Analysis and Computation of Google's PageRank

Ilse Ipsen

North Carolina State University

Joint work with Rebecca M. Wills

Google

The heart of our software is PageRank™,
a system for ranking web pages developed by our founders Larry Page and Sergey Brin at Stanford University.

And while we have dozens of engineers working to improve every aspect of Google on a daily basis, PageRank continues to provide the basis for all of our web search tools.

http://www.google.com/technology/index.html (May 2005)

PageRank

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

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

Overview

- Simple Web Model
- Google Matrix
- Stability of PageRank
- Eigenvalue Problem: Power Method
- Linear System: Jacobi Method
- Dangling Nodes
- Krylov Spaces

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

Matrix S

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

Left eigenvector: $\omega^T S = \omega^T \quad \omega \geq 0 \quad \|\omega\|_1 = 1$

Ranking: ω_i is probability that surfer visits page i

But:

- S does not model surfing behaviour properly
- Rank sinks, and pages with zero rank
- Several eigenvalues with magnitude 1
  $\Rightarrow$ power method does not converge

Remedy: Change the matrix

Google Matrix

Convex combination

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

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

Properties of G:

- stochastic $\Rightarrow$ G has eigenvalue 1
- primitive $\Rightarrow$ spectral radius 1 unique

Page Rank

Unique left eigenvector:

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

Power method converges to π

ith entry of π: PageRank of page i

PageRank $\doteq$ largest left eigenvector of G

Stability of PageRank

How sensitive is PageRank π to

- Round off errors
- Changes in damping factor α
- Changes in personalization vector v
- Addition/deletion of links

Perturbation Theory

For Markov chains

Schweizer 1968, Meyer 1980
Haviv & van Heyden 1984
Funderlic & Meyer 1986
Seneta 1988, 1991
Ipsen & Meyer 1994
Kirkland, Neumann & Shader 1998
Cho & Meyer 2000, 2001
Kirkland 2003, 2004

Perturbation Theory

For Google matrix

Chien, Dwork, Kumar & Sivakumar 2001
Ng, Zheng & Jordan 2001
Bianchini, Gori & Scarselli 2003
Boldi, Santini & Vigna 2004
Langville & Meyer 2004
Golub & Greif 2004
Kirkland 2005
Chien, Dwork, Kumar, Simon & Sivakumar 2005

Changes in the Matrix S

Exact:

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

Perturbed:

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

Error:

$$\tilde{\pi}^T - \pi^T = \alpha \tilde{\pi}^T E (I - \alpha S)^{-1}$$

$$\|\tilde{\pi} - \pi\|_1 \leq \frac{\alpha}{1-\alpha} \|E\|_\infty$$

[Kirkland 2005]

Changes in Damping Factor α

Exact:

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

Perturbed:

$$\tilde{\pi}^T \tilde{G} = \tilde{\pi}^T \qquad \tilde{G} = (\alpha + \mu)S + (1 - (\alpha + \mu))\,\mathbf{1}v^T$$

Error:

$$\|\tilde{\pi} - \pi\|_1 \leq \frac{2}{1-\alpha}\,\mu$$

[Langville & Meyer 2004]

Changes in Vector v

Exact:

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

Perturbed:

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

Error:

$$\|\tilde{\pi} - \pi\|_1 \leq \|f\|_1$$

Sensitivity of PageRank π

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

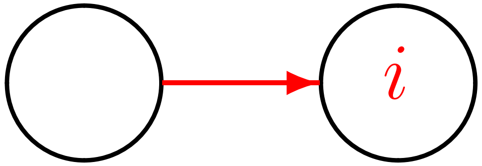
Changes in

- S: condition number $\alpha/(1-\alpha)$
- α: condition number $2/(1-\alpha)$
- f: condition number 1

$\alpha = .85$: condition numbers ≤ 14
$\alpha = .99$: condition numbers ≤ 200

PageRank insensitive to perturbations

Adding an In-Link



$$\tilde{\pi}_i \geq \pi_i$$

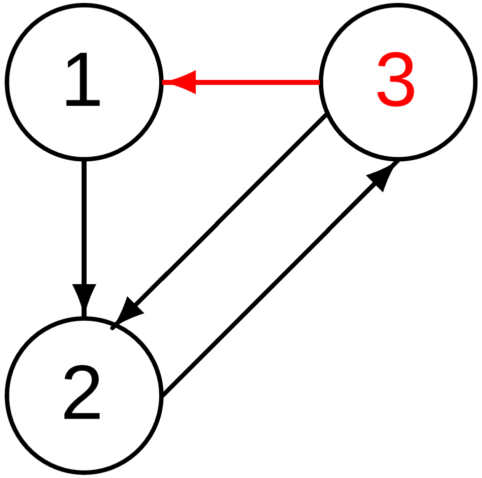
Adding an in-link can only increase PageRank (monotonicity)

Removing an in-link can only decrease PageRank

[Chien, Dwork, Kumar & Sivakumar 2001]

[Chien, Dwork, Kumar, Simon & Sivakumar 2005]

Adding an Out-Link



$$\tilde{\pi}_3 = \frac{1+\alpha+\alpha^2}{3(1+\alpha+\alpha^2/2)} < \pi_3 = \frac{1+\alpha+\alpha^2}{3(1+\alpha)}$$

Adding an out-link may decrease PageRank

PageRank Computation

Eigenvector problem: $\pi^T G = \pi^T$

Power method:

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

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

until $\|x^{(k+1)} - x^{(k)}\| \leq \tau$

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

Why is an Iteration Cheap?

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

$$S = H + \underbrace{dw^T}_{\text{dangling nodes}} \qquad w \geq 0 \quad \|w\|_1 = 1$$

often: $w = \frac{1}{n}\,\mathbf{1}$

Matrix H: models webgraph
substochastic
dimension: several billion
very sparse

Matrix Vector Multiplication

Vector $x > 0 \quad \|x\|_1 = 1$

$$\begin{aligned} x^T G &= x^T \left[\alpha(H + dw^T) + (1-\alpha)\mathbf{1}v^T\right] \\ &= \alpha\, x^T H + \underbrace{\alpha\, x^T d}_{\text{scalar}}\, w^T + (1-\alpha)v^T \end{aligned}$$

Cost: # non-zeros in H

Asymptotic Convergence Rate

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

Power method convergence rate:

$$|\lambda_2(G)|/|\lambda_1(G)| \leq \alpha$$

Eigenvalues of G:

$$\lambda_1(G) = 1 \qquad \lambda_i(G) = \alpha\, \lambda_i(S) \quad i > 1$$

[Haveliwala & Kamvar 2003] [Langville & Meyer 2003] [Elden 2003]

Error in Power Method

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

Error in iteration k

- Forward: $\quad e_k \equiv x^{(k)} - \pi$

$$e_k^T = \alpha^k \, e_0^T S^k \qquad \|e_k\|_1 \leq 2\,\alpha^k$$

  [Bianchini, Gori & Scarselli 2003]

- Backward: $\quad r_k^T = [x^{(k)}]^T G - [x^{(k)}]^T$

$$r_k^T = \alpha^k \, r_0^T S^k \qquad \|r_k\|_1 \leq 2\,\alpha^k$$

Termination

Residual norm $\|r_k\|_1 \leq 2\,\alpha^k$

Stop when $\|r_k\|_1 \leq 10^{-8}$

For $\alpha = .85$: $k \geq 119$

n	2293	2947	3468	5757	281903	683446
k	86	85	90	90	91	92

Thanks to Chen Greif!

One-norm too stringent?

Infinity-Norm Bounds

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

Error in iteration k

- Forward:

$$\max_i |\pi_i - x_i^{(k)}| \leq \alpha^k \max_i |\pi_i - x_i^{(0)}|$$

- Backward:

$$\max_i |x_i^{(k+1)} - x_i^{(k)}| \leq \alpha^k \max_i |x_i^{(1)} - x_i^{(0)}|$$

Iteration Counts

n	1-N	∞-N	Disagrees	%
2293	86	75	19	.8
2947	85	76	0	0
3468	90	83	3	.1
5757	90	79	18	.3
281903	91	69	24064	8.5
683446	92	65	50292	7.4

Disagrees: # pages with different rankings

$\|r_k\|_{1,\infty} \leq 10^{-8}$, $\alpha = .85$

Termination Criterion

Old: $\|x^{(k+1)} - x^{(k)}\|_1 \leq \tau$

- Bound becomes more stringent as n grows

New: $\|x^{(k+1)} - x^{(k)}\|_\infty \leq \tau$

- Reduces iteration count
- Disagreements in ranking $\leq 10\%$

Iteration Counts for Different α

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

α	$n = 281903$	$n = 683446$	bound
.85	69	65	119
.90	107	102	166
.95	219	220	415
.99	1114	1208	2075

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

Fewer iterations than predicted by bound

Properties of Power Method

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

But: can be slow

PageRank Computation

- Power method
  Page, Brin, Motwani & Winograd 1999
- Acceleration of power method
  Brezinski & Redivo-Zaglia 2004
  Kamvar, Haveliwala, Manning & Golub 2003
  Haveliwala, Kamvar, Klein, Manning & Golub 2003
- 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 from Linear System

Eigenvector problem:

$$\pi^T \underbrace{(\alpha S + (1-\alpha)\mathbf{1}v^T)}_{G} = \pi^T \qquad \pi \geq 0 \quad \|\pi\|_1 = 1$$

Linear system:

$$\pi^T(I - \alpha S) = (1-\alpha)v^T$$

$I - \alpha S$ nonsingular M-matrix

[Arasu, Novak, Tomkins & Tomlin 2002]

[Bianchini, Gori & Scarselli 2003]

Stationary Iterative Methods

$$\pi^T(I - \alpha S) = (1 - \alpha)v^T$$

- Can be faster than power method
- Can be faster than Krylov space methods
- Predictable, monotonic convergence
- Can converge even for $\alpha \approx 1$
- Less storage than Krylov space methods
- Accurate (no subtractions)

[Gleich, Zhulov & Berkhin 2005]

Example [Gleich, Zhukov & Berkhin 2005]

Web graph: 1.4 billion nodes
6.6 billion edges
Beowulf cluster with 140 processors
Stopping criterion: residual norm $\leq 10^{-7}$

BiCGSTAB: 28.2 minutes (preconditioner?)

Power method: 35.5 minutes

Jacobi Method

Assume no page has a link to itself

$$\pi^T(I-\alpha S) = (1-\alpha)v^T \qquad I-\alpha S = D - O$$

$$[x^{(k+1)}]^T = [x^{(k)}]^T O D^{-1} + (1-\alpha)v^T D^{-1}$$

- $I-\alpha S$ is M-matrix
- Jacobi converges
- No dangling nodes: $\quad D = I \quad O = \alpha S$
  Jacobi method = power method

Dangling Nodes

$S = H + dw^T$ is dense

What to do about dangling nodes?

- Remove [Brin, Page, Motwani & Winograd 1998]
  No PageRank for dangling nodes
  Biased PageRank for other nodes
- Lump into single state [Lee, Golub & Zenios 2003]
  As above
- Remove dw^T [Langville & Meyer 2004]
  [Arasu, Novak, Tomkins & Tomlin 2002]
  H is not stochastic
  What is being computed?

Use v for Dangling Nodes

$$\pi^T(I - \alpha S) = (1-\alpha)v^T \qquad S = H + dw^T$$

Choose $w = v$

$$\pi^T(I - \alpha H) = \underbrace{(1 - \alpha + \alpha\pi^T d)\, v^T}_{\text{multiple of } v^T}$$

Solve $\quad \delta^T(I - \alpha H) =$ multiple of v^T

Then δ is multiple of π

[Gleich, Zhukov & Berkhin 2005]

Iteration Counts for $w = v$

n	Power	Jacobi
2293	75	74
2947	76	74
3468	83	82
5757	79	78
281903	69	69
683446	65	65

$\|r_k\|_\infty \leq 10^{-8}$, $\alpha = .85$

Jacobi: same # iterations as power method

Extension to Arbitrary w

$$\pi^T(I - \alpha S) = (1-\alpha)v^T \qquad S = H + dw^T$$

Rank-one update: $I - \alpha S = (I - \alpha H) - \alpha d w^T$

1. Solve $\quad \delta^T(I - \alpha H) = (1-\alpha)v^T$
2. Solve $\quad \omega^T(I - \alpha H) = w^T$
3. Update $\quad \pi^T = \delta^T + \frac{\alpha\, \delta^T d}{1 - \alpha\, \omega^T d}\, \omega^T$

This requires only two sparse solves

Solve with $I - \alpha H$

After similarity permutation:

$$H = \begin{pmatrix} H_1 & H_2 \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} x_1^T & x_2^T \end{pmatrix} \begin{pmatrix} I - \alpha H_1 & -\alpha H_2 \\ 0 & I \end{pmatrix} = \begin{pmatrix} b_1^T & b_2^T \end{pmatrix}$$

1. Sparse solve $x_1^T (I - \alpha H_1) = b_1$
2. Set $x_2^T = \alpha \, x_1^T H_2 + b_2^T$

PageRank via Linear System

Rank one update: $S = \begin{pmatrix} H_1 & H_2 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \mathbf{1} \end{pmatrix} w^T$

- General dangling node vector $w \geq 0$, $\|w\|_1 = 1$
  Traditional: $w = \frac{1}{n}\mathbf{1}, \quad w = v$
- Cost:
  Two sparse solves with $I - \alpha H_1$ via Jacobi
  Two matrix vector multiplications with H_2
  Inner products and vector additions
- More dangling nodes $\Rightarrow$ cheaper

Krylov Spaces

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

- Eigen problem: $\pi^T G = \pi^T \qquad \|\pi\|_1 = 1$
- Linear system: $\pi^T(I - \alpha S) = (1-\alpha)\, v^T$
- Solution:

$$\pi^T = (1-\alpha)\sum_{j=0}^{\infty} \alpha^j \, v^T S^j$$

in $\mathcal{K}_\infty(v^T, S) \equiv \mathrm{span}\{v^T, v^T S, v^T S^2, \ldots\}$

Krylov Space Approximations

Contribution of π in $\mathcal{K}_k(v^T, S)$:

$$\pi_k^T \equiv (1-\alpha)\sum_{j=0}^{k} \alpha^j\, v^T S^j$$

Error: $\|\pi_k - \pi\|_1 \leq \ \alpha^{k+1}$

Iteration k of Power method with $x_0 = v$:

$$[x^{(k)}]^T = \pi_k^T + \alpha^{k+1}\, v^T S^k \quad \text{in} \quad \mathcal{K}_k(v^T, S)$$

Error: $\|x^{(k)} - \pi\|_1 \leq 2\, \alpha^{k+1}$

Power method produces good approximation

Modified Power Method

Compute $\pi_k^T \equiv (1-\alpha) \sum_{j=0}^{k} \alpha^j \, v^T S^j$

Recursion $\quad \pi_{k+1}^T = \alpha \pi_k^T S + \pi_0^T$

Norm $\quad \|\pi_k\|_1 = 1 - \alpha^{k+1}$

Residual $\quad \pi_k^T G - \pi_k^T = \pi_{k+1}^T - \pi_k^T - \alpha^{k+1} \pi_0^T$

Set $\pi_0 = (1-\alpha) v^T$

Repeat $\quad \pi_{k+1}^T = \alpha \pi_k^T S + \pi_0^T \quad$ until

$\|\pi_{k+1} - \pi_k - \alpha^{k+1} \pi_0\| / (1 - \alpha^{k+1}) \leq \tau$

Iteration Counts for $\alpha = .85$

n	Power	MPower	Disagrees	%
2293	75	90	32	1.4
2947	76	90	0	0
3468	83	87	3	.09
5757	79	82	5	.09
281903	64	74	40091	14.2
683446	65	72	130885	19

Disagrees: # pages with different rankings

Iteration Counts for $\alpha = .99$

n	Power	MPower	Disagrees	%
2293	1188	1274	15	.7
2947	1165	1253	0	0
3468	1327	1190	3	.09
5757	1258	1162	0	0
281903	1114	1008	58967	21
683446	1208	910	284962	42

Disagrees: # pages with different rankings

Modified Power Method

- Computes contribution of π in $\mathcal{K}_k(v^T, S)$

Compared to power method:

- Fewer iterations for $\alpha \approx 1$ and larger n
- High rank disagreements for "very" large n

Potential?

Summary

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

- PageRank = left eigenvector of G
- PageRank insensitive to perturbations in G
- Adding in-links can only increase PageRank
- Adding out-links may decrease PageRank
- PageRank π in $\mathcal{K}_\infty(v^T, S)$
- Iterate k of power method in $\mathcal{K}_k(v^T, S)$
- Forward, backward errors $\leq 2\, \alpha^k$
- Infinity-norm termination criterion

Summary, ctd

Modified Power Method

- Computes contribution of π in $\mathcal{K}_k(v^T, S)$
- May be faster for large α

Jacobi Method

- Computes PageRank from linear system
- Can be competitive with Krylov methods
- Rank one update for dangling nodes
- More dangling nodes $\Rightarrow$ cheaper