Spectral analysis of ranking algorithms

Social and Technological Networks

Rik Sarkar

University of Edinburgh, 2018.
Recap: HITS algorithm

• Evaluate hub and authority scores
• Apply Authority update to all nodes:
  – auth(p) = sum of all hub(q) where q -> p is a link
• Apply Hub update to all nodes:
  – hub(p) = sum of all auth(r) where p->r 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

\[ h \leftarrow M a \]
• 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:

\[ a^{(1)} = M^T h^{(0)} \]

\[ h^{(1)} = M a^{(1)} = M M^T h^{(0)} \]

• Over \( k \) rounds:

\[ h^{(k)} = (M M^T)^k h^{(0)} \]
Convergence

• Remember that $h$ keeps increasing
• We want to show that the normalized value
  \[ \frac{h^{(k)}}{c^k} \]
  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

\[(M M^T) h^{(*)} = c h^{(*)}\]

• Implies that for matrix \((M M^T)\)
• \(c\) is an eigen value, with
• \(h^{(*)}\) as the corresponding eigen vector
Proof of convergence to eigen vectors

• Useful Theorem: A symmetric matrix has orthogonal eigen vectors.
  – They form a basis of n-D space
  – Any vector can be written as a linear combination
• \((MM^T)\) is symmetric
• 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 $y$
Now to prove convergence:

- Suppose sorted eigen values are:
  \[ |c_1| \geq |c_2| \geq \cdots \geq |c_n| \]

- Corresponding eigen vectors are:
  \[ z_1, z_2, \ldots, z_n, \]

- We can write any vector \( x \) as
  \[
  x = p_1 z_1 + p_2 z_2 + \cdots + p_n z_n
  \]

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

- 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

\[ r \leftarrow N^T r \]
• 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
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