Data Mining Project 3

P76061425 林聖軒

Usage

link_analysis.py

```
$ python3 link_analysis.py [-h]
```

optional Options	Description
-h --help	show this help message and exit
-f GRAPH_FILE	graph file,(default="./hw3dataset/graph_1.txt")
-mode MODE	ha=HubsAuthorities, pr=PageRank, sr=SimRank, all=all above, (default=all)
-d D	PageRank d, (default=0.1)
-c C	SimRank c, (default=0.8)

Implementation detail

三種演算法都寫在link_analysis.py檔案中。

- HITS
  依照投影片所寫的演算法，如下

  利用兩層for迴圈進行計算，第一層回圈對每一個node迭代，第二層迴圈則計算單個node的
  authorites值和hub值，authorites用該node每個parent的hub值相加，hub則用該node每個
  chid的authorites值相加，再對所有的authorites值和hub值除以2norm來做normalization，
  一直迭代到authorites值和hub值前一次結果差值的2norm加總小於epsilon(這邊設為1e-10)則
  結束迭代。

- PageRank
  同樣依照投影片所寫的公式計算，如下

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

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

對每個node做迭代，用上面的公式計算pageRank值，D值設定為0.1，並做2norm normalization，一直迭代到和前一次結果差值的2norm加總小於epsilon(這邊設為1e-10)則結束迭代。

- SimRank

$$S(a,b) = \frac{C}{|I(a)||I(b)|} \sum_{i=1}^{|I(a)|} \sum_{j=1}^{|I(b)|} S(I_i(a), I_j(b))$$

依照上面的公式定義，對每個點與其它所有的計算相似度，給定初始值後，用一個二維矩陣來做計算，對每個點迭代計算與其它所有不同點的結果，迭代到與上次誤差不大則結束迭代。

Result analysis and discussion

以下呈現 graph 1~6及IBM的data的directed、bidirected結果

graph_1.txt

- HITS

```
HubsAuthorities:
authorities:
best node: 2 value: 0.447213595499958
{1: 0.0, 2: 0.447213595499958, 3: 0.447213595499958, 4: 0.447213595499958, 5: 0.447213595499958, 6: 0.447213595499958}
hub:
best node: 1 value: 0.447213595499958
{1: 0.447213595499958, 2: 0.447213595499958, 3: 0.447213595499958, 4: 0.447213595499958, 5: 0.447213595499958, 6: 0.0}
```

- PageRank

```
PageRank:
best node: 6 value: 0.8098604354579024
{1: 0.02987343505244505, 2: 0.07806422962433922, 3: 0.15580395647032097, 4: 0.2812110011932284, 5: 0.4835133182936731, 6: 0.8098
604354579024}
```

- SimRank

```
SimRank:
===========================
1 simRank:
  1 : 1.0
===========================
2 simRank:
  2 : 1.0
===========================
3 simRank:
  3 : 1.0
===========================
4 simRank:
  4 : 1.0
===========================
5 simRank:
  5 : 1.0
===========================
6 simRank:
  6 : 1.0
===========================
```

graph_2.txt

對每個node做迭代，用上面的公式計算pageRank值，D值設定為0.1，並做2norm

- HITS

```
HubsAuthorities:
authorities:
best node: 1 value: 0.447213595499958
{1: 0.447213595499958, 2: 0.447213595499958, 3: 0.447213595499958, 4: 0.447213595499958, 5: 0.447213595499958}
hub:
best node: 1 value: 0.447213595499958
{1: 0.447213595499958, 2: 0.447213595499958, 3: 0.447213595499958, 4: 0.447213595499958, 5: 0.447213595499958}
```

- PageRank

```
PageRank:
best node: 1 value: 0.4472135954999579
{1: 0.4472135954999579, 2: 0.4472135954999579, 3: 0.4472135954999579, 4: 0.4472135954999579, 5: 0.4472135954999579}
```

- SimRank

```
SimRank:
==============================
1 simRank:
  1 : 1.0
==============================
2 simRank:
  2 : 1.0
==============================
3 simRank:
  3 : 1.0
==============================
4 simRank:
  4 : 1.0
==============================
5 simRank:
  5 : 1.0
==============================
```

graph_3.txt

- HITS

```
HubsAuthorities:
authorities:
best node: 2 value: 0.6015009550106639
{1: 0.37174803445513915, 2: 0.6015009550106639, 3: 0.6015009550106639, 4: 0.37174803445513915}
hub:
best node: 2 value: 0.6015009550106639
{1: 0.37174803445513915, 2: 0.6015009550106639, 3: 0.6015009550106639, 4: 0.37174803445513915}
```

- PageRank

```
PageRank:
best node: 2 value: 0.6288503045177238
{1: 0.3233377406180201, 2: 0.6288503045177238, 3: 0.6288503045177238, 4: 0.3233377406180201}
```

- SimRank

```
SimRank:
==============================
1 simRank:
  1 : 1.0
  3 : 0.666665950838784
==============================
2 simRank:
  2 : 1.0
  4 : 0.6666663803355136
==============================
3 simRank:
  1 : 0.666665950838784
  3 : 1.0
==============================
4 simRank:
  2 : 0.6666663803355136
  4 : 1.0
==============================
```

graph_4.txt

- HITS

```
HubsAuthorities:
authorities:
best node: 5 value: 0.500635020035182
{1: 0.34668186714993793, 2: 0.44219353423699814, 3: 0.49913837843536446, 4: 0.34840643183576686, 5: 0.500635020035182, 7: 0.2089
9872237286285, 6: 0.13940770946290396}
hub:
best node: 1 value: 0.6464257201947676
{1: 0.6464257201947676, 2: 0.11208722834189842, 3: 0.2550547508483144, 4: 0.4662086257344565, 5: 0.43118315727820944, 7: 0.16186
24944799001, 6: 0.27394972282179847}
```

- PageRank

```
PageRank:
best node: 1 value: 0.6897307449446073
{1: 0.6897307449446073, 2: 0.3809396414535482, 3: 0.32583077962000323, 4: 0.24718966745450782, 5: 0.4206993927446729, 7: 0.14680
844458168535, 6: 0.11553078354758817}
```

- SimRank

  部分結果

```
SimRank:
=============================
1 simRank:
  1 : 1.0
  2 : 0.36026281492356466
  3 : 0.34895923549390523
  4 : 0.353732790334621
  5 : 0.3376569689372631
  7 : 0.292390591672489
  6 : 0.4150753327709355
=============================
2 simRank:
  1 : 0.36026281492356466
  2 : 1.0
  3 : 0.4067901545437175
  4 : 0.3697456432140367
  5 : 0.4121804875460285
  7 : 0.45405120688760703
  6 : 0.28544007954046635
=============================
3 simRank:
  1 : 0.34895923549390523
  2 : 0.4067901545437175
  3 : 1.0
  4 : 0.4495651492700736
  5 : 0.39005261515889406
  7 : 0.4510372691391691
  6 : 0.4480930294009781
=============================
4 simRank:
  1 : 0.353732790334621
```

graph_5.txt

- HITS

```
HubsAuthorities:
authorities:
best node: 61 value: 0.4913507493678007
{1: 0.0, 8: 5.0551588352906e-21, 11: 5.0551588352906e-21, 168: 9.581214346542798e-21
37547726e-21, 264: 9.974639126503919e-21, 307: 8.624809002850715e-20, 2: 0.0, 9: 1.0
-36, 13: 1.0303114758841021e-36, 14: 1.0303114758841021e-36, 3: 0.0, 6: 8.0538964851
: 1.6711069155232595e-24, 235: 2.4716118411437983e-24, 296: 2.4716118411437983e-24,
47607530715013e-13, 136: 1.946695083121902e-13, 217: 1.3527743382664407e-13, 265: 4.
-13, 300: 3.2806262036016433e-13, 344: 1.3527743382664407e-13, 351: 1.35277433826664
674467060311323e-13, 457: 3.806972617554789e-13, 5: 0.0, 7: 1.5856304994682065e-12, 1
82065e-12, 46: 1.6297473212007835e-11, 187: 1.6297473212007835e-11, 191: 1.353156518
: 3.3146008735269256e-12, 436: 1.802644358606656e-11, 444: 1.802644358606656e-11, 26
```

```
hub:
best node: 274 value: 0.19194388446558547
{1: 1.231867931739032
9e-20, 8: 0.0, 11: 1.1064176421022567e-20, 168: 0.0, 227: 3.5154175655997557e-
, 264: 5.792616904009446e-25, 307: 5.421469103985068e-20, 2: 7.769692186900134e-37, 9: 0.0, 10: 0.0
14: 0.0, 3: 1.3968880625423045e-24, 6: 1.5015218013521525e-24, 219: 1.766623469598047e-49, 223: 1.
8479505468704896e-39, 296: 0.0, 336: 0.0, 4: 3.2942403866017905e-13, 40: 0.0, 217: 0.0, 2
87: 0.0, 300: 3.0488515835869463e-13, 344: 3.095502310032329e-13, 351: 2.7283037668010906e-13, 363:
2.6650010870749144e-13, 457: 0.0, 5: 6.625657820428537e-12, 7: 1.1193244187505134e-11, 12: 1.05502
0.0, 187: 0.0, 191: 0.0, 244: 0.0, 306: 1.0884517999544842e-11, 436: 0.0, 444: 0.0, 26: 0.0, 448:
.7048779425144836e-43, 143: 3.7048779425144836e-43, 256: 0.0, 258: 9.466163877704537e-12, 380: 1.00
449: 0.0, 459: 0.0, 468: 4.56563031613582e-16, 199: 4.224502959165661e-21, 414: 4.684863268578041e
```

- PageRank

```
PageRank:
best node: 96 value: 0.42529688179019226
{1: 0.0003838391336762150 3, 8: 0.00047268061944264985, 11: 0.00047268061944264985, 168: 0.00062584662774
317253511, 253: 0.000911008576494035, 264: 0.001563429435756411, 307: 0.0010755521075624286, 2: 0.0003838
539311733767476, 10: 0.000539311733767476, 13: 0.000539311733767476, 14: 0.000539311733767476, 3: 0.00038
0048748753373705574, 219: 0.0007392657830702924, 223: 0.0006003189596915332, 235: 0.0007392657830702924,
924, 336: 0.0007392657830702924, 4: 0.00038383913367621503, 40: 0.003525892176968348, 136: 0.0010728327760
644683477815, 265: 0.0023321703465915993, 287: 0.0035258921769683348, 300: 0.0018901087898008544, 344: 0.0
: 0.0009091644683477815, 363: 0.0028934340409779115, 454: 0.002863192058449047, 457: 0.0025198666922375489
21503, 7: 0.0004460281737127194 5, 12: 0.0004460281737127194 5, 15: 0.0004460281737127194 5, 46: 0.000976701
7670508563154 14, 191: 0.0008585478053000892, 244: 0.0007033368062918031, 306: 0.0005496214477813752, 436
444: 0.0010802983597001973, 26: 0.0004966705596306926, 448: 0.0004966705596306926, 124: 0.0005595354810
4130159532, 256: 0.0007608006870846 09, 258: 0.0006631287551679 49, 380: 0.0007971100673166266, 398: 0.0001
.0004354568586158453, 459: 0.00091451604559503 68, 468: 0.0010181093196636925, 199: 0.00053700514197487 51
8751, 60: 0.0008424129722406145, 162: 0.0008207315198459613, 469: 0.0008207315198459613, 27: 0.0005227859
418605101132 36, 86: 0.00042841860510113236, 107: 0.00042841860510113236, 180: 0.00042841860510113236, 19
6, 239: 0.00042841860510113236, 271: 0.00042841860510113236, 292: 0.00042841860510113236, 305: 0.0004284
042841860510113236, 364: 0.00042841860510113236, 370: 0.00042841860510113236, 397: 0.00042841860510113230
113236, 437: 0.00042841860510113236, 446: 0.00042841860510113236, 447: 0.00042841860510113236, 462: 0.000
```

- SimRank

  部分結果

```
SimRank:
===========================
1 simRank:
  1 : 1.0
===========================
8 simRank:
  8 : 1.0
  11 : 0.8
  168 : 0.4
  227 : 0.4
  253 : 0.26666666666666666
  264 : 0.26666666666666666
  307 : 0.16
===========================
11 simRank:
  8 : 0.8
  11 : 1.0
  168 : 0.4
  227 : 0.4
  253 : 0.26666666666666666
  264 : 0.26666666666666666
  307 : 0.16
===========================
168 simRank:
  8 : 0.4
  11 : 0.4
  168 : 1.0
  227 : 0.2
  253 : 0.13333333333333333
  264 : 0.26666666666666666
  307 : 0.16
  199 : 0.4
```

graph_6.txt

- HITS

  部分結果

```
HubsAuthorities:
authorities:
best node: 761 value: 0.27506602043419465
{1: 0.0, 6: 0.010177770169822317, 68: 0.02752099908896793, 95: 0.03114541
06947384117, 273: 0.04062637731637039, 298: 0.02958980196848592, 367: 0.0
4419376493018139, 387: 0.04419376493018139, 410: 0.045262460483570105, 41
54: 0.04403932924876352, 578: 0.0399288443048292, 635: 0.0178936096637067
211, 747: 0.0009057381270866483, 748: 0.006931605055739387, 848: 0.040047
301437860955, 897: 0.02155625697145912, 946: 0.04265502389105189, 951: 0.
```

```
hub:
best node: 171 value: 0.15626346514487824
{1: 0.0260417167399385, 6: 0.0, 68: 0.03914212350351502, 95: 0.037128520671219314, 14
.037614247440977015, 298: 0.0, 367: 0.0, 374: 0.0, 387: 0.03072705680926536, 410: 1.5
2e-74, 415: 0.0, 554: 0.0351673802867363, 578: 0.03528431067031816, 635: 0.0, 725: 0.
48: 0.04053340686801797, 848: 0.039107393846534715, 856: 0.0, 897: 0.0398451257615606
9738457967447, 951: 0.0, 955: 0.039398143730637436, 1021: 0.0, 1058: 0.03645904182603
3838068154698983, 7: 0.09528565756995332, 62: 0.08726049011433283, 78: 0.096924841834
```

- PageRank

  部分結果

```
PageRank:
best node: 410 value: 0.2279418820340828
{1: 0.0001804719368261024, 6: 0.05396533804586473, 68: 0.11911735603844277, 95: 0.1426250909284221,
73: 0.1886196014896138, 298: 0.1385005825535572, 367: 0.1757064272346964, 374: 0.2141624240324738, 3
: 0.2279418820340828, 415: 0.2045476854291009, 554: 0.2117013297695143, 578: 0.1836290376121718, 635
 0.03452580516612473, 747: 0.0001943167432122305, 748: 0.04495102564194828, 848: 0.1745598827193314,
897: 0.09580865151423427, 946: 0.19214378561638826, 951: 0.0001943167432122305, 955: 0.1295646448028
8, 1058: 0.030668136656590024, 1084: 0.1975317203837002, 7: 0.0005411799338277944, 62: 0.0043845735
0685, 180: 0.002909838056413321, 225: 0.002346068304347421, 370: 0.002288025616399507, 394: 0.004575
663590261634, 501: 0.003945987710044221, 528: 0.001991467324902724, 609: 0.0005164226215504499, 761:
0.0002403996060773334, 1003: 0.000673941768814534, 1089: 0.001180104909088555, 1121: 0.0005869490199
51108, 1151: 0.004698752266386397, 1227: 0.002763905686121427, 8: 0.0002463189198311188, 79: 0.00071
21479268229111, 139: 0.0004293056618189392, 202: 0.0006732443700952347, 386: 0.0020154684509788474, 5
```

- SimRank

  部分結果

```
--------------------------------
SimRank:
============================
1 simRank:
  1 : 1.0
============================
6 simRank:
  6 : 1.0
  68 : 0.09757035435557052
  95 : 0.09878742351777717
  142 : 0.09508440747535006
  273 : 0.08433220277442827
  298 : 0.09492760507977256
  367 : 0.07334497517895733
  374 : 0.0878705223514029
  387 : 0.0878705223514029
  410 : 0.089067359878046
  415 : 0.08460904358169692
  554 : 0.08968576721302333
  578 : 0.08250744778924138
  635 : 0.10725380944809841
  725 : 0.12745661183132975
  747 : 0.07272727272727274
  748 : 0.1284625826063654
  848 : 0.08189854588433737
  856 : 0.0920092627450845
```

ibmData_directed.txt

- HITS

  部分結果

```
HubsAuthorities:
authorities:
best node: 67992 value: 0.5421788490264103
{4499: 0.0, 38752: 1.7753646570467023e-238, 47063: 6.7776037074e-3
 3.3888018537e-314, 64853: 1.870775370661633e-128, 66214: 1.775364
5514928995e-214, 78463: 6.7776037074e-314, 84054: 1.24164251280752
7738036494772777, 73402: 3.114864620615391e-197, 32435: 6.77760370
314, 53831: 8.831039970316398e-06, 58239: 7.101776198453915e-06, 5
62605: 3.3888018537e-314, 2851: 0.0, 4078: 0.0, 4884: 0.0, 17007:
```

```
hub:
best node: 63977 value: 0.6389067619253529
{4499: 5.707161826550339e-238, 38752: 3.631261612e-314, 47063: 7.2625232233e-314, 50619: 3.631261612e-314, 5547
755167826e-238, 72113: 7.882691696642997e-214, 76897: 3.631261612e-314, 78463: 1.330478731639877e-263, 84054:
402: 0.16017681720216384, 32435: 5.582917388177691e-203, 46235: 3.631261612e-314, 52204: 5.41625073037205e-06,
: 7.2625232233e-314, 62605: 0.0, 2851: 0.0, 4078: 0.0, 4884: 0.0, 17007: 0.0, 30955: 0.0, 33957: 0.0, 40549: 0.
5e-214, 55145: 1.652790566122906e-176, 76342: 3.631261612e-314, 78588: 0.0, 81809: 3.631261612e-314, 86834: 0.0
38327: 7.2625232233e-314, 63127: 0.0, 95903: 0.0, 96296: 0.0, 99081: 0.0, 6263: 0.0, 11305: 0.0, 14137: 0.0, 17
04812218515e-06, 90735: 0.0, 14859: 1.2313941241283674e-140, 20082: 0.0, 22209: 3.631261612e-314, 40098: 1.3304
```

- PageRank

  部分結果

```
hub:
best node: 63977 value: 0.6389067619253529
{4499: 5.707161826550339e-238, 38752: 3.631261612e-314, 47063: 7.2625232233e-31.
314, 63637: 7.66474312914447e-129, 64853: 5.707161826550339e-238, 66214: 1.9023
 76897: 3.631261612e-314, 78463: 1.3304787316398477e-263, 84054: 0.0, 5835: 7.8
, 51713: 2.1160327946174892e-197, 73402: 0.16017681720216384, 32435: 5.58291738
1625073037205e-06, 53831: 2.3175576588198347e-05, 58239: 1.267844344831805e-166
2233e-314, 62605: 0.0, 2851: 0.0, 4078: 0.0, 4884: 0.0, 17007: 0.0, 30955: 0.0,
69296443185e-116, 53853: 4.617573887887693e-214, 54627: 4.617573887887675e-214,
2e-314, 78588: 0.0, 81809: 3.631261612e-314, 86834: 0.0, 89659: 0.0, 90145: 3.6
 30902: 3.631261612e-314, 38327: 7.2625232233e-314, 63127: 0.0, 95903: 0.0, 962
: 0.0, 17989: 0.0, 43590: 2.132484685901079e-203, 63617: 3.3377249982039767e-19
```

ibmData_bidirected.txt

- HITS

  部分結果

HubsAuthorities:
authorities:
best node: 63977 value: 0.4720471944945436
{4499: 2.1823624276636164e-07, 38752: 8.497514104307469e-07, 47063: 4.085286212576949e-06, 50619: 9.49
7, 66214: 0.00010136130397488126, 72113: 4.108894723929904e-05, 76897: 9.388277206909576e-06, 78463: 6
51713: 0.06240944779842597, 73402: 0.07525022928311888, 32435: 0.00031835726806414636, 46235: 0.000330
04: 0.0012452828552891, 60310: 0.00027948030042328225, 62605: 5.5184742716263354e-05, 2851: 2.0509193
5: 1.1934440137020769e-08, 33957: 5.7985028964800645e-08, 40549: 2.817277664323154e-07, 42779: 1.36881
55145: 6.617949585064728e-05, 76342: 8.275698913155277e-05, 78588: 0.000352939033043174, 81809: 7.2982
97239: 4.2710415942943655e-07, 8480: 2.422127310359189e-190, 11464: 6.795790981154236e-190, 30902: 9

hub:
best node: 63977 value: 0.4720471944945436
{4499: 2.1823624276636164e-07, 38752: 8.497514104307469e-07, 47063: 4.085286212576949e-06, 50619: 9.495865679017517e-06, 5547
7, 66214: 0.00010136130397488126, 72113: 4.108894723929904e-05, 76897: 9.388277206909576e-06, 78463: 6.457501377569796e-06, 8
51713: 0.06240944779842597, 73402: 0.07525022928311888, 32435: 0.00031835726806414636, 46235: 0.0003305795389183676, 52204: 0
04: 0.0012452828552891, 60310: 0.00027948030042328225, 62605: 5.5184742716263354e-05, 2851: 2.050919301250246e-11, 4078: 1.0
5: 1.1934440137020769e-08, 33957: 5.7985028964800645e-08, 40549: 2.817277664323154e-07, 42779: 1.368810801259135e-06, 43211:
55145: 6.617949585064728e-05, 76342: 8.275698913155277e-05, 78588: 0.000352939033043174, 81809: 7.298258899846174e-05, 86834:
97239: 4.2710415942943655e-07, 8480: 2.422127310359189e-190, 11464: 6.795790981154236e-190, 30902: 9.591094977003596e-190, 3
190, 96296: 3.3168116442633003e-190, 99081: 1.8762958113230832e-190, 6263: 1.1352059807862953e-09, 11305: 5.7491924926113416e
911le-07, 63617: 3.2095339693645255e-06, 70689: 2.7624988400824085e-06, 81504: 1.0780999785676666e-05, 90735: 5.18373218831292
8250034e-05, 40098: 7.103489061553227e-06, 41710: 1.925275125440299e-06, 41899: 2.6469681108853823e-06, 44487: 4.376670111478
79709693957653, 83941: 0.0004998488892566864, 85459: 0.00011421228758701473, 5572: 1.3594867357251557e-221, 6677: 2.639965043
1571156040213e-221, 37227: 5.639307191991125e-221, 38594: 5.639307191991125e-221, 44889: 5.311571156040213e-221, 57125: 4.675
0: 1.3594867357251557e-221, 3932: 4.651161256802376e-07, 5975: 9.712528235650583e-08, 8885: 2.676988680691616e-08, 36783: 2.6
66: 1.7028185733762545e-06, 84242: 3.362298000475737e-07, 674: 4.128340034372763e-12, 5425: 2.0907766463087206e-11, 6003: 2.2
4: 2.133672445683577e-09, 58240: 1.0366581014664757e-08, 59586: 5.036734565668722e-08, 63001: 2.4471625896412907e-07, 75431:
88217: 5.265173144673408e-07, 95105: 1.0396340097112792e-07, 527: 6.045599180817594e-09, 3192: 1.7069648787696414e-08, 10497:
19009: 1.4517158704960894e-06, 20947: 6.991594752763574e-06, 23130: 4.063876506114333e-06, 27400: 1.3589699472279013e-05, 463

- PageRank

部分結果

PageRank:
best node: 24816 value: 0.07222031992484003
{4499: 0.012928871518141058, 38752: 0.008602905857847568, 47063: 0.01716121005071939, 50619: 0.008569815664814963, 5
085169837268885315, 72113: 0.02548563656546946, 76897: 0.008518016764889388, 78463: 0.017069218426641636, 84054: 0.01
98941896, 73402: 0.021186502645173036, 32435: 0.021311385459434593, 46235: 0.0085066261156887995, 52204: 0.0296885818
, 60310: 0.017343512230078767, 62605: 0.004345205221123988, 2851: 0.004462719405186021, 4078: 0.008907134878784528,
.008782627310417512, 40549: 0.008738213535976161, 42779: 0.008088244200984137, 43211: 0.008632575703103531 09, 53853: 0
1398204412, 78588: 0.012803639621249993, 81809: 0.008551499655845379, 86834: 0.01284169090976727, 99659: 0.008577800
3, 11464: 0.011667071118540618, 30902: 0.011656053253872575, 38327: 0.023296360096380804, 63127: 0.011165288335653526
: 0.004375769502845898, 11305: 0.008732975812932723, 14137: 0.008708841442143983, 17989: 0.008679063931137961, 43590
5679932676887563, 90735: 0.008541821591747918, 14859: 0.02985026828953624, 45382: 0.00849308480250739, 22209: 0.00852
44037, 44487: 0.012740871398049315, 45382: 0.008493084802505739, 56749: 0.0169719377425713, 81313: 0.033948182088999
7: 0.011579402652409986, 9351: 0.011567992880071571, 16182: 0.011553794146106008, 37227:
59465187708482, 63320: 0.01156799288071571, 69313: 0.011579402652409986, 96610: 0.0057968642618272375, 3932: 0.01278
4337, 38212: 0.008522807407204557, 45973: 0.012763767408782949, 76266: 0.029723468662354757, 84242: 0.0042556704567

discussion

- 透過上面呈現的結果可以觀察到，像圖1這種直接從1連續連連到5也沒有cycle的圖，authorities會在起始node(0)的位置值為0，因為沒有父節點可以計算出值，hub則是會在結束點(6)位置為0，因為沒有子節點能夠計算出值，而PageRank則會在起始點比較低。
- 在實作SimRank的過程中，發現若依照遞迴式直接coding，在遇到有cycle的圖片時會無法結束，所以會用給予每個node對應其他node的相似度初始值，再依照公式計算，直到誤差夠小就結束迭代的這種計算方式來實作此演算法。

Computation performance analysis

HITS

- time

graph	time
graph_1	0m0.091s
graph_2	0m0.092s
graph_3	0m0.091s
graph_4	0m0.093s
graph_5	0m0.129s
graph_6	0m0.813s
ibmData_directed	0m5.744s
ibmData_bidirected	0m5.446s

PageRank

- time

graph	time
graph_1	0m0.090s
graph_2	0m0.091s
graph_3	0m0.090s
graph_4	0m0.091s
graph_5	0m0.157s
graph_6	0m0.322s
ibmData_directed	0m1.942s
ibmData_bidirected	2m19.967s

SimRank

- time

graph	time
graph_1	0m0.092s
graph_2	0m0.092s
graph_3	0m0.092s
graph_4	0m0.096s
graph_5	0m9.897s
graph_6	0m39.897s

analysis

- 上面的執行時間結果可以觀察到，在圖1~4這種很小的圖時間差距不大觀察不出什麼，而5、6和ibmData_directed及ibmData_bidirected開始，就會因為點數的不同和link的特性而有不同的執行時間，而SimRank時間複雜度較高，在點數多圖複雜時，時間差距就會很明顯。

Discussion

- 在這個project中要我們實作HITS、PageRank及SimRank三種不同的演算法，此三種方法概念上略有一些差異，但都對搜尋引擎有很大的幫助，可以應用於含有元素之間相互參照的情況，而且不只是要考慮經度問題，還要將計算的時間複雜度考量進去，因此在寫程式時上網搜尋作法也會發現一些演算法變體。

Find a way (e.g., add/delete some links) to increase hub, authority,and PageRank of Node 1 in first 3 graphs respectively

- hub的計算方式是child node的authority值相加出來的，所以若要增加hub，以圖1為例，要增加結束點6(無child或少child)之node的child link數，或是增加影響權重，圖2及圖3也是同理。

- authority的方法也類似，authority的計算方法是parent node的hub值相加出來的，因此要增加authority擇要增加起始點(無或少parent)之node的parent link數，，或是增加影響權重，圖2及圖3也是依此類推。

Questions & Discussion

More limitations about link analysis algorithms

- 大部分的演算法，都沒有辦法在圖中很好的找到每個node之間最佳的相關性，評分的標準只用連結束來判定可能有些不足，連結數多寡的可能有太多變因，網頁質量和連結數其實相關性是不太足夠的。

Can link analysis algorithms really find the "important" pages from Web?

- 如上題所述，沒有辦法找到很好的important pages，在實際情況中的連結可能也有很多相干度不高的網頁，甚至是廣告蓋版的問題等等，更舊的網頁分數也會因為演算反可能分數高，但實際重要程度可能不及新網頁的質量。

What are practical issues when implement these algorithms in a real Web?

- 最常見的就是用在搜尋引擎，做網頁排名，像PageRank是google早期用來對搜尋引擎的搜尋結果中做網頁排名的演算法，而像google這種資料量如此龐大的公司，不僅僅是要考量到演算法的精準度，還要顧及時間複雜度不能夠太高，以免造成效能不佳導致使用者體驗不好的問題，因此也有了許多的演算法變體。

Any new idea about the link analysis algorithm?

- 可能可以多考慮幾層的關係而不只是一層，但時間複雜度也要有所取捨，或為不同的網頁判斷不同的權重，不然就是加入一些使用者偏好的因素在裡面，如瀏覽紀錄或書籤網站等等，藉此來設計新的演算法。

What is the effect of "C" parameter in SimRank?

- C在SimRank的演算法中代表著阻尼常數，有衰退的效用，較近的共同父節點有比較強的影響力，而比較遠的會因為此係數的關係影響遞減。