Solution Using Excel

The final PageRank algorithm

- $(1\text{-}d)\boldsymbol{E}/n + d\boldsymbol{A}^T$ is a **stochastic matrix** (transposed). It is also **irreducible** and **aperiodic**
- If we scale Equation (25) so that $\boldsymbol{e}^T\boldsymbol{P} = n$,

$$\boldsymbol{P} = (1-d)\boldsymbol{e} + d\boldsymbol{A}^T\boldsymbol{P} \tag{27}$$

- PageRank for each page i is

$$P(i) = (1-d) + d\sum_{j=1}^{n} A_{ji}P(j) \tag{28}$$

The final PageRank (cont …)

- (28) is equivalent to the formula given in the PageRank paper

$$P(i) = (1-d) + d \sum_{(j,i)\in E} \frac{P(j)}{O_j}$$

- The parameter *d* is called the **damping factor** which can be set to between 0 and 1. *d* = 0.85 was used in the PageRank paper.

Compute PageRank

- Use the power iteration method

PageRank-Iterate(G)

$P_0 \leftarrow e/n$

$k = 1$

repeat

$P_{k+1} \leftarrow (1-d)e + dA^T P_k$;

$k = k + 1$;

until $||P_{k+1} - P_k||_1 < \varepsilon$

return P_{k+1}

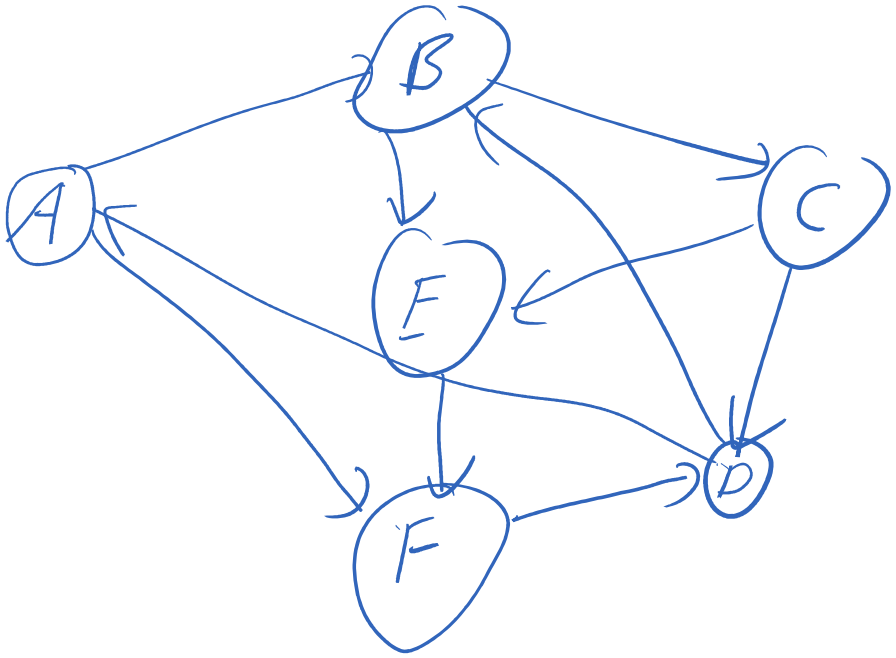
Fig. 6. The power iteration method for PageRank

PageRank using excel

- Without damping factor

e=1, d=1

- With damping factor



HITS

The link graph G

- HITS works on the pages in *S*, and assigns every page in *S* an **authority score** and a **hub score**.
- Let the number of pages in *S* be *n*.
- We again use *G* = (*V*, *E*) to denote the hyperlink graph of *S*.
- We use ***L*** to denote the adjacency matrix of the graph.

$$L_{ij} = \begin{cases} 1 & if\ (i,j) \in E \\ 0 & otherwise \end{cases}$$

The HITS algorithm

- Let the authority score of the page *i* be *a*(*i*), and the hub score of page *i* be *h*(*i*).
- The mutual reinforcing relationship of the two scores is represented as follows:

$$a(i) = \sum_{(j,i)\in E} h(j) \qquad (31)$$

$$h(i) = \sum_{(i,j)\in E} a(j) \qquad (32)$$

HITS in matrix form

- We use $\boldsymbol{a}$ to denote the column vector with all the authority scores,

  $$\boldsymbol{a} = (a(1), a(2), \ldots, a(n))^T, \text{ and}$$

- use $\boldsymbol{h}$ to denote the column vector with all the authority scores,

  $$\boldsymbol{h} = (h(1), h(2), \ldots, h(n))^T,$$

- Then,

  $$\boldsymbol{a} = \boldsymbol{L}^T\boldsymbol{h} \quad (33)$$

  $$\boldsymbol{h} = \boldsymbol{L}\boldsymbol{a} \quad (34)$$

Computation of HITS

- The computation of authority scores and hub scores is the same as the computation of the PageRank scores, using power iteration.
- If we use $\boldsymbol{a}_k$ and $\boldsymbol{h}_k$ to denote authority and hub vectors at the kth iteration, the iterations for generating the final solutions are

$$\boldsymbol{a}_k = \boldsymbol{L}^T \boldsymbol{L} \boldsymbol{a}_{k-1} \quad (35)$$

$$\boldsymbol{h}_k = \boldsymbol{L} \boldsymbol{L}^T \boldsymbol{h}_{k-1} \quad (36)$$

starting with

$$\boldsymbol{a}_0 = \boldsymbol{h}_0 = (1, 1, \ldots, 1), \quad (37)$$

The algorithm

HITS-Iterate(G)

$\boldsymbol{a}_0 = \boldsymbol{h}_0 = (1, 1, \ldots, 1)$;

$k = 1$

Repeat

$\boldsymbol{a}_k = \boldsymbol{L}^T \boldsymbol{L} \boldsymbol{a}_{k-1}$;

$\boldsymbol{h}_k = \boldsymbol{L}\boldsymbol{L}^T \boldsymbol{h}_{k-1}$;

normalize $\boldsymbol{a}_k$;

normalize $\boldsymbol{h}_k$;

$k = k + 1$;

until $\boldsymbol{a}_k$ and $\boldsymbol{h}_k$ do not change significantly;

return $\boldsymbol{a}_k$ and $\boldsymbol{h}_k$

. **Fig. 9.** The HITS algorithm based on power iteration

HITS using excel

- 3 Hubs
- 5 Authorities