Authority and Co-cite, Hub and Co-reference

- Given the adjacency matrix A (with entries 0 or 1)
  $a_i = A^T A a_{i-1} \quad a_i = (A^T A)^i a_0$
  $h_i = A A^T h_{i-1} \quad h_i = (A A^T)^i h_0$
- Co-citation: the number of pages co-cite P_i and P_j
  Co-reference: the number of pages co-referenced by P_i and P_j.
- $A^T A = D + C$ where C is the matrix of co-citation and $D = diag(d_1, d_2, \cdots, d_j)$
- $$\begin{aligned} C_{ij} &= \sum_k A_{ki} A_{kj} = (A^T A)_{ij} \\ C_{ii} &= 0 \\ d_i &= \sum_k A_{ki} = \sum_k A_{ki} A_{ki} = (A^T A)_{ii} \end{aligned}$$

$$A^T A = D + C, \; A A^T = O + R$$

Authority and Co-cite, Hub and Co-reference

- Given the adjacency matrix A (with entries 0 or 1)
  $a_i = A^T A a_{i-1} \quad a_i = (A^T A)^i a_0$
  $h_i = A A^T h_{i-1} \quad h_i = (A A^T)^i h_0$
- Co-citation: the number of pages co-cite P_i and P_j
  Co-reference: the number of pages co-referenced by P_i and P_j.
- $A^T A = D + C$ where C is the matrix of co-citation and $D = diag(d_1, d_2, \cdots, d_j)$
-

$$
\begin{aligned}
C_{ij} &= \sum_k A_{ki} A_{kj} = (A^T A)_{ij} \\
C_{ii} &= 0 \\
d_i &= \sum_k A_{ki} = \sum_k A_{ki} A_{ki} = (A^T A)_{ii}
\end{aligned}
$$

$$A^T A = D + C, \; A A^T = O + R$$

Authority and Co-cite, Hub and Co-reference

- Given the adjacency matrix A (with entries 0 or 1)
  $a_i = A^T A a_{i-1} \quad a_i = (A^T A)^i a_0$
  $h_i = A A^T h_{i-1} \quad h_i = (A A^T)^i h_0$
- Co-citation: the number of pages co-cite P_i and P_j
  Co-reference: the number of pages co-referenced by P_i and P_j.
- $A^T A = D + C$ where C is the matrix of co-citation and $D = diag(d_1, d_2, \cdots, d_j)$
-

$$
\begin{aligned}
C_{ij} &= \sum_k A_{ki} A_{kj} = (A^T A)_{ij} \\
C_{ii} &= 0 \\
d_i &= \sum_k A_{ki} = \sum_k A_{ki} A_{ki} = (A^T A)_{ii}
\end{aligned}
$$

$A^T A = D + C,\ AA^T = O + R$

Authority and Co-cite, Hub and Co-reference

- Given the adjacency matrix A (with entries 0 or 1)
  $a_i = A^T A a_{i-1} \quad a_i = (A^T A)^i a_0$
  $h_i = A A^T h_{i-1} \quad h_i = (A A^T)^i h_0$
- Co-citation: the number of pages co-cite P_i and P_j
  Co-reference: the number of pages co-referenced by P_i and P_j.
- $A^T A = D + C$ where C is the matrix of co-citation and $D = diag(d_1, d_2, \cdots, d_j)$
-

$$
\begin{aligned}
C_{ij} &= \sum_k A_{ki} A_{kj} = (A^T A)_{ij} \\
C_{ii} &= 0 \\
d_i &= \sum_k A_{ki} = \sum_k A_{ki} A_{ki} = (A^T A)_{ii}
\end{aligned}
$$

$$A^T A = D + C,\ AA^T = O + R$$

Probabilistic analysis

Expected value of co-citation/co-reference

For **a fixed degree sequence random graphs**

$$E(C_{ik}) = \frac{d_i d_k}{n-1}$$
$$E(R_{ik}) = \frac{o_i o_k}{n-1}$$

The node with large indegree d_i tend to have large co-citations with other nodes.

$$E(A^T A) = E(D) + E(C) = diag(h_1, h_2, \cdots, h_n) + \mathbf{d}\mathbf{d}^T / n - 1$$

where $h_i \equiv d_i - d_i^2/(n-1)$ and $\mathbf{d} = (d_1, d_2, \cdots, d_n)^T$.

Probabilistic analysis

Expected value of co-citation/co-reference

For **a fixed degree sequence random graphs**

$$E(C_{ik}) = \frac{d_i d_k}{n-1}$$
$$E(R_{ik}) = \frac{o_i o_k}{n-1}$$

The node with large indegree d_i tend to have large co-citations with other nodes.

$$E(A^T A) = E(D) + E(C) = diag(h_1, h_2, \cdots, h_n) + \mathbf{d}\mathbf{d}^T / n - 1$$

where $h_i \equiv d_i - d_i^2/(n-1)$ and $\mathbf{d} = (d_1, d_2, \cdots, d_n)^T$.

Spectral Decomposition of Diagonal Plus Rank-1 matrices

Let $M = D + cc^T$, D is a diagonal $n \times n$ matrix of the block form:

$$D = diag(\tau_1 I_1, \tau_2 I_2, \cdots, \tau_l I_l)$$

where I_k is the identity matrix of size n_k, $\tau_1 > \tau_2 > \cdots > \tau_l$
Then, the eigenvalues of M are given by

$$\hat{\tau}_1 > \underbrace{\tau_1 = \cdots = \tau_1} > \hat{\tau}_2 > \underbrace{\tau_2 = \cdots = \tau_2} > \cdots > \hat{\tau}_l > \underbrace{\tau_l = \cdots = \tau_l}$$

and the eigenvector of A corresponds to the eigenvalue $\hat{\tau}_k$ is

$$\left(\frac{c_1^T}{\hat{\tau}_1 - \tau_1}, \frac{c_2^T}{\hat{\tau}_2 - \tau_2}, \cdots, \frac{c_l^T}{\hat{\tau}_l - \tau_l}\right)^T.$$

The eigenvector corresponds to τ_k is of the form

$$(0 \cdots 0, u_k^T, 0 \cdots 0)^T$$

where u_k is a vector of n_k satisfying $c_k^T u_k = 0$.

Average Analysis of HITS

$E(A^T A) = E(D) + E(C) = diag(h_1, h_2, \cdots, h_n) + \mathbf{d}\mathbf{d}^T/n - 1$
where $h_i \equiv d_i - d_i^2/(n-1)$ and $\mathbf{d} = (d_1, d_2, \cdots, d_n)^T$.

If $h_1 > h_2 > \cdots > h_m \geq h_{m+1} \geq \cdots \geq h_n$, Then, the m largest eigenvalues λ_i satisfying

$$\lambda_1 > h_1 > \lambda_2 > h_2 > \cdots > \lambda_m > h_m$$

the corresponding eigenvectors are

$$\mathbf{u}_k = (\frac{d_1}{\lambda_k - h_1}, \frac{d_2}{\lambda_k - h_2}, \cdots, \frac{d_n}{\lambda_k - h_n})$$

Prerequisite

$h_i - h_j = (d_i - d_j)[1 - (d_i + d_j)/(n-1)] > 0$
as long as $d_1 > \cdots > d_m > d_{m+1} \geq d_{m+1} \geq d_{m+2} \cdots \geq d_n$
and $d_i + d_j < n - 1$ for $\forall i, j$

Average Analysis of HITS

$E(A^T A) = E(D) + E(C) = diag(h_1, h_2, \cdots, h_n) + \mathbf{d}\mathbf{d}^T/n - 1$
where $h_i \equiv d_i - d_i^2/(n-1)$ and $\mathbf{d} = (d_1, d_2, \cdots, d_n)^T$.

If $h_1 > h_2 > \cdots > h_m \geq h_{m+1} \geq \cdots \geq h_n$, Then, the m largest eigenvalues λ_i satisfying

$$\lambda_1 > h_1 > \lambda_2 > h_2 > \cdots > \lambda_m > h_m$$

the corresponding eigenvectors are

$$\mathbf{u}_k = (\frac{d_1}{\lambda_k - h_1}, \frac{d_2}{\lambda_k - h_2}, \cdots, \frac{d_n}{\lambda_k - h_n})$$

Prerequisite

$h_i - h_j = (d_i - d_j)[1 - (d_i + d_j)/(n-1)] > 0$
as long as $d_1 > \cdots > d_m > d_{m+1} \geq d_{m+1} \geq d_{m+2} \cdots \geq d_n$
and $d_i + d_j < n - 1$ for $\forall i, j$

Average Analysis of HITS

$E(A^T A) = E(D) + E(C) = diag(h_1, h_2, \cdots, h_n) + \mathbf{d}\mathbf{d}^T / n - 1$
where $h_i \equiv d_i - d_i^2/(n-1)$ and $\mathbf{d} = (d_1, d_2, \cdots, d_n)^T$.

If $h_1 > h_2 > \cdots > h_m \geq h_{m+1} \geq \cdots \geq h_n$, Then, the m largest eigenvalues λ_i satisfying

$$\lambda_1 > h_1 > \lambda_2 > h_2 > \cdots > \lambda_m > h_m$$

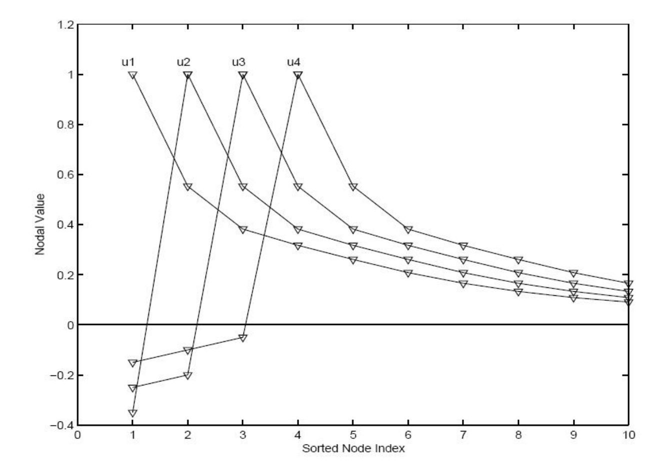
the corresponding eigenvectors are

$$\mathbf{u}_k = (\frac{d_1}{\lambda_k - h_1}, \frac{d_2}{\lambda_k - h_2}, \cdots, \frac{d_n}{\lambda_k - h_n})$$

Prerequisite

$h_i - h_j = (d_i - d_j)[1 - (d_i + d_j)/(n-1)] > 0$
as long as $d_1 > \cdots > d_m > d_{m+1} \geq d_{m+1} \geq d_{m+2} \cdots \geq d_n$
and $d_i + d_j < n - 1$ for $\forall i, j$

Eigenvectors



HITS = ranking according to indegrees??

For any $i < j$

$$\begin{aligned}
\mathbf{u_1}(i) - \mathbf{u_1}(j) &= \frac{d_i}{\lambda_1 - h_i} - \frac{d_j}{lambda_1 - h_j} \\
&= \frac{(d_i - d_j)[\lambda_1 - d_i d_j/(n-1)]}{(\lambda_1 - h_i)(\lambda - h_j)} \\
&\geq 0
\end{aligned}$$

as $\lambda_1 - d_i d_j/(n-1) > h_i - d_i d_j/(n-1) = d_i(1 - \frac{d_i + d_j}{n-1}) > 0$

What's the nature of AVERAGE?

The authority ranking is, ON AVERAGE, identical to the ranking according to web page indegrees.

HITS = ranking according to indegrees??

For any $i < j$

$$\begin{aligned} \mathbf{u_1}(i) - \mathbf{u_1}(j) &= \frac{d_i}{\lambda_1 - h_i} - \frac{d_j}{lambda_1 - h_j} \\ &= \frac{(d_i - d_j)[\lambda_1 - d_i d_j/(n-1)]}{(\lambda_1 - h_i)(\lambda - h_j)} \\ &\geq 0 \end{aligned}$$

as $\lambda_1 - d_i d_j/(n-1) > h_i - d_i d_j/(n-1) = d_i(1 - \frac{d_i + d_j}{n-1}) > 0$

What's the nature of AVERAGE?

The authority ranking is, ON AVERAGE, identical to the ranking according to web page indegrees.