

# CSC4008 - Assignment 3

Yohandi

November 26, 2023

## 1 Implementing PageRank and HITS [90 pts]

### 1.1 PageRank Implementation

- List the top 5 node ids with the highest PageRank scores.

Node	PageRank
263	0.002020291181518219
537	0.0019433415714531492
965	0.0019254478071662631
243	0.001852634016241731
285	0.0018273721700645144

- List the bottom 5 node ids with the lowest PageRank scores.

Node	PageRank
558	0.0003286018525215297
93	0.0003513568937516577
62	0.00035314810510596274
424	0.00035481538649301454
408	0.00038779848719291705

### 1.2 HITS Implementation

- List the 5 node ids with the highest hubbiness score.

Node	PageRank
840	1.0
155	0.9499618624906541
234	0.8986645288972263
389	0.863417110184379
408	0.8632841092495217

- List the 5 node ids with the lowest hubbiness score.

Node	PageRank
23	0.04206685489093653
835	0.05779059354433016
141	0.0645311764622518
539	0.06602659373418493
889	0.07678413939216454

- List the 5 node ids with the highest authority score.

Node	PageRank
893	1.0
16	0.9635572849634397
799	0.9510158161074016
146	0.9246703586198443
473	0.8998661973604049

- List the 5 node ids with the lowest authority score.

Node	PageRank
19	0.05608316377607618
135	0.06653910487622794
462	0.075442286246419
24	0.08171239406816944
910	0.08571673456144875

## 2 Bloom Filter [10 pts]

The table below shows the corresponding hashed value for each string  $\in S$ .

string	$h_1(\text{string})$	$h_2(\text{string})$
hi	1	6
big	1	2
data	1	5
spark	4	1

- Initially, the state is:

state	0	0	0	0	0	0	0
-------	---	---	---	---	---	---	---

- After inserting **hi**, the state is:

state	0	1	0	0	0	0	1
-------	---	---	---	---	---	---	---

- After inserting **big**, the state is:

state	0	1	1	0	0	0	1
-------	---	---	---	---	---	---	---

- After inserting **data**, the state is:

state	0	1	1	0	0	1	1
-------	---	---	---	---	---	---	---

- The final result, which is obtained after inserting **spark**, the state is:

state	0	1	1	0	1	1	1
-------	---	---	---	---	---	---	---

### 3 Appendix

- 120040025\_1a.ipynb's Output:

```
Top 5 nodes by PageRank:
Node 263 has a PageRank of 0.002020291181518219
Node 537 has a PageRank of 0.0019433415714531492
Node 965 has a PageRank of 0.0019254478071662631
Node 243 has a PageRank of 0.001852634016241731
Node 285 has a PageRank of 0.0018273721700645144
Bottom 5 nodes by PageRank:
Node 558 has a PageRank of 0.0003286018525215297
Node 93 has a PageRank of 0.0003513568937516577
Node 62 has a PageRank of 0.00035314810510596274
Node 424 has a PageRank of 0.00035481538649301454
Node 408 has a PageRank of 0.00038779848719291705
```

- 120040025\_1b.ipynb's Output:

```
Top 5 nodes by hub score:
(840, 1.0)
(155, 0.9499618624906541)
(234, 0.8986645288972263)
(389, 0.863417110184379)
(472, 0.8632841092495217)
Bottom 5 nodes by hub score:
(23, 0.04206685489093653)
(835, 0.05779059354433016)
(141, 0.0645311764622518)
(539, 0.06602659373418493)
(889, 0.07678413939216454)
Top 5 nodes by authority score:
(893, 1.0)
(16, 0.9635572849634397)
(799, 0.9510158161074016)
(146, 0.9246703586198443)
(473, 0.8998661973604049)
Bottom 5 nodes by authority score:
(19, 0.05608316377607618)
(135, 0.06653910487622794)
(462, 0.075442286246419)
(24, 0.08171239406816944)
(910, 0.08571673456144875)
```