

Data Mining Project3 作業注意事項

提醒

- 本次作業需自行實作 PageRank、HITS 與 SimRank 演算法
- 需額外提供 project1 使用的 IBM dataset，格式請參考 project1 補交使用的 IBM testing dataset
- 其餘請詳細閱讀 project3 作業說明

輸出規範

- 若輸入的 graph 為 `graph_1.txt`，則必須將 PageRank, HITS 與 SimRank 的結果，輸出成以下檔案
 - `graph_1_PageRank.txt`
 - `graph_1_HITS_authority.txt`
 - `graph_1_HITS_hub.txt`
 - `graph_1_SimRank.txt`
- 每個於作業中有使用到的 graph 都必須產生出以上 4 個檔案，檔案內容範例如下所示
- 寫檔可參考 `numpy.savetxt` (<https://numpy.org/doc/stable/reference/generated/numpy.savetxt.html>)

PageRank

必須在報告與程式碼內提供的值

- Damping Factor d

範例

STDOUT

```
PageRank:
[0.19808502 0.19863 0.20326078 0.19853369 0.20149051]
```

Contents in `PageRank.txt`

- array 元素以 `whitespace` 分隔

0.19808502 0.19863 0.20326078 0.19853369 0.20149051

HITS

範例

STDOUT

```
Authority:
[0.  0.2 0.2 0.2 0.2 0.2]
Hub:
[0.2 0.2 0.2 0.2 0.2 0. ]
```

Contents in HITS_authority.txt

- array 以 whitespace 分隔

```
0.  0.2 0.2 0.2 0.2 0.2
```

Contents in HITS_hub.txt

- array 以 whitespace 分隔

```
0.2 0.2 0.2 0.2 0.2 0.
```

SimRank

必須在報告與程式碼內提供的值

- Decay Factor C

範例

STDOUT

```
SimRank:
[[1. 0. 0. 0. 0.]
 [0. 1. 0. 0. 0.]
 [0. 0. 1. 0. 0.]
 [0. 0. 0. 1. 0.]
 [0. 0. 0. 0. 1.]]
```

Contents in SimRank .txt

```
1 0 0 0 0
0 1 0 0 0
0 0 1 0 0
0 0 0 1 0
0 0 0 0 1
```