**HW2 Community Detection**

**M10602141\_李弈鴻**

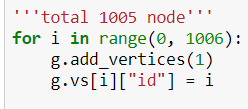
**Development :**

Language: Python( lib: igraph )

Tools: jupyter

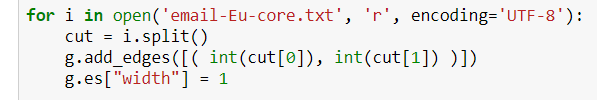
**Code :**

Step1



知道點的編號是流水編號從0到1005，創出這些node

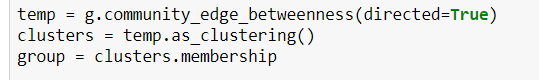
Step2



讀取email-Eu-core.txt，將每筆檔案切開，source\_id、target\_id

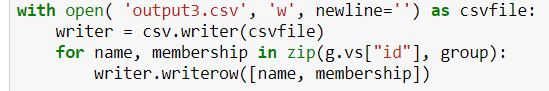
進行node跟node之間的連接

Step3

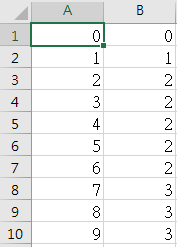


這邊用igraph裡，communuty\_edge\_betweenness() 函式，透過betweenness來進行分群，再以mebership顯示是哪一群的

Step4



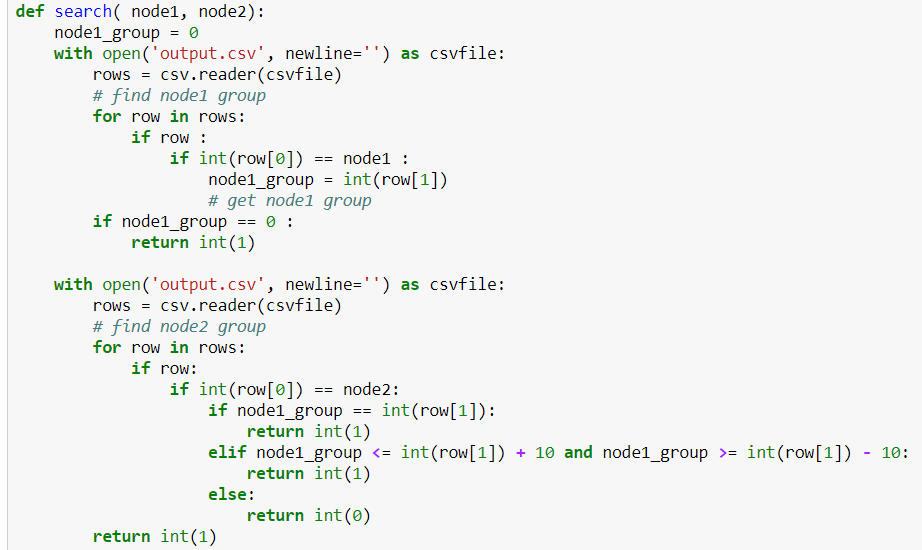
寫檔寫出來，每一個node所屬於哪一個group



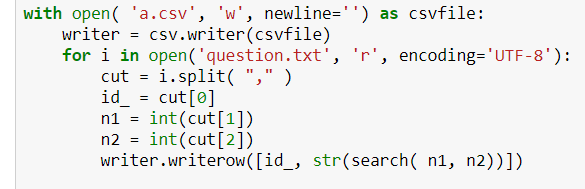
->output檔案樣子A欄是id，B欄是group

Step5

寫另一個比較程式



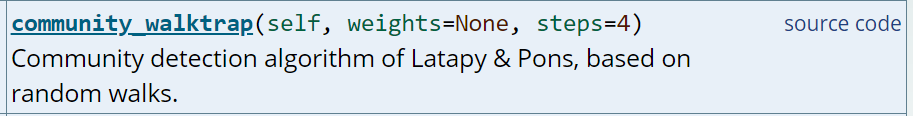
->search function



寫入a.csv，讀取question.txt 每一行，將node1，node2丟進去search()比較

**Problems:**

這邊用communuty\_edge\_betweenness() 跑出來的準確率只能到達53%，之後用了



這個演算法才能達到7X%的準確率。

**Learn in this homework:**

每個dataset 可能適合不同的演算法，要去嘗試才可以知道哪個準確率比較高。

還有也知道python跑的速度真的是蠻慢的，創node跟edges connect速度很快，可是跑演算法時就要跑一段時間，大概要3~4個小時才會跑完。