**社群媒體資料分析實務 HW2**

**M10915078 資工碩一 游紹宏**

* **工具：Python**
* **框架：NetworkX**
* **演算法：Louvain Algorithm**

1. **模組度定義：社群內部邊權重和減去與社群相連的邊權重和**

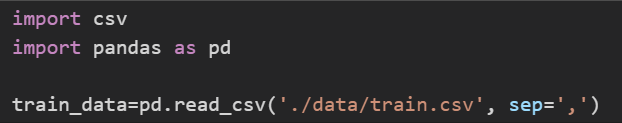
評估一個社群網路劃分好壞的度量方法，物理含義是計算社群內節點的連邊數與隨機情況下的邊數之差距，取值範圍是 。

1. **Louvain Algo.：基於模組度的圖演算法模型**
2. 初始時將每個頂點當作一個社群，社群個數與頂點個數相同。
3. 依次將每個頂點與之相鄰頂點合併在一起，計算它們的模組度增益是否大於0，如果大於0，就將該結點放入該相鄰結點所在社群。
4. 迭代第二步，直至演算法穩定，即所有頂點所屬社群不再變化。
5. 將各個社群所有節點壓縮成為一個結點，社群內點的權重轉化為新結點環的權重，社群間權重轉化為新結點邊的權重。
6. 重複步驟1-3，直至演算法穩定。
7. **程式實作：**

**透過NetworX 的擴充函式 python-louvain 來實作前頁所提及的Louvain Algorithm，藉此來計算出每個不同的社群。**

**並以此為基礎來回答Kaggle題目上所需的問題。**

1. **讀入訓練資料**



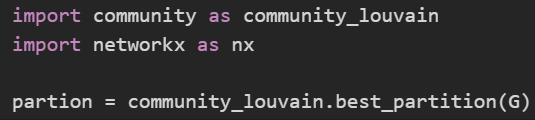
1. **將訓練資料紀載的節點關係載入基於NetworkX的圖形物件來建立模型基礎**



1. **利用 python-louvain 模組來進行社群分析，**

**將原本緊密連結的圖形分割出許多獨立的社群。**

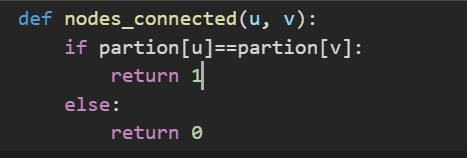
**並將之儲存於名為partion的物件中**



1. **建立判斷節點間連線是否成立的函式：**

**若兩節點位在同一個社群，則回傳值為1**

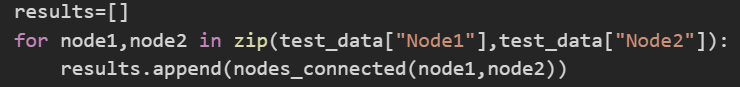
**反之則回傳0 (代表兩者各自處在不同社群)**



1. **讀入要進行測試的資料**



1. **針對測試資料所記載的節點關係進行分析，並儲存結果道明為results的一維陣列**



1. **將結果輸出並儲存為符合Kaggle上傳格式的csv檔案形式**

