

Spectral analysis of ranking algorithms

Social and Technological Networks

Rik Sarkar

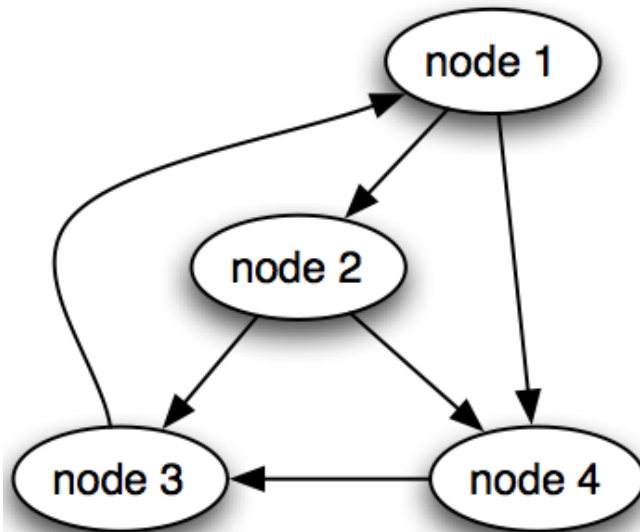
University of Edinburgh, 2016.

Recap: HITS algorithm

- Evaluate hub and authority scores
- Apply Authority update to all nodes:
 - $\text{auth}(p) = \text{sum of all } \text{hub}(q) \text{ where } q \rightarrow p \text{ is a link}$
- Apply Hub update to all nodes:
 - $\text{hub}(p) = \text{sum of all } \text{auth}(r) \text{ where } p \rightarrow r \text{ is a link}$
- Repeat for k rounds

Adjacency matrix

- Example



$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

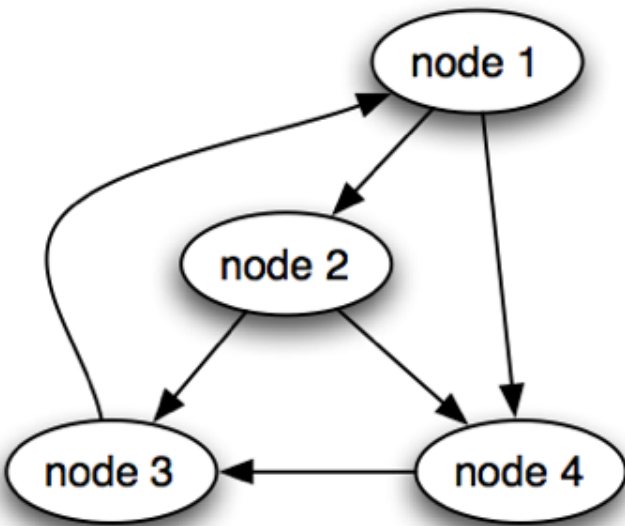
Hubs and authority scores

- Can be written as vectors h and a
- The dimension (number of elements) of the vectors are n

Update rules

- Are matrix multiplications

$$h \leftarrow Ma$$



$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 6 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 7 \\ 2 \\ 4 \end{bmatrix}$$

- Hub rule for i : sum of a -values of *nodes that i points to*:

$$h \leftarrow Ma$$

- Authority rule for i : sum of h -values of *nodes that point to i* :

$$a \leftarrow M^T h$$

Iterations

- After one round:

$$a^{\langle 1 \rangle} = M^T h^{\langle 0 \rangle}$$

$$h^{\langle 1 \rangle} = M a^{\langle 1 \rangle} = M M^T h^{\langle 0 \rangle}$$

- Over k rounds:

$$h^{\langle k \rangle} = (M M^T)^k h^{\langle 0 \rangle}$$

Convergence

- Remember that h keeps increasing
- We want to show that the normalized value

$$\frac{h^{(k)}}{c^k} = \frac{(MM^T)^k h^{(0)}}{c^k}$$

- Converges to a vector of finite real numbers as k goes to infinity
- If convergence happens, then there is a c :

$$(MM^T)h^{(*)} = ch^{(*)}$$

Eigen values and vectors

$$(MM^T)h^{\langle * \rangle} = ch^{\langle * \rangle}$$

- Implies that for matrix (MM^T)
- c is an eigen value, with
- $h^{\langle * \rangle}$ as the corresponding eigen vector

Proof of convergence to eigen vectors

- Useful Theorem: A symmetric matrix has orthogonal eigen vectors. (see sample problems from lecture 1)
 - They form a basis of n-D space
 - Any vector can be written as a linear combination
- (MM^T) is symmetric

Now to prove convergence:

- Suppose sorted eigen values are:

$$|c_1| \geq |c_2| \geq \cdots \geq |c_n|$$

- Corresponding eigen vectors are:

$$z_1, z_2, \dots, z_n,$$

- We can write any vector x as

$$x = p_1 z_1 + p_2 z_2 + \cdots + p_n z_n$$

- So: $(MM^T)x = (MM^T)(p_1 z_1 + p_2 z_2 + \cdots + p_n z_n)$
 $= p_1 MM^T z_1 + p_2 MM^T z_2 + \cdots + p_n MM^T z_n$
 $= p_1 c_1 z_1 + p_2 c_2 z_2 + \cdots + p_n c_n z_n,$

$$\begin{aligned}
(MM^T)x &= (MM^T)(p_1z_1 + p_2z_2 + \cdots + p_nz_n) \\
&= p_1MM^Tz_1 + p_2MM^Tz_2 + \cdots + p_nMM^Tz_n \\
&= p_1c_1z_1 + p_2c_2z_2 + \cdots + p_nc_nz_n,
\end{aligned}$$

- After k iterations:

$$(MM^T)^k x = c_1^k p_1 z_1 + c_2^k p_2 z_2 + \cdots + c_n^k p_n z_n$$

- For hubs: $h^{(k)} = (MM^T)^k h^{(0)} = c_1^k q_1 z_1 + c_2^k q_2 z_2 + \cdots + c_n^k q_n z_n$

- So: $\frac{h^{(k)}}{c_1^k} = q_1 z_1 + \left(\frac{c_2}{c_1}\right)^k q_2 z_2 + \cdots + \left(\frac{c_n}{c_1}\right)^k q_n z_n$

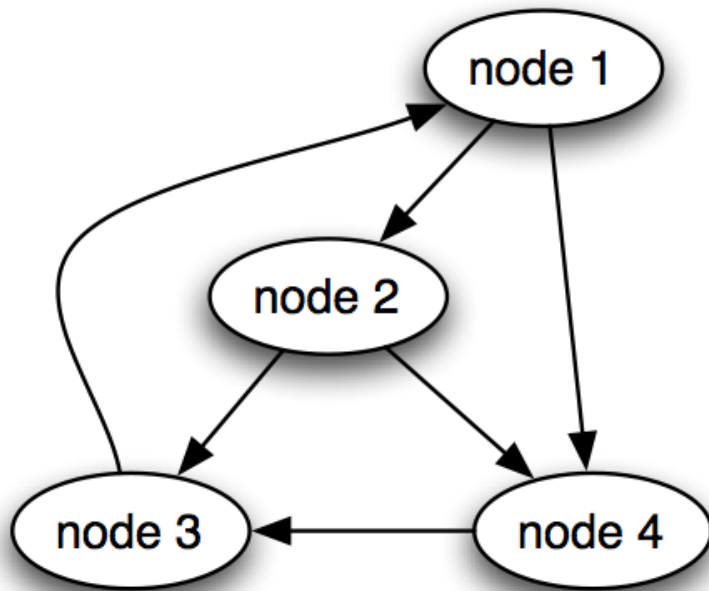
- If $|c_1| > |c_2|$, only the first term remains.

- So, $\frac{h^{(k)}}{c_1^k}$ converges to $q_1 z_1$

Properties

- The vector $q_1 z_1$ is a simple multiple of z_1
 - A vector essentially similar to the first eigen vector
 - Therefore independent of starting values of h
- q_1 can be shown to be non-zero always, so the scores are not zero
- Authority score analysis is analogous

Pagerank Update rule as a matrix derived from adjacency



$$\begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$r \leftarrow N^T r$$

- Scaled pagerank:

$$r \leftarrow \tilde{N}^T r.$$

- Over k iterations:

$$r^{(k)} = (\tilde{N}^T)^k r^{(0)}$$

- Pagerank does not need normalization.

$$\tilde{N}^T r^{(*)} = r^{(*)}$$

- We are looking for an eigen vector with eigen value=1

- For matrix P with all positive values, Perron's theorem says:
 - A unique positive real valued largest eigen value c exists
 - Corresponding eigen vector y is unique and has positive real coordinates
 - If $c=1$, then $P^k x$ converges to y

Random walks

- A random walker is moving along random directed edges
- Suppose vector b shows the probabilities of walker currently being at different nodes
- Then vector $N^T b$ gives the probabilities for the next step

Random walks

- Thus, pagerank values of nodes after k iterations is equivalent to:
 - The probabilities of the walker being at the nodes after k steps
- The final values given by the eigen vector are the steady state probabilities
 - Note that these depend only on the network and are independent of the starting points

History of web search

- YAHOO: A directory (hierarchic list) of websites
 - Jerry Yang, David Filo, Stanford 1995
- 1998: Authoritative sources in hyperlinked environment (HITS), symposium on discrete algorithms
 - Jon Kleinberg, Cornell
- 1998: Pagerank citation ranking: Bringing order to the web
 - Larry Page, Sergey Brin, Rajeev Motwani, Terry Winograd, Stanford techreport

Spectral graph theory

- Undirected graphs
- Diffusion operator
 - Describes diffusion of stuff — step by step
 - Stuff at a vertex uniformly distributed to neighbors — in every step