Google PageRank

Angela Chung, Nancy Che Mahan, Conor Whelan
with Dr. Jen-Mei Chang

California State University Long Beach, Department of Mathematics & Statistics

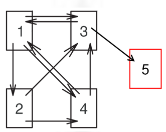
INTRODUCTION

Search engines such as Google do three basic things:

1. Find all pages with public access
2. Index the data
3. ***Rate the importance of each page***

Definitions:

* **Dangling Node**
  Page with no outgoing links
  Here, 5 is a dangling node



* **Column-stochastic matrix**
  Example:

$$\begin{bmatrix} \frac{1}{3} & 0 & \frac{1}{2} \\ \frac{1}{3} & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 \end{bmatrix}$$

* **Non-unique rankings:** $\dim(V(A)) > 1$
  where V(A) is the eigenspace of A

METHOD

Power Method

- Is an iterative method that finds the dominant eigenvalue and the corresponding eigenvector for any square matrix.
- Guarantees convergence to the unique eigenvector
- At every iteration, the vector x_k is multiplied by the matrix A and normalized.

$$x_{k+1} = \frac{Ax_k}{\|Ax_k\|}$$

Requirements to use the Power Method for PageRank:

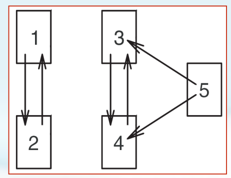
- ✓ No dangling nodes
- ✓ Column stochastic matrix
- ✓ Unique ranking: *dim (V(A)) = 1*

APPLICATION & RESULT

Example: A small scale web of only five pages

- ***Step 1:* *Find the link matrix A:***

$$\vec{x} = A\vec{x}$$



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

$$\vec{x}_1 = [\tfrac{1}{2}, \tfrac{1}{2}, 0, 0, 0]^T \qquad \vec{x}_2 = [0, 0, \tfrac{1}{2}, \tfrac{1}{2}, 0]^T$$

Problem: $\dim(V(A)) = 2$

- ***Step 2:* *Modify matrix A to ensure unique-ranking***

Modification formula: $M = (1-m)A + mS$

$0 \le m \le 1$, S = nxn matrix with all entries 1/n

$$M = \begin{bmatrix} 0.03 & 0.88 & 0.03 & 0.03 & 0.03 \\ 0.88 & 0.03 & 0.03 & 0.03 & 0.03 \\ 0.03 & 0.03 & 0.03 & 0.88 & 0.455 \\ 0.03 & 0.03 & 0.88 & 0.03 & 0.455 \\ 0.03 & 0.03 & 0.03 & 0.03 & 0.03 \end{bmatrix}$$

$$\dim(V(M)) = 1$$

- ***Step 3:* *Apply Power Method to solve for eigenvector x.***

$$\vec{x} = [\,0.2,\ 0.2,\ 0.285,\ 0.285,\ 0.03\,]^T$$

Ranking: $x_3 = x_4 > x_1 = x_2 > x_5$

→ Google will show either **page 3** or **page 4** at the top!

DISCUSSIONS

- **Our example above only consisted of 5 pages and thus the eigenvector was easily calculated. In reality, Google is dealing with link matrices that are much greater in size, and thus the Power Method is a powerful tool.**
- The rated importance of web pages is not the only factor in how links are presented, but it is a significant one.
- The power method is not the only method used to find the eigenvector, however Google uses this method because of the following motivations:
  - ✓ Simple to implement
  - ✓ Requires minimal storage
  - ✓ Robust and predictable convergence behavior
  - ✓ Numerically stable

CONCLUSION

Google's PageRank algorithm ranks the importance of web pages based on the eigenvector of a weighted link matrix.

The power method finds this unique eigenvector with eigenvalue 1, and by construction, appropriately ranks the web of pages.

ACKNOWLEDGEMENT / REFERENCES

- *The $25,000,000,000 Eigenvector: The Linear Algebra behind Google* (Bryan & Leise, 2006)
- *Ordinal Ranking for Google's Pagerank* (Wills & Ipsen, 2009)
- Images from:
  http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/
  http://www.raffertypendery.com/internet-marketing-articles/page-rank.php
  http://www.csulb.edu/~jchang9/images/img121250747.jpg