INFSCI 2725: DATA ANALYTICS

ASSIGNMENT- 8 Submitted By: Nidhi Agarwal

Solution#1 HITS Algorithm:

It helps to rank web pages. It is query- dependent and has two scores: Authority score and Hub score.
The number of in-links to a webpage is its authority score, whereas the number of out-links is the Hub score.

The algorithm is implemented in Python.

Screenshot for Authority and HUB Scores:

```
Authority scores:  [ 0.17241379  0.06896552  0.10344828  0.05172414  0.0862069   0.06896552
  0.10344828  0.06896552  0.10344828  0.12068966  0.          0.
  0.05172414  0.          0.        ]

Hub scores:  [ 0.03846154  0.15384615  0.          0.11538462  0.03846154  0.13461538
  0.07692308  0.          0.          0.          0.03846154  0.11538462
  0.07692308  0.07692308  0.13461538]
```

Figure-1

It's mentioned in assignment, that there is not one correct answer. We take the nodes with high **Authority** scores as authority nodes. These are:
Node 1, with score= 0.17
Node 10, with score= 0.12

The nodes with high **Hub** scores are the Hubs, these are:
Node 2, with score=0.15
Nodes 6 and 15, with score=0.13

Solution#2:

We find the Hubs and Authorities using HITS Algorithm.

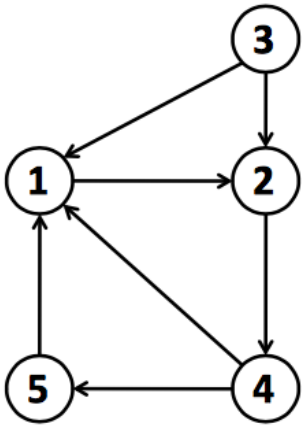

Figure-3

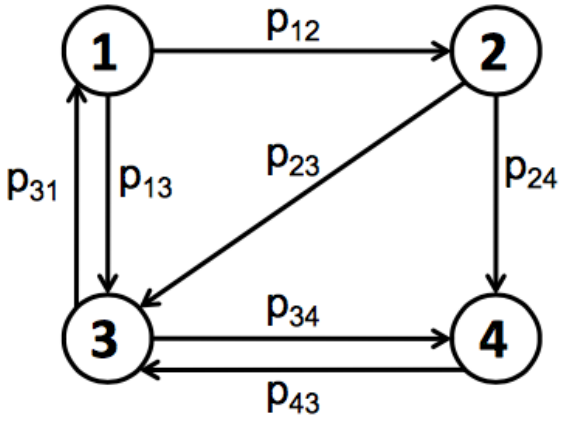

Figure-4

Screenshot for Figure-3:

```
hits(A1)
Authority scores:  [ 0.45454545  0.27272727  0.          0.09090909  0.18181818]
Hub scores:  [ 0.13333333  0.06666667  0.33333333  0.26666667  0.2       ]
```

Authority → node 2

Hub → node 3

By looking at the graph, node 1 appears to be an Authority. When we use the algorithm, it points towards node 2 as the authority. It may be because node 2 has an in-link from node 1,which has high number of in-links.

Here the results are not consistent with the notion of Authority and hubs.

Screenshot for Figure-4:

```
hits(A2)
Authority scores:  [ 0.15384615  0.15384615  0.38461538  0.30769231]

Hub scores:  [ 0.26666667  0.33333333  0.2         0.2         ]
```

Authority → node 3
Hub → node 2
Nodes 1,2 and 3 have two out-links each. So it ambiguous to judge which one is hub. The algorithm points towards node 2.
Here the results are consistent with the notion of Authority and hubs. This is because node 2 has 2 out-links and 1 in-link. There is no in-link and out-link towards the same node.

References:

1. https://en.wikipedia.org/wiki/HITS_algorithm
2. https://cs7083.wordpress.com/2013/01/31/demystifying-the-pagerank-and-hits-algorithms