Exercise 8

1.0 Given the following Document Term matrix calculate the Item/Item matrix:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Term1 | Term2 | Term3 | Term4 | Term5 | Term6 | Term7 | Term8 |
| Item 1 | 3 | 0 | 2 | 2 | 0 | 0 | 4 | 3 |
| Item 2 | 0 | 0 | 4 | 3 | 2 | 0 | 0 | 2 |
| Item 3 | 2 | 2 | 0 | 2 | 2 | 1 | 0 | 0 |
| Item 4 | 0 | 1 | 0 | 2 | 2 | 0 | 1 | 0 |
| Item 5 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| Item 6 | 2 | 1 | 3 | 4 | 2 | 2 | 0 | 2 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 | Item 6 |
| Item 1 |  |  |  |  |  |  |
| Item 2 |  |  |  |  |  |  |
| Item 3 |  |  |  |  |  |  |
| Item 4 |  |  |  |  |  |  |
| Item 5 |  |  |  |  |  |  |
| Item 6 |  |  |  |  |  |  |

a. Determine the Item Relationship matrix using a threshold of 8 or higher

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 | Item 6 |
| Item 1 |  |  |  |  |  |  |
| Item 2 |  |  |  |  |  |  |
| Item 3 |  |  |  |  |  |  |
| Item 4 |  |  |  |  |  |  |
| Item 5 |  |  |  |  |  |  |
| Item 6 |  |  |  |  |  |  |

b. Determine the clusters using the clique technique

c. Determine the clusters using the single link technique

d. Determine the clusters using the star technique where the item selected for the new seed for the next star is the smallest number item nor already part of a class.

e. If for question a.if you used a threshold of 11 or higher discuss how it would changes the results fromclique, single link, star from above.Generate a new Item relationship matrix based upon 11 as threshold.

The item relationship matrix would look like:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 | Item 6 |
| Item 1 |  |  |  |  |  |  |
| Item 2 |  |  |  |  |  |  |
| Item 3 |  |  |  |  |  |  |
| Item 4 |  |  |  |  |  |  |
| Item 5 |  |  |  |  |  |  |
| Item 6 |  |  |  |  |  |  |

Clique – now is

Single link

Star :

1. K-means clustering
   1. Using the data from problem 1.0 above, assuming K=3 for K-means clustering approach (using existing clusters) – with CL1 = Item1; CL2 = Item3, CL3 = Item5 calculate the new clusters. If there are multiple classes with same number select the class it was in if it’s an option.

Centroid of Class 1 =

Centroid of Class 2 =

Centroid of Class 3 =

SIM(CL1, I1) =

SIM(CL2, I1) =

SIM(CL3, I1) =

I1 belongs to .

SIM(CL1, I2) =

SIM(CL2, I2) =

SIM(CL3, I2) =

I2 belongs to .

SIM(CL1, I3) =

SIM(CL2, I3) =

SIM(CL3, I3) =

I3 belongs to .

SIM(CL1, I4) =

SIM(CL2, I4) =

SIM(CL3, I4) =

I4 belongs to .

SIM(CL1, I5) =

SIM(CL2, I5) =

SIM(CL3, I5) =

I5 belongs to .

SIM(CL1, I6) =

SIM(CL2, I6) =

SIM(CL3, I6) =

I6 belongs to

New Classes are:

Class 1=

Class 2 =

Class 3 =

**REPEAT above until items do not move between classes**

* 1. Compare the results to the clique cluster from a. Why are they different.