

**科目：1071\_資料探勘 DATA MINING**

**學生：楊沛霖**

**學號：Q36071156**

**日期：2018-11-18**

**作業要求：**

1. 利用課程給的八個 Graphs 來計算 HITS、PageRank 和 SimRank。
2. 運算所需的參數為 random jumping probability
3. 討論和分析每個演算法和結果。

## 1. 各圖間的比較

### graph\_1:

是一條線單一方向的圖由  $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6$

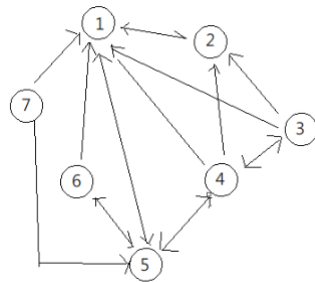
### graph\_2:

是一條線但是方向是雙向的由  $1 \leftrightarrow 2 \leftrightarrow 3 \leftrightarrow 4$

### graph\_3:

是一條線單一方向的圖由  $1 > 2 > 3 \dots > 6$

### graph\_4:



### graph\_5,6:

已過於複雜

### graph\_7,8:

project1 transaction data，這裡取的點與點的連結是利用第 2 行和第三行做連結（第一行永遠和第二行一樣）

ccat hw3dataset/data.ntrans_1.ascii.tlen_5.nitems_1.npats_2		
1	1	0
1	1	6
1	1	7
1	1	8
1	1	9
2	2	4
2	2	5
2	2	6
2	2	8
3	3	0
3	3	7
4	4	4
4	4	6
4	4	7
4	4	8
5	5	0
5	5	5
5	5	7
5	5	8
5	5	9
6	6	0
6	6	1
6	6	3

## 2. Hits

實作方式:

```
HubsAuthorities(G)
1   $\mathbf{1} \leftarrow [1, \dots, 1] \in \mathbb{R}^{|V|}$ 
2   $\mathbf{a}_0 \leftarrow \mathbf{h}_0 \leftarrow \mathbf{1}$ 
3   $t \leftarrow 1$ 
4  repeat
5      for each  $v$  in  $V$ 
6          do  $a_t(v) \leftarrow \sum_{w \in \text{pa}[v]} h_{t-1}(w)$ 
7           $h_t(v) \leftarrow \sum_{w \in \text{ch}[v]} a_{t-1}(w)$ 
8           $\mathbf{a}_t \leftarrow \mathbf{a}_t / \|\mathbf{a}_t\|$ 
9           $\mathbf{h}_t \leftarrow \mathbf{h}_t / \|\mathbf{h}_t\|$  normalization
10          $t \leftarrow t + 1$ 
11 until  $\|\mathbf{a}_t - \mathbf{a}_{t-1}\| + \|\mathbf{h}_t - \mathbf{h}_{t-1}\| < \epsilon$ 
12 return  $(\mathbf{a}_t, \mathbf{h}_t)$ 
```

1. 資料初始化,看有幾個點就創幾個  $\mathbf{a}$  和  $\mathbf{h}$  而他們的值皆為 1(1.2.3 行)
2. 對每個有連接到  $\mathbf{a}_t(v)$  的點的  $\mathbf{h}_{t-1}(w)$  作加總(第 6 行)
3. 對每個有連接到  $\mathbf{h}_t(v)$  的點的  $\mathbf{a}_{t-1}(w)$  作加總(第 7 行)
4. 正規化(第 8.9 行)
5. 重複 2-5 直到  $\mathbf{a}$  和  $\mathbf{h}$  與前項相減絕對值得加總小於誤差值
6. 取得  $\mathbf{a}_t$  和  $\mathbf{h}_t$

D 設定為：0.15

正規化方式為：平方後加總在開根號（類似於 M S E 或是空間距離的算

法）

**graph\_1:**

時間與 cpu：

0.46s user 0.19s system 14% cpu 4.626 total

結果：

收斂執行次數 : 3

收斂結束時 Error : 0.9190116821894447

```
authority page : {'1': 0.5, '2': 0.5, '3': 0.5, '4': 0.5, '5': 0.0, '6': 0}  
hub : {'1': 0.5, '2': 0.5, '3': 0.5, '4': 0.5, '5': 0.0, '6': 0}
```

## graph\_2:

時間與 cpu :

0.46s user 0.15s system 4% cpu 14.856 total

結果 :

收斂執行次數 : 3

收斂結束時 Error : 0.0

authority page :

```
{'1': 0.4472135954999579, '2': 0.4472135954999579, '3':  
0.4472135954999579, '4': 0.4472135954999579, '5':  
0.4472135954999579}
```

hub page :

```
{'1': 0.4472135954999579, '2': 0.4472135954999579, '3':  
0.4472135954999579, '4': 0.4472135954999579, '5':  
0.4472135954999579}
```

## graph\_3:

時間與 cpu :

0.46s user 0.15s system 4% cpu 14.856 total

結果 :

收斂執行次數 : 3

收斂結束時 Error : 0.24854976622531513

authority page :

```
{'1': 0.3922322702763681, '2': 0.5883484054145521, '3':
```

```
0.5883484054145521, '4': 0.3922322702763681}
```

hub page :

```
{'1': 0.3922322702763681, '2': 0.5883484054145521, '3':  
0.5883484054145521, '4': 0.3922322702763681}
```

## graph\_4:

時間與 cpu :

0.44s user 0.15s system 48% cpu 1.256 total

結果 :

收斂執行次數 : 3

收斂結束時 Error: 0.5012605876917594

authority page :

```
{'1': 0.5063160398440778, '2': 0.23014365447458082, '3':  
0.276172385369497, '4': 0.3222011162644131, '5':  
0.552344770738994, '6': 0.4142585780542455, '7':  
0.18411492357966464}
```

hub page :

```
{'1': 0.5063160398440778, '2': 0.23014365447458082, '3':  
0.276172385369497, '4': 0.3222011162644131, '5':  
0.552344770738994, '6': 0.4142585780542455, '7':  
0.18411492357966464}
```

## graph\_5:

時間與 cpu :

0.45s user 0.15s system 55% cpu 1.056 total

結果 :

收斂執行次數 : 3

收斂結束時 Error: 1.4092964448407446

authority page :

authority page : { '3': 0.04119823789241145, '6': 0.043487028886434305, '1': 0.07781889379677717, '396': 0.03204307391632001, '223': 0.0709525208147086, '412': 0.4463142438344573, '38': 0.25863338232458294, '95': 0.12130592268321146, '179': 0.14648262361746292, '236': 0.12130592268321146, '43': 0.05950856584459431, '109': 0.027465491928274296, '369': 0.05950856584459431, '152': 0.13961625063539435, '204': 0.12130592268321146, '273': 0.10070680373700575, '277': 0.12130592268321146, '321': 0.10070680373700575, '327': 0.11901713168918861, '386': 0.10070680373700575, '439': 0.08468526677884575, '94': 0.12130592268321146, '47': 0.029754282922297155, '192': 0.03204307391632001, '399': 0.013732745964137148, '58': 0.06866372982068572, '103': 0.06408614783264002, '443': 0.06866372982068572, '458': 0.1533489965995315, '5': 0.08468526677884576, '306': 0.1602153695816001, '380': 0.13503866864734862, '7': 0.14877141461148577, '60': 0.14877141461148577, '84': 0.16708174256366864, '12': 0.13961625063539435, '258': 0.14419383262344004, '70': 0.038909446898388586, '184': 0.0823964757848229, '11': 0.1373274596413715, '113': 0.029754282922297155, '119': 0.05493098385654859, '323': 0.018310327952182855, '463': 0.038909446898388586, '37': 0.0869740577728686, '85': 0.04119823789241145, '134': 0.048064610874480024, '176': 0.0869740577728686, '182': 0.0869740577728686, '186': 0.05721977485057143, '224': 0.08468526677884575, '249': 0.013732745964137148, '254': 0.07553010280275431, '274': 0.13274987765332574, '285': 0.11901713168918862, '303': 0.06408614783264002, '325': 0.057219774850571455, '440': 0.057219774850571455, '348': 0.1304610866593029, '115': 0.09155163976091432, '185': 0.061797356838617166, '212': 0.061797356838617166, '283': 0.12359471367723432, '293': 0.10986196771309717, '297': 0.09384043075493717, '299': 0.09384043075493717, '315': 0.0709525208147086, '326': 0.07553010280275431, '371': 0.10528438572505146, '416': 0.07553010280275431, '426': 0.10528438572505146, '450': 0.07553010280275431, '456': 0.07553010280275431, '56': 0.09155163976091434, '284': 0.05035340186850288, '130': 0.01144395497011429, '148': 0.03662065590436573, '381': 0.01144395497011429, '2': 0.004577581988045716, '202': 0.018310327952182855, '227': 0.12359471367723433, '253': 0.013732745964137148, '259': 0.006866372982068574, '264': 0.02859911894620572, '4': 0.11672834069516574, '344': 0.09384043075493717, '363': 0.09384043075493717, '421': 0.09384043075493717, '454': 0.07781889379677717, '265': 0.05493098385654859, '300': 0.0869740577728686, '413': 0.038909446898388586, '351': 0.04577581988045717, '307': 0.05493098385654859, '235': 0.027465491928274296, '174': 0.027465491928274296, '13': 0.03662065590436573, '27': 0.0, '31': 0.0, '106': 0.0, '222': 0.0, '34': 0.0, '72': 0.0, '81': 0.0, '141': 0.0, '261': 0.0, '96': 0.0, '124': 0.0, '131': 0.0, '143': 0.0, '162': 0.0, '199': 0.0, '414': 0.0, '219': 0.0, '221': 0.0, '345': 0.0, '448': 0.0, '468': 0.0, '8': 0, '168': 0, '9': 0, '10': 0, '14': 0, '296': 0, '336': 0, '40': 0, '136': 0, '217': 0, '287': 0, '457': 0, '15': 0, '46': 0, '187': 0, '191': 0, '244': 0, '436': 0, '444': 0, '26': 0, '256': 0, '398': 0, '449': 0, '459': 0, '469': 0, '59': 0, '86': 0, '107': 0, '180': 0, '195': 0, '239': 0, '271': 0, '292': 0, '305': 0, '312': 0, '364': 0, '370': 0, '397': 0, '409': 0, '437': 0, '446': 0, '447': 0, '462': 0, '465': 0, '140': 0, '183': 0, '188': 0, '225': 0, '238': 0, '245': 0, '334': 0, '338': 0, '342': 0, '358': 0, '360': 0, '368': 0, '379': 0, '382': 0, '432': 0, '433': 0, '453': 0, '65': 0, '207': 0, '232': 0, '276': 0, '407': 0, '466': 0, '61': 0, '122': 0, '282': 0, '25': 0, '88': 0, '104': 0, '133': 0, '154': 0, '203': 0, '314': 0, '404': 0, '24': 0, '35': 0, '92': 0, '167': 0, '126': 0, '267': 0, '372': 0, '68': 0, '145': 0, '343': 0, '55': 0, '79': 0, '156': 0, '164': 0, '193': 0, '216': 0, '229': 0, '247': 0, '266': 0, '278': 0, '353': 0, '22': 0, '137': 0, '422': 0, '389': 0,

hub page :

hub : { '3': 0.04119823789241142, '5': 0.08468526677884572, '1': 0.07781889379677713, '2': 0.004577581988045714, '223': 0.07095252081470857, '227': 0.12359471367723428, '235': 0.027465491928274285, '254': 0.07553010280275428, '412': 0.44631424383445706, '70': 0.03890944689838857, '94': 0.1213059226832114, '152': 0.13961625063539426, '273': 0.10070680373700569, '277': 0.1213059226832114, '369': 0.05950856584459428, '386': 0.10070680373700569, '192': 0.03204307391632, '202': 0.018310327952182855, '38': 0.25863338232458283, '184': 0.08239647578482284, '458': 0.1533489965995314, '12': 0.13961625063539426, '84': 0.1670817425636686, '258': 0.14419383262344, '306': 0.16021536958160001, '380': 0.13503866864734856, '43': 0.05950856584459428, '109': 0.027465491928274285, '58': 0.06866372982068572, '103': 0.06408614783264, '443': 0.06866372982068572, '56': 0.09155163976091427, '204': 0.1213059226832114, '321': 0.10070680373700569, '327': 0.11901713168918855, '439': 0.0846852667788457, '264': 0.02059911894620571, '307': 0.05493098385654857, '7': 0.1487714146114857, '60': 0.1487714146114857, '148': 0.03662065590436571, '274': 0.1327498776533257, '37': 0.08697405777286857, '115': 0.09155163976091428, '176': 0.08697405777286857, '182': 0.08697405777286857, '186': 0.05721977485057143, '224': 0.08468526677884572, '285': 0.11901713168918857, '293': 0.10986196771309714, '325': 0.05721977485057143, '348': 0.13046108665930284, '440': 0.05721977485057143, '85': 0.04119823789241143, '174': 0.027465491928274285, '284': 0.05035340186850286, '303': 0.06408614783264, '134': 0.048064610874479996, '179': 0.14648262361746284, '249': 0.013732745964137143, '13': 0.03662065590436571, '212': 0.06179735683861714, '297': 0.09384043075493714, '299': 0.09384043075493714, '326': 0.0755301028027543, '416': 0.0755301028027543, '450': 0.0755301028027543, '456': 0.0755301028027543, '47': 0.02975428292229714, '11': 0.1373274596413714, '399': 0.013732745964137143, '185': 0.06179735683861714, '283': 0.12359471367723428, '315': 0.07095252081470857, '371': 0.10528438572505143, '426': 0.10528438572505143, '6': 0.043487028886434284, '396': 0.03204307391632, '130': 0.011443954970114283, '381': 0.011443954970114283, '259': 0.006866372982068571, '253': 0.013732745964137143, '113': 0.02975428292229714, '4': 0.11672834069516569, '300': 0.08697405777286855, '344': 0.09384043075493713, '351': 0.04577581988045714, '363': 0.09384043075493713, '421': 0.09384043075493713, '454': 0.07781889379677713, '119': 0.05493098385654857, '323': 0.018310327952182855, '463': 0.038909446898388565, '95': 0.12130592268321141, '236': 0.12130592268321141, '265': 0.05493098385654857, '413': 0.038909446898388565, '27': 0.0, '31': 0.0, '106': 0.0, '222': 0.0, '34': 0.0, '72': 0.0, '81': 0.0, '141': 0.0, '261': 0.0, '96': 0.0, '124': 0.0, '131': 0.0, '143': 0.0, '162': 0.0, '199': 0.0, '414': 0.0, '219': 0.0, '221': 0.0, '345': 0.0, '448': 0.0, '468': 0.0, '8': 0, '168': 0, '9': 0, '10': 0, '14': 0, '296': 0, '336': 0, '40': 0, '136': 0, '217': 0, '287': 0, '457': 0, '15': 0, '46': 0, '187': 0, '191': 0, '244': 0, '436': 0, '444': 0, '26': 0, '256': 0, '398': 0, '449': 0, '459': 0, '469': 0, '59': 0, '86': 0, '107': 0, '180': 0, '195': 0, '239': 0, '271': 0, '292': 0, '305': 0, '312': 0, '364': 0, '370': 0, '397': 0, '409': 0, '437': 0, '446': 0, '447': 0, '462': 0, '465': 0, '140': 0, '183': 0, '188': 0, '225': 0, '238': 0, '245': 0, '334': 0, '338': 0, '342': 0, '358': 0, '360': 0, '368': 0, '379': 0, '382': 0, '432': 0, '433': 0, '453': 0, '65': 0, '207': 0, '232': 0, '276': 0, '407': 0, '466': 0, '61': 0, '122': 0, '282': 0, '25': 0, '88': 0, '104': 0, '133': 0, '154': 0, '203': 0, '314': 0, '404': 0, '24': 0, '35': 0, '92': 0, '167': 0, '126': 0, '267': 0, '372': 0, '68': 0, '145': 0, '343': 0, '55': 0, '79': 0, '156': 0, '164': 0, '193': 0, '216': 0,

graph\_6:

時間與 cpu :

0.48s user 0.25s system 57% cpu 1.456 total

結果 :

收斂執行次數 : 3

收斂結束時 Error : 0.8876169361023691

資料量太大其他結果放在: some\_output\hits\_G6.txt



## graph\_7(Bidirected graph):

每一點都會互相指

時間與 cpu :

0.44s user 0.14s system 51% cpu 1.127 total

結果 :

收斂執行次數 :3

收斂結束時 Error :0.2246391705315527

authority page :

```
authority page : {'1': 0.23118372816008875, '6': 0.37567355826014426, '8': 0.3708572305901424, '5': 0.3178776262201221, '3': 0.255265366510098, '10': 0.22636740049008694, '11': 0.21191841748008136, '12': 0.22155107282008507, '13': 0.10595920874004067, '0': 0.28897966020011095, '7': 0.3082449708801184, '9': 0.26971434952010365, '2': 0.21673474515008323, '4': 0.21673474515008323}
```

hub page:

```
hub : {'0': 0.28897966020011095, '6': 0.3756735582601442, '7': 0.3082449708801183, '8': 0.37085723059014236, '9': 0.26971434952010354, '1': 0.23118372816008873, '5': 0.31787762622012206, '3': 0.255265366510098, '10': 0.2263674004900869, '11': 0.2119184174800813, '12': 0.22155107282008502, '13': 0.10595920874004065, '2': 0.21673474515008315, '4': 0.21673474515008315}
```

## graph\_8(directed graph):

時間與 cpu :

0.39s user 0.11s system 74% cpu 0.672 total

結果 :

收斂執行次數 :3

收斂結束時 Error :0.41622469097770454

authority page :

```
authority page : {'6': 0.2970145510386751, '2': 0.32530165113759657, '4': 0.3394452011870573, '7': 0.3677323012859787, '8': 0.353588751236518, '10': 0.28287100098921436, '12': 0.2404403508408322, '3': 0.08486130029676432, '5': 0.2545839008902929, '6': 0.4101629514343609, '11': 0.2121532507419108, '13': 0.08486130029676431, '9': 0.014143550049460718, '0': 0}
```

hub page:

```
hub : {'6': 0.4101629514343609, '7': 0.3677323012859787, '8': 0.353588751236518, '11': 0.21215325074191077, '12': 0.2404403508408322, '13': 0.08486130029676431, '2': 0.32530165113759657, '4': 0.3394452011870573, '5': 0.2545839008902929, '10': 0.28287100098921436, '1': 0.2970145510386751, '3': 0.08486130029676432, '9': 0.014143550049460718, '0': 0}
```

## 3. PageRank

實作方式:

$$PR(P_i) = \frac{(d)}{n} + (1-d) \times \sum_{l_{j,i} \in E} PR(P_j) / \text{Outdegree}(P_j)$$

D(damping factor)=0.1~0.15  
n=|page set|

因此為 Quick reference 所以只要照著這個公式去實作即可。

$\sum_{l_{j,i} \in E} PR(P_j) / \text{Outdegree}(P_j)$  為  $P_j$  的 P R 除上所有射入  $P_j$  點的數量

D 設定為 : 0.15

**graph\_1:**

時間與 cpu :

0.41s user 0.13s system 49% cpu 1.071 total

結果 :

```
{'1': 0.024999999999999998, '2': 0.16666666666666666, '3':  
0.16666666666666666, '4': 0.16666666666666666, '5':  
0.16666666666666666, '6': 0.16666666666666666}
```

**graph\_2:**

時間與 cpu :

0.34s user 0.07s system 92% cpu 0.471 total

結果 :

```
{'1': 0.2, '2': 0.2, '3': 0.2, '4': 0.2, '5': 0.2}
```

**graph\_3:**

時間與 cpu :

0.38s user 0.12s system 72% cpu 0.711 total



結果：

```
{'1': 0.14375, '2': 0.35624999999999996, '3': 0.35624999999999996, '4':  
0.14375}
```

#### **graph\_4:**

時間與 cpu：

```
0.33s user 0.06s system 92% cpu 0.381 total
```

結果：

```
{'1': 0.2946428571428571, '2': 0.1469047619047619, '3':  
0.11654761904761904, '4': 0.07607142857142857, '5':  
0.2683333333333333, '6': 0.05178571428571428, '7':  
0.045714285714285714}
```

#### **graph\_5:**

時間與 cpu：

```
0.41s user 0.05s system 92% cpu 0.781 total
```

結果： 由於數量眾多,只取前面常出現的數值,以下格式為:

(數值,出現次數)

```
(0.0005787389582698751, 12), (0.00044065387348969436, 15),  
(0.00044928419128845566, 17), (0.00041521714734597684, 19),  
(0.00045924225028702635, 20), (0.0005463752665245202, 20),  
(0.0004264392324093816, 20), (0.00039534470504619753, 20),  
(0.0003923240938166311, 24), (0.00036752328582650653, 25),  
(0.0004331023454157782, 30), (0.0003663003663003663, 39)
```

#### **graph\_6:**

時間與 cpu：

```
0.45s user 0.04s system 92% cpu 0.981 total
```

結果： 由於數量眾多,只取前面常出現的數值,以下格式為:

(數值,出現次數)

(0.00015099077090119435, 21), (0.00013248091788613933, 21),  
(0.00015361267397098017, 24), (0.0001452225841476656, 25),  
(0.00013788125555226532, 28), (0.00014312506169183694, 29),  
(0.00015511090429657206, 32), (0.00014983713355048862, 35),  
(0.00014687063750581667, 39), (0.00014378053745928339, 43),  
(0.0001425081433224756, 46), (0.00015224472454326583, 48),  
(0.00014447830198591993, 48), (0.00014877223753445253, 51),  
(0.00013903233494875664, 62)

**graph\_7**(Bidirected graph):

時間與 cpu :

0.45s user 0.17s system 44% cpu 1.380 total

結果：

(0.04283459595959596, 1), (0.10682552566481138, 1),  
(0.11226834673263245, 1), (0.09337945784374355, 1),  
(0.11559536693465265, 1), (0.08672979797979798, 1),  
(0.046207611832611836, 1), (0.07099927849927849, 1),  
(0.07470959595959596, 1), (0.07976911976911975, 1),  
(0.04367784992784993, 1), (0.045988584827870535, 1),  
(0.044762033601319307, 1), (0.036252834467120174, 1)

**graph\_8**(directed graph):

時間與 cpu :

0.44s user 0.17s system 54% cpu 1.114 total

結果：

(0.018303571428571426, 1), (0.020833333333333332, 1),  
 (0.08306547619047619, 1), (0.07901785714285714, 1),  
 (0.07800595238095238, 1), (0.09773809523809524, 1),  
 (0.11038690476190477, 1), (0.11949404761904761, 1),  
 (0.14985119047619047, 1), (0.13973214285714283, 1),  
 (0.010714285714285714, 4)

## 4. SimRank

實作方式：

- (1) 当  $a = b$  时  $s(a, b) = 1$ .
- (2) 当  $\mathcal{I}(a) = \emptyset$  或者  $\mathcal{I}(b) = \emptyset$  时  $s(a, b) = 0$ .
- (3) 其他情况下：

$$s(a, b) = \frac{C}{|\mathcal{I}(a)| |\mathcal{I}(b)|} \sum_{i=1}^{|\mathcal{I}(a)|} \sum_{j=1}^{|\mathcal{I}(b)|} s(\mathcal{I}_i(a), \mathcal{I}_j(b))$$

$$\sum_{i=1}^{|\mathcal{I}(a)|} \sum_{j=1}^{|\mathcal{I}(b)|} s(\mathcal{I}_i(a), \mathcal{I}_j(b))$$

意思為執行每個射入的 node 兩兩做 SimRank 並且加總

C 為：0.6

搜尋深度為：10

**graph\_1:**

時間與 cpu：0.45s user 0.12s system 20% cpu 2.814 total

執行結果：除了 node 相同的為 1 以外的 node 與 node 都為 0

討論：

由於 G\_1 是一個由 1>2>3....>6 一條線的圖，因此除了自己和自

己其他的射入點都為 0

**graph\_2:**

時間與 cpu : 0.37s user 0.83s system 70% cpu 0.605 total

執行結果：除了 node 相同的為 1 以外的 node 與 node 都為 0

### graph\_3:

時間與 cpu : 7.88s user 0.11s system 90% cpu 8.607 total

執行結果：

除了點(1,3)(3,1)(2,4)(4,2)為 0.42856、node 相同的為 1 外，其

他都為 0

### graph\_4:

時間與 cpu : 0.35s user 0.06s system 95% cpu 0.432 total

執行結果：

1 1 1.00000	2 1 0.14265	3 1 0.13302
1 2 0.12686	2 2 1.00000	3 2 0.18629
1 3 0.11582	2 3 0.18811	3 3 1.00000
1 4 0.12226	2 4 0.15064	3 4 0.23882
1 5 0.10949	2 5 0.18843	3 5 0.17004
1 6 0.18525	2 6 0.06686	3 6 0.24574
1 7 0.05927	2 7 0.23441	3 7 0.23191

4 1 0.13715	5 1 0.12743	6 1 0.19490	7 1 0.07940
4 2 0.14390	5 2 0.19681	6 2 0.04250	7 2 0.24530
4 3 0.23338	5 3 0.17861	6 3 0.22250	7 3 0.24427
4 4 1.00000	5 4 0.13604	6 4 0.31125	7 4 0.33181
4 5 0.12268	5 5 1.00000	6 5 0.04347	7 5 0.20189
4 6 0.33181	5 6 0.06712	6 6 1.00000	7 6 0.06361
4 7 0.31125	5 7 0.20497	6 7 0.02250	7 7 1.00000

### graph\_5:

時間與 cpu : 0.35s user 0.06s system 95% cpu 0.432 total

執行結果：除了 node 相同的為 1 以外的 node 與 node 都為 0

## 5. 三種演算法的比較

### Hits:

由於 Hits 的算法就算被多個 node 指但沒有對其它 node 指的點 hub 值會是 0 而如果都指向人而沒被指的話 authority 也會是 0，而相較於 PageRank Hits 是比較容易被作弊的，比如作弊者可以建立一個網頁，頁面內容增加很多指向高質量網頁或者著名網站的網址，這就是一個很好的 Hub 頁面，之後作弊者再將這個網頁連結指向作弊網頁，於是可以提升作弊網頁的 Authority 得分。

### RankPage:

相對於 Hits PageRank 的算法是比較全面性的,因為 Hits 只需要負責兩個網站之間的關係,但是 PageRank 則會牽扯到整個網路(會去抓指向該 node 的 node 的資訊)

### SimRank:

對於 SimRank 的算法比起 Hits 更接近 PageRank,因為他們的關聯度都不單單只是兩個網頁,而是在整個網路上,而 SimRank 更是

將迭代的概念用的更為貫徹.

## 6. 增加 node 1 的 hub, authority, and PageRank

### 1. graph\_1

- Hits:可以再增加一個點在 node1 的前面指向 node1 即可增加他的 hub 和 authority
- PageRank:可以拿掉 node6 即可

### 2. graph\_2

- Hits:只要再隨意找一個 node 並且將它指向 node1 即可
- PageRank: 只要再隨意找一個 node 並且將它指向 node1 即可

### 3. graph\_3

- Hits:只要將 node4 指向 node1 即可
- PageRank: 只要將 node4 指向 node1 即可

## 7. 問題討論

### 1. More limitations about link analysis algo

- 在這些演算法中都必須要設定最大深度才行，否則就會出現永遠跑不完的情況出現，而這個深度的其實也不是越多越好，在這些 dataSet 中常常只要跑個 3.4 次就可以知道有沒有循環的輪迴了。

## 2. What are practical issues when use in real web

- 在真實的 web 中應該沒辦法像我們的 dataset 一樣可以看到他的完整有向圖，因為你在逛網頁的時候那個網頁通常只知道你去了哪裡，而沒辦法知道你是從哪裡來(in\_edges)

## 3. What do the result say for your actor/movie graph

- 我們可以從演算法裡面得知這個 graph 的哪個 node 有著比其他 node 都還要高的權重(重要性),比起人用肉眼看還要準確並且有明確性

## 4. Any new idea about the link analysis algorithm?

- 對於六度空間分隔理論，這種人與人資間的 social network 就很適用利用 link analysis algorithm 來做優化，雖然就目前而言該理論一直飽受爭議，但是現在有了 FB 等等的社群網站多多少少都能夠利用此方法來預測了。

## 5. What is the effect of 'C' in SimRank



- C 的值影響性其實沒有我一開始想的還要大，由於在同一個圖中 C 的值基本上是一樣的，因此對其他 node 來說反而沒意義，除非拿來跟不同的圖的 node 做比較才會顯示出他的用意。不過在同一個圖中也會因為 C 所以不段的滾動某個較長的邊的 node 值

## 6. Design a new link-based similarity measurement

- 就像第四題所說的我們可以建立一個與人群有關的 social network 利用每個人可以走 6 步的方式來記錄此人與周遭的人要透過多少人才能有關聯，這應該是一個非常有趣的例子。

## 8. 心得

這次的 **project** 和前幾次的很不一樣,前面的都可以用肉眼稍微計算一次就可以得到答案,但這次的需要大量的運算,真的很佩服當初可以想出這些演算法的科學家們。而在這次的 **project** 中學到最大的應該就是圖形的應用了吧!原本還在思考要用怎麼樣的方式去把圖給實做出來,但幸好後來找到了 **networkx** 讓圖這邊可以快速地找到 **inoutedge**,真的方便了許多。

還有在 **HITS** 和 **PageRank** 的部分因為要加入 **Project 1** 的 **dataset**，而那個 **dataset** 的 **parse** 方法又與原本的不一樣因此還有稍微做修改，而比較有趣的地方是第七張圖要求的是雙向圖，因此在放點的連線的時候左邊連到右邊，同時右邊還要連到左邊，也就是 **key** 變 **value** 而 **value** 變成 **key**。