Mutually Recursive Definition

- A good hub links to many good authorities
- A good authority is linked from many good hubs
- Model using two scores for each node:
  - **Hub** score and **Authority** score
  - Represented as vectors $\boldsymbol{h}$ and $\boldsymbol{a}$

Hubs and Authorities

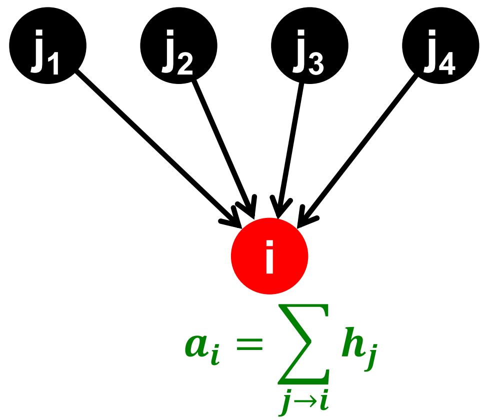
[Kleinberg '98]

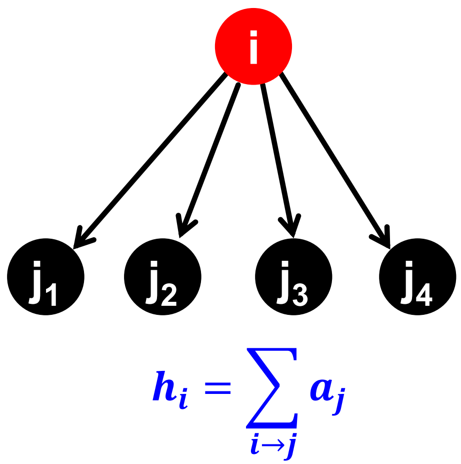
- **Each page i has 2 scores:**
  - Authority score(权威度值): $\boldsymbol{a_i}$
  - Hub score(导航度值): $\boldsymbol{h_i}$

HITS algorithm:

- Initialize: $a_j^{(0)} = 1/\sqrt{\mathrm{N}},\ \ \mathrm{h}_\mathrm{j}^{(0)} = 1/\sqrt{\mathrm{N}}$
- Then keep iterating until **convergence**:
  - $\forall \boldsymbol{i}$: Authority: $a_i^{(t+1)} = \sum_{\boldsymbol{j \to i}} h_j^{(t)}$
  - $\forall \boldsymbol{i}$: Hub: $h_i^{(t+1)} = \sum_{\boldsymbol{i \to j}} a_j^{(t)}$
  - $\forall \boldsymbol{i}$: Normalize:

$$\sum_i \left(a_i^{(t+1)}\right)^2 = 1, \sum_j \left(h_j^{(t+1)}\right)^2 = 1$$





[Kleinberg '98]

Hubs and Authorities

- **HITS converges to a single stable point**
- **Notation:**
  - Vector $\boldsymbol{a} = (a_1 \dots, a_n)$, $\boldsymbol{h} = (h_1 \dots, h_n)$
  - Adjacency matrix $\boldsymbol{A}$ (NxN): $\boldsymbol{A_{ij}} = 1$ if $i \rightarrow j$, 0 otherwise
- **Then** $\boldsymbol{h_i = \sum_{i \rightarrow j} a_j}$ **can be rewritten as** $\boldsymbol{h_i = \sum_j A_{ij} \cdot a_j}$

  So: $\boldsymbol{h = A \cdot a}$
- **Similarly,** $\boldsymbol{a_i = \sum_{j \rightarrow i} h_j}$ **can be rewritten as** $\boldsymbol{a_i = \sum_j A_{ji} \cdot h_j = A^T \cdot h}$

Hubs and Authorities

- **HITS algorithm in vector notation:**
  - Set: $\boldsymbol{a_i} = \boldsymbol{h_i} = \frac{\mathbf{1}}{\sqrt{\boldsymbol{n}}}$

  Repeat until convergence:
  - $\boldsymbol{h} = \boldsymbol{A} \cdot \boldsymbol{a}$
  - $\boldsymbol{a} = \boldsymbol{A}^{\boldsymbol{T}} \cdot \boldsymbol{h}$
  - Normalize $\boldsymbol{a}$ and $\boldsymbol{h}$
- **Then:** $\boldsymbol{a} = \underbrace{\boldsymbol{A}^{\boldsymbol{T}} \cdot (\underbrace{\boldsymbol{A} \cdot \boldsymbol{a}}_{\text{new } \boldsymbol{h}})}_{\text{new } \boldsymbol{a}}$

Convergence criterion:

$$\sum_i \left(h_i^{(t)} - h_i^{(t-1)}\right)^2 < \varepsilon$$

$$\sum_i \left(a_i^{(t)} - a_i^{(t-1)}\right)^2 < \varepsilon$$

$\boldsymbol{a}$ is updated (in 2 steps):

$$a = A^T(A\,a) = (A^T A)\,a$$

$\boldsymbol{h}$ is updated (in 2 steps):

$$h = A\,(A^T h) = (A\,A^T)\,h$$

Repeated matrix powering

Existence and Uniqueness

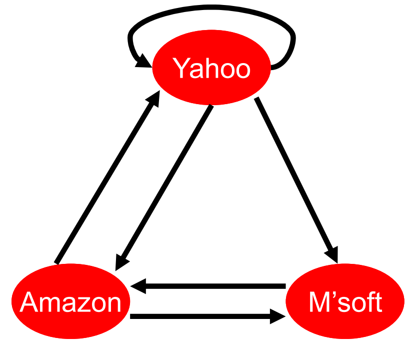
- $\boldsymbol{h = \lambda A a}$
- $\boldsymbol{a = \mu A^T h}$
- $\boldsymbol{h = \lambda \mu A A^T h}$
- $\boldsymbol{a = \lambda \mu A^T A a}$

$\lambda = 1 / \sum h_i$
$\mu = 1 / \sum a_i$

- Under reasonable assumptions about **A**, HITS **converges to vectors h^* and a^***:
  - h^* is the **principal eigenvector** (主特征向量) of matrix $A A^T$
  - a^* is the **principal eigenvector** of matrix $A^T A$

Example of HITS

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



$h = \lambda A a$	h(yahoo)	=	.58	.80	.80	.79	. . .	.788
	h(amazon)	=	.58	.53	.53	.57	. . .	.577
	h(m'soft)	=	.58	.27	.27	.23	. . .	.211
$a = \mu A^T h$	a(yahoo)	=	.58	.58	.62	.62	. . .	.628
	a(amazon)	=	.58	.58	.49	.49	. . .	.459
	a(m'soft)	=	.58	.58	.62	.62	. . .	.628

PageRank and HITS

- **PageRank and HITS are two solutions to the same problem:**
  - **What is the value of an in-link from *u* to *v*?**
  - In the PageRank model, the value of the link depends on the links **into** *u*
  - In the HITS model, it depends on the value of the other links **out of** *u*
- **The destinies of PageRank and HITS were very different**