Hubs and Authorities

HITS (Hyperlink-Induced Topic Search)

→ It is developed by John Kleinberg in 1997.

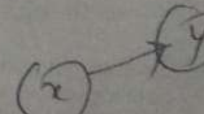
→ Every web page is assigned two scores, one is called Hub score and other is its authority score

→ For any query, we compute two ranked lists i.e. hub score list & authority score list

→ A good hub page is one which points to many good authorities & a good authority page is one which is pointed to by many good hub pages.

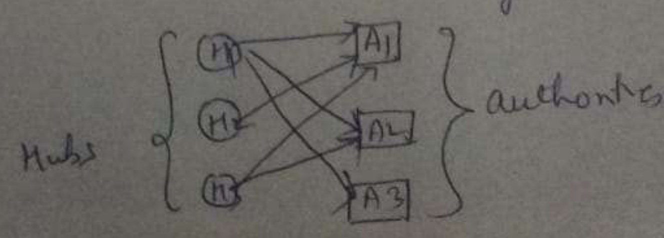
→ Let us take a subset of webgraph and then compute its hub & authority score.

for a webpage x, x has a hyperlink that points to y.



Hub score of node $x = h(x) = \sum_{x \to y} a(y)$

authority score of node $x = a(x) = \sum_{y \to x} h(y)$



Algorithm :-

1. Assign each node an authority and hub score of webgraph.

   i.e. initially $h(x) = a(x) = 1$ or $h(x) = a(x) = \frac{1}{\sqrt{M}}$

2. Apply the authority update rule : each node's authority score is the sum of hub scores of each node that it points to.

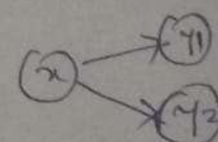
   Indegree

   $$a(x) = \sum_{y \to x} h(y)$$

3. Apply the hub update rule : each node's hub score is sum of authority score of each node that it points to

   Outdegree

   $$h(x) = \sum_{x \to y} a(y)$$

4. Normalize authority and hub scores :

   $$auth(j) = \frac{auth(j)}{\sum_{i \in M} auth(i)}$$

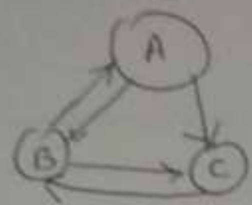
   $$hub(j) = \frac{hub(j)}{\sum_{i \in M} hub(i)}$$

5. Repeat 2, 3, 4 until convergence condition is met. i.e. scores become constant.

Example :-

Step 1

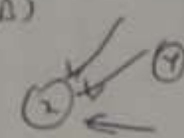
Let $a_{1,0} = h_{1,0} = 1$



Step 2

	old A	old H	New A	New H
A	1	1	1 (B)	2 (B & C)
B	1	1	2 (A & C)	2 (A & C)
C	1	1	2 (A & B)	1 (B)

authority (indegree) $(x) = \sum_{y \to x} h(y)$



hub $(x) = \sum_{x \to y} a(y)$

Normalize :- authority $(A) = \frac{A(A)}{A(A)+A(B)+A(C)} = \frac{1}{5}$

Similarly for other nodes

	new A	new H	new A	new H
A	1/5	2/5	H(B) 2/5	A(B)+A(C) 4/5
B	2/5	2/5	H(A)+H(C) 3/5	A(A)+A(C) 3/5
C	2/5	1/5	H(A)+H(B) 4/5	A(B) 2/5

and now normalize and do it iteratively till stopping condition is met

Matrix Notation :-

Let vector $a = (a_1, a_2 - a_n)$ be a authority vector

$h = (h_1, h_2 - - h_n)$ be a hub vector

n = no of nodes in the web graph.

Step1 : initialize $a_i = h_i = \frac{1}{\sqrt{n}}$ or

$a_i = h_i = 1$

Step2 : Repeat until stopping condition is met (or it converges)

(i) $h = A.a \Rightarrow h = (AA^T).h$

(ii) $a = A^T.h \Rightarrow a = (A^TA).a$

(iii) Normalize a & h. (Use L_1 or L_2 norm)

Step3 :- If convergence criterion is met, stop else go to step2.

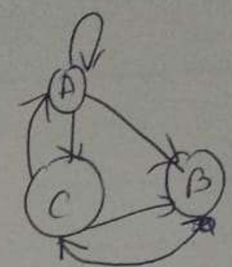
L2 Norm is :- $\sum_{i \in N} (a_i)^2 = 1$ & $\sum_{i \in N} (h_i)^2 = 1$

Hence, at every iteration. normalize each value of a & h by

$$a_i^{(t)} = \frac{a_i^{(t)}}{\sqrt{\sum_{j=1}^{n} a_j^2}} \text{ and } h_i^{(t)} = \frac{h_i^{(t)}}{\sqrt{\sum_{j=1}^{n} h_j^2}}$$

$$a_i^{(t)} = \frac{a_i^{(t)}}{\sqrt{a_1^2 + a_2^2 + \cdots a_n^2}} \text{ and } h_i^{(t)} = \frac{h_i^{(t)}}{\sqrt{h_1^2 + h_2^2 \cdots h_n^2}}$$

Example :- $[A] = \begin{array}{c|ccc} & A & B & C \\ \hline A & 1 & 1 & 1 \\ B & 0 & 0 & 1 \\ C & 1 & 1 & 0 \end{array}$



$$a = (A^T \times A \cdot a)$$

$$h = (A \times A^T \cdot h)$$

$$A^T A = \begin{bmatrix} 2 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

$$A A^T = \begin{bmatrix} 3 & 1 & 2 \\ 1 & 1 & 0 \\ 2 & 0 & 2 \end{bmatrix}$$

$$a = [A^T A] \cdot a$$

$$= \begin{bmatrix} 2 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \\ 4 \end{bmatrix}$$

$$h = [A A^T] h = \begin{bmatrix} 3 & 1 & 2 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \\ 4 \end{bmatrix}$$

next iteration

$$a = \begin{bmatrix} 24 \\ 24 \\ 18 \end{bmatrix}$$

$$h = \begin{bmatrix} 28 \\ 8 \\ 20 \end{bmatrix}$$

and so on till it converges.