Link Analysis

Stony Brook University
CSE545, Spring 2019
The Web, circa 1998
The Web, circa 1998

Match keywords, language (information retrieval)

Explore directory
The Web, circa 1998

Easy to game with “term spam”

Match keywords, language (information retrieval)

Explore directory

Time-consuming; Not open-ended
Enter PageRank

The Anatomy of a Large-Scale Hypertextual Web Search Engine

Sergey Brin and Lawrence Page

Computer Science Department,
Stanford University, Stanford, CA 94305, USA
sergey@cs.stanford.edu and page@cs.stanford.edu

Abstract
In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure and produce much text and hyperlink

The PageRank Citation Ranking:
Bringing Order to the Web

January 29, 1998

... Abstract

The importance of a Web page is an inherently subjective matter, which depends on the readers' interests, knowledge and attitudes. But there is still much that can be said objectively in assessing the importance of Web pages. This paper introduces PageRank, a method for...
PageRank

Key Idea: Consider the citations of the website.
PageRank

**Key Idea:** Consider the *citations* of the website.

Who links to it? and what are their citations?
PageRank

Key Idea: Consider the citations of the website.

Who links to it? and what are their citations?

Innovation 1: What pages would a “random Web surfer” end up at?

Innovation 2: Not just own terms but what terms are used by citations?
Innovation 1: What pages would a “random Web surfer” end up at?

Innovation 2: Not just own terms but what terms are used by citations?
PageRank

Innovation 1: What pages would a “random Web surfer” end up at?

Innovation 2: Not just own terms but what terms are used by citations?
PageRank

View 1: Flow Model:

- in-links (citations) as votes
- but, citations from important pages should count more.

=> Use recursion to figure out if each page is important.

Innovation 1: What pages would a “random Web surfer” end up at?

Innovation 2: Not just own terms but what terms are used by citations?
PageRank

View 1: Flow Model:

How to compute?

Each page \((j)\) has an importance (i.e. rank, \(r_j\))

\[
vote_j = \frac{r_i}{n_j} \\
(\text{\(n_j\) is \(|\text{out-links}|\)})
\]

\[
\sum_{i \in \text{inLinks}(j)} vote_i
\]

\[
r_j = \sum_{i \in \text{inLinks}(j)} vote_i
\]
PageRank

View 1: Flow Model:

Each page \((j)\) has an importance (i.e. rank, \(r_j\))

\[
vote_j = \frac{r_i}{n_j}
\]

\[
r_j = \sum_{i \in \text{inLinks}(j)} vote_i
\]

\(n_j\) is |out-links|
PageRank

View 1: Flow Model:

How to compute?

Each page \((j)\) has an importance (i.e. rank, \(r_j\))

\[
vote_j = \frac{r_i}{n_j} \quad (n_j \text{ is } |\text{out-links}|)
\]

\[
r_j = \sum_{i \in \text{inLinks}(j)} vote_i
\]
PageRank

View 1: Flow Model:
A System of Equations:

\[ r_A = \frac{r_B}{2} + \frac{r_C}{1} \]

How to compute?

Each page \((j)\) has an importance (i.e. rank, \(r_j\))

\[ vote_j = \frac{r_i}{n_j} \quad \text{(} n_j \text{ is } |\text{out-links}|) \]

\[ r_j = \sum_{i \in \text{inLinks}(j)} vote_i \]
PageRank

View 1: Flow Model:

A System of Equations:

\[
\begin{align*}
    r_A &= \frac{r_B}{2} + \frac{r_C}{1} \\
    r_B &= \frac{3}{r_A} + \frac{2}{r_D} \\
    r_C &= \frac{3}{r_A} + \frac{2}{r_D} \\
    r_D &= \frac{3}{r_A} + \frac{2}{r_B}
\end{align*}
\]

How to compute?

Each page \((j)\) has an importance (i.e. rank, \(r_j\))

\[
\text{vote}_j = \frac{r_j}{n_j}
\]

\(n_j\) is \(|\text{out-links}|\)

\[
 r_j = \sum_{i \in \text{inLinks}(j)} \text{vote}_i
\]
PageRank

View 1: Flow Model: Solve

1 = r_A + r_B + r_C + r_D

How to compute?

Each page (j) has an importance (i.e. rank, r_j)

\[
vote_j = \frac{r_j}{n_j}
\]

\[
r_j = \sum_{i \in \text{inLinks}(j)} vote_i
\]

(n_j is |out-links|)
PageRank

\[ 1 = r_A + r_B + r_C + r_D \]

\[
\begin{align*}
    r_A &= \frac{r_B}{2} + \frac{r_C}{1} \\
    r_B &= \frac{r_A}{3} + \frac{r_D}{2} \\
    r_C &= \frac{r_A}{3} + \frac{r_D}{2} \\
    r_D &= \frac{r_A}{3} + \frac{r_B}{2}
\end{align*}
\]

Transition Matrix, \( M \)

<table>
<thead>
<tr>
<th>to \ from</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1/2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>C</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>D</td>
<td>1/3</td>
<td>1/2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
View 2: Matrix Formulation

1 = r_A + r_B + r_C + r_D

Transition Matrix, M

<table>
<thead>
<tr>
<th>to \ from</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1/2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>C</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>D</td>
<td>1/3</td>
<td>1/2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Innovation: What pages would a “random Web surfer” end up at?

View 2: Matrix Formulation

\[ 1 = r_A + r_B + r_C + r_D \]

\[
\begin{align*}
  r_A &= \frac{r_B}{2} + \frac{r_C}{1} \\
  r_B &= \frac{2r_A}{3} + \frac{1}{r_D} \\
  r_C &= \frac{3r_A}{2} + \frac{r_D}{2} \\
  r_D &= \frac{3}{2} + \frac{r_B}{2}
\end{align*}
\]

Transition Matrix, \( M \): 

<table>
<thead>
<tr>
<th></th>
<th>( A )</th>
<th>( B )</th>
<th>( C )</th>
<th>( D )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>0</td>
<td>1/2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( B )</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>( C )</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>( D )</td>
<td>1/3</td>
<td>1/2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
View 2: Matrix Formulation

\[
1 = r_A + r_B + r_C + r_D
\]

\[
\begin{align*}
 r_A &= \frac{r_B}{2} + \frac{r_C}{1} \\
 r_B &= \frac{r_A}{3} + \frac{r_D}{2} \\
 r_C &= \frac{r_A}{3} + \frac{r_D}{2} \\
 r_D &= \frac{r_A}{3} + \frac{r_B}{2}
\end{align*}
\]

### Transition Matrix, M

<table>
<thead>
<tr>
<th>to \ from</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1/2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>C</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>D</td>
<td>1/3</td>
<td>1/2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Innovation: What pages would a “random Web surfer” end up at?
To Start, all are equally likely at ¼: ends up at D

View 2: Matrix Formulation

\[ 1 = r_A + r_B + r_C + r_D \]

\[
egin{align*}
r_A &= \frac{r_B}{2} + \frac{r_C}{3} \\
r_B &= \frac{r_A}{2} + \frac{1}{2} \\
r_C &= \frac{r_A}{3} + \frac{r_D}{2} \\
r_D &= \frac{r_A}{3} + \frac{r_B}{2}
\end{align*}
\]

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>to \ from</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>1/2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>C</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>D</td>
<td>1/3</td>
<td>1/2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Transition Matrix, M
Innovation: What pages would a “random Web surfer” end up at?
To Start, all are equally likely at $\frac{1}{4}$: ends up at $D$
$C$ and $B$ are then equally likely: $\rightarrow D \rightarrow B = \frac{1}{4} \times \frac{1}{2}$; $\rightarrow D \rightarrow C = \frac{1}{4} \times \frac{1}{2}$

View 2: Matrix Formulation

$$1 = r_A + r_B + r_C + r_D$$

$$r_A = \frac{r_B}{2} + \frac{r_C}{1}$$
$$r_B = \frac{3}{2} r_A + \frac{1}{2} r_D$$
$$r_C = \frac{3}{2} r_A + \frac{2}{2} r_D$$
$$r_D = \frac{3}{2} r_A + \frac{1}{2} r_B$$

<table>
<thead>
<tr>
<th>to \ from</th>
<th>$A$</th>
<th>$B$</th>
<th>$C$</th>
<th>$D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>0</td>
<td>1/2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$B$</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>$C$</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>$D$</td>
<td>1/3</td>
<td>1/2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Transition Matrix, $M$
Innovation: What pages would a “random Web surfer” end up at?

To Start, all are equally likely at $\frac{1}{4}$: ends up at D

C and B are then equally likely: $\rightarrow D \rightarrow B = \frac{1}{4} \times \frac{1}{2}$; $\rightarrow D \rightarrow C = \frac{1}{4} \times \frac{1}{2}$

Ends up at C: then A is only option: $\rightarrow D \rightarrow C \rightarrow A = \frac{1}{4} \times \frac{1}{2} \times 1$

View 2: Matrix Formulation

$$1 = r_A + r_B + r_C + r_D$$

$$r_A = \frac{r_B}{2} + \frac{r_C}{3}$$

$$r_B = \frac{r_A}{3} + \frac{r_D}{2}$$

$$r_C = \frac{r_A}{3} + \frac{r_D}{2}$$

$$r_D = \frac{r_A}{3} + \frac{r_B}{2}$$

Transition Matrix, $M$

<table>
<thead>
<tr>
<th>to \ from</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1/2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>C</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>D</td>
<td>1/3</td>
<td>1/2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Innovation: What pages would a “random Web surfer” end up at?

View 2: Matrix Formulation

\[ 1 = r_A + r_B + r_C + r_D \]

\begin{align*}
    r_A &= \frac{r_B}{2} + \frac{r_C}{1} \\
    r_B &= \frac{r_A}{3} + \frac{r_D}{2} \\
    r_C &= \frac{r_A}{3} + \frac{r_D}{2} \\
    r_D &= \frac{r_A}{3} + \frac{r_B}{2}
\end{align*}

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
</table>
| **to**
| **from** | A  | B  | C  | D  |
| A    | 0  | 1/2| 1  | 0  |
| B    | 1/3| 0  | 0  | 1/2|
| C    | 1/3| 0  | 0  | 1/2|
| D    | 1/3| 1/2| 0  | 0  |

Transition Matrix, $M$
Innovation: What pages would a “random Web surfer” end up at? …

View 2: Matrix Formulation

\[ 1 = r_A + r_B + r_C + r_D \]

\[
\begin{align*}
    r_A &= \frac{r_B}{2} + \frac{r_C}{1} \\
    r_B &= \frac{r_A}{3} + \frac{r_D}{2} \\
    r_C &= \frac{r_A}{3} + \frac{r_B}{2} \\
    r_D &= \frac{r_A}{3} + \frac{r_D}{2}
\end{align*}
\]

\[
\begin{array}{c|cccc}
\text{to} & A & B & C & D \\
\hline
A & 0 & 1/2 & 1 & 0 \\
B & 1/3 & 0 & 0 & 1/2 \\
C & 1/3 & 0 & 0 & 1/2 \\
D & 1/3 & 1/2 & 0 & 0 \\
\end{array}
\]

Transition Matrix, M
Innovation: What pages would a “random Web surfer” end up at?

View 2: Matrix Formulation

\[ 1 = r_A + r_B + r_C + r_D \]

\begin{align*}
  r_A &= \frac{r_B}{2} + \frac{r_C}{1} \\
  r_B &= \frac{r_A}{3} + \frac{r_D}{2} \\
  r_C &= \frac{r_A}{3} + \frac{r_D}{2} \\
  r_D &= \frac{r_A}{3} + \frac{r_B}{2}
\end{align*}

\[
\begin{array}{|c|cccc|}
\hline
\text{to} & \text{from} & A & B & C & D \\
\hline
A & 0 & 1/2 & 1 & 0 \\
B & 1/3 & 0 & 0 & 1/2 \\
C & 1/3 & 0 & 0 & 1/2 \\
D & 1/3 & 1/2 & 0 & 0 \\
\hline
\end{array}
\]

Transition Matrix, M
Innovation: What pages would a “random Web surfer” end up at?
To start: N=4 nodes, so \( r = \left[ \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right] \)

View 2: Matrix Formulation

\[
1 = r_A + r_B + r_C + r_D
\]

\[
\begin{align*}
  r_A &= \frac{r_B}{2} + \frac{r_C}{2} \\
  r_B &= \frac{r_A}{3} + \frac{r_D}{2} \\
  r_C &= \frac{r_A}{3} + \frac{r_D}{2} \\
  r_D &= \frac{r_A}{3} + \frac{r_B}{2}
\end{align*}
\]

<table>
<thead>
<tr>
<th>to</th>
<th>from</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td>0</td>
<td>1/2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>C</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>D</td>
<td>1/3</td>
<td>1/2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Transition Matrix, M
Innovation: What pages would a “random Web surfer” end up at?
To start: N=4 nodes, so \( r = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}] \)
after 1st iteration: \( M \cdot r = [\frac{3}{8}, \frac{5}{24}, \frac{5}{24}, \frac{5}{24}] \)

View 2: Matrix Formulation

\[
1 = r_A + r_B + r_C + r_D
\]

\[
\begin{align*}
  r_A &= \frac{r_B}{2} + \frac{r_C}{3} \\
  r_B &= \frac{r_A}{3} + \frac{r_D}{2} \\
  r_C &= \frac{r_A}{3} + \frac{r_D}{2} \\
  r_D &= \frac{r_A}{3} + \frac{r_B}{2}
\end{align*}
\]

Transition Matrix, \( M \)

<table>
<thead>
<tr>
<th>to \ from</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>1/2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>C</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>D</td>
<td>1/3</td>
<td>1/2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Innovation: What pages would a “random Web surfer” end up at?

To start: N=4 nodes, so \( r = \left[ \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right] \)

after 1st iteration: \( M \cdot r = \left[ \frac{3}{8}, \frac{5}{24}, \frac{5}{24}, \frac{5}{24} \right] \)

after 2nd iteration: \( M(M \cdot r) = M^2 \cdot r = \left[ \frac{15}{48}, \frac{11}{48}, \ldots \right] \)

**View 2: Matrix Formulation**

\[ 1 = r_A + r_B + r_C + r_D \]

\[
\begin{align*}
  r_A &= \frac{r_B}{2} + \frac{r_C}{3} \\
  r_B &= \frac{2}{r_A} + \frac{1}{r_D} \\
  r_C &= \frac{3}{r_A} + \frac{2}{r_D} \\
  r_D &= \frac{3}{r_A} + \frac{2}{r_B}
\end{align*}
\]

<table>
<thead>
<tr>
<th></th>
<th>to ( A )</th>
<th>to ( B )</th>
<th>to ( C )</th>
<th>to ( D )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A )</td>
<td>0</td>
<td>1/2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>( B )</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>( C )</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/2</td>
</tr>
<tr>
<td>( D )</td>
<td>1/3</td>
<td>1/2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Transition Matrix, \( M \)
Innovation: What pages would a “random Web surfer” end up at?
To start: N=4 nodes, so \( r = \left[ \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right] \)
after 1st iteration: \( M \cdot r = \left[ \frac{3}{8}, \frac{5}{24}, \frac{5}{24}, \frac{5}{24} \right] \)
after 2nd iteration: \( M(M \cdot r) = M^2 \cdot r = \left[ \frac{15}{48}, \frac{11}{48}, \ldots \right] \)

**Power iteration algorithm**

initialize: \( r[0] = \left[ \frac{1}{N}, \ldots, \frac{1}{N} \right], \)
\( r[-1]=\left[0,\ldots,0\right] \)
while (err_norm(\( r[t] \),\( r[t-1] \))>min_err):

\[
\text{err}_\text{norm}(v1, v2) = |v1 - v2| \text{ #L1 norm}
\]
Innovation: What pages would a “random Web surfer” end up at?

To start: N=4 nodes, so \( r = \left[ \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \right] \)

after 1st iteration: \( M \cdot r = \left[ \frac{3}{8}, \frac{5}{24}, \frac{5}{24}, \frac{5}{24} \right] \)

after 2nd iteration: \( M(M \cdot r) = M^2 \cdot r = \left[ \frac{15}{48}, \frac{11}{48}, \ldots \right] \)

Power iteration algorithm

initialize: \( r[0] = [1/N, \ldots, 1/N], \) 
\( r[-1]=[0, \ldots, 0] \)

while (err_norm(\( r[t] \), \( r[t-1] \))>min_err):
  \( r[t+1] = M \cdot r[t] \)
  \( t+=1 \)

solution = \( r[t] \)

err_norm(\( v1 \), \( v2 \)) = \|v1 - v2\| #L1 norm

\[ \text{ “Transition Matrix”, } M \]
As err_norm gets smaller we are moving toward: $r = M \cdot r$

View 3: Eigenvectors:

**Power iteration algorithm**

initialize: $r[0] = [1/N, \ldots, 1/N]$,  
$r[-1]= [0, \ldots, 0]$

while (err_norm($r[t], r[t-1]$)>min_err):  
  $r[t+1] = M \cdot r[t]$
  $t+=1$

solution = $r[t]$

err_norm($v1, v2$) = $|v1 - v2|$ #L1 norm
As \( \text{err\_norm} \) gets smaller we are moving toward: \( r = M \cdot r \)

**View 3: Eigenvectors:**
We are actually just finding the *eigenvector* of \( M \).

---

**Power iteration algorithm**

initialize: \( r[0] = [1/N, \ldots, 1/N] \)
\( r[-1]=[0,\ldots,0] \)

while (\( \text{err\_norm}(r[t],r[t-1])>\text{min\_err} \)):
  \( r[t+1] = M \cdot r[t] \)
  \( t+=1 \)

solution = \( r[t] \)

\( \text{err\_norm}(v1, v2) = |v1 - v2| \) #L1 norm

---

(Leskovec at al., 2014; http://www.mmds.org/)
As err_norm gets smaller we are moving toward: $r = M \cdot r$

View 3: Eigenvectors:
We are actually just finding the *eigenvector* of M.

**Power iteration algorithm**

initialize: $r[0] = [1/N, \ldots, 1/N]$
$r[-1]=[0,\ldots,0]$

while (err_norm$(r[t],r[t-1])>\text{min\_err}$):

$r[t+1] = M \cdot r[t]$
$t+=1$
solution = $r[t]$

err_norm$(v1, v2) = \text{sum}(|v1 - v2|)$
#L1 norm

$x$ is an *eigenvector* of A if:
$A \cdot x = \lambda \cdot x$

$\lambda = 1$ (eigenvalue for 1st principal eigenvector)
since columns of M sum to 1.
Thus, if $r$ is $x$, then $Mr=1r$
View 4: Markov Process

Where is surfer at time $t+1$? $p(t+1) = M \cdot p(t)$

Suppose: $p(t+1) = p(t)$, then $p(t)$ is a stationary distribution of a random walk.

Thus, $r$ is a stationary distribution. Probability of being at given node.
View 4: Markov Process

Where is surfer at time t+1? \[ p(t+1) = M \cdot p(t) \]

Suppose: \( p(t+1) = p(t) \), then \( p(t) \) is a stationary distribution of a random walk.

Thus, \( r \) is a stationary distribution. Probability of being at given node.

aka 1st order Markov Process

- Rich probabilistic theory. One finding:
  - Stationary distributions have a unique distribution if:
    - No "dead-ends": a node can’t propagate its rank
    - No "spider traps": set of nodes with no way out.

Also known as being stochastic, irreducible, and aperiodic.
aka 1st order Markov Process

- Rich probabilistic theory. One finding:
  - Stationary distributions have a unique distribution if:
    - No "dead-ends": a node can’t propagate its rank
    - No "spider traps": set of nodes with no way out.
  - Also known as being stochastic, irreducible, and aperiodic.

What would \( r \) converge to?
View 4: Markov Process - Problems for vanilla PI

aka 1st order Markov Process

- Rich probabilistic theory. One finding:
  - Stationary distributions have a unique distribution if:
    - No "dead-ends": a node can’t propagate its rank
    - No "spider traps": set of nodes with no way out.
  - Also known as being stochastic, irreducible, and aperiodic.

<table>
<thead>
<tr>
<th>to \ from</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>1/3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

What would \( r \) converge to?
aka 1st order Markov Process

- Rich probabilistic theory. One finding:
  - Stationary distributions have a unique distribution if:
    - Columns sum to 1
    - Same node doesn’t repeat at regular intervals
    - Non-zero chance of going to any other node
  - Also known as being stochastic, irreducible, and aperiodic.
Goals:
No “dead-ends”
No “spider traps”

The “Google” PageRank Formulation
Add teleportation: At each step, two choices
1. Follow a random link (probability, $\beta = \sim .85$)
2. Teleport to a random node (probability, $1-\beta$)
Goals:
No “dead-ends”
No “spider traps”

The “Google” PageRank Formulation
Add teleportation: At each step, two choices
1. Follow a random link (probability, $\beta = \sim .85$)
2. Teleport to a random node (probability, $1-\beta$)

<table>
<thead>
<tr>
<th>to \ from</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>$\frac{1}{3}$</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>$\frac{1}{3}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>$\frac{1}{3}$</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
**Goals:**
- No “dead-ends”
- No “spider traps”

---

**The “Google” PageRank Formulation**

Add teleportation: At each step, two choices

1. Follow a random link (probability, $\beta = \sim.85$)
2. Teleport to a random node (probability, $1-\beta$)

---

<table>
<thead>
<tr>
<th>to \ from</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0+.15*¼</td>
<td>1</td>
<td>0+.15*¼</td>
</tr>
<tr>
<td>B</td>
<td>⅓</td>
<td>0+.15*¼</td>
<td>0</td>
<td>.85<em>1+.15</em>¼</td>
</tr>
<tr>
<td>C</td>
<td>⅓</td>
<td>0+.15*¼</td>
<td>0</td>
<td>0+.15*¼</td>
</tr>
<tr>
<td>D</td>
<td>⅓</td>
<td>.85<em>1+.15</em>¼</td>
<td>0</td>
<td>0+.15*¼</td>
</tr>
</tbody>
</table>
Goals:
No “dead-ends”
No “spider traps”

The “Google” PageRank Formulation
Add teleportation: At each step, two choices
1. Follow a random link (probability, $\beta = \sim .85$)
2. Teleport to a random node (probability, $1-\beta$)

```
 to \ from | A | B | C | D \\
-----------|---|---|---|---|
  A         | 0+.15*1/4 | 0+.15*1/4 | 85*1+.15*1/4 | 0+.15*1/4 |
  B         | .85*1/3+.15*1/4 | 0+.15*1/4 | 0+.15*1/4 | .85*1+.15*1/4 |
  C         | .85*1/3+.15*1/4 | 0+.15*1/4 | 0+.15*1/4 | 0+.15*1/4 |
  D         | .85*1/3+.15*1/4 | .85*1+.15*1/4 | 0+.15*1/4 | 0+.15*1/4 |
```
**Goals:**
- No “dead-ends”
- No “spider traps”

**The “Google” PageRank Formulation**

Add teleportation: At each step, two choices

1. Follow a random link (probability, $\beta = \sim .85$)
2. Teleport to a random node (probability, $1-\beta$)

### Transition Probabilities

<table>
<thead>
<tr>
<th>to \ from</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>B</strong></td>
<td>$\frac{1}{3}$</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>C</strong></td>
<td>$\frac{1}{3}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>D</strong></td>
<td>$\frac{1}{3}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
**Goals:**

No "dead-ends"
No "spider traps"

**The “Google” PageRank Formulation**

Add teleportation: At each step, two choices
1. Follow a random link (probability, $\beta = \sim .85$)
2. Teleport to a random node (probability, $1-\beta$)

<table>
<thead>
<tr>
<th>$to$ \ $from$</th>
<th>$A$</th>
<th>$B$</th>
<th>$C$</th>
<th>$D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>0</td>
<td>$\frac{1}{4}$</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$B$</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{4}$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$C$</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{4}$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$D$</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{4}$</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Goals:
No “dead-ends”
No “spider traps”

The “Google” PageRank Formulation
Add teleportation: At each step, two choices
1. Follow a random link (probability, $\beta = \sim.85$)
2. Teleport to a random node (probability, $1-\beta$)

<table>
<thead>
<tr>
<th>to \ from</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0.85<em>⅓+.15</em>⅓</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>⅓</td>
<td>0.85<em>⅓+.15</em>⅓</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>⅓</td>
<td>0.85<em>⅓+.15</em>⅓</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>⅓</td>
<td>0.85<em>⅓+.15</em>⅓</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
The “Google” PageRank Formulation
Add teleportation: At each step, two choices
1. Follow a random link (probability, $\beta = \sim .85$)
2. Teleport to a random node (probability, $1-\beta$)
   (Teleport from a dead-end has probability 1)

Goals:
No “dead-ends”
No “spider traps”
Goals:
No “dead-ends”
No “spider traps”

Teleportation, as Flow Model:

\[
 r_j = \sum_{i \rightarrow j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{N}
\]

(Brin and Page, 1998)
Teleportation, as Flow Model:

\[ r_j = \sum_{i \rightarrow j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{N} \]

(Brin and Page, 1998)

Teleportation, as Matrix Model:

\[ M' = \beta M + (1 - \beta) \left[ \frac{1}{N} \right]_{N \times N} \]
**Goals:**
No “dead-ends”
No “spider traps”

**Teleportation, as Flow Model:**

\[
 r_j = \sum_{i \rightarrow j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{N}
\]

*(Brin and Page, 1998)*

**Teleportation, as Matrix Model:**

\[
 M' = \beta M + (1 - \beta) \left[ \frac{1}{N} \right]_{N \times N}
\]

<table>
<thead>
<tr>
<th>to \ from</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0+.15*¼</td>
<td>.85<em>¼+.15</em>¼</td>
<td>85<em>1+.15</em>¼</td>
<td>0+.15*¼</td>
</tr>
<tr>
<td>B</td>
<td>.85<em>¾+.15</em>¼</td>
<td>.85<em>¼+.15</em>¼</td>
<td>0+.15*¼</td>
<td>.85<em>1+.15</em>¼</td>
</tr>
<tr>
<td>C</td>
<td>.85<em>¾+.15</em>¼</td>
<td>.85<em>¼+.15</em>¼</td>
<td>0+.15*¼</td>
<td>0+.15*¼</td>
</tr>
<tr>
<td>D</td>
<td>.85<em>¾+.15</em>¼</td>
<td>.85<em>¼+.15</em>¼</td>
<td>0+.15*¼</td>
<td>0+.15*¼</td>
</tr>
</tbody>
</table>
Goals:
No “dead-ends”
No “spider traps”

Teleportation, as Flow Model:
\[ r_j = \sum_{i \to j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{N} \]

(Brin and Page, 1998)

Teleportation, as Matrix Model:
\[ M' = \beta M + (1 - \beta) \left[ \frac{1}{N} \right]_{N \times N} \]

To apply:
run power iterations over \( M' \) instead of \( M \).

<table>
<thead>
<tr>
<th>to \ from</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0+.15*¼</td>
<td>1*¼</td>
<td>85<em>1+.15</em>¼</td>
<td>0+.15*¼</td>
</tr>
<tr>
<td>B</td>
<td>.85<em>⅓+.15</em>¼</td>
<td>1*¼</td>
<td>0+.15*¼</td>
<td>.85<em>1+.15</em>¼</td>
</tr>
<tr>
<td>C</td>
<td>.85<em>⅓+.15</em>¼</td>
<td>1*¼</td>
<td>0+.15*¼</td>
<td>0+.15*¼</td>
</tr>
<tr>
<td>D</td>
<td>.85<em>⅓+.15</em>¼</td>
<td>1*¼</td>
<td>0+.15*¼</td>
<td>0+.15*¼</td>
</tr>
</tbody>
</table>
**Goals:**
No “dead-ends”
No “spider traps”

**Teleportation, as Flow Model:**

\[ r_j = \sum_{i \to j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{N} \]

(Brin and Page, 1998)

**Steps:**
1. Compute M
2. Add 1/N to all dead-ends.
3. Convert M to M’
4. Run Power Iterations.

<table>
<thead>
<tr>
<th>to \ from</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0+.15*1( \frac{1}{4} )</td>
<td>1*1( \frac{1}{4} )</td>
<td>85<em>1+.15</em>1( \frac{1}{4} )</td>
<td>0+.15*1( \frac{1}{4} )</td>
</tr>
<tr>
<td>B</td>
<td>.85<em>1( \frac{1}{3} )+.15</em>1( \frac{1}{4} )</td>
<td>1*1( \frac{1}{4} )</td>
<td>0+.15*1( \frac{1}{4} )</td>
<td>.85<em>1+.15</em>1( \frac{1}{4} )</td>
</tr>
<tr>
<td>C</td>
<td>.85<em>1( \frac{1}{3} )+.15</em>1( \frac{1}{4} )</td>
<td>1*1( \frac{1}{4} )</td>
<td>0+.15*1( \frac{1}{4} )</td>
<td>0+.15*1( \frac{1}{4} )</td>
</tr>
<tr>
<td>D</td>
<td>.85<em>1( \frac{1}{3} )+.15</em>1( \frac{1}{4} )</td>
<td>1*1( \frac{1}{4} )</td>
<td>0+.15*1( \frac{1}{4} )</td>
<td>0+.15*1( \frac{1}{4} )</td>
</tr>
</tbody>
</table>
**Goals:**
No “dead-ends”
No “spider traps”

Teleportation, as Flow Model:

\[ r_j = \sum_{i \rightarrow j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{N} \]

(Brin and Page, 1998)

Teleportation, as Matrix Model:

\[ M' = \beta M + (1 - \beta) \left[ \frac{1}{N} \right]_{N \times N} \]

**Steps:**
1. Compute \( M \)
2. Add \( \frac{1}{N} \) to all dead-ends.
3. Convert \( M \) to \( M' \)
4. Run Power Iterations.

In Practice, Just store \( \beta M \) as sparse matrix and distribute according to above.
Teleportation, as Flow Model:

\[ r_j = \sum_{i \rightarrow j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{N} \]

(Brin and Page, 1998)

In other words, you only need to store \( M \) (as a sparse matrix) and \( r \) (as a vector), but never store \( M' \). Use this function within the inner loop of power iterations to achieve the same result as if using \( M' \).

Teleportation, as Matrix Model:

\[ M' = \beta M + (1 - \beta) \left[ \frac{1}{N} \right]_{N \times N} \]

Steps:
1. Compute \( M \)
2. Add \( 1/N \) to all dead-ends.
3. Convert \( M \) to \( M' \)
4. Run Power Iterations.

In Practice, Just store \( \beta M \) as a sparse matrix and distribute according to above.
Summary

- Flow View: Link Voting
- Matrix View: Linear Algebra
  - Eigenvectors View
- Markov Process View
- How to remove:
  - Dead Ends
  - Spider Traps

In practice, sparse matrix, implement teleportation functionally rather than update $M'$. 