Programming Assignment 1
Star War Social Network Analysis

Dataset

- The Star War dataset consists of the social network of Star War characters extracted from movie scripts
    - https://www.kaggle.com/ruchi798/star-wars

- `starwars-full-interactions-allCharacters.json`
    - Contain the social network extracted from all episodes
    - The links between characters are defined by the times the characters speak within the same scene

Dataset

- 112 nodes
  - Character name
  - The number of scenes the character appeared in

```
"nodes" :  112 items
  ▼ [ 100 items
    ▼ 0 : { 3 items
        "name" : string "R2-D2"
        "value" : int 171
        "colour" : string "#bde0f6"
      }
    ▼ 1 : { 3 items
        "name" : string "CHEWBACCA"
        "value" : int 145
```

- 450 links
  - Undirected
  - Number of scenes where the two characters of the link appeared together

```
▼ [ 100 items
  ▼ 0 : { 3 items
      "source" : int 1
      "target" : int 0
      "value" : int 17
    }
  ▼ 1 : { 3 items
      "source" : int 3
```

Task

- Load the social network data and construct the corresponding adjacent matrix
  - Binary weight

- Compute the betweenness values the nodes
  - Note that there can be more than one shortest path of two nodes

- $C_{between}(v_i) = \sum_{s \neq t \neq v_i} \dfrac{\sigma_{st}(v_i)}{\sigma_{st}}$

  the number of shortest paths from node *s* to *t* that pass through v_i

  the number of shortest paths from node *s* to *t*

- (No normalization)

Output

- Save the adjacent matrix as a text file

e.g.

node 0 1 2 3 4 5 ....

0 | 0,1,0,1,0,1,1 …..
1 | 1,0,0,0,1,0,1 …..
2
3
…

matrix.csv

- Show the top 10 betweenness characters

e.g.

top1 | {node編號},{分數}
top2 | {node編號},{分數}
top3 | …
…

rank.csv

- Two weeks to do this programming assignment

Grading

- Matrix (30%)

- Betweenness score and rank (30%)

- Code and one-page report (40%)

- No late submission will be accepted.

- Only when obtaining shortest paths can you use packages.

- 上傳的 matrix.csv 或 rank.csv 若格式有誤，則該項次不予計分。但可於一週內重新上傳結果與程式碼，分數會9折計算。（遲交者不適用）