Exercises on Clustering

ROCK Example

- Suppose we have four verses contains some subjects , as follows:
- P1={ judgment, faith, prayer, fair}
- P2={ fasting, faith, prayer}
- P3={ fair, fasting, faith}
- P4={ fasting, prayer, pilgrimage}
- **the similarity threshold = 0.3, and number of required cluster is 2.**

Using Jaccard coefficient as a similarity measure, we obtain the following similarity table

	P1	P2	P3	P4
P1	1	0.4	0.4	0.17
P2		1	0.5	0.5
P3			1	0.2
P4				1

ROCK Example

	P1	P2	P3	P4
P1	1	0.4	0.4	0.17
P2		1	0.5	0.5
P3			1	0.2
P4				1

- Since we have a similarity threshold equal to 0.3, then we derive the adjacency table: →

	P1	P2	P3	P4
P1	1	1	1	0
P2		1	1	1
P3			1	0
P4				1

- By multiplying the adjacency table with itself, we derive the following table which shows the number of links (or common neighbors): →

	P1	P2	P3	P4
P1	-	3	3	1
P2		-	3	2
P3			-	1
P4				-

ROCK Example

- we compute the goodness measure for all adjacent points ,assuming that $f(\theta) =$ $1-\theta \: / \: 1+\theta$ = 1-0.3 /1+0.3 = 0.54

$$g(P_i, P_j) = \frac{link[P_i, P_j]}{(n+m)^{1+2f(\theta)} - n^{1+2f(\theta)} - m^{1+2f(\theta)}}$$

- we obtain the following table➔ :

- we have an equal goodness measure for merging ((P1,P2), (P2,P3), (P3,P1))

Pair	Goodness measure
P1,P2	1.35
P1,P3	1.35
P1,P4	0.45
P2,P3	1.35
P2,P4	0.90
P3,P4	0.45

ROCK Example

- Now, we start the hierarchical algorithm by merging, say P1 and P2.

- A new cluster (let's call it C(P1,P2)) is formed.

- It should be noted that for some other hierarchical clustering techniques, we will not start the clustering process by merging P1 and P2, since Sim(P1,P2) = 0.4,which is not the highest. But, ROCK uses the number of links as the similarity measure rather than distance.

ROCK Example

- Now, after merging P1 and P2, we have only three clusters. The following table shows the number of common neighbors for these clusters:→

	C(P1,P2)	P3	P4
C(P1,P2)	-	3+3	2+1
P3		-	1
P4			-

- Then we can obtain the following goodness measures for all adjacent clusters:→

Pair	Goodness measure
C(P1,P2),P3	1.31
C(P1,P2),P4	0.66
P3,P4	0.45

ROCK Example

- Since the number of required clusters is 2, then we finish the clustering algorithm by merging C(P1,P2) and P3, obtaining a new cluster C(P1,P2,P3) which contains {P1,P2,P3} leaving P4 alone in a separate cluster.