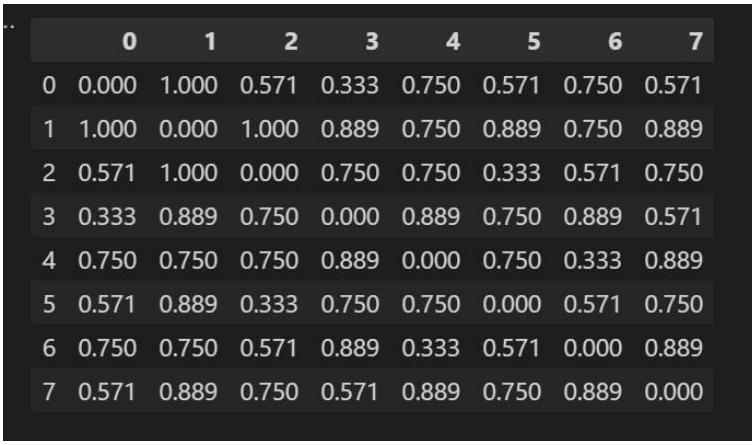


	0	1	2	3	4	5	6	7
0	0.000	1.000	0.571	0.333	0.750	0.571	0.750	0.571
1	1.000	0.000	1.000	0.889	0.750	0.889	0.750	0.889
2	0.571	1.000	0.000	0.750	0.750	0.333	0.571	0.750
3	0.333	0.889	0.750	0.000	0.889	0.750	0.889	0.571
4	0.750	0.750	0.750	0.889	0.000	0.750	0.333	0.889
5	0.571	0.889	0.333	0.750	0.750	0.000	0.571	0.750
6	0.750	0.750	0.571	0.889	0.333	0.571	0.000	0.889
7	0.571	0.889	0.750	0.571	0.889	0.750	0.889	0.000

Figure 1: Pairwise Jaccard distances

2 Task 2

2.1 Task2.a

The code of Jaccard distances compution are below:

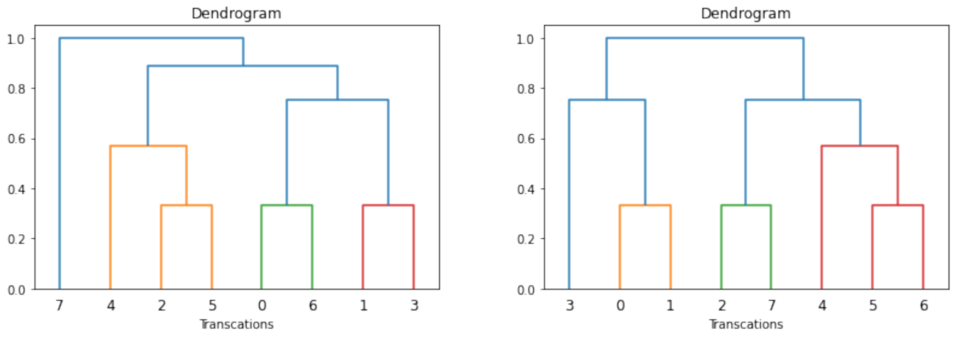
```python
def jaccard_dis(arr1, arr2):
    a = set(arr1)
    b = set(arr2)
    c = a.intersection(b)
    return round(1- float(len(c)) / (len(a) + len(b) - len(c)),3)

def get_all_jdis(num,data):
    j_dis=np.zeros((num,num))
    oned=[]
    for i in range(num):
        temp=[]
        for j in range(i+1,num):
            j_dis[i][j]=j_dis[j][i]=jaccard_dis(data[i],data[j])
            oned.append(j_dis[i][j])
    return oned,pd.DataFrame(j_dis)
```

The pairwise Jaccard distances of data are shown in Fig.1.

2.2 Task2.b

we simulated the agglomerative hierarchical clustering algorithm with the complete linkage metric. It's also possible to obtain 2 different clusterings depending on data order. After shuffling the data several times, we can get the corresponding cluster dendrogram in Fig.2a and Fig.2b.



(a) Dendrogram 1

(b) Dendrogram 2

Figure 2: Dendrograms on complete-link metric

Explain why clusters merged, using Fig.2a as an example:

Firstly, every point will be seen as a cluster in the beginning, then based on the pairwise jaccard distances of data, we can find the smallest distance in the points and merged them into a new cluster. After that, we recompute the distances between clusters and merge the nearest clusters iteratively until there are only one cluster left.

$$\begin{bmatrix} 0 & 6 & 0.333 \\ 1 & 3 & 0.333 \\ 2 & 5 & 0.333 \\ 4 & 10 & 0.571 \\ 8 & 9 & 0.75 \\ 11 & 12 & 0.889 \\ 7 & 13 & 1. \end{bmatrix}$$

From the computation result above we can know, [0,6] firstly merged into new cluster 8, then $[1,3] \rightarrow 9, [2,5] \rightarrow 10$. Now we can see 4 and 10 are merged into new cluster 11, the cluster 10 are also newly formed from the merges before. Repeat this process until there is only one cluster, the algorithm terminated. we now obtain the clustering results in dendrgrams.

2.3 Task2.c

Now, we repeated the task above only with the single linkage metric. The results didn't change and irrelevant to data order. The dendrogram is shown in Fig.3.

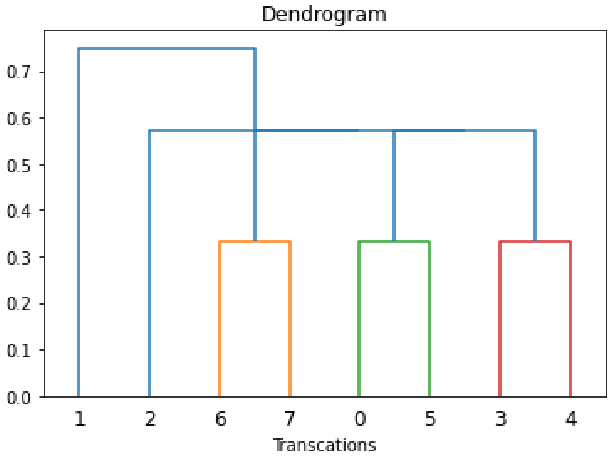


Figure 3: Dendrogram on single-link metric