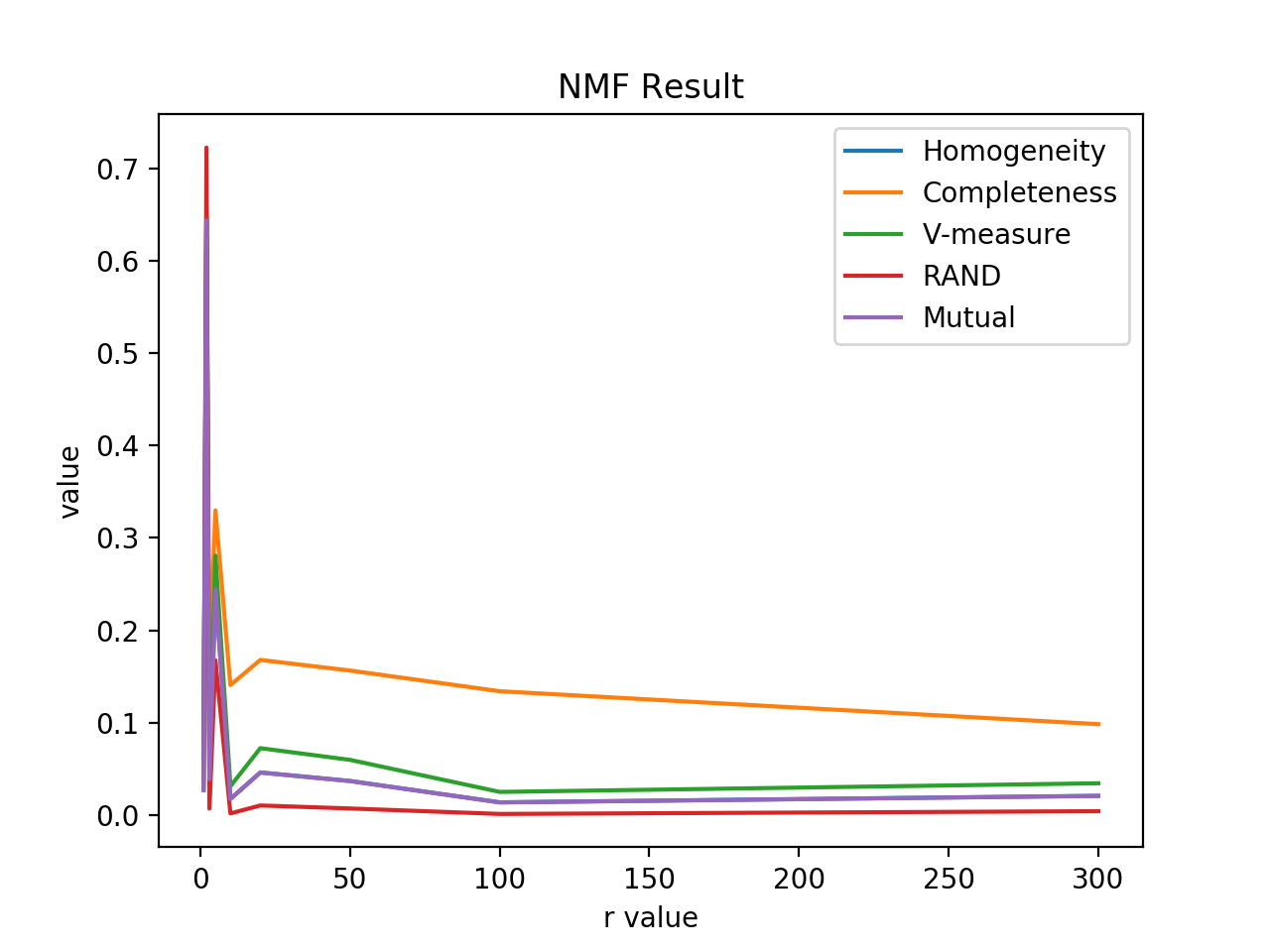
SVD

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Homogeneity | Completeness | V-measure | Rand Score | Mutual Info |
| r=1 | 0.791 | 0.791 | 0.791 | 0.872 | 0.791 |
| r=2 | 0.027 | 0.028 | 0.028 | 0.036 | 0.027 |
| r=3 | 0.644 | 0.651 | 0.647 | 0.722 | 0.644 |
| r=5 | 0.039 | 0.169 | 0.063 | 0.007 | 0.039 |
| r=10 | 0.244 | 0.330 | 0.281 | 0.168 | 0.244 |
| r=20 | 0.750 | 0.752 | 0.751 | 0.835 | 0.750 |
| r=50 | 0.742 | 0.744 | 0.743 | 0.826 | 0.742 |
| r=100 | 0.775 | 0.775 | 0.775 | 0.858 | 0.775 |
| r=300 | 0.773 | 0.774 | 0.773 | 0.857 | 0.773 |



NMF

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Homogeneity | Completeness | V-measure | Rand Score | Mutual Info |
| r=1 | 0.027 | 0.029 | 0.028 | 0.036 | 0.027 |
| r=2 | 0.662 | 0.666 | 0.664 | 0.753 | 0.662 |
| r=3 | 0.684 | 0.686 | 0.685 | 0.779 | 0.684 |
| r=5 | 0.248 | 0.332 | 0.284 | 0.172 | 0.248 |
| r=10 | 0.711 | 0.714 | 0.712 | 0.797 | 0.711 |
| r=20 | 0.046 | 0.168 | 0.072 | 0.010 | 0.046 |
| r=50 | 0.037 | 0.156 | 0.060 | 0.007 | 0.037 |
| r=100 | 0.014 | 0.134 | 0.025 | 0.001 | 0.014 |
| r=300 | 0.021 | 0.098 | 0.034 | 0.004 | 0.021 |



for the result above we can see that:

Best r for SVD: 100

Best r for NMF: 2

The non-monotomic behavior of the measures as r increases is because as r increases the information contained is more intact, which means that more information and noises will be included. This also means the data is less compressed. There should be a balance between information and noise, which will be our best r.

Confusion Matrix:

svd with r=1

[[1114 1229]

[ 685 1704]]

nmf with r=1

[[1232 1111]

[1705 684]]

svd with r=2

[[2095 248]

[ 65 2324]]

nmf with r=2

[[ 313 2030]

[2347 42]]

svd with r=3

[[ 200 2143]

[2311 78]]

nmf with r=3

[[2166 177]

[2389 0]]

svd with r=5

[[2342 1]

[1384 1005]]

nmf with r=5

[[ 1 2342]

[ 993 1396]]

svd with r=10

[[2139 204]

[ 50 2339]]

nmf with r=10

[[2262 81]

[2389 0]]

svd with r=20

[[2190 153]

[ 51 2338]]

nmf with r=20

[[2121 222]

[2387 2]]

svd with r=50

[[2176 167]

[ 48 2341]]

nmf with r=50

[[2162 181]

[2387 2]]

svd with r=100

[[ 113 2230]

[2328 61]]

nmf with r=100

[[2279 64]

[2389 0]]

svd with r=300

[[2226 117]

[ 59 2330]]

nmf with r=300

[[ 146 2197]

[ 13 2376]]

**Question4**:

==================================================

normal svd

==================================================

Confusion Matrix:

[[1396 947]

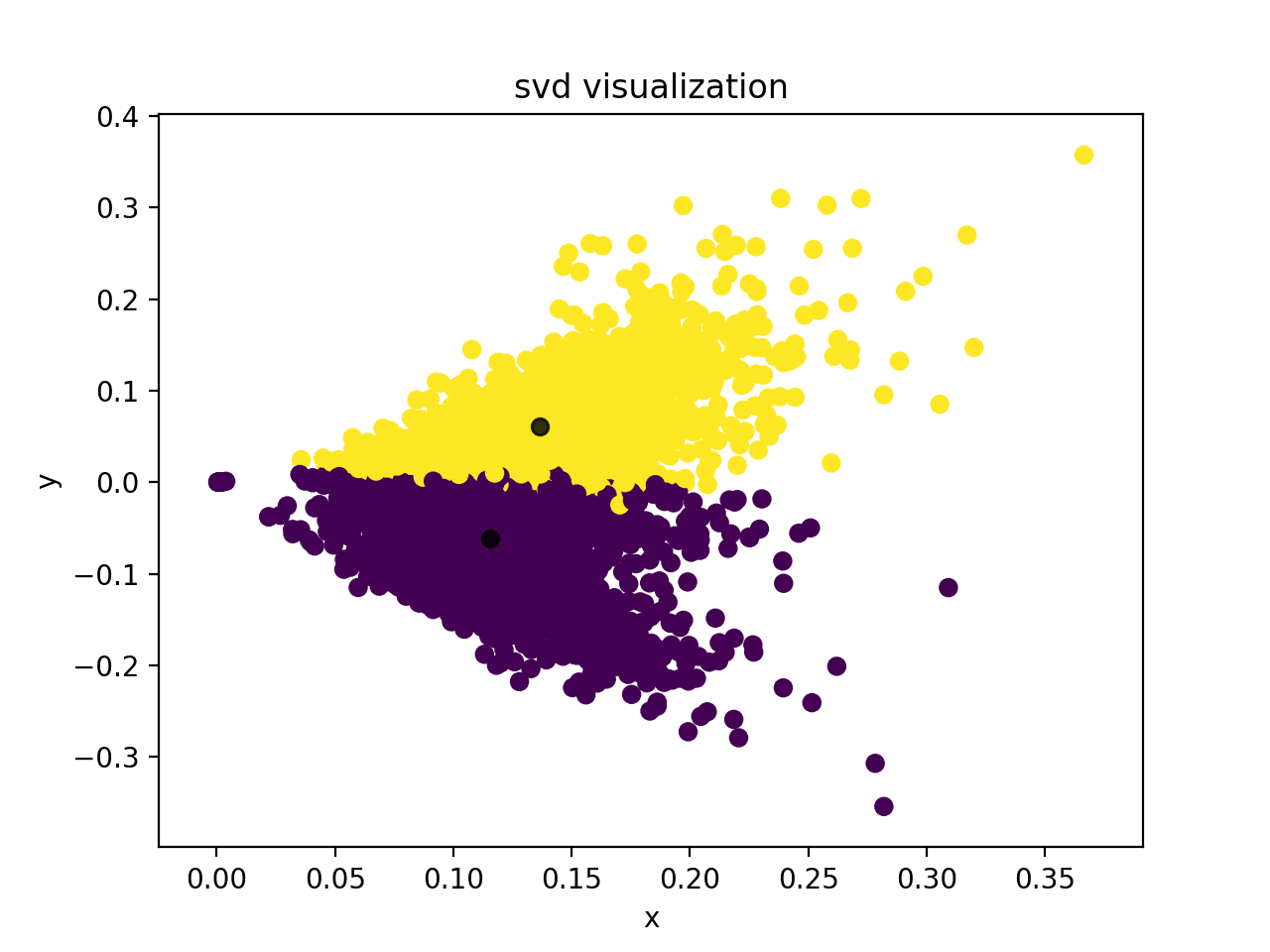
[ 119 2270]]

Homogeneity Score: 0.278

Completeness Score: 0.308

V-measure: 0.292

Adjusted Rand Score: 0.302

Adjusted Mutual Info Score: 0.278

==================================================

normal nmf

==================================================

Confusion Matrix:

[[ 139 2204]

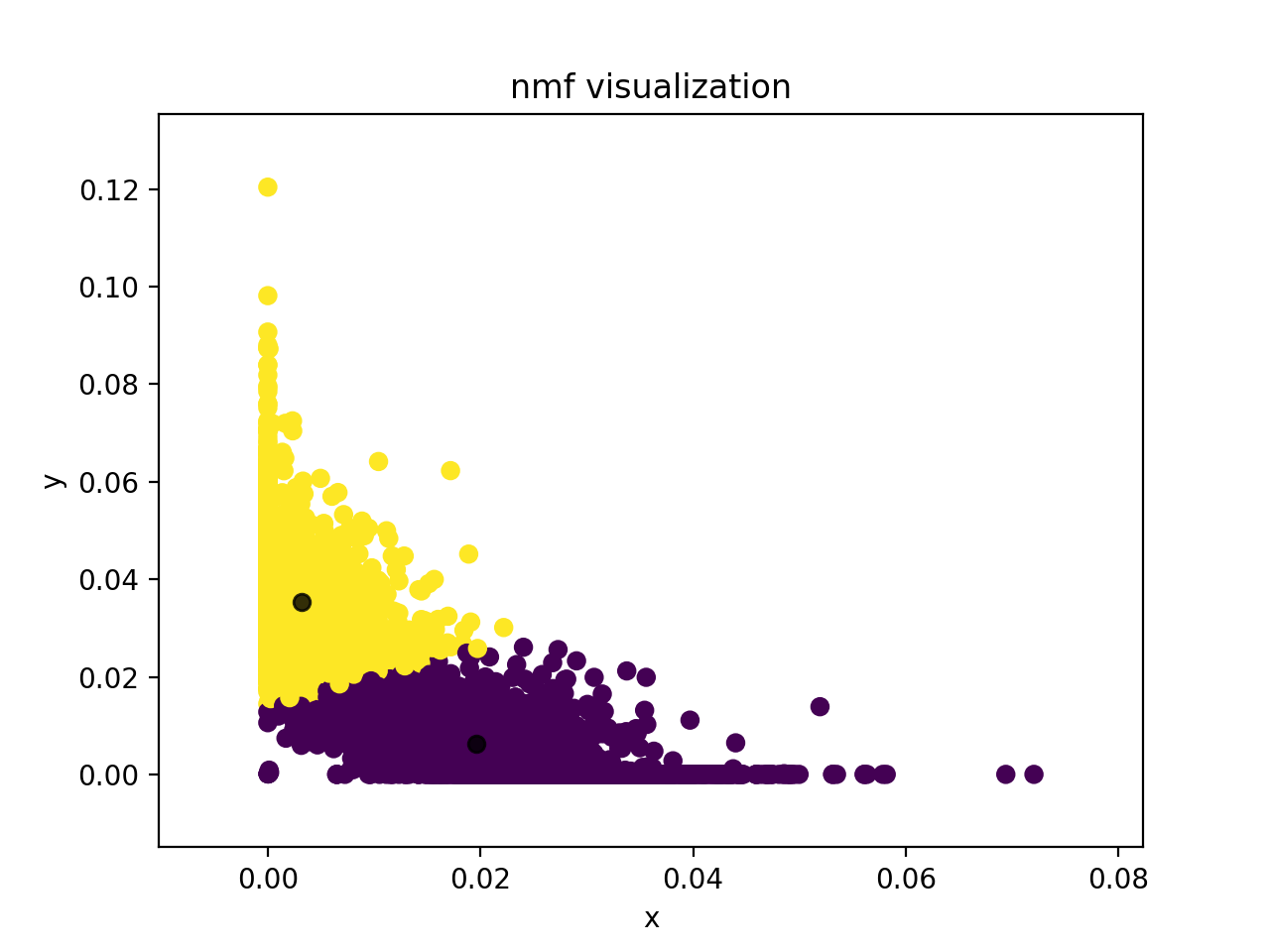
[2296 93]]

Homogeneity Score: 0.719

Completeness Score: 0.719

V-measure: 0.719

Adjusted Rand Score: 0.813

Adjusted Mutual Info Score: 0.719

==================================================

normalized svd

==================================================

Confusion Matrix:

[[1396 947]

[ 119 2270]]

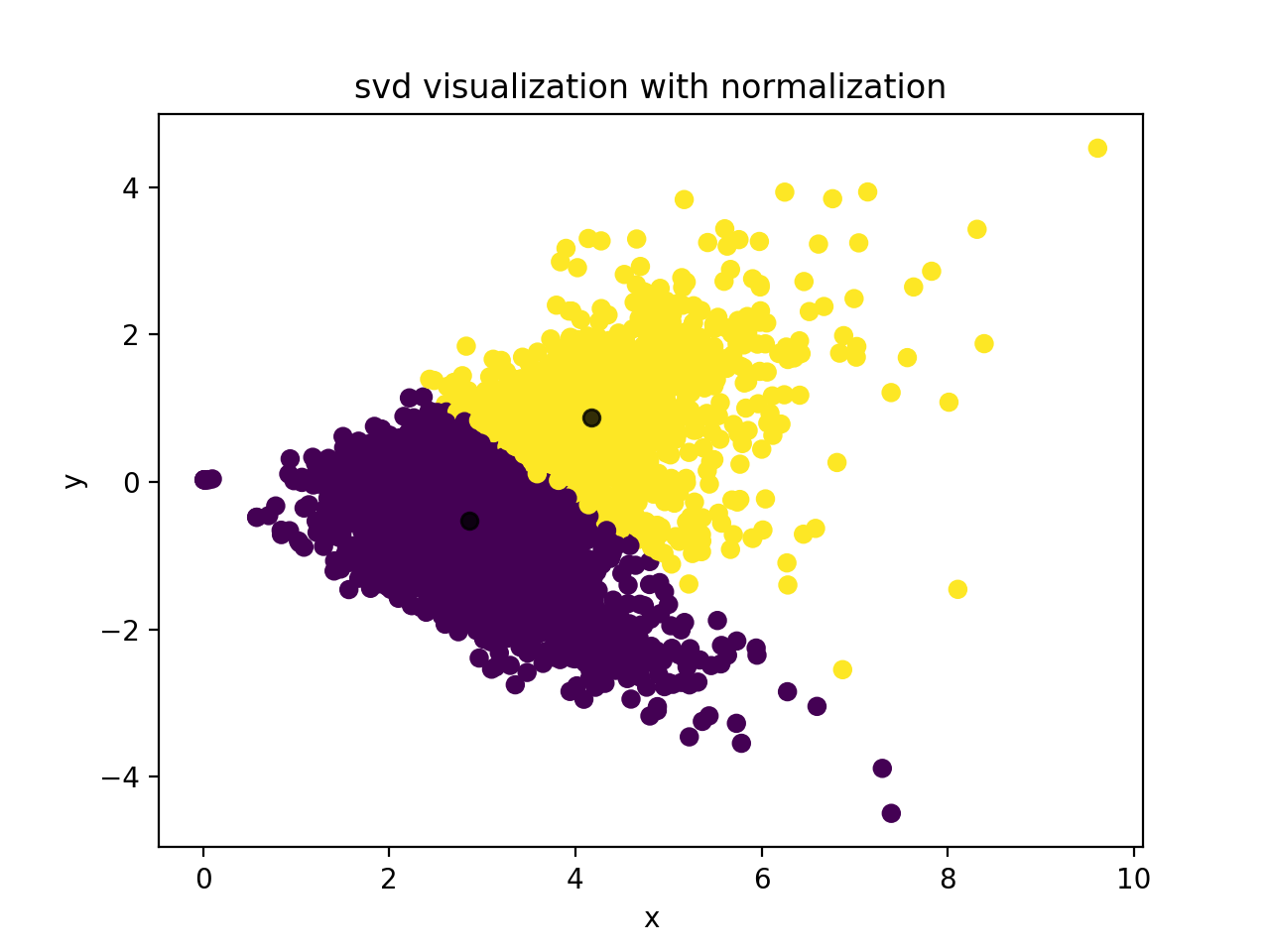
Homogeneity Score: 0.278

Completeness Score: 0.308

V-measure: 0.292

Adjusted Rand Score: 0.302

Adjusted Mutual Info Score: 0.278



==================================================

normalized nmf

==================================================

Confusion Matrix:

[[ 139 2204]

[2296 93]]

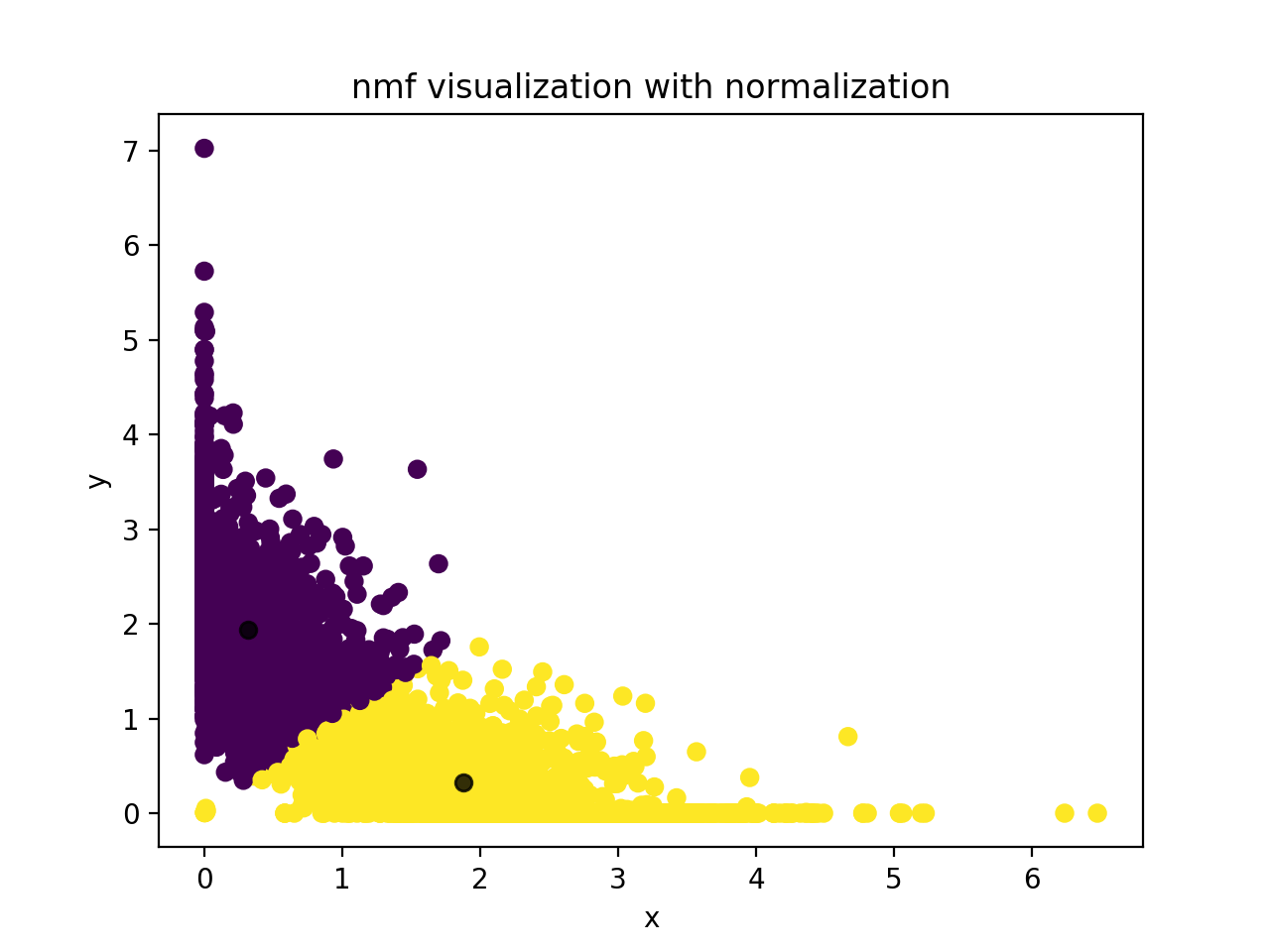
Homogeneity Score: 0.719

Completeness Score: 0.719

V-measure: 0.719

Adjusted Rand Score: 0.813

Adjusted Mutual Info Score: 0.719



==================================================

non-linear transformation for nmf

==================================================

Confusion Matrix:

[[ 301 2042]

[2347 42]]

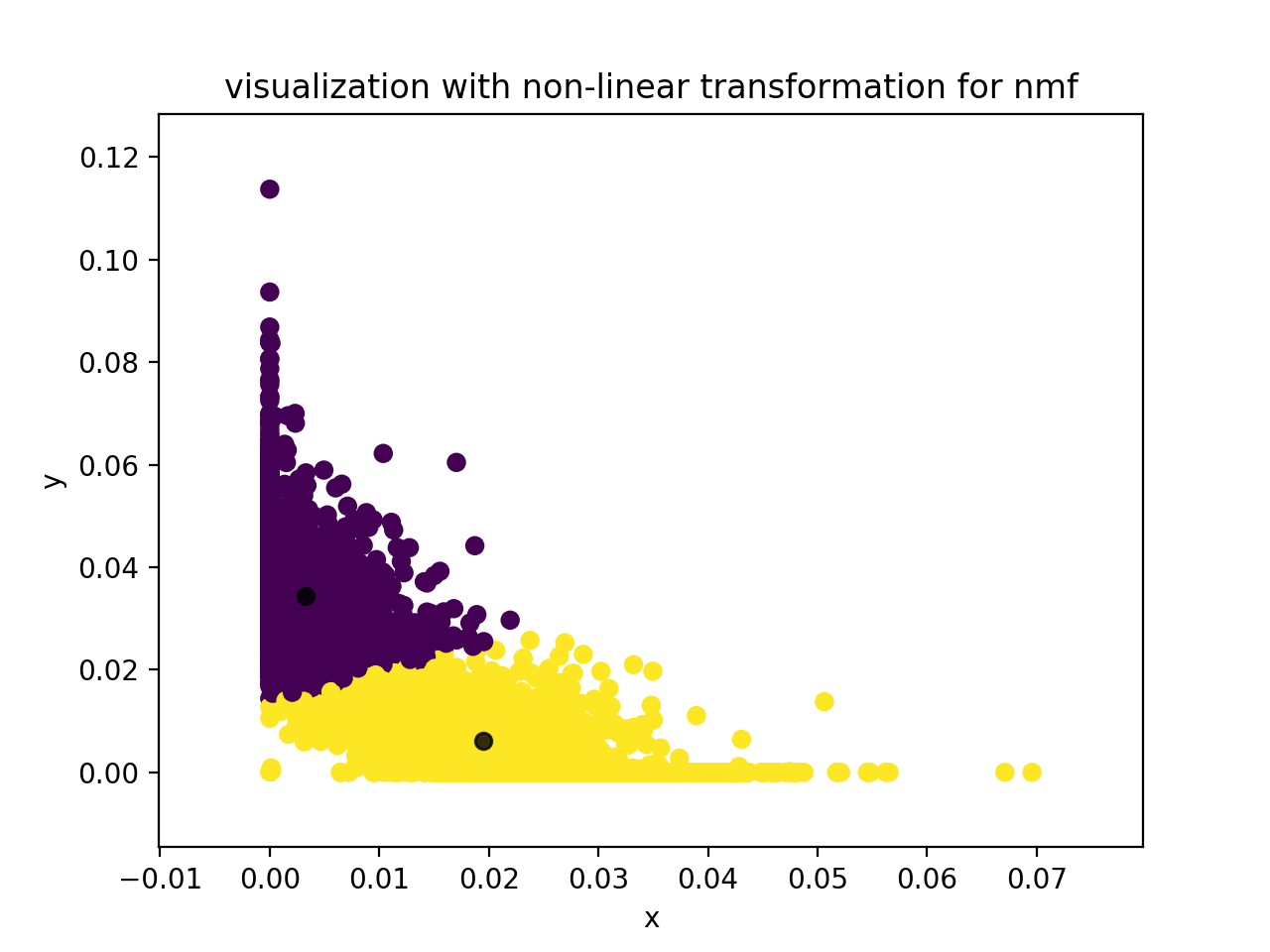
Homogeneity Score: 0.651

Completeness Score: 0.658

V-measure: 0.655

Adjusted Rand Score: 0.731

Adjusted Mutual Info Score: 0.651



==================================================

non-linear first and then normalization

==================================================

Confusion Matrix:

[[2225 118]

[ 101 2288]]

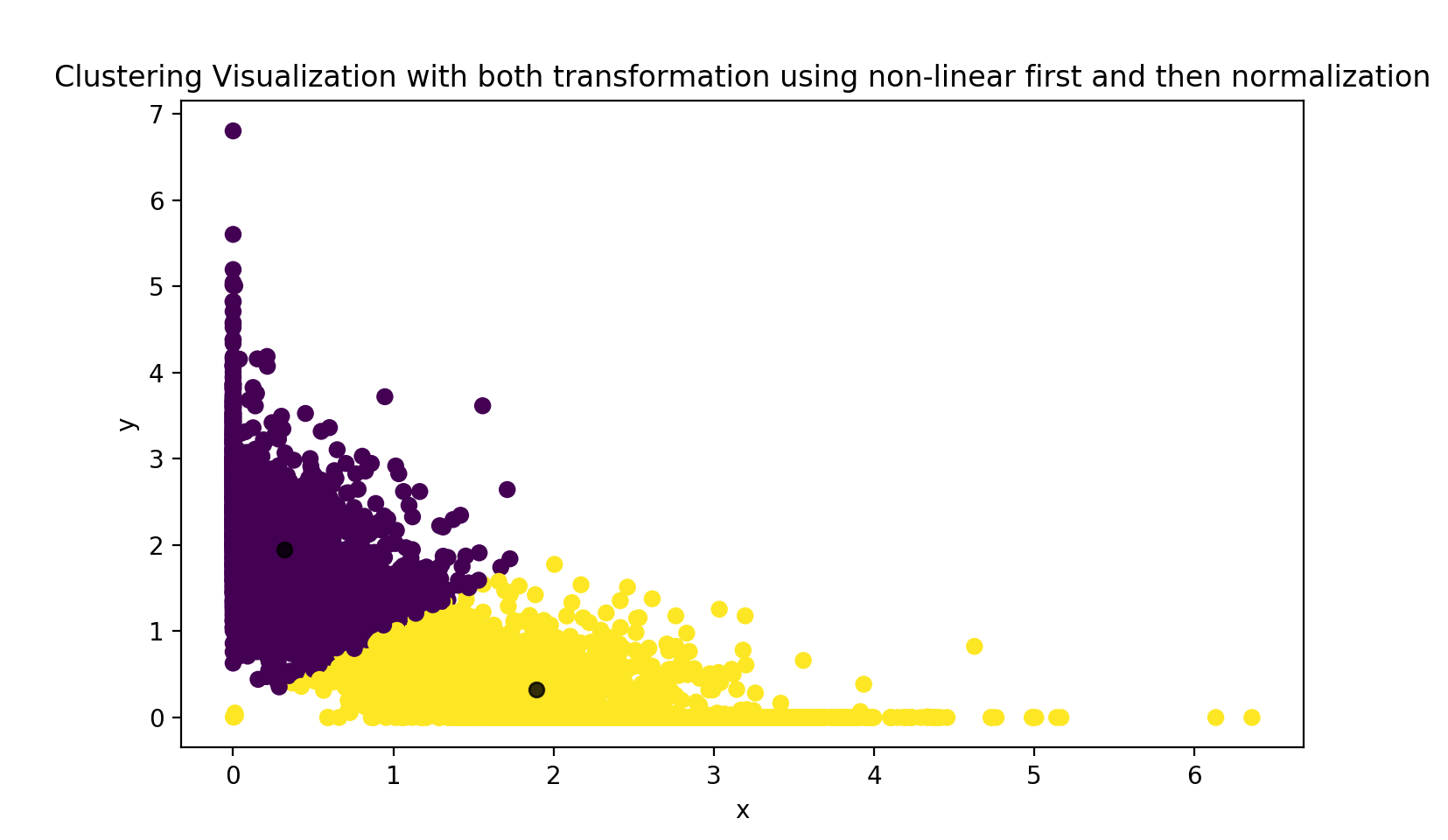
Homogeneity Score: 0.730

Completeness Score: 0.730

V-measure: 0.730

Adjusted Rand Score: 0.823

Adjusted Mutual Info Score: 0.730



==================================================

normalization first and then non-linear

==================================================

Confusion Matrix:

[[ 119 2224]

[2288 101]]

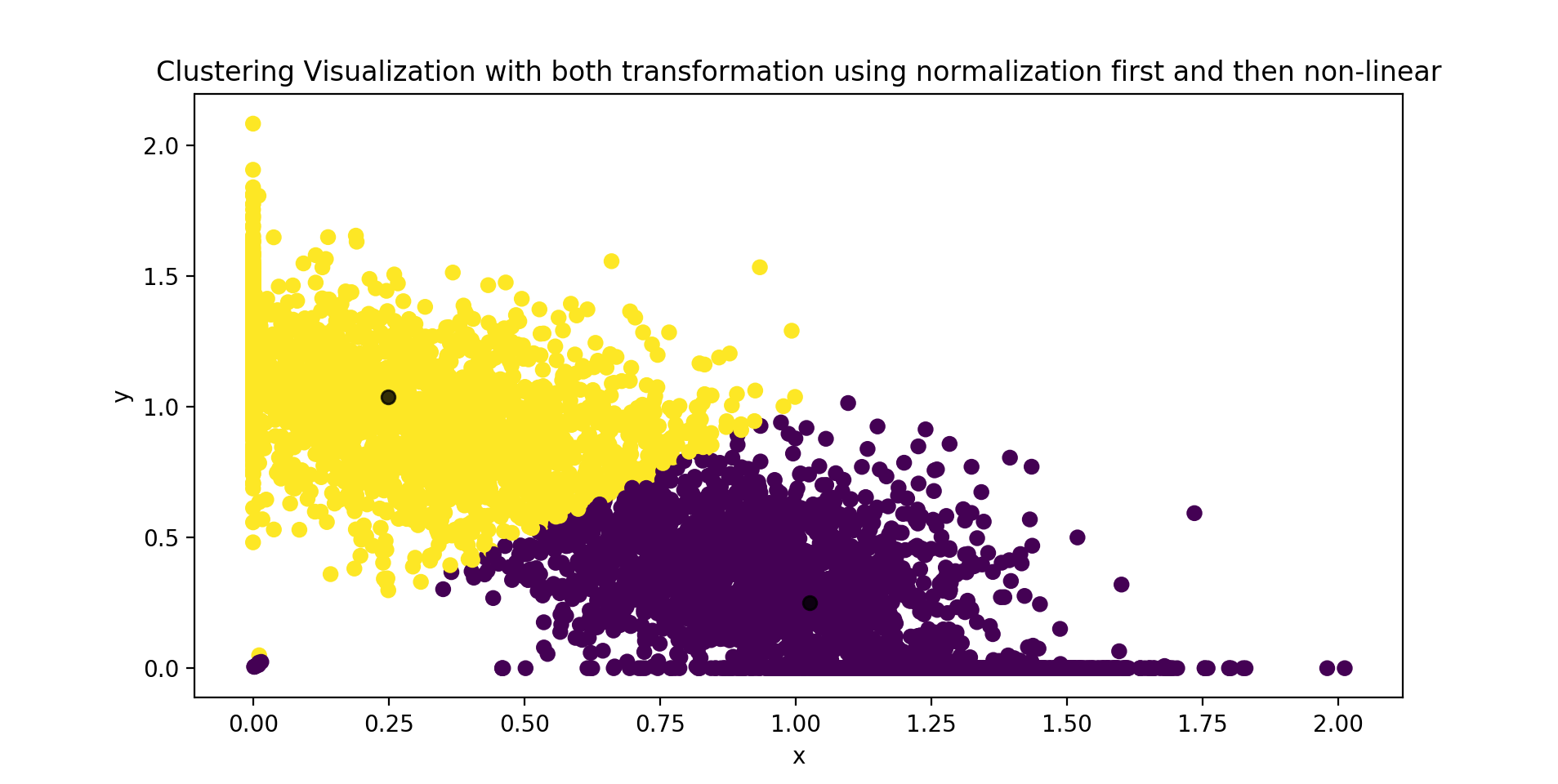
Homogeneity Score: 0.729

Completeness Score: 0.729

V-measure: 0.729

Adjusted Rand Score: 0.823

Adjusted Mutual Info Score: 0.729



Logarithm transformation will increase the clustering result because this transformation makes the data more organized, so that less noise is contained in the dataset. This fact makes clustering much easier.

20 Categories:

==================================================

================== Question 1 ==================

==================================================

dimensions: (11314, 34062)

==================================================

================== Question 2 ==================

==================================================

==================================================

quesiton 2

==================================================

Confusion Matrix:

[[ 1 44 0 1 0 0 101 0 17 0 0 2 175 139 0 0 0 0

0 0]

[ 7 1 0 0 46 0 0 51 47 0 0 1 341 1 0 0 0 47

2 40]

[ 7 0 0 0 67 0 0 14 16 0 0 0 146 0 0 0 10 306

10 15]

[ 13 0 0 0 96 0 0 80 32 0 3 1 174 0 0 0 146 32

0 13]

[ 44 1 0 0 41 0 0 308 13 0 3 0 106 0 0 0 55 2

0 5]

[ 20 0 0 0 4 1 0 1 26 0 0 0 167 0 0 0 0 17

3 354]

[ 15 0 3 0 24 0 0 61 11 12 5 28 372 0 0 0 27 17

1 9]

[ 1 20 0 0 0 0 0 1 16 0 101 311 138 0 0 0 1 2

0 3]

[ 0 6 0 0 0 0 0 0 41 1 44 347 158 0 0 0 0 0

1 0]

[ 3 0 0 0 0 0 0 0 10 349 6 0 227 0 0 0 0 0

0 2]

[ 0 0 16 0 0 0 0 1 9 490 0 0 84 0 0 0 0 0

0 0]

[ 0 36 0 0 0 366 0 6 15 0 1 0 162 0 0 0 0 4

1 4]

[ 25 0 0 0 8 0 0 42 33 1 80 7 377 0 0 0 5 2

6 5]

[ 2 5 51 0 0 0 0 1 15 0 7 1 433 1 0 77 0 1

0 0]

[ 1 5 0 0 0 0 0 0 4 0 25 2 167 0 0 0 0 0

388 1]

[ 3 12 0 0 0 0 0 0 1 0 12 0 216 355 0 0 0 0

0 0]

[ 0 422 0 0 3 2 0 0 10 0 14 0 94 0 0 0 0 1

0 0]

[ 3 28 0 253 0 0 0 0 6 0 0 0 143 2 128 0 0 0

0 1]

[ 1 237 4 0 0 1 1 0 4 2 8 1 202 3 0 0 0 0

0 1]

[ 1 37 0 0 0 0 32 0 10 0 4 1 177 114 0 1 0 0

0 0]]

Homogeneity Score: 0.420

Completeness Score: 0.520

V-measure: 0.465

Adjusted Rand Score: 0.142

Adjusted Mutual Info Score: 0.417

==================================================

================== Question 3 ==================

==================================================

==================================================

svd with r=1

==================================================

Confusion Matrix:

[[ 8 43 38 6 28 58 25 8 27 38 0 51 14 35 2 38 19 0 28 14]

[ 3 40 62 35 17 64 33 1 41 49 1 34 3 23 1 53 48 16 52 8]

[ 3 67 54 20 17 70 44 0 36 63 10 57 7 27 0 41 13 10 46 6]

[ 4 50 50 19 24 48 44 3 26 65 0 50 8 34 0 66 35 7 37 20]

[ 2 56 59 20 21 66 40 2 30 51 0 50 2 35 0 58 24 10 42 10]

[ 1 42 61 43 10 60 23 5 64 66 1 26 4 13 1 44 49 15 55 10]

[ 0 33 71 56 8 49 17 0 62 51 1 29 1 19 0 39 65 16 63 5]

[ 5 59 58 29 21 64 32 0 41 54 0 38 11 25 1 51 32 6 52 15]

[ 2 47 64 28 7 74 24 0 60 58 0 39 1 9 0 45 44 9 80 7]

[ 3 53 64 35 8 52 27 1 44 61 2 36 6 18 0 45 51 20 60 11]

[ 2 40 46 61 13 51 23 0 60 77 8 25 1 18 0 39 42 28 61 5]

[21 66 40 1 48 44 70 5 16 43 1 50 18 62 2 50 8 3 14 33]

[ 0 49 63 32 12 64 27 0 40 68 0 41 1 23 0 48 41 11 68 3]

[ 2 37 50 42 11 54 19 0 73 66 0 29 3 18 0 41 56 17 71 5]

[ 3 38 63 38 12 57 34 2 60 51 1 40 3 20 0 40 55 14 52 10]

[18 42 47 17 39 46 38 8 30 49 0 46 30 41 1 45 19 14 41 28]

[10 57 32 14 36 39 49 6 28 36 0 48 23 42 2 45 22 2 26 29]

[ 8 30 58 27 13 52 40 4 45 60 2 37 5 23 3 50 36 21 42 8]

[12 39 43 9 21 39 46 4 30 41 0 41 11 20 2 35 15 3 32 22]

[ 9 29 25 9 14 32 34 5 30 31 1 23 9 23 1 28 17 10 28 19]]

Homogeneity Score: 0.024

Completeness Score: 0.026

V-measure: 0.025

Adjusted Rand Score: 0.004

Adjusted Mutual Info Score: 0.019

==================================================

nmf with r=1

==================================================

Confusion Matrix:

[[32 16 46 5 27 13 50 2 52 22 40 0 8 36 9 0 16 29 51 26]

[55 8 43 29 33 3 58 1 34 32 55 14 1 24 3 1 48 59 64 19]

[47 6 69 15 46 6 66 0 60 29 64 9 0 25 4 10 16 49 51 19]

[32 19 51 19 41 9 51 0 51 31 64 5 3 39 4 0 29 45 72 25]

[38 11 66 16 43 2 58 0 48 34 55 10 2 35 2 0 16 57 65 20]

[58 9 47 35 22 5 56 2 27 68 62 14 4 15 1 1 40 62 55 10]

[68 6 35 48 19 1 55 0 30 62 56 14 0 19 0 1 60 67 37 7]

[52 15 64 22 35 11 63 1 39 34 60 6 0 25 5 0 33 57 51 21]

[67 7 48 22 25 1 68 0 44 63 62 7 0 9 2 0 40 75 51 7]

[54 11 55 30 28 6 60 0 33 47 63 18 1 20 3 2 42 65 50 9]

[64 5 39 54 24 1 55 0 27 51 79 25 0 18 2 8 44 46 44 14]

[15 33 73 2 71 19 45 2 49 11 48 2 5 64 21 1 7 31 49 47]

[69 4 54 33 29 1 64 0 40 37 78 8 0 21 0 0 35 55 50 13]

[75 5 49 38 19 3 61 0 29 58 64 16 0 17 2 0 56 53 36 13]

[56 11 39 35 34 4 52 0 41 50 49 13 2 19 3 1 50 69 53 12]

[38 28 47 17 41 29 43 1 44 25 52 13 8 41 19 0 16 48 48 41]

[23 27 57 12 50 25 45 2 46 29 28 2 6 46 10 0 16 37 48 37]

[42 8 36 24 41 5 46 3 34 40 65 19 4 22 8 2 35 57 57 16]

[35 20 42 7 49 12 37 2 39 25 50 2 4 18 13 0 13 36 37 24]

[31 19 33 8 33 8 31 2 22 27 29 9 4 25 10 1 14 21 36 14]]

Homogeneity Score: 0.024

Completeness Score: 0.027

V-measure: 0.026

Adjusted Rand Score: 0.004

Adjusted Mutual Info Score: 0.019

==================================================

svd with r=2

==================================================

Confusion Matrix:

[[ 0 9 117 124 0 1 0 23 2 61 8 0 0 55 0 19 0 0

36 25]

[ 58 30 0 0 20 27 126 0 17 0 0 114 71 0 70 2 31 4

1 13]

[ 59 12 0 0 73 18 120 0 5 0 0 71 136 0 34 0 52 11

0 0]

[ 68 4 0 0 86 9 127 0 13 0 0 56 119 0 32 0 48 26

0 2]

[116 17 0 0 16 13 129 0 10 1 0 75 82 0 70 1 41 3

0 4]

[ 56 20 0 0 23 27 169 0 7 2 0 123 81 0 38 0 37 8

1 1]

[ 40 33 0 0 44 40 119 0 10 4 0 155 67 0 49 2 10 4

2 6]

[ 10 136 2 2 0 46 11 0 59 24 2 33 0 0 122 27 0 0

27 93]

[ 5 155 0 7 0 43 8 0 26 71 0 44 0 0 63 10 0 0

28 138]

[ 3 107 0 7 0 72 4 0 15 109 4 16 1 0 19 23 0 0

67 150]

[ 1 80 1 8 0 108 2 0 5 141 2 9 0 0 15 21 0 0

70 137]

[ 32 51 1 4 1 8 12 0 143 25 18 10 2 2 76 38 6 0

61 105]

[ 47 86 0 0 2 43 49 0 35 11 0 104 16 0 154 0 7 0

1 36]

[ 1 102 0 16 0 71 2 0 10 138 1 24 1 0 33 19 1 1

62 112]

[ 9 142 0 3 0 75 6 0 41 52 4 40 0 0 54 10 4 0

35 118]

[ 0 12 128 113 0 18 2 49 2 72 8 2 0 119 5 17 0 0

26 26]

[ 2 27 34 51 0 14 0 2 15 75 33 5 0 10 5 70 0 0

128 75]

[ 1 16 117 153 0 38 0 8 0 115 14 2 0 33 1 8 0 0

34 24]

[ 2 23 32 58 0 13 3 0 9 90 25 2 0 5 1 44 0 0

86 72]

[ 0 15 61 64 0 17 0 27 1 63 5 2 0 48 2 14 0 0

31 27]]

Homogeneity Score: 0.230

Completeness Score: 0.245

V-measure: 0.237

Adjusted Rand Score: 0.074

Adjusted Mutual Info Score: 0.226

==================================================

nmf with r=2

==================================================

Confusion Matrix:

[[ 1 33 0 5 124 0 1 62 26 6 3 3 0 0 0 66 31 0

119 0]

[ 31 3 9 30 0 52 84 0 0 5 9 44 3 73 102 0 14 92

0 33]

[ 27 0 45 7 0 95 51 0 0 10 1 19 3 119 77 0 2 114

0 21]

[ 36 0 62 9 0 94 49 0 0 2 2 9 4 106 95 0 2 94

0 26]

[ 59 1 9 16 0 42 56 0 0 0 1 24 1 92 98 0 7 119

0 53]

[ 25 1 18 16 0 46 106 0 0 2 0 42 6 78 115 0 1 121

0 16]

[ 16 2 20 17 0 43 127 1 0 9 9 68 0 53 106 0 10 84

0 20]

[ 5 53 0 102 20 0 32 25 1 37 51 76 0 0 15 12 110 2

10 43]

[ 1 78 0 55 11 0 34 65 0 69 22 59 0 1 19 4 144 2

19 15]

[ 1 71 0 22 49 0 13 123 3 68 11 40 0 2 4 18 83 2

80 7]

[ 1 33 0 18 63 0 7 153 1 96 4 26 0 0 5 15 54 0

122 2]

[ 18 95 2 78 28 3 11 25 6 12 102 15 0 4 12 31 75 10

20 48]

[ 22 8 0 107 0 6 98 8 0 18 7 76 0 24 75 0 41 23

0 78]

[ 1 61 2 26 39 0 24 111 0 79 7 62 0 1 7 6 89 1

65 13]

[ 8 71 1 50 20 1 33 44 3 79 28 79 0 1 16 5 117 5

14 18]

[ 0 37 0 7 128 0 4 65 52 24 11 10 0 0 2 100 28 1

127 3]

[ 1 66 0 13 127 0 3 68 19 20 14 15 0 0 1 82 35 0

80 2]

[ 1 24 0 3 113 0 2 112 15 40 2 17 0 0 0 49 30 0

156 0]

[ 0 55 0 7 90 0 2 79 14 19 12 7 0 3 2 48 33 0

94 0]

[ 0 31 0 5 77 0 2 65 25 20 2 13 0 0 0 51 24 0

62 0]]

Homogeneity Score: 0.210

Completeness Score: 0.221

V-measure: 0.216

Adjusted Rand Score: 0.065

Adjusted Mutual Info Score: 0.206

==================================================

svd with r=3

==================================================

Confusion Matrix:

[[ 0 0 0 43 26 149 0 6 6 1 63 92 0 0 0 1 68 12

5 8]

[ 0 0 113 0 0 0 14 70 53 157 1 0 0 73 0 84 5 0

1 13]

[ 0 0 155 0 0 0 58 19 34 113 0 0 0 149 0 62 0 0

0 1]

[ 0 2 148 0 0 1 96 16 12 111 0 0 0 119 0 76 1 0

0 8]

[ 0 1 127 0 0 0 13 43 25 137 0 0 0 87 0 140 2 0

1 2]

[ 0 2 174 0 0 1 24 24 48 189 0 0 0 75 0 52 0 0

0 4]

[ 0 4 77 0 0 1 19 82 100 177 0 0 0 55 0 55 3 0

1 11]

[ 0 15 0 15 0 0 0 216 113 23 0 0 3 0 0 25 44 0

15 125]

[ 0 18 0 12 0 7 0 151 130 30 0 1 0 0 0 17 127 0

5 100]

[ 8 205 1 10 0 0 0 47 106 9 0 0 76 0 1 6 52 0

13 63]

[ 9 292 0 5 0 0 0 27 105 5 0 0 110 0 1 2 26 0

5 13]

[211 86 2 2 0 0 0 34 14 10 0 0 73 2 78 24 15 0

15 29]

[ 0 15 14 0 0 1 0 198 88 130 0 0 0 12 0 79 16 0

0 38]

[ 0 45 1 53 0 0 2 70 148 20 0 3 1 0 0 5 175 0

4 67]

[ 1 52 3 11 0 0 0 129 151 21 0 0 9 1 0 17 81 0

10 107]

[ 0 0 1 10 94 179 0 8 26 3 133 49 0 0 0 1 61 27

3 4]

[ 1 36 0 112 0 1 0 22 38 0 0 9 48 0 0 2 108 0

93 76]

[ 0 12 0 228 0 7 0 3 57 1 0 17 5 0 0 1 180 2

42 9]

[ 1 48 0 100 1 5 0 11 24 4 1 9 31 0 0 3 121 0

53 53]

[ 0 1 0 30 32 87 0 8 27 1 49 40 0 0 0 0 77 12

5 8]]

Homogeneity Score: 0.297

Completeness Score: 0.324

V-measure: 0.310

Adjusted Rand Score: 0.106

Adjusted Mutual Info Score: 0.294

==================================================

nmf with r=3

==================================================

Confusion Matrix:

[[ 0 3 0 0 0 6 84 72 13 0 126 39 0 1 23 2 1 105

4 1]

[ 1 2 125 80 92 7 0 0 47 0 0 5 15 2 0 120 72 1

1 14]

[ 0 1 140 31 165 2 0 0 29 0 0 0 69 0 0 90 63 0

0 1]

[ 1 1 133 43 143 1 0 0 14 0 0 1 85 1 0 82 76 1

0 8]

[ 1 0 133 75 94 5 0 0 24 0 0 0 12 2 0 96 111 1

0 24]

[ 0 2 201 22 108 5 0 0 33 0 0 0 25 0 0 155 38 2

0 2]

[ 8 0 57 136 58 46 0 0 77 0 0 2 18 15 0 54 62 1

2 49]

[107 0 0 34 0 108 0 1 37 18 0 12 0 139 0 8 6 0

20 104]

[ 76 0 0 32 0 174 0 1 20 7 0 27 0 195 0 4 4 1

6 51]

[196 0 0 10 0 90 0 0 18 79 0 2 0 174 0 3 3 0

2 20]

[217 0 0 6 0 77 0 0 13 101 0 3 0 166 0 0 0 0

0 17]

[ 1 226 27 24 8 7 10 31 27 0 9 53 3 2 1 41 16 91

9 9]

[ 5 3 32 158 13 70 0 0 68 0 0 10 1 21 0 81 53 1

0 75]

[ 40 0 1 38 1 131 0 12 124 3 0 100 2 76 0 13 3 10

20 20]

[ 47 0 3 38 1 165 0 1 69 1 0 34 1 126 0 16 7 1

18 65]

[ 0 8 0 6 0 4 135 38 26 1 165 33 0 1 69 4 1 103

5 0]

[ 14 3 0 9 0 52 6 85 29 0 12 150 0 41 1 2 1 43

86 12]

[ 0 1 1 0 0 12 55 59 53 0 147 60 0 2 4 2 0 152

14 2]

[ 13 0 1 3 0 46 6 61 20 1 13 125 0 50 0 2 3 50

61 10]

[ 0 1 0 4 0 13 61 40 26 0 62 50 0 4 30 3 0 77

5 1]]

Homogeneity Score: 0.306

Completeness Score: 0.318

V-measure: 0.312

Adjusted Rand Score: 0.112

Adjusted Mutual Info Score: 0.302

==================================================

svd with r=5

==================================================

Confusion Matrix:

[[ 0 0 21 0 0 0 93 0 0 4 23 0 0 0 0 210 3 80

9 37]

[ 15 0 0 0 2 18 1 0 0 116 1 1 0 65 59 0 246 19

40 1]

[ 40 0 0 0 9 23 0 0 0 55 0 0 3 141 110 0 201 4

5 0]

[ 5 0 0 0 108 148 0 34 0 46 0 4 71 8 69 1 74 1

21 0]

[ 0 1 0 0 88 250 0 4 0 77 1 2 35 5 30 0 52 4

29 0]

[ 53 0 0 0 0 2 0 0 0 60 0 2 0 218 13 0 234 4

6 1]

[ 2 22 0 0 48 130 0 1 2 232 1 1 22 4 23 1 56 7

33 0]

[ 0 12 0 0 1 16 0 0 1 175 16 0 0 4 0 0 15 76

254 24]

[ 0 17 0 0 0 8 0 0 0 188 3 0 0 0 0 7 13 144

200 18]

[ 0 314 0 0 0 0 0 0 118 99 2 0 0 0 0 0 7 17

38 2]

[ 0 305 0 0 0 1 0 0 228 53 0 0 0 0 0 0 2 6

5 0]

[ 0 0 0 149 0 7 0 0 0 11 10 281 0 6 1 0 29 42

15 44]

[ 0 1 0 0 9 95 0 0 0 207 0 2 1 4 5 1 65 60

138 3]

[ 1 62 0 0 0 0 0 0 1 132 8 0 0 0 1 1 24 231

70 63]

[ 1 1 0 0 0 2 0 0 0 175 9 1 0 3 0 0 45 169

146 41]

[ 0 0 84 0 0 1 166 0 0 23 11 0 0 0 0 226 4 72

7 5]

[ 0 0 0 0 0 2 0 0 0 33 103 6 0 0 0 4 1 137

50 210]

[ 0 0 2 0 0 0 3 0 0 35 78 1 0 1 0 15 0 194

6 229]

[ 0 5 0 0 0 0 2 0 0 21 66 3 0 1 0 8 4 162

45 148]

[ 0 0 32 0 0 0 64 0 0 27 21 0 0 0 0 111 0 83

14 25]]

Homogeneity Score: 0.361

Completeness Score: 0.404

V-measure: 0.382

Adjusted Rand Score: 0.143

Adjusted Mutual Info Score: 0.358

==================================================

nmf with r=5

==================================================

Confusion Matrix:

[[158 1 5 0 0 17 53 0 0 0 127 0 1 23 84 4 0 7

0 0]

[ 0 189 4 1 1 0 56 0 0 18 0 0 3 10 1 2 0 190

4 105]

[ 0 185 0 0 10 0 31 0 0 59 0 0 0 2 0 0 2 101

13 188]

[ 1 91 4 3 86 0 19 0 0 20 0 33 6 5 0 0 55 114

107 46]

[ 0 80 5 2 51 0 46 0 0 2 0 4 4 14 0 1 29 159

164 17]

[ 1 229 0 3 0 0 20 0 0 33 1 0 4 4 0 0 0 147

1 150]

[ 1 70 9 1 26 0 142 0 0 4 0 1 4 34 0 1 16 202

60 14]

[ 0 8 150 0 2 0 121 0 0 0 3 0 8 212 0 51 0 37

2 0]

[ 2 5 144 0 0 0 110 0 0 0 8 0 2 276 0 16 0 34

1 0]

[ 0 5 223 0 0 0 88 0 0 0 0 0 1 185 0 75 0 19

0 1]

[ 0 1 227 0 0 0 74 0 0 0 0 0 0 203 0 83 0 12

0 0]

[ 0 18 6 184 0 0 14 53 160 1 1 0 125 8 0 3 0 16

4 2]

[ 1 32 14 1 6 0 157 0 0 0 1 0 49 97 0 0 1 201

24 7]

[ 0 3 113 0 0 0 181 0 0 2 17 0 3 210 0 10 0 54

0 1]

[ 0 12 107 0 0 0 179 0 0 1 2 0 14 206 0 15 0 55

1 1]

[204 2 2 0 0 81 57 0 0 0 50 0 0 12 171 1 0 18

1 0]

[ 0 1 186 2 0 0 49 0 1 0 23 0 47 138 0 92 0 7

0 0]

[ 1 1 139 0 0 2 103 0 0 0 70 0 12 204 0 29 0 3

0 0]

[ 5 5 123 1 0 0 44 0 0 0 31 0 37 169 2 46 0 2

0 0]

[ 86 0 12 1 0 33 65 0 0 0 71 0 4 36 60 4 0 5

0 0]]

Homogeneity Score: 0.296

Completeness Score: 0.348

V-measure: 0.320

Adjusted Rand Score: 0.096

Adjusted Mutual Info Score: 0.292

==================================================

svd with r=10

==================================================

Confusion Matrix:

[[ 3 151 186 0 0 0 42 0 0 0 0 0 58 0 2 32 6 0

0 0]

[199 30 1 0 0 115 0 0 1 1 0 29 2 0 6 141 0 17

42 0]

[234 6 0 0 0 36 0 0 0 11 0 50 1 0 0 70 0 30

153 0]

[ 36 12 1 0 1 52 0 0 4 110 0 196 0 45 4 50 0 8

71 0]

[ 26 15 0 0 0 70 0 0 2 69 0 275 1 8 6 78 0 2

26 0]

[331 7 0 0 0 36 0 0 2 0 0 3 0 0 1 82 0 126

5 0]

[ 11 7 1 0 8 326 0 0 1 36 0 54 1 2 14 115 0 1

8 0]

[ 7 29 0 0 0 34 0 0 0 0 0 0 26 0 359 138 0 1

0 0]

[ 3 75 3 0 1 32 0 0 0 0 0 1 8 0 326 149 0 0

0 0]

[ 2 60 0 0 350 18 0 0 0 0 0 0 1 0 3 163 0 0

0 0]

[ 0 8 0 0 493 14 0 0 0 0 0 0 0 0 0 85 0 0

0 0]

[ 20 67 0 0 0 12 0 149 286 0 0 7 18 0 1 35 0 0

0 0]

[ 19 77 1 0 1 95 0 0 2 7 0 63 0 0 59 264 0 1

2 0]

[ 2 260 1 76 0 29 0 0 0 0 0 0 27 0 3 195 0 1

0 0]

[ 7 185 0 0 0 74 0 0 1 0 0 0 16 0 33 277 0 0

0 0]

[ 0 63 296 0 0 21 142 0 0 0 0 1 10 0 0 65 1 0

0 0]

[ 1 210 0 0 0 3 0 0 3 0 0 4 248 0 22 50 5 0

0 0]

[ 1 70 2 0 0 3 2 0 1 0 112 0 21 0 0 57 191 0

0 104]

[ 1 236 4 0 3 6 1 0 1 0 0 0 139 0 11 56 7 0

0 0]

[ 0 116 117 0 0 3 48 0 1 0 0 0 35 0 0 55 2 0

0 0]]

Homogeneity Score: 0.399

Completeness Score: 0.460

V-measure: 0.427

Adjusted Rand Score: 0.175

Adjusted Mutual Info Score: 0.396

==================================================

nmf with r=10

==================================================

Confusion Matrix:

[[ 97 176 0 0 0 0 1 0 15 39 0 41 0 0 1 3 0 106

1 0]

[ 69 1 1 0 0 16 0 0 169 0 22 3 49 1 110 126 0 16

1 0]

[ 48 0 0 0 0 78 0 0 82 0 51 1 49 12 35 228 1 6

0 0]

[ 38 1 5 0 0 10 0 48 92 0 69 1 133 109 43 32 0 9

0 0]

[ 61 0 3 0 0 0 0 7 133 0 42 0 174 70 53 18 1 15

1 0]

[ 50 0 5 0 0 122 0 0 72 0 2 0 4 0 41 294 0 2

1 0]

[ 31 1 1 0 2 2 0 2 191 0 12 1 36 32 234 19 12 9

0 0]

[122 0 0 0 0 1 0 0 126 0 0 78 0 2 32 5 0 228

0 0]

[163 4 0 0 0 0 0 0 116 0 0 33 0 0 22 4 1 255

0 0]

[ 67 0 0 0 88 0 0 0 59 0 0 1 1 0 18 1 342 20

0 0]

[ 21 0 0 0 239 0 0 0 22 0 0 0 0 0 13 0 303 2

0 0]

[ 54 0 268 0 0 1 0 0 26 0 0 11 2 0 10 9 0 39

0 175]

[165 1 3 0 0 0 0 0 209 0 2 0 40 6 68 13 2 82

0 0]

[187 1 0 64 0 1 0 0 92 0 0 33 0 0 19 2 0 156

39 0]

[179 0 0 0 0 0 0 0 166 0 0 30 0 0 39 1 0 178

0 0]

[ 87 303 0 0 0 0 0 0 37 130 0 7 0 0 10 0 0 25

0 0]

[ 71 1 3 0 0 1 0 0 32 0 2 217 3 0 1 0 0 215

0 0]

[139 2 1 0 0 0 350 0 22 2 0 11 0 0 2 1 0 34

0 0]

[105 6 1 0 0 0 1 0 22 1 0 116 0 0 6 1 5 201

0 0]

[110 117 1 0 0 0 0 0 20 47 0 18 0 0 3 0 1 59

1 0]]

Homogeneity Score: 0.354

Completeness Score: 0.407

V-measure: 0.379

Adjusted Rand Score: 0.135

Adjusted Mutual Info Score: 0.350

==================================================

svd with r=20

==================================================

Confusion Matrix:

[[ 0 209 2 129 0 0 0 0 0 0 1 0 24 0 0 0 108 0

0 7]

[ 0 52 2 1 0 0 41 0 36 0 0 0 0 0 1 20 0 1

45 385]

[ 11 39 1 0 0 0 29 1 279 0 0 0 0 0 1 13 0 0

60 157]

[ 87 38 4 0 0 0 153 58 29 0 0 0 0 0 9 10 0 0

89 113]

[ 52 37 0 0 0 0 321 19 1 0 0 0 1 0 11 7 0 0

31 98]

[ 0 43 0 0 0 1 14 0 16 0 0 0 0 0 1 237 0 2

3 276]

[ 33 76 12 0 6 1 19 1 11 0 0 0 0 0 333 5 0 1

11 76]

[ 2 128 398 0 0 0 0 0 2 0 0 0 11 0 7 3 0 0

0 43]

[ 0 234 308 0 1 0 0 0 0 0 0 0 3 0 13 0 0 0

0 39]

[ 0 211 3 0 347 0 0 0 0 0 0 0 0 0 1 0 0 0

0 35]

[ 0 76 0 0 496 0 0 0 0 0 0 0 0 0 1 0 0 0

0 27]

[ 1 141 0 0 0 352 15 0 5 0 0 0 36 0 0 2 0 1

0 42]

[ 4 227 32 0 1 0 118 0 2 0 0 0 2 0 5 4 0 7

4 185]

[ 0 388 1 0 0 0 4 0 1 76 0 0 4 50 1 0 0 0

0 69]

[ 0 204 1 0 0 0 0 0 0 0 0 0 5 0 1 1 0 339

0 42]

[ 0 231 0 338 0 0 1 0 0 0 0 0 10 0 0 0 0 0

0 19]

[ 0 139 1 0 0 2 0 0 1 0 0 0 386 0 0 0 0 0

3 14]

[ 0 188 0 2 0 0 0 0 0 0 233 117 19 0 0 1 0 0

0 4]

[ 0 283 4 3 5 1 0 0 0 0 0 0 159 0 0 1 1 1

0 7]

[ 0 188 1 113 0 0 0 0 0 0 0 0 32 0 0 0 34 0

0 9]]

Homogeneity Score: 0.444

Completeness Score: 0.534

V-measure: 0.485

Adjusted Rand Score: 0.179

Adjusted Mutual Info Score: 0.441

==================================================

nmf with r=20

==================================================

Confusion Matrix:

[[ 23 0 146 0 0 1 0 2 0 1 254 0 36 14 0 0 0 0

0 3]

[ 1 2 1 0 0 0 0 1 0 7 37 2 0 391 58 0 44 0

0 40]

[ 0 0 0 0 0 0 0 1 0 4 19 12 0 171 297 0 65 12

0 10]

[ 0 0 1 0 0 0 0 3 2 1 21 1 0 244 32 0 128 140

0 17]

[ 1 0 0 0 0 0 0 1 0 2 41 0 0 338 2 0 111 65

0 17]

[ 0 4 0 0 1 0 0 0 0 20 16 70 0 307 28 0 6 0

0 141]

[ 1 1 1 0 1 0 0 23 4 10 14 1 0 432 15 0 34 29

10 9]

[ 16 0 0 0 0 0 0 405 0 0 38 1 0 129 2 0 0 1

0 2]

[ 3 0 3 0 0 0 0 372 0 0 95 1 0 122 0 0 0 0

1 1]

[ 1 0 0 0 0 0 0 0 0 1 68 0 0 158 0 0 0 0

366 3]

[ 0 0 0 0 0 0 0 0 17 0 23 0 0 72 0 0 0 0

488 0]

[ 36 1 0 0 345 0 0 0 0 0 121 0 0 86 3 0 1 0

0 2]

[ 0 7 1 0 0 0 0 33 0 1 108 0 0 413 2 0 13 5

0 8]

[ 7 0 1 76 0 0 50 2 1 2 284 0 0 170 1 0 0 0

0 0]

[ 7 348 0 0 0 0 0 2 0 1 109 0 0 125 0 0 0 0

0 1]

[ 8 0 313 0 0 0 0 0 0 1 88 0 128 61 0 0 0 0

0 0]

[361 0 0 0 2 0 0 2 0 0 147 0 0 31 0 0 3 0

0 0]

[ 19 0 2 0 0 231 0 0 0 1 131 0 2 63 0 114 0 0

0 1]

[196 0 5 0 1 0 0 2 11 1 209 0 1 36 0 0 0 0

2 1]

[ 32 0 104 0 0 0 0 1 0 0 152 0 46 42 0 0 0 0

0 0]]

Homogeneity Score: 0.403

Completeness Score: 0.511

V-measure: 0.451

Adjusted Rand Score: 0.154

Adjusted Mutual Info Score: 0.399

==================================================

svd with r=50

==================================================

Confusion Matrix:

[[ 45 0 0 103 0 190 1 0 9 0 0 0 3 0 2 0 0 0

101 26]

[ 0 44 0 0 0 45 0 0 406 42 10 0 1 0 35 0 0 0

1 0]

[ 0 270 0 0 0 36 0 9 178 71 7 1 1 0 6 0 12 0

0 0]

[ 0 33 0 0 0 45 0 127 209 104 56 0 3 0 13 0 0 0

0 0]

[ 0 2 0 0 0 44 0 50 151 38 276 0 0 0 17 0 0 0

0 0]

[ 0 21 1 0 0 33 0 0 317 4 2 0 0 0 150 0 65 0

0 0]

[ 0 13 0 0 7 64 0 21 93 19 338 3 14 0 11 0 1 0

0 1]

[ 0 2 0 0 0 171 0 0 58 0 4 0 342 0 2 0 1 0

0 14]

[ 0 0 0 0 1 518 0 0 38 0 12 0 24 0 1 0 1 0

0 3]

[ 0 0 0 0 345 203 0 0 41 0 0 5 0 0 3 0 0 0

0 0]

[ 0 0 0 1 504 60 0 0 34 0 1 0 0 0 0 0 0 0

0 0]

[ 0 4 356 0 0 130 0 0 61 0 3 0 0 0 1 0 0 0

0 40]

[ 0 2 0 0 1 210 0 3 300 6 25 9 26 0 6 0 0 0

0 3]

[ 0 1 0 0 0 363 0 0 91 0 2 7 0 50 0 0 0 76

0 4]

[ 0 0 0 0 0 534 0 0 50 0 0 0 0 0 2 0 0 0

0 7]

[205 0 0 0 0 178 0 0 35 0 0 17 0 0 0 0 0 0

153 11]

[ 0 1 2 0 0 95 0 0 16 3 1 0 3 0 0 0 0 0

0 425]

[ 0 0 0 0 0 170 229 0 17 0 0 2 0 0 1 120 0 0

2 23]

[ 0 0 1 1 3 293 0 0 6 0 0 0 1 0 1 0 0 0

3 156]

[ 97 0 0 33 0 169 0 0 10 0 0 0 1 0 0 0 0 0

36 31]]

Homogeneity Score: 0.402

Completeness Score: 0.517

V-measure: 0.452

Adjusted Rand Score: 0.153

Adjusted Mutual Info Score: 0.399

==================================================

nmf with r=50

==================================================

Confusion Matrix:

[[103 0 0 0 0 0 0 163 0 81 0 0 0 0 28 84 3 0

0 18]

[ 1 3 0 76 1 0 0 0 0 399 0 0 7 0 0 50 47 0

0 0]

[ 0 0 0 338 0 12 0 1 10 189 0 1 4 1 0 18 17 0

0 0]

[ 1 0 0 146 1 103 0 0 0 237 0 2 1 44 0 28 27 0

0 0]

[ 0 0 0 77 0 62 0 0 0 358 0 1 2 13 0 37 28 0

0 0]

[ 0 4 0 24 12 0 0 0 0 334 0 7 19 0 0 21 172 0

0 0]

[ 1 1 0 38 1 42 0 0 0 452 1 0 10 0 0 18 17 4

0 0]

[ 0 0 0 2 0 2 0 0 0 453 0 0 0 0 0 121 4 12

0 0]

[ 4 0 0 1 0 0 0 0 0 154 0 0 0 0 0 44 5 390

0 0]

[ 0 0 0 0 0 0 0 1 0 489 0 0 1 0 0 103 3 0

0 0]

[ 0 0 0 0 0 0 0 5 0 506 0 0 0 0 0 89 0 0

0 0]

[ 0 1 0 3 0 1 0 0 0 160 212 128 0 0 0 88 2 0

0 0]

[ 1 0 0 9 0 4 0 0 0 523 0 0 1 0 0 37 9 0

7 0]

[ 1 0 0 1 0 0 75 0 0 340 0 0 2 0 0 174 1 0

0 0]

[ 0 277 0 0 0 0 0 0 0 201 0 0 1 0 0 77 2 0

35 0]

[297 0 0 0 0 0 0 1 0 134 0 0 1 0 96 70 0 0

0 0]

[ 0 0 0 4 0 0 0 2 0 322 1 0 0 0 0 216 1 0

0 0]

[ 1 0 210 0 0 0 0 98 0 175 0 0 1 0 2 76 1 0

0 0]

[ 5 0 0 0 0 0 0 0 0 229 1 0 1 0 1 227 1 0

0 0]

[ 87 0 0 0 0 0 0 7 0 146 0 0 0 0 27 76 0 0

0 34]]

Homogeneity Score: 0.264

Completeness Score: 0.432

V-measure: 0.327

Adjusted Rand Score: 0.051

Adjusted Mutual Info Score: 0.260

==================================================

svd with r=100

==================================================

Confusion Matrix:

[[ 0 0 0 44 0 0 0 0 0 0 1 0 87 97 156 0 0 92

0 3]

[ 0 0 0 0 0 0 0 0 71 0 0 0 409 0 6 58 0 1

38 1]

[ 0 0 0 0 0 0 0 0 366 0 0 1 175 0 0 21 12 0

15 1]

[ 1 0 0 0 0 0 0 0 53 2 0 50 175 0 3 193 101 0

11 1]

[ 0 0 0 0 0 0 0 0 3 0 0 20 157 0 2 341 51 0

4 0]

[ 0 0 0 0 1 0 0 0 25 0 0 0 283 0 1 9 0 0

274 0]

[ 8 0 1 0 0 0 5 0 18 4 0 1 257 0 7 227 32 1

7 17]

[ 0 0 3 0 0 0 11 0 2 0 0 0 197 0 29 2 2 0

3 345]

[ 1 0 0 0 0 0 422 0 1 0 0 0 146 0 16 3 0 0

0 9]

[325 0 0 0 0 0 0 31 0 0 0 0 224 0 17 0 0 0

0 0]

[489 0 0 0 0 0 0 0 0 18 0 0 90 0 2 1 0 0

0 0]

[ 0 0 5 0 358 0 0 0 3 0 0 0 138 0 85 2 1 0

3 0]

[ 1 0 0 0 0 0 0 0 3 0 0 0 491 0 5 56 6 0

4 25]

[ 0 0 0 0 0 76 0 0 1 1 0 0 371 0 144 1 0 0

0 0]

[ 0 0 1 0 0 0 0 0 0 0 0 0 537 0 54 0 0 0

1 0]

[ 0 0 0 246 0 0 0 0 0 0 0 0 162 0 61 0 0 130

0 0]

[ 0 0 223 0 2 0 0 0 0 0 0 0 89 0 228 2 0 0

0 2]

[ 0 121 0 0 0 0 0 0 0 0 233 0 137 0 70 0 0 2

1 0]

[ 2 0 1 0 1 0 0 0 0 11 0 0 92 1 352 0 0 3

1 1]

[ 0 0 1 97 0 0 0 0 0 0 0 0 124 35 89 0 0 30

0 1]]

Homogeneity Score: 0.413

Completeness Score: 0.546

V-measure: 0.470

Adjusted Rand Score: 0.131

Adjusted Mutual Info Score: 0.410

==================================================

nmf with r=100

==================================================

Confusion Matrix:

[[ 0 0 9 95 0 0 0 18 0 78 0 0 0 0 0 1 0 0

104 175]

[ 0 17 0 1 7 28 0 39 0 0 0 0 0 0 1 2 0 0

34 455]

[ 0 16 0 0 3 58 0 43 0 0 0 0 0 0 1 1 0 0

26 443]

[ 0 60 0 0 8 27 0 48 0 0 0 0 0 0 6 12 0 0

31 398]

[ 0 121 0 1 15 21 0 50 1 0 0 0 0 0 9 17 0 0

34 309]

[ 0 0 0 0 8 302 0 31 3 0 0 0 0 0 0 1 0 0

13 235]

[ 0 6 0 0 27 4 2 57 0 0 0 0 0 0 277 28 0 0

14 170]

[ 0 1 0 0 13 11 10 92 0 0 0 0 1 1 5 7 0 0

115 338]

[ 0 0 0 0 9 8 0 81 0 0 0 0 0 0 10 2 0 0

91 397]

[ 0 0 0 0 20 0 0 44 0 0 27 0 0 0 0 0 0 0

96 410]

[ 0 0 0 0 19 0 0 70 0 3 0 0 0 0 0 5 0 0

112 391]

[ 0 0 0 3 2 3 0 30 350 0 0 0 0 0 0 1 0 0

60 146]

[ 0 1 0 0 5 1 18 33 0 0 0 0 0 0 4 156 0 0

36 337]

[ 0 0 0 11 4 2 0 57 0 0 0 50 0 0 1 0 0 75

93 301]

[ 0 0 0 0 4 1 2 64 0 0 0 0 1 0 0 4 0 0

108 409]

[ 0 0 56 169 4 0 0 39 0 0 0 0 0 3 0 1 0 0

101 226]

[ 0 0 0 5 1 0 0 25 1 0 0 0 137 110 2 0 0 0

68 197]

[181 0 0 3 2 9 0 20 0 74 0 0 0 0 0 0 30 0

66 179]

[ 0 0 0 8 25 1 0 62 1 0 0 0 1 1 0 0 0 0

152 214]

[ 0 0 1 80 2 0 0 24 0 18 0 0 1 18 0 0 0 1

68 164]]

Homogeneity Score: 0.207

Completeness Score: 0.334

V-measure: 0.256

Adjusted Rand Score: 0.032

Adjusted Mutual Info Score: 0.203

==================================================

svd with r=300

==================================================

Confusion Matrix:

[[297 0 0 0 21 0 0 0 98 0 1 0 2 0 0 61 0 0

0 0]

[ 8 168 0 6 0 0 0 0 1 0 0 0 0 0 1 397 0 0

3 0]

[ 0 362 0 4 0 0 0 11 0 0 0 1 1 0 26 186 0 0

0 0]

[ 1 107 1 1 0 0 2 153 0 0 0 0 3 0 24 298 0 0

0 0]

[ 2 38 0 2 0 0 0 67 0 0 0 0 0 0 14 455 0 0

0 0]

[ 1 300 0 20 0 1 0 0 0 0 0 0 0 0 8 260 0 0

3 0]

[ 6 31 13 10 0 1 4 35 0 0 0 0 18 0 0 465 0 1

1 0]

[ 19 4 0 0 0 0 0 1 0 0 0 0 356 1 0 210 0 3

0 0]

[ 20 1 1 0 0 0 0 0 0 0 0 0 29 0 0 547 0 0

0 0]

[ 6 0 409 1 0 0 2 0 0 0 0 0 0 0 0 179 0 0

0 0]

[ 2 0 348 0 0 0 150 0 0 0 0 0 0 0 0 100 0 0

0 0]

[ 62 9 0 0 0 369 0 0 0 0 0 0 0 0 0 149 0 5

1 0]

[ 1 12 0 1 0 0 0 5 0 0 0 8 31 0 0 526 0 0

7 0]

[145 2 0 2 0 0 1 0 0 76 0 7 0 0 0 360 1 0

0 0]

[ 13 1 0 1 0 0 0 0 0 0 0 0 1 0 0 195 0 1

381 0]

[ 91 0 0 1 0 0 0 0 332 0 0 17 0 3 0 155 0 0

0 0]

[ 72 0 0 0 0 2 0 0 0 0 0 0 1 125 0 107 5 234

0 0]

[ 67 1 0 1 0 0 0 0 2 0 239 0 0 0 0 131 0 0

0 123]

[326 1 2 1 0 1 5 0 3 0 0 0 1 2 0 95 27 1

0 0]

[115 0 0 0 0 0 0 0 111 0 0 0 2 20 0 127 0 2

0 0]]

Homogeneity Score: 0.382

Completeness Score: 0.552

V-measure: 0.451

Adjusted Rand Score: 0.108

Adjusted Mutual Info Score: 0.379

==================================================

nmf with r=300

==================================================

Confusion Matrix:

[[ 2 55 0 0 30 2 3 1 0 0 0 0 238 0 24 0 125 0

0 0]

[ 0 14 0 0 35 16 125 17 0 0 0 0 256 0 1 5 65 0

0 50]

[ 0 12 0 0 61 6 117 6 2 0 0 0 278 0 0 4 74 0

0 31]

[ 0 7 0 0 46 20 133 2 0 0 0 0 230 0 1 60 51 0

0 40]

[ 0 16 0 0 58 24 77 101 3 0 0 0 208 0 4 2 52 0

0 33]

[ 0 12 0 0 65 4 115 0 0 0 0 0 302 0 1 2 49 0

0 43]

[ 0 5 0 0 105 46 52 7 9 0 1 0 309 0 1 3 35 0

0 12]

[ 0 33 0 0 92 24 53 1 2 0 0 0 253 0 1 1 118 0

0 16]

[ 0 23 0 0 84 14 42 0 2 0 0 0 300 0 0 0 121 0

0 12]

[ 0 32 0 0 42 16 38 0 6 0 0 51 304 0 0 0 91 0

0 17]

[ 0 28 0 0 57 28 22 0 1 0 34 1 321 0 4 0 97 0

0 7]

[ 0 27 0 0 78 7 12 1 0 0 0 0 324 0 4 0 123 16

0 3]

[ 0 14 0 0 52 7 89 3 1 0 0 0 283 0 1 2 103 0

0 36]

[ 0 20 74 0 63 14 52 1 0 0 0 0 249 0 4 0 106 0

1 10]

[ 0 23 0 0 68 13 29 2 1 0 0 0 345 0 0 0 102 0

0 10]

[ 19 43 0 14 53 18 21 0 0 0 0 0 241 0 62 0 122 0

0 6]

[ 0 48 0 0 45 5 16 0 0 0 0 0 262 38 10 0 117 0

2 3]

[ 0 23 0 0 43 3 5 0 0 70 0 0 287 0 16 0 117 0

0 0]

[ 0 35 0 0 76 14 2 0 10 0 0 0 186 0 4 0 108 0

27 3]

[ 1 28 1 0 30 3 3 0 2 0 0 0 172 4 42 0 91 0

0 0]]

Homogeneity Score: 0.075

Completeness Score: 0.125

V-measure: 0.094

Adjusted Rand Score: 0.006

Adjusted Mutual Info Score: 0.069

==================================================

quesiton 4b1 with normal svd

==================================================

Confusion Matrix:

[[100 0 38 18 0 0 3 0 17 67 1 44 0 133 15 32 4 8

0 0]

[ 0 84 3 1 7 19 24 77 0 0 13 5 103 0 54 10 24 0

101 59]

[ 0 83 0 0 13 77 19 132 0 0 5 0 58 0 15 4 7 0

106 72]

[ 0 72 1 0 31 79 6 124 0 0 11 1 58 0 9 2 15 0

116 65]

[ 0 116 3 1 3 20 14 70 0 0 21 0 59 0 23 4 24 0

122 98]

[ 0 87 0 0 13 18 23 84 0 0 10 1 114 0 22 12 10 0

152 47]

[ 1 66 1 2 3 43 35 64 0 0 8 3 136 0 29 23 12 0

133 26]

[ 6 37 47 18 0 0 28 0 0 1 27 35 26 2 154 94 96 1

18 4]

[ 26 16 86 7 0 0 24 0 0 0 7 23 40 1 156 136 59 0

16 1]

[ 33 6 127 18 0 0 51 1 0 0 5 52 18 0 115 134 27 1

8 1]

[ 53 5 117 9 0 0 88 0 0 0 3 58 10 0 87 156 9 0

5 0]

[ 12 32 71 43 2 2 4 2 0 1 78 76 3 0 98 32 109 7

16 7]

[ 1 69 12 0 0 2 29 18 0 0 19 1 76 0 124 55 87 0

72 26]

[ 44 12 101 9 2 0 54 0 0 0 4 47 28 0 105 159 23 0

5 1]

[ 19 19 65 10 1 0 48 1 0 0 19 35 37 0 131 122 66 2

17 1]

[110 2 40 17 0 0 27 0 57 116 1 30 2 135 8 38 6 8

2 0]

[ 56 3 116 66 0 0 14 0 2 23 5 109 6 31 34 57 15 9

0 0]

[155 1 41 12 0 0 45 0 3 55 0 34 2 127 15 61 2 11

0 0]

[ 79 3 91 47 0 0 11 1 1 8 3 86 3 29 35 48 13 7

0 0]

[ 73 0 39 12 0 0 16 0 28 53 0 34 5 63 16 32 1 5

0 0]]

Homogeneity Score: 0.215

Completeness Score: 0.230

V-measure: 0.222

Adjusted Rand Score: 0.067

Adjusted Mutual Info Score: 0.210

==================================================

normal svd

==================================================

Confusion Matrix:

[[100 0 38 18 0 0 3 0 17 67 1 44 0 133 15 32 4 8

0 0]

[ 0 84 3 1 7 19 24 77 0 0 13 5 103 0 54 10 24 0

101 59]

[ 0 83 0 0 13 77 19 132 0 0 5 0 58 0 15 4 7 0

106 72]

[ 0 72 1 0 31 79 6 124 0 0 11 1 58 0 9 2 15 0

116 65]

[ 0 116 3 1 3 20 14 70 0 0 21 0 59 0 23 4 24 0

122 98]

[ 0 87 0 0 13 18 23 84 0 0 10 1 114 0 22 12 10 0

152 47]

[ 1 66 1 2 3 43 35 64 0 0 8 3 136 0 29 23 12 0

133 26]

[ 6 37 47 18 0 0 28 0 0 1 27 35 26 2 154 94 96 1

18 4]

[ 26 16 86 7 0 0 24 0 0 0 7 23 40 1 156 136 59 0

16 1]

[ 33 6 127 18 0 0 51 1 0 0 5 52 18 0 115 134 27 1

8 1]

[ 53 5 117 9 0 0 88 0 0 0 3 58 10 0 87 156 9 0

5 0]

[ 12 32 71 43 2 2 4 2 0 1 78 76 3 0 98 32 109 7

16 7]

[ 1 69 12 0 0 2 29 18 0 0 19 1 76 0 124 55 87 0

72 26]

[ 44 12 101 9 2 0 54 0 0 0 4 47 28 0 105 159 23 0

5 1]

[ 19 19 65 10 1 0 48 1 0 0 19 35 37 0 131 122 66 2

17 1]

[110 2 40 17 0 0 27 0 57 116 1 30 2 135 8 38 6 8

2 0]

[ 56 3 116 66 0 0 14 0 2 23 5 109 6 31 34 57 15 9

0 0]

[155 1 41 12 0 0 45 0 3 55 0 34 2 127 15 61 2 11

0 0]

[ 79 3 91 47 0 0 11 1 1 8 3 86 3 29 35 48 13 7

0 0]

[ 73 0 39 12 0 0 16 0 28 53 0 34 5 63 16 32 1 5

0 0]]

Homogeneity Score: 0.215

Completeness Score: 0.230

V-measure: 0.222

Adjusted Rand Score: 0.067

Adjusted Mutual Info Score: 0.210

==================================================

quesiton 4b1 with normal nmf

==================================================

Confusion Matrix:

[[ 7 0 0 129 47 48 0 99 13 1 1 89 19 23 2 0 2 0

0 0]

[ 6 58 55 0 1 0 109 0 36 58 37 0 11 0 9 3 22 77

10 92]

[ 1 33 101 0 0 0 102 0 12 45 35 0 0 0 11 3 7 133

48 60]

[ 2 25 99 0 0 0 101 0 6 49 42 0 2 0 2 4 18 110

64 66]

[ 1 32 42 0 0 1 123 1 13 81 71 0 4 0 5 1 21 98

10 74]

[ 1 65 51 0 0 1 146 0 16 48 29 0 0 0 1 6 11 89

18 111]

[ 7 107 46 0 0 1 93 1 32 57 21 1 4 0 16 0 11 62

20 106]

[ 45 41 0 19 9 44 4 27 134 45 6 17 96 1 44 0 49 1

0 12]

[ 26 42 0 22 2 95 5 63 110 24 1 9 101 0 56 0 23 1

0 18]

[ 24 19 0 77 11 121 2 109 54 11 1 30 60 2 63 0 5 3

0 5]

[ 7 16 0 74 7 151 1 147 31 5 1 45 29 0 79 0 3 0

0 4]

[101 6 3 38 27 28 13 45 65 31 17 28 98 3 7 0 72 4

2 7]

[ 3 63 7 0 0 16 38 3 107 100 29 0 36 0 28 0 66 24

1 70]

[ 7 39 1 56 5 141 2 91 66 22 1 26 57 0 63 0 7 0

2 8]

[ 32 45 1 20 4 75 4 42 104 24 8 11 97 2 78 0 25 3

1 17]

[ 9 4 0 116 83 59 1 102 16 4 0 111 31 37 19 0 5 0

0 2]

[ 21 10 0 115 58 60 0 91 20 2 0 100 37 11 16 0 5 0

0 0]

[ 3 9 0 133 30 85 0 139 16 0 1 79 21 13 34 0 0 0

0 1]

[ 23 3 0 75 37 62 1 111 14 0 0 78 31 9 15 0 1 3

0 2]

[ 8 9 0 75 39 58 0 58 14 1 0 67 15 16 17 0 0 0

0 0]]

Homogeneity Score: 0.206

Completeness Score: 0.217

V-measure: 0.212

Adjusted Rand Score: 0.063

Adjusted Mutual Info Score: 0.202

==================================================

normal nmf

==================================================

Confusion Matrix:

[[ 7 0 0 129 47 48 0 99 13 1 1 89 19 23 2 0 2 0

0 0]

[ 6 58 55 0 1 0 109 0 36 58 37 0 11 0 9 3 22 77

10 92]

[ 1 33 101 0 0 0 102 0 12 45 35 0 0 0 11 3 7 133

48 60]

[ 2 25 99 0 0 0 101 0 6 49 42 0 2 0 2 4 18 110

64 66]

[ 1 32 42 0 0 1 123 1 13 81 71 0 4 0 5 1 21 98

10 74]

[ 1 65 51 0 0 1 146 0 16 48 29 0 0 0 1 6 11 89

18 111]

[ 7 107 46 0 0 1 93 1 32 57 21 1 4 0 16 0 11 62

20 106]

[ 45 41 0 19 9 44 4 27 134 45 6 17 96 1 44 0 49 1

0 12]

[ 26 42 0 22 2 95 5 63 110 24 1 9 101 0 56 0 23 1

0 18]

[ 24 19 0 77 11 121 2 109 54 11 1 30 60 2 63 0 5 3

0 5]

[ 7 16 0 74 7 151 1 147 31 5 1 45 29 0 79 0 3 0

0 4]

[101 6 3 38 27 28 13 45 65 31 17 28 98 3 7 0 72 4

2 7]

[ 3 63 7 0 0 16 38 3 107 100 29 0 36 0 28 0 66 24

1 70]

[ 7 39 1 56 5 141 2 91 66 22 1 26 57 0 63 0 7 0

2 8]

[ 32 45 1 20 4 75 4 42 104 24 8 11 97 2 78 0 25 3

1 17]

[ 9 4 0 116 83 59 1 102 16 4 0 111 31 37 19 0 5 0

0 2]

[ 21 10 0 115 58 60 0 91 20 2 0 100 37 11 16 0 5 0

0 0]

[ 3 9 0 133 30 85 0 139 16 0 1 79 21 13 34 0 0 0

0 1]

[ 23 3 0 75 37 62 1 111 14 0 0 78 31 9 15 0 1 3

0 2]

[ 8 9 0 75 39 58 0 58 14 1 0 67 15 16 17 0 0 0

0 0]]

Homogeneity Score: 0.206

Completeness Score: 0.217

V-measure: 0.212

Adjusted Rand Score: 0.063

Adjusted Mutual Info Score: 0.202

==================================================

non-linear transformation for nmf

==================================================

Confusion Matrix:

[[ 0 5 14 0 5 30 0 124 124 64 0 0 1 0 1 28 26 55

3 0]

[ 97 8 5 56 39 2 120 0 0 0 10 40 60 84 34 17 0 0

9 3]

[ 72 11 0 106 11 0 113 0 0 0 48 31 29 136 26 4 0 0

1 3]

[ 66 2 1 103 10 1 124 0 0 0 65 39 18 113 36 3 0 0

5 4]

[ 87 1 0 47 22 1 132 0 0 0 10 67 30 108 60 9 0 0

3 1]

[111 1 0 53 19 1 172 0 0 0 19 27 56 93 25 6 0 1

3 6]

[125 12 3 45 28 2 110 0 0 0 21 33 98 67 15 18 0 1

7 0]

[ 23 43 31 0 103 50 4 16 10 9 0 38 53 1 5 128 1 21

58 0]

[ 22 60 17 0 56 95 5 6 20 3 0 18 48 1 1 156 0 61

29 0]

[ 10 65 15 0 19 80 4 48 86 18 0 6 25 3 0 82 3 122

11 0]

[ 6 91 6 0 19 48 1 64 126 15 0 2 17 0 1 49 1 152

2 0]

[ 10 11 67 3 67 83 15 24 23 27 2 38 10 4 18 64 6 20

103 0]

[ 99 22 0 9 115 13 49 0 0 0 1 79 73 24 20 54 0 10

23 0]

[ 15 72 19 1 30 63 3 31 71 6 2 12 48 0 1 95 0 114

11 0]

[ 27 78 22 1 54 80 4 13 15 4 1 18 57 3 8 129 3 44

32 0]

[ 1 24 13 0 8 38 3 128 128 101 0 3 7 0 0 26 51 61

7 0]

[ 1 17 20 0 12 61 0 120 87 81 0 3 9 0 1 37 19 65

13 0]

[ 1 37 3 0 2 34 0 118 154 49 0 0 10 0 1 29 15 108

3 0]

[ 3 19 21 0 5 52 1 87 99 48 0 0 3 3 0 32 14 69

9 0]

[ 1 19 10 0 6 35 0 75 63 52 0 0 9 0 0 19 24 62

2 0]]

Homogeneity Score: 0.211

Completeness Score: 0.222

V-measure: 0.217

Adjusted Rand Score: 0.068

Adjusted Mutual Info Score: 0.207

==================================================

non-linear first and then normalization

==================================================

Confusion Matrix:

[[ 20 0 126 4 1 98 0 14 0 1 23 0 90 6 0 1 51 43

0 2]

[ 8 80 0 6 57 0 90 31 3 63 0 55 0 7 112 40 0 0

10 22]

[ 0 134 0 10 46 0 58 11 3 29 0 105 0 1 100 39 0 0

48 7]

[ 2 112 0 1 46 0 56 5 4 20 0 100 0 2 114 45 0 0

65 18]

[ 4 102 0 2 80 1 72 11 1 31 0 43 0 1 121 77 1 0

10 21]

[ 0 93 0 1 42 0 105 13 6 57 0 52 0 1 156 35 1 0

19 12]

[ 5 66 0 11 62 0 108 27 0 99 0 46 1 7 97 23 1 0

20 12]

[ 87 1 17 44 52 26 13 137 0 51 1 0 17 45 4 6 33 9

0 51]

[104 1 21 61 26 49 20 116 0 46 0 0 8 23 4 3 90 2

0 24]

[ 62 3 73 70 10 99 8 62 0 20 2 0 30 22 2 1 118 10

0 5]

[ 27 0 72 93 6 143 4 35 0 21 0 0 43 6 1 1 138 7

0 3]

[ 95 4 33 10 29 45 7 69 0 9 3 3 26 101 13 20 27 27

2 72]

[ 33 23 0 23 101 1 77 106 0 66 0 8 0 3 40 30 12 0

1 67]

[ 58 0 49 69 22 86 7 67 0 50 0 1 25 7 3 1 133 5

2 9]

[ 89 3 20 82 27 34 19 112 0 53 2 1 11 32 4 8 65 4

1 26]

[ 36 0 113 20 4 98 1 16 0 7 37 0 107 9 3 0 60 83

0 5]

[ 43 0 113 19 2 85 1 21 0 9 11 0 99 19 0 1 63 56

0 4]

[ 20 0 135 39 0 126 1 19 0 11 13 0 75 3 0 1 91 30

0 0]

[ 35 3 72 20 0 102 3 17 0 2 9 0 77 22 1 0 66 36

0 0]

[ 15 0 72 19 1 54 0 16 0 10 15 0 65 8 0 0 61 40

0 1]]

Homogeneity Score: 0.207

Completeness Score: 0.218

V-measure: 0.212

Adjusted Rand Score: 0.064

Adjusted Mutual Info Score: 0.203

==================================================

normalization first and then non-linear

==================================================

Confusion Matrix:

[[128 5 0 33 0 2 26 1 0 0 136 0 54 1 1 0 12 50

4 27]

[ 0 33 91 1 52 31 1 4 59 106 0 51 0 51 33 43 18 0

9 1]

[ 0 15 64 0 81 7 0 10 47 149 0 23 0 46 32 114 2 0

1 0]

[ 0 5 81 0 68 22 1 0 38 144 0 21 0 43 32 130 3 0

2 0]

[ 0 18 75 1 101 31 1 0 64 93 0 22 0 67 65 32 7 0

1 0]

[ 0 20 122 0 63 15 2 0 81 119 0 48 0 46 20 54 2 0

1 0]

[ 0 46 96 1 42 13 8 4 67 87 0 78 0 58 15 51 8 1

9 1]

[ 15 103 7 47 2 57 56 17 7 2 12 25 2 45 6 0 97 5

61 28]

[ 7 89 7 90 2 29 107 25 16 1 14 26 1 20 1 0 91 23

28 21]

[ 44 48 4 99 2 8 85 36 6 1 78 14 11 8 1 0 43 75

13 21]

[ 57 25 0 75 1 4 75 65 5 0 118 13 6 4 0 0 20 114

6 12]

[ 17 28 2 65 13 80 29 3 11 4 18 4 22 20 16 2 78 11

113 59]

[ 0 93 38 10 29 84 18 12 68 17 0 40 0 86 25 4 55 2

9 1]

[ 18 66 2 84 0 15 101 36 11 2 53 21 1 18 0 2 48 70

9 37]

[ 10 89 6 65 5 28 89 48 17 2 10 20 4 22 7 2 88 14

41 26]

[146 13 1 40 0 6 33 17 2 0 135 3 104 2 0 0 8 54

11 24]

[130 18 0 71 0 5 26 8 0 0 80 6 51 2 0 0 28 50

17 54]

[115 14 0 50 0 0 37 25 1 0 167 10 33 0 1 0 9 85

3 14]

[ 85 8 1 67 0 0 38 8 3 3 90 1 41 0 0 0 18 52

18 32]

[ 91 11 0 45 0 0 30 13 1 0 59 2 48 1 0 0 12 43

3 18]]

Homogeneity Score: 0.217

Completeness Score: 0.221

V-measure: 0.219

Adjusted Rand Score: 0.068

Adjusted Mutual Info Score: 0.213