Experiment No:-10

19_Sanket Chandrashekhar Harvande

PAGE No.
DATE / /

Experiment No. - 10

Aim :- Implementation of HITS Algorithm in python.

Theory :-

The biggest difference between PageRank & HITS

- HITS calculate weights based on the hubness & authority value.
- PageRank calculate the ranks based on the proportional rank passed around the sites.

Hyperlink Induced Topic Search (HITS) Algorithm is a link analysis algorithm that rates webpages developed by Jon Kleinberg. This algorithm is used to the web-link structures to discover and rank the webpages relevant for particular search

- Given a query to a search engine the set of Highly relevant web pages are called Roots.
- They are potential authorities.
- Pages that are not very relevant but point to pages in the Root are called Hubs. Thus an Authority is page that many hub link to a webpage authority.

Algorithm :-

Steps :-

- Initialize the hub & authority of each node with a value 1.
- For each iteration, update the hub & authority of

every node in the graph.

- The new authority y is the sum of hub of its parents
- The new hub is the sum of the authority of its children.
- Normalize the new authority & hub.

Initially

Hub scores :-

Authority scores :-

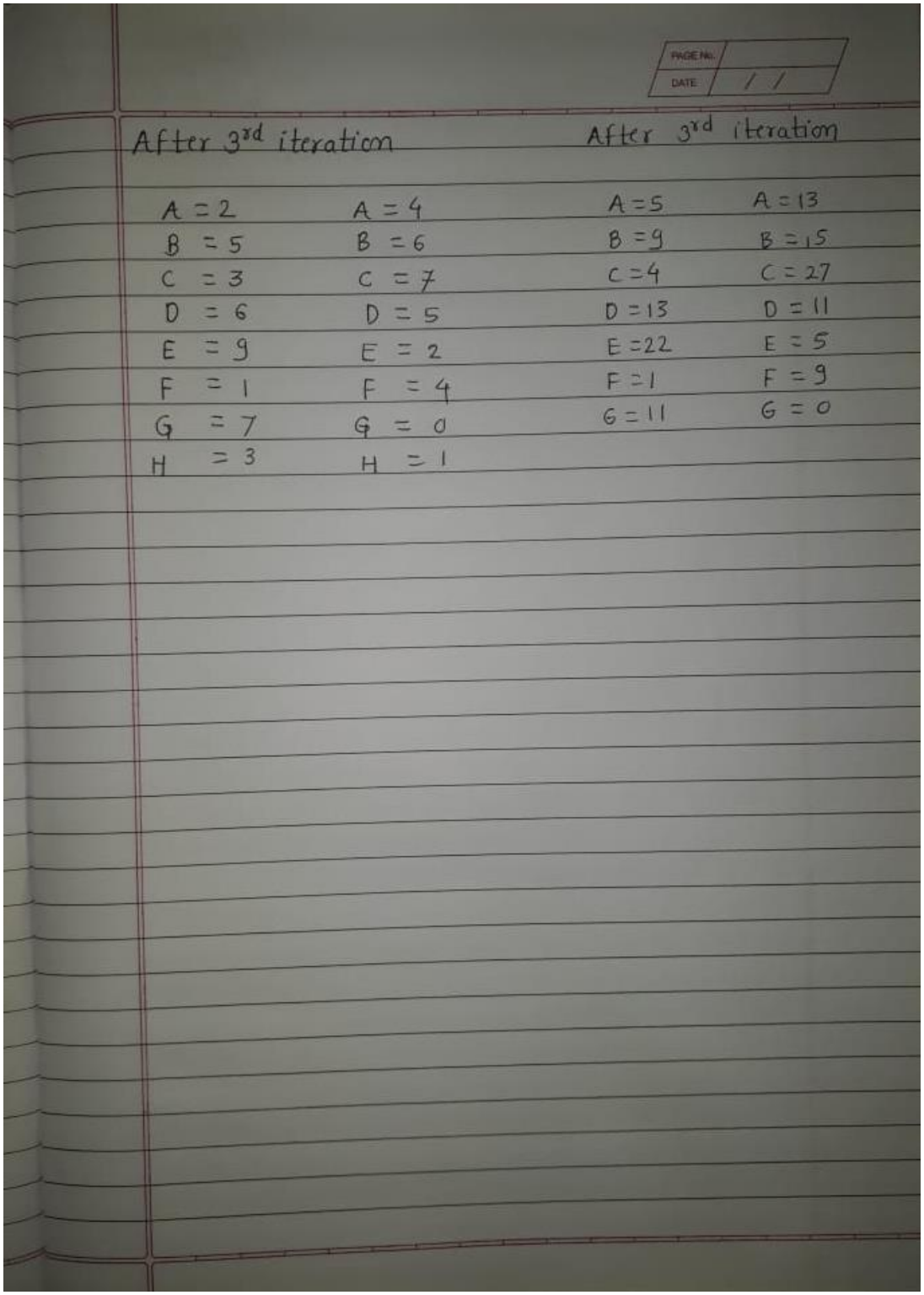
A = 1	A = 1
B = 1	B = 1
C = 1	C = 1
D = 1	D = 1
E = 1	E = 1
F = 1	F = 1
G = 1	G = 1
H = 1	H = 1

After 1st iteration

Hub scores

Authority scores :-

A = 1	A = 3
B = 2	B = 2
C = 1	C = 4
D = 2	D = 2
E = 4	E = 1
F = 1	F = 1
G = 2	G = 0
H = 1	H = 1



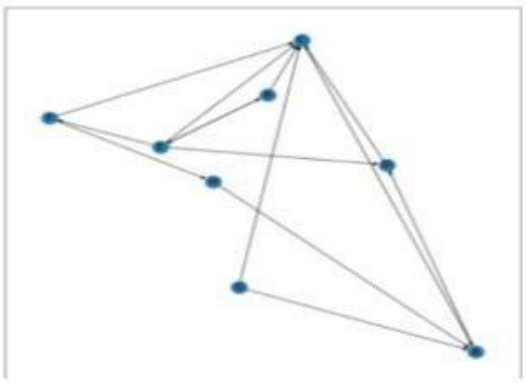
After 3rd iteration

A = 2	A = 4
B = 5	B = 6
C = 3	C = 7
D = 6	D = 5
E = 9	E = 2
F = 1	F = 4
G = 7	G = 0
H = 3	H = 1

After 3rd iteration

A = 5	A = 13
B = 9	B = 15
C = 4	C = 27
D = 13	D = 11
E = 22	E = 5
F = 1	F = 9
G = 11	G = 0

Output:-



Hub Scores: {'A': 0.04642540386472174, 'D': 0.133660375232863, 'B': 0.15763599440595596, 'C': 0.037389132480584515, 'E': 0.2588144594158868, 'F': 0.15763599440595596, 'H': 0.037389132480584515, 'G': 0.17104950771344754}

Authority Scores: {'A': 0.10864044085687284, 'D': 0.13489685393050574, 'B': 0.11437974045401585, 'C': 0.3883728005172019, 'E': 0.06966521189369385, 'F': 0.11437974045401585, 'H': 0.06966521189369385, 'G': 0.0}