PageRank

lecture 12 (October 9, 2008)

Edith Law
Page Rank

What’s the big deal?
Life before PageRank

Found It!!!

Congratulations, it only took you 65,298 seconds
Life after PageRank

I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions. I will use Google before asking dumb questions.
Evolution of the web

Centralization
Idea #1: Centralization
Evolution of the web

Centralization

Relevancy

Centralization
Idea #2: Relevancy

Given a query, how do we know what to retrieve?

1. Web directories

2. More sophisticated indexing methods

Filename  →  Description  →  Content
The index size war
Evolution of the web

- Centralization
- Relevancy
- Ranking
Idea #3: Ranking
Page Rank
How it works ...
A page is **important** if it is pointed to by other **important** pages.
Importance

\[ r^{(t+1)}(P_i) = \sum_{j \in E(i)} r^{(t)}(P_j) / l(P_j) \]
The Algorithm

Given a web graph with \( n \) nodes, where the nodes are pages and edges are hyperlinks

- Assign each node an initial page rank
- repeat until convergence
  - calculate the page rank of each node (using the equation in the previous slide)
Example

<table>
<thead>
<tr>
<th></th>
<th>Iteration 0</th>
<th>Iteration 1</th>
<th>Iteration 2</th>
<th>Page Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_1$</td>
<td>1/5</td>
<td>1/20</td>
<td>1/40</td>
<td>5</td>
</tr>
<tr>
<td>$P_2$</td>
<td>1/5</td>
<td>5/20</td>
<td>3/40</td>
<td>4</td>
</tr>
<tr>
<td>$P_3$</td>
<td>1/5</td>
<td>1/10</td>
<td>5/40</td>
<td>3</td>
</tr>
<tr>
<td>$P_4$</td>
<td>1/5</td>
<td>5/20</td>
<td>15/40</td>
<td>2</td>
</tr>
<tr>
<td>$P_5$</td>
<td>1/5</td>
<td>7/20</td>
<td>16/40</td>
<td>1</td>
</tr>
</tbody>
</table>

\[ r_1(P_5) = \frac{1}{5} + \frac{1}{5} \times \frac{1}{4} + \frac{1}{5} \times \frac{1}{2} = \frac{7}{20} \]
Matrix representation

\[
\begin{bmatrix}
0.05 & 0 & 0.25 & 0 & 0 \\
1 & 0 & 0.25 & 0 & 0 \\
0 & 0 & 0 & 0.5 & 0 \\
0 & 0 & 0.25 & 0 & 1 \\
0 & 1 & 0.25 & 0.5 & 0
\end{bmatrix}
\begin{bmatrix}
1/5 \\
1/5 \\
1/5 \\
1/5 \\
1/5
\end{bmatrix}
\]

\[r(t+1) = \begin{bmatrix} H \end{bmatrix} \begin{bmatrix} r(t) \end{bmatrix}\]
Three Questions

\[ r^{(t+1)} = H \cdot r^{(t)} \]

Also known as the \textit{power method}

- Does this converge?
- Does it converge to what we want?
- Are the results reasonable?
Does it converge?

<table>
<thead>
<tr>
<th></th>
<th>Iteration 0</th>
<th>Iteration 1</th>
<th>Iteration 2</th>
<th>Iteration 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P_1)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>(P_2)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Diagram:

\[\begin{array}{c}
\text{1} \\
\text{2}
\end{array}\]

\[\text{1} \rightarrow \text{2} \quad \text{and} \quad \text{2} \rightarrow \text{1}\]
Does it converge to what we want?

<table>
<thead>
<tr>
<th></th>
<th>Iteration 0</th>
<th>Iteration 1</th>
<th>Iteration 2</th>
<th>Iteration 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_1 )</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( P_2 )</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Does it converge to what we want?

Page ranks to converge to 0.
Looks a lot like ...

\[ r^{(t+1)} = H \ r^{(t)} \]

Markov Chains

Set of states \( X \)

Transition matrix \( P \) where \( P_{ij} = P(X_t=j \mid X_{t-1}=i) \)

\( \pi \) specifying the probability of being at each state \( x \in X \)

Goal is to find \( \pi \) such that \( \pi = P \pi \)
Why is this analogy useful?

There exists a theory about Markov chains that says that for any start vector, the power method applied to a Markov transition matrix $P$ will converge to a unique positive stationary vector as long as $P$ is stochastic, irreducible and aperiodic.
Make $H$ stochastic

$$S = H + a\left(\frac{1}{n} e^T\right)$$
Make $H$ aperiodic

A chain is periodic if there exists $k > 1$ such that the interval between two visits to some state $s$ is always a multiple of $k$. 
Make $H$ irreducible

From any state, there is a non-zero probability of going from one state to another.
The Google Matrix

\[ G = \alpha S + (1-\alpha) \frac{1}{n} ee^T \]

The Random Surfer Model: for each page, time spent \( \propto \) importance.
Are the results reasonable?

\[ G = \alpha S + (1-\alpha) \frac{1}{n} ee^T \]

\[ G \text{ is stochastic, aperiodic and irreducible.} \]

\[ r^{(t+1)} = G \ r^{(t)} \]

\[ G \text{ is dense but computable using the sparse matrix } H. \]

\[ G = \alpha S + (1-\alpha) \frac{1}{n} ee^T \]
\[ = \alpha(H + \frac{1}{nae^T}) + (1-\alpha) \frac{1}{n} ee^T \]
\[ = \alpha H + (\alpha a + (1-\alpha)e) \frac{1}{n} e^T \]
Page Rank

The problems
The Rich Gets Richer

(Cho et al, 04)
Google Bombs

Google's (and Inktomi's) Miserable Failure
A search for miserable failure on Google brings up the official George W. Bush biography from the US White House web site. Dismissed by Google as not a ...
Google Bombs
Google Bombs

Did you mean: french military defeats

No standard web pages containing all your search terms were found.

Your search - french military victories - did not match any documents.

Suggestions:
- Make sure all words are spelled correctly.
- Try different keywords
- Try more general keywords
- Try fewer keywords.

Also, you can try Google Answers for expert help with your search.

These Weapons of Mass Destruction cannot be displayed

The weapons you are looking for are currently unavailable. The country might be experiencing technical difficulties, or you may need to adjust your weapons inspectors mandate.

Please try the following:
- Click the Regime change button, or try again later.
- If you are George Bush and typed the country's name in the address bar, make sure that it is spelled correctly. (IRAQ).
- To check your weapons inspector settings, click the UN menu, and then click Weapons Inspector Options. On the Security Council tab, click Consensus. The settings should match those provided by your government or NATO.
- If the Security Council has enabled it, The United States of America can examine your country and automatically discover Weapons of Mass Destruction.
- If you would like to use the CIA to try and discover them, click Expose weapons.
- Some countries require 120 thousand troops to liberate them. Click the Panic menu and then click About US foreign policy to determine what regime they will install.
- If you are an Old European Country trying to protect your interests, make sure your options are left wide open as long as possible. Click the Tools menu, and then click on
Link Farms
Take-home

Ranking is important.

Relationship between links and the importance of pages.

Why PageRank converges (to the right answer).

How link-based ranking methods can be manipulated.
g2g

ttyl