

# Graph Partitioning Clustering

Mehrdad Mohammadian 39700248

## Data

	T1	T2	T3	T4
	0.1	0.4	0.1	0.3
	0.1	0.2	0.5	0.4
	0.4	0.1	0.1	0.3
	0.2	0.4	0.1	0.3
	0.2	0.1	0.5	0.4
	0.1	0.1	0.1	0.2

## Calculate Similarity Matrix

$$\text{sim}(T_i, T_j) = \sum_{k=1}^N (w_{ik} \times w_{jk})$$

```
import numpy as np
import pandas as pd
terms = ['t1', 't2', 't3', 't4']
sim_matrix = np.zeros(shape=(len(terms), len(terms)), dtype=float)
data = pd.DataFrame(data={
    't1': [0.1, 0.1, 0.4, 0.2, 0.2, 0.1],
    't2': [0.4, 0.2, 0.1, 0.4, 0.1, 0.1],
    't3': [0.1, 0.5, 0.1, 0.1, 0.5, 0.1],
    't4': [0.3, 0.4, 0.3, 0.3, 0.4, 0.2]
})
for xkey, xterm in enumerate(terms):
    for ykey, yterm in enumerate(terms):
        if yterm != xterm:
            sim_matrix[xkey][ykey] = np.dot(x[xterm], x[yterm])
print(sim_matrix)
```

```
----- OUTPUT -----
array([[0.  , 0.21, 0.23, 0.35],
       [0.21, 0.  , 0.25, 0.41],
       [0.23, 0.25, 0.  , 0.51],
       [0.35, 0.41, 0.51, 0.  ]])
```

**output:**

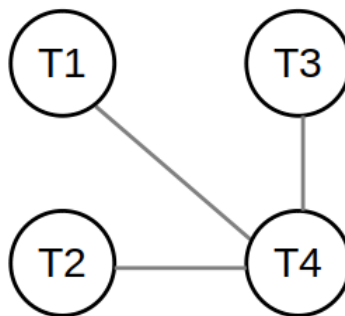
note:  $T1 \text{ dotproduct } T1 = 0$ ,  $T2 \text{ dotproduct } T2 = 0$ , ...

$$\begin{vmatrix} 0 & & & \\ 0.21 & 0 & & \\ 0.23 & 0.25 & 0 & \\ 0.35 & 0.41 & 0.51 & 0 \end{vmatrix}$$

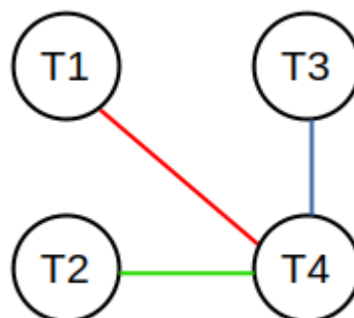
**set threshold:**

threshold = 0.35

$$\begin{vmatrix} 0 & & & \\ 0 & 0 & & \\ 0 & 0 & 0 & \\ 1 & 1 & 1 & 0 \end{vmatrix}$$

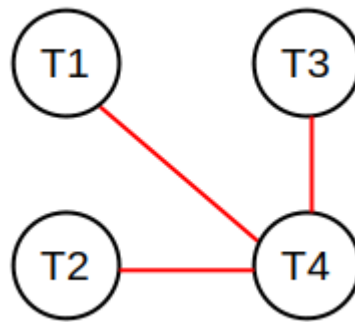


**Apply Clique Method**



$[T1, T4]$   
 $[T3, T4]$   
 $[T2, T4]$

## Apply Single Link Method



$[T4, T3, T2, T1]$

The End  
1400/7/22