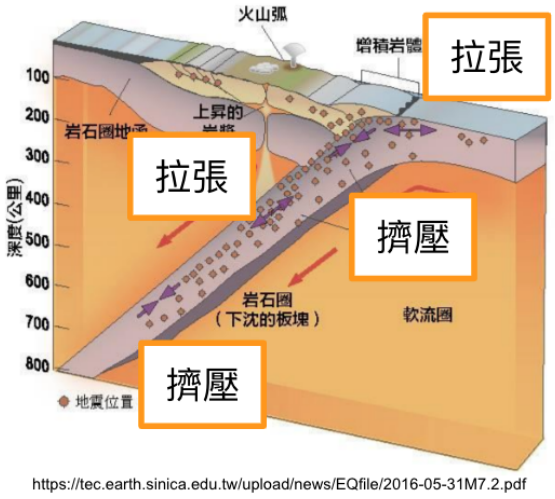
Earthquakes Cluster Applied Ricci Curvature

r09224220 童靖惠 d07921004 張麒仙

Motivation

It’s important to observe earthquake distribution and source properties like extension or collision in subduction zone. Well cluster ways can let us understand the regional properties.

Ricci curvature is one piece of information to classify these. Our objective is using Ricci curvature to classify earthquake happen in Earth to know the regional properties.



https://tec.earth.sinica.edu.tw/upload/news/EQfile/2016-05-31M7.2.pdf

Method

Curvatures on surface

Ricci curvature

Ricci curvature can compute the curvature of graph by using nodes and edges. Given a metric space (X, d). $x, y \in X$. m_x, m_y are two probability measures on X. W is Wasserstein distance.

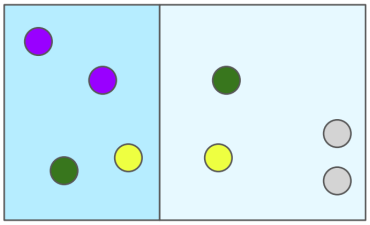
Ricci curvature of x and y is defined by $1-\frac{W(m_x,m_y)}{d(x,y)}$. For m_x, we set probability α to stay the same node, and probability $1-\alpha$ to move to neighbors, and α is a parameter

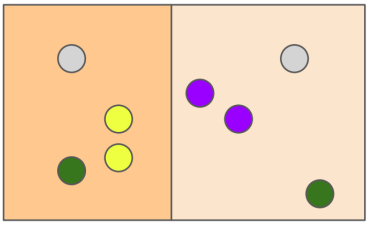
we can decide.

ARI and Modularity

Adjusted Rand Index(ARI)

Before understanding ARI, we can know RI first.





$$RI = \frac{a+d}{a+b+c+d}$$

Given two partitions U_1(blue: ground truth) and U_2(orange: result of clustering).

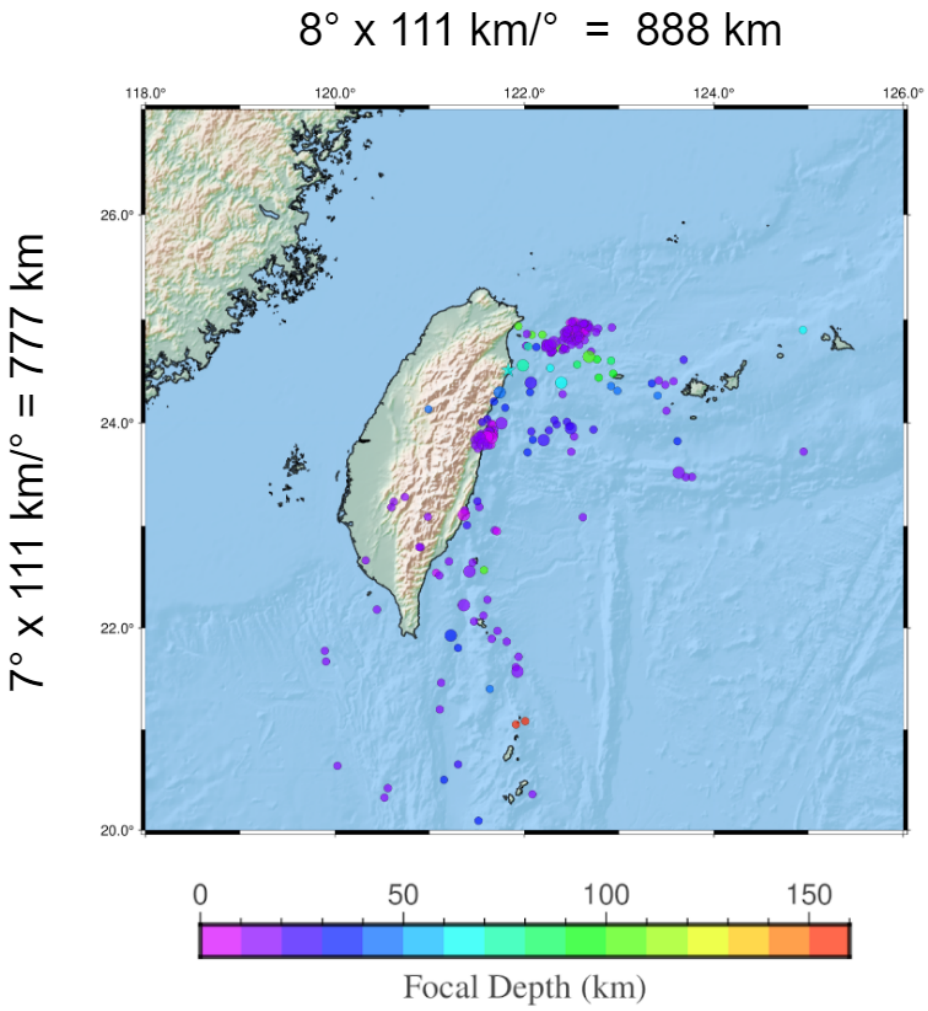
Let a be the number of pairs of objects that are placed in the same class in U_1 and in the same cluster in U_2, b be the number of pairs of objects in the same class in U_1 but not in the same cluster in U_2 , c be the number of pairs of objects in the same cluster in U_1 but not in the same class in U_2, and d be the number of pairs of objects in different classes and different clusters in both partitions. The quantities a and d can be interpreted as agreements, and c and d as disagreements. The Rand index is $\frac{a+d}{a+b+c+d}$. The expected value of RI of two random partitions does not take a constant value. To solve the problem , ARI is introduced. The adjusted Rand index is the corrected-for-chance version of the Rand index. Such a correction for chance establishes a baseline by using the expected similarity of all pair-wise comparisons between clusterings specified by a random model.

Modularity

Modularity is used to detect networks. Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules.

Dataset

We use Taiwan’s earthquake data during 2021. We select the magnitude that is greater than 4 for our data. The number of events are 229. We let longtitude, latitude and depth be our 3-dimension data for clustering.



https://s3-us-west-2.amazonaws.com/secure.notion-static.com/bea93127-f9c4-47c4-860d-e9d504d68244/Event.txt

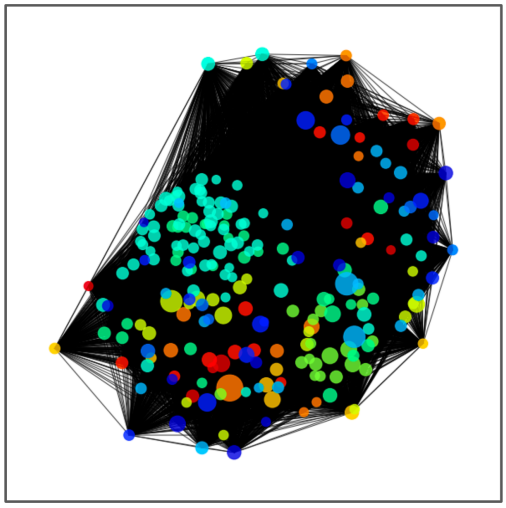
Data pre-processing

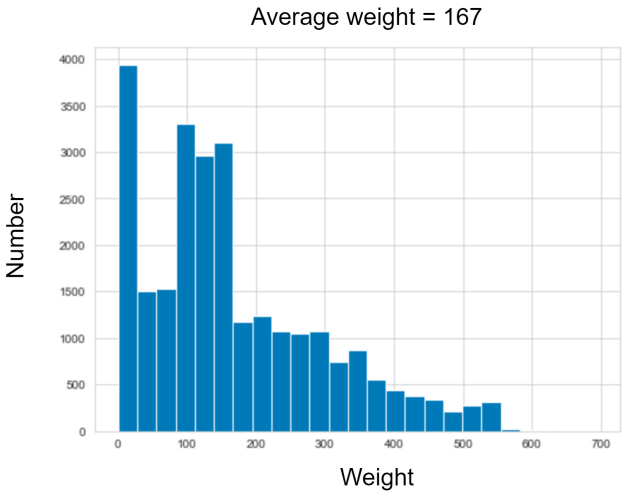
In order to let data to graph structure, we make events as nodes and the distance(in km) between 2 events as edges.

We consider two cases of our experiment:

1. complete graph: the number of edges is 26106 and the average weight is 167

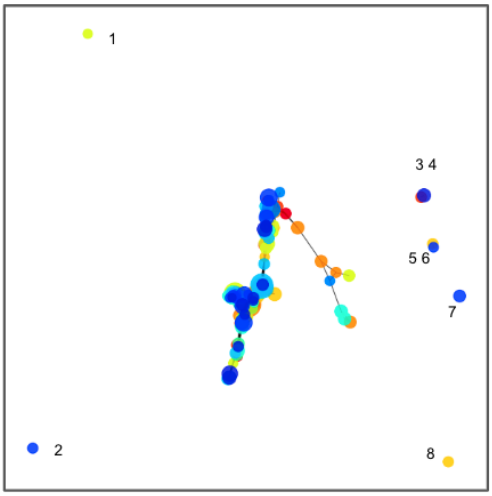
   https://s3-us-west-2.amazonaws.com/secure.notion-static.com/31ee6354-1f88-4419-9d1d-f322e4be8c14/Gtest4.gpickle

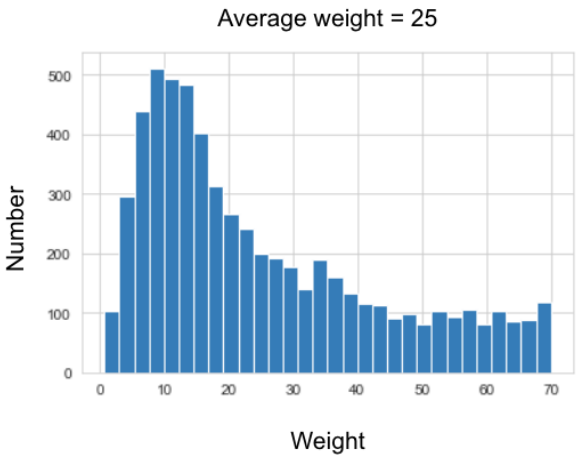




2. the distance between two nodes that is greater than 70 km, not connect edge for the two nodes: the number of edges is 6021 and the average weight is 25

   https://s3-us-west-2.amazonaws.com/secure.notion-static.com/0dbf30c6-52b1-41f0-89f5-77a3663589cd/Gtest7.gpickle





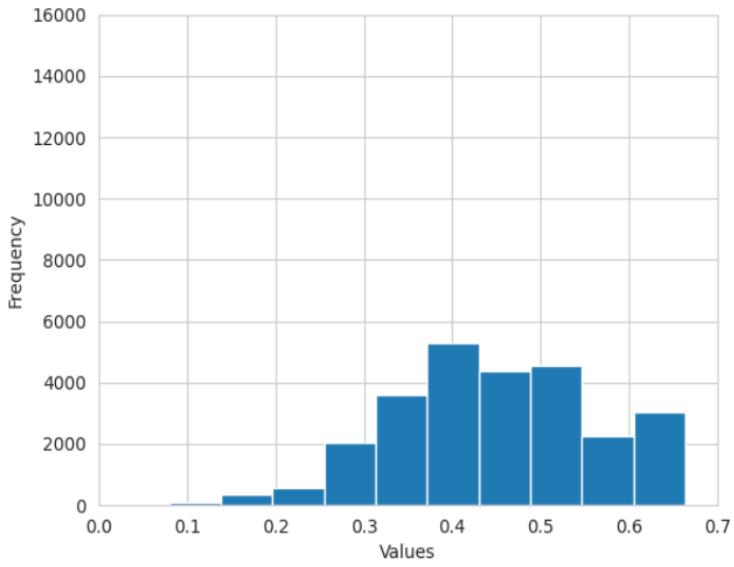
Result and Discussion

Case1: complete graph

We try to adjust the m_x for the neighbors, hope that the nearest neighbors have the larger weight. We consider the m_x for the neighbor is $\frac{max_w - w_i}{sum_w} + \frac{(max_w)(N_{nbr}) - 2sum_w}{(sum_w)(N_{nbr})}$ or e^{-w_i}, surely we confirm m_x for all neighbors is $1 - \alpha$ totally.

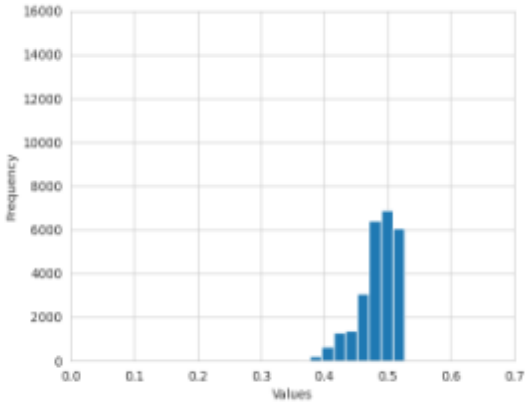
However, this case is still fail to cluster, the number of clustering is one. We try to find why the cluster is fail.

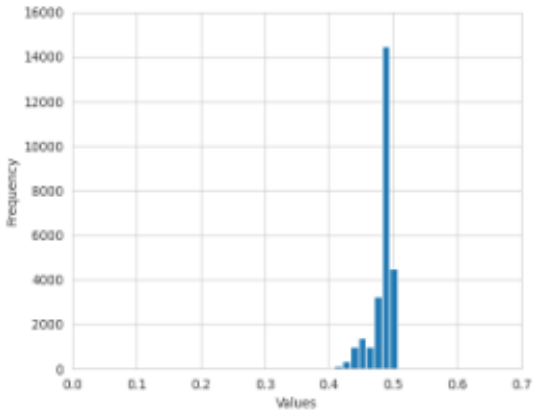
Frrm original ricci curvature historgram, the values is positive which means they are one cluster.



Moreover, we try to apply ricci flow which is an iterative process that aims to smooth out the curvatures of the input graph by adjusting the edges weight, it stretches edges of large negative ricci curvature and shrinks edges of large positive ricci curvature over time.

So, after 5(left figure) and 10(right figure) iterations, ricci curvature is more convergent.

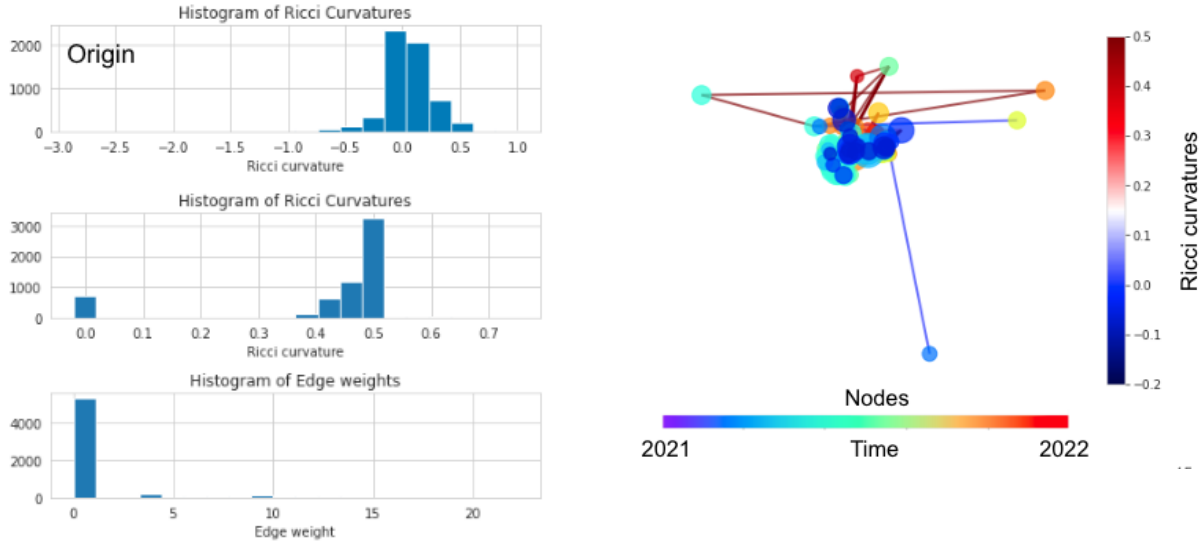




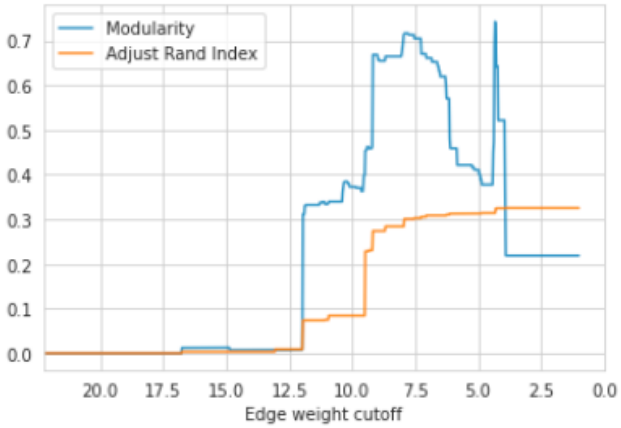
The situation represents this graph is still one cluster over time. We consider the reason is that’s why we can’t sucessful divide them into many clusters.

Case2: adding knowledge

In this case, the range of origin ricci curvatures is -1 - 0.5. After applying ricci flow, the ricci curvatures are divided into 2 clusters in the histogram. The weight after iterations means ricci curvatures variations over time. Although we can't obviously the cluster effect on the right figure, we can set the threshold to detect the graph.

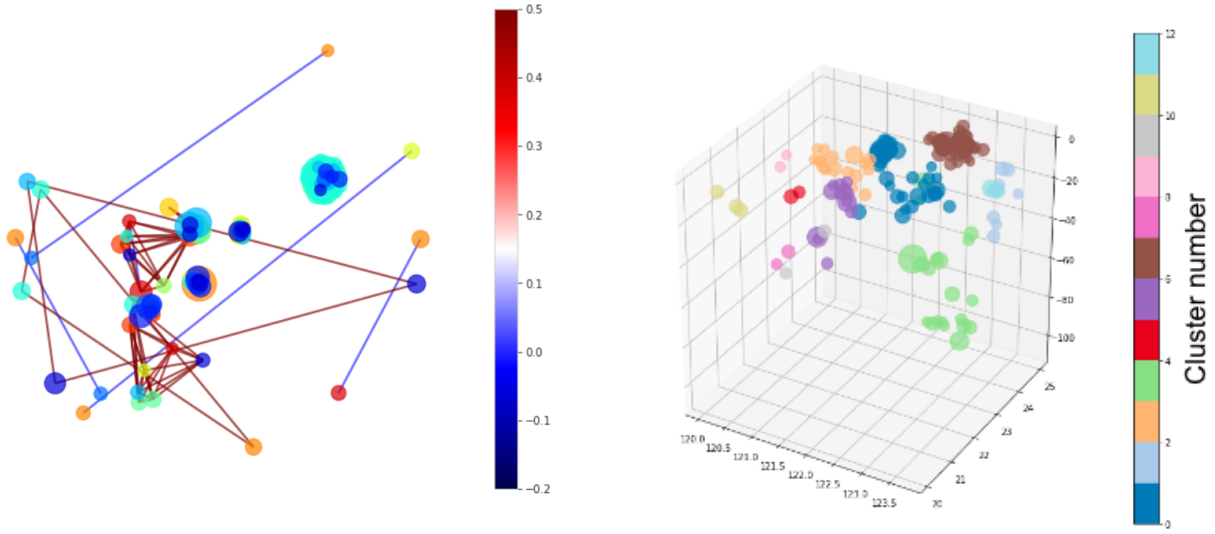


In order to detect the community of graph, we set the threshold and also refer to ARI & Modularity. Although we hope that the 2 values as high as possible, we still think about clustering is subjective. We still decide the final result.



The clustering result of case 2 shows below, the number of clustering is 12 when setting the threshold as 7.95. We can clearly observe the well clustering results in the 3-D view.

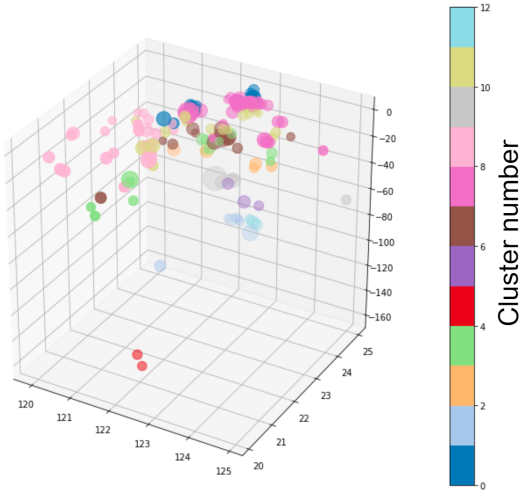
https://s3-us-west-2.amazonaws.com/secure.notion-static.com/39dc28ee-8109-4044-8786-b7a54b2b0453/Gtest7_goodcut.gpickle



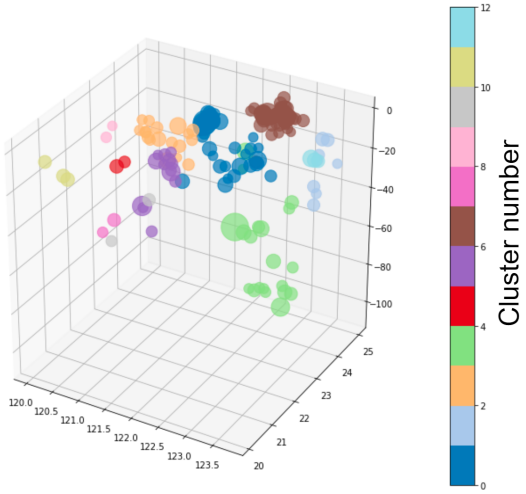
We compare the result with Kmeans clustering, the results are shown in below. We think the result of Ricci curvature is better than Kmeans.

However, if some of nodes don’t have any edge, after calculating ricci curvature, the nodes will be disappear in the final result.

Kmean



Ricci curvature



Code

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/3cc5dcb6-b8dc-4173-b5c6-d2f264e69e38/01_CreateGraph.ipynb

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/778e30df-db59-4f63-882e-011774a760e8/02_RicciCurvature.ipynb

PPT

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/7dee53ca-9908-44c5-bea1-1feeea846290/DAML_Final_Project.pdf