Analysis and Computation of Google's PageRank

Ilse Ipsen

North Carolina State University, USA

Joint work with Rebecca S. Wills

PageRank

An objective measure of the citation importance of a web page [Brin & Page 1998]

. . . continues to provide the basis for all of our web search tools http://www.google.com/technology/

- Assigns a rank to every web page
- Influences the order in which Google displays search results
- Based on link structure of the web graph

Overview

- PageRank
- Computation of PageRank
- Termination Criterion
- Ranking

Simple Web Model

Construct matrix S

- Page i has $d \geq 1$ outgoing links:

  If page i has link to page j then $s_{ij} = 1/d$

  else $s_{ij} = 0$

- Page i has 0 outgoing links: $s_{ij} = 1/n$

  (dangling node)

s_{ij}: probability that surfer moves

from page i to page j

Google Matrix

S is stochastic: $0 \le s_{ij} \le 1 \quad S\mathbf{1} = \mathbf{1}$

Google Matrix: convex combination

$$G = \alpha S + (1 - \alpha)\mathbf{1}v^T$$

Damping factor $0 < \alpha < 1$, e.g. $\alpha = .85$
Personalization vector $v \ge 0 \quad \|v\|_1 = 1$

TrustRank: $v_i = 0$ if page i is spam page

[Gyöngyi, Garcia-Molina & Pedersen 2004]

PageRank

$$G = \alpha S + (1-\alpha)\mathbf{1}v^T$$

Unique left eigenvector:

$$\pi^T G = \pi^T \qquad \pi \geq 0 \qquad \|\pi\|_1 = 1$$

ith entry of π: PageRank of page i

PageRank $\doteq$ largest left eigenvector of G

Properties of PageRank and extensions:

[Serra-Capizzano 2004, Horn & Serra-Capizzano 2006]

PageRank Computation

- Power method
  Page, Brin, Motwani & Winograd 1999
- Acceleration of power method
  Kamvar, Haveliwala, Manning & Golub 2003
  Haveliwala, Kamvar, Klein, Manning & Golub 2003
  Brezinski & Redivo-Zaglia 2004
  Brezinski, Redivo-Zaglia & Serra-Capizzano 2005
- Aggregation/Disaggregation
  Langville & Meyer 2002, 2003, 2004
  Ipsen & Kirkland 2004

PageRank Computation

- Methods that adapt to web graph
  Broder, Lempel, Maghoul & Pedersen 2004
  Kamvar, Haveliwala & Golub 2004
  Haveliwala, Kamvar, Manning & Golub 2003
  Lee, Golub & Zenios 2003
  Lu, Zhang, Xi, Chen, Liu, Lyu & Ma 2004
- Krylov methods
  Golub & Greif 2004

PageRank Computation

- Linear System Solution
  Arasu, Novak & Tomkins 2003
  Bianchini, Gori & Scarselli 2003
  Gleich, Zukov & Berkhin 2004
  Del Corso, Gullì & Romani 2004
- Survey
  Berkhin 2005

Power Method

Want: π such that $\pi^T G = \pi^T$

Power method: $[x^{(k)}]^T = [x^{(0)}]^T G^k, \quad k \geq 1$

- Converges to unique vector
- Vectorizes
- Storage for only a single vector
- Matrix vector multiplication with very sparse matrix
- Accurate (no subtractions)
- Simple (few decisions)

Error in Power Method

$$\pi^T G = \pi^T \qquad G = \alpha S + (1-\alpha) 1 v^T$$

Forward error in iteration k:

$$\|x^{(k)} - \pi\| \leq 2\,\alpha^k$$

[Bianchini, Gori & Scarselli 2003]

Residual in iteration k:

$$\|[x^{(k)}]^T G - [x^{(k)}]^T\| \leq 2\,\alpha^k$$

Norms: $1, \infty$

Termination Criterion

Residual:

$$[x^{(k)}]^T G - [x^{(k)}]^T = [x^{(k+1)}]^T - [x^{(k)}]^T$$

Power method:

Pick $x^{(0)} \geq 0 \quad \|x^{(0)}\|_1 = 1$

Repeat $\quad [x^{(k+1)}]^T = [x^{(k)}]^T G$

until $\|x^{(k+1)} - x^{(k)}\|$ small enough

After k iterations: $\quad \|x^{(k+1)} - x^{(k)}\| \leq 2\,\alpha^k$

Iteration Counts for Different α

$$G = \alpha S + (1 - \alpha) 1 v^T$$

α	$n = 281903$	$n = 683446$	bound
.85	69	65	118
.90	107	102	166
.95	219	220	415
.99	1114	1208	2075

bound: k such that $2\ \alpha^k \leq 10^{-8}$

Iteration counts independent of matrix size n

Ranking

After k iterations of power method:

Residual: $\|[x^{(k)}]^T G - [x^{(k)}]^T\| \leq 2\,\alpha^k$

Error: $\|x^{(k)} - \pi\| \leq 2\,\alpha^k$

But: Do the components of $x^{(k)}$ have the same ranking as those of π?

Rank-stability, rank-similarity: [Lempel & Moran, 2005] [Borodin, Roberts, Rosenthal & Tsaparas 2005]

Absolute Errors versus Ranking

$$\pi^T = \begin{pmatrix} .23 & .24 & .26 & .27 \end{pmatrix}$$

- $$[x^{(k)}]^T = \begin{pmatrix} .27 & .26 & .24 & .23 \end{pmatrix}$$

  $$\|x^{(k)} - \pi\|_\infty = .04$$

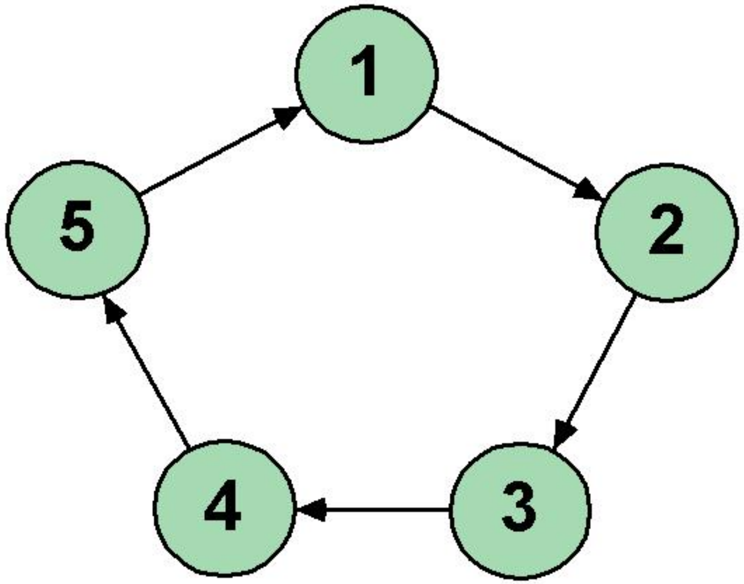
  Small error, but incorrect ranking

- $$[x^{(k)}]^T = \begin{pmatrix} 0 & .001 & .002 & .997 \end{pmatrix}$$

  $$\|x^{(k)} - \pi\|_\infty = .727$$

  Large error, but correct ranking

Web Graph is a Ring



$$S = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

All Pages are Trusted

S is circulant of order n, $\quad v = \frac{1}{n}\mathbf{1}$

- PageRank: $\quad \pi = \frac{1}{n}\mathbf{1}$

  All pages have same PageRank

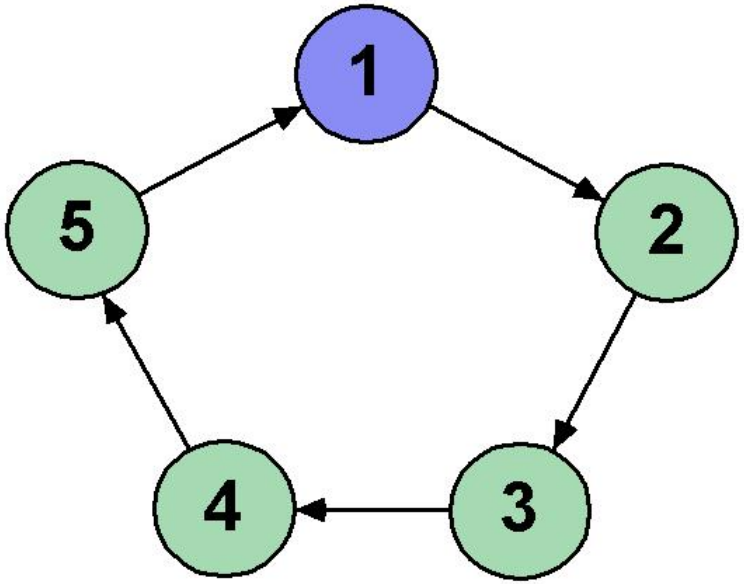
- Power method

  $x^{(0)} = v$: $\quad x^{(0)} = \pi \qquad$ correct ranking

  $x^{(0)} \neq v$: $\quad [x^{(k)}]^T = \frac{1}{n}\mathbf{1}^T + \alpha^k \left([x^{(0)}]^T S^k - \frac{1}{n}\mathbf{1}\right)$

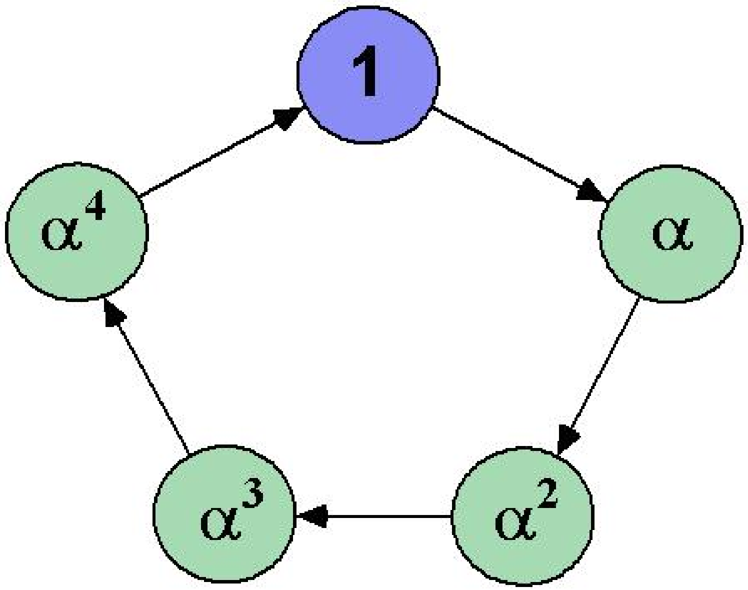
  Ranking does not converge (in exact arithmetic)

Only One Page is Trusted



$$v^T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Only One Page is Trusted



$$\pi^T \sim \begin{pmatrix} 1 & \alpha & \alpha^2 & \alpha^3 & \alpha^4 \end{pmatrix}$$

PageRank decreases with distance from page 1

Only One Page is Trusted

S is circulant of order n, $\quad v = e_1$

- PageRank: $\quad \pi^T \sim \begin{pmatrix} 1 & \alpha & \dots & \alpha^{n-1} \end{pmatrix}$

- Power method with $x^{(0)} = v$:

  $[x^{(k)}]^T \sim \begin{pmatrix} 1 & \alpha & \dots & \alpha^{k-1} & \frac{\alpha^k}{1-\alpha} & 0 & \dots & 0 \end{pmatrix}$

  $[x^{(n)}]^T \sim \begin{pmatrix} 1 + \frac{\alpha^n}{1-\alpha} & \alpha & \alpha^2 & \dots & \alpha^{n-1} \end{pmatrix}$

  Rank convergence in n iterations

Too Many Iterations

Power method with $x^{(0)} = v = e_1$:

- After n iterations:
  $[x^{(n)}]^T \sim \begin{pmatrix} 1 + \frac{\alpha^n}{1-\alpha} & \alpha & \alpha^2 & \dots & \alpha^{n-1} \end{pmatrix}$

- After $n + 1$ iterations:
  $[x^{(n+1)}]^T \sim \begin{pmatrix} 1 + \alpha^n & \alpha + \frac{\alpha^{n+1}}{1-\alpha} & \alpha^2 & \dots & \alpha^{n-1} \end{pmatrix}$

  If $\alpha = .85$, $n = 10$: $\quad \alpha + \frac{\alpha^{n+1}}{1-\alpha} > 1 + \alpha^n$

  Additional iterations can destroy a converged ranking

Recovery of Ranking

S is circulant of order n

- After k iterations:

$$[x^{(k)}]^T = \alpha^k \, [x^{(0)}]^T S^k + (1-\alpha) v^T \sum_{j=0}^{k-1} \alpha^j S^j$$

- After $k+n$ iterations:

$$[x^{(k+n)}]^T = \alpha^n \, [x^{(k)}]^T + (1-\alpha^n) \pi^T$$

If $x^{(k)}$ has correct ranking, so does $x^{(k+n)}$

Any Personalization Vector

S is circulant of order n

- PageRank: $\pi^T \sim v^T \sum_{j=0}^{n-1} \alpha^j S^j$
- Power method with $x^{(0)} = \frac{1}{n}\mathbf{1}$

$$[x^{(n)}]^T = \underbrace{(1-\alpha^n)}_{\text{scalar}}\pi^T + \underbrace{\frac{\alpha^n}{n}\mathbf{1}^T}_{\text{constant vector}}$$

For any v: rank convergence after n iterations

Summary

- Google matrix $G = \alpha S + (1 - \alpha)\, \mathbf{1} v^T$
- PageRank π: left eigenvector of G
- Computation of π via power method
- Convergence of residual/error norms:
  Depends only on α
- Convergence of rank:
  May never happen
  Depends on: α, v, starting vector, matrix size,
  graph of S
  A converged ranking can be destroyed