CME 200: Workshop Week 5
PageRank, Jacobi & Gauss-Seidell

Institute for Computational and Mathematical Engineering
(ICME)

October 30, 2013

PageRank

Jacobi & Gauss-Seidel

How does Google google

Page rank and searching

INSTITUTE for COMPUTATIONAL & MATHEMATICAL ENGINEERING

Pages, outgoing links and incoming links

Linked by what-we'd-like-to-think-are-important pages

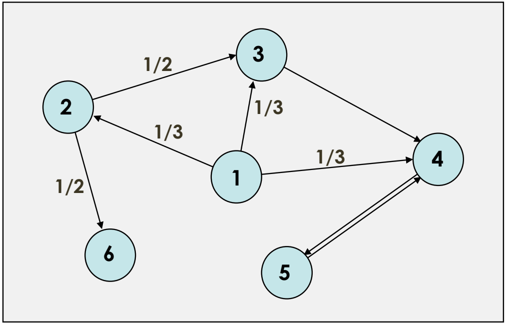
Linked by rather more, and more important, pages

STANFORD UNIVERSITY

Link counting: a simple example

The importance of a page is determined by the importance of the pages that link to it:

$$x_1 = 0$$
$$x_2 = \frac{1}{3}x_1$$
$$x_3 = \frac{1}{3}x_1 + \frac{1}{2}x_2$$
$$x_4 = \frac{1}{3}x_1 + x_3 + x_5$$
$$x_5 = x_4$$
$$x_6 = \frac{1}{2}x_2$$



A very small internet "graph"

STANFORD UNIVERSITY

Rewrite relations in convenient mathematical way

$$
\begin{aligned}
x_1 &= 0x_1 + 0x_2 + 0x_3 + 0x_4 + 0x_5 + 0x_6 \\
x_2 &= \frac{1}{3}x_1 + 0x_2 + 0x_3 + 0x_4 + 0x_5 + 0x_6 \\
x_3 &= \frac{1}{3}x_1 + \frac{1}{2}x_2 + 0x_3 + 0x_4 + 0x_5 + 0x_6 \\
x_4 &= \frac{1}{3}x_1 + 0x_2 + 1x_3 + 0x_4 + 1x_5 + 0x_6 \\
x_5 &= 0x_1 + 0x_2 + 0x_3 + 1x_4 + 0x_5 + 0x_6 \\
x_6 &= 0x_1 + \frac{1}{2}x_2 + 0x_3 + 0x_4 + 0x_5 + 0x_6
\end{aligned}
$$

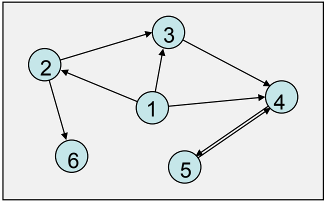
And a bit more convenient

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix}$$

or

$$\mathbf{x} = \mathbf{P}\mathbf{x}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/2 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix}$$



Element k in column m = "probability" of going from node m to node k

Non zero column adds up to 1

Zero column m node m hangs (node 6)

Zero row k node k is not linked to (node 1)

A simple example to introduce the idea of iterative solves

Feeling sorry for page 1

$$x_1 = 0$$

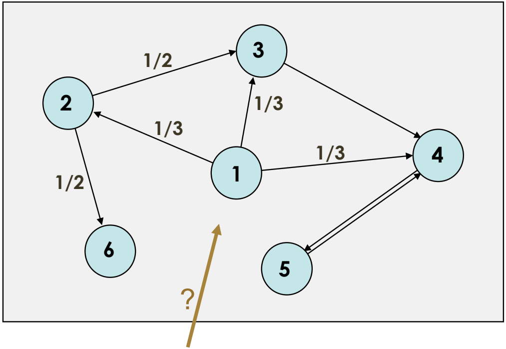
$$x_2 = \frac{1}{3}x_1$$

$$x_3 = \frac{1}{3}x_1 + \frac{1}{2}x_2$$

$$x_4 = \frac{1}{3}x_1 + x_3 + x_5$$

$$x_5 = x_4$$

$$x_6 = \frac{1}{2}x_2$$



Let's give all pages a vote (albeit small)

$$x = \alpha Px + \begin{bmatrix} 1/n \\ \vdots \\ 1/n \end{bmatrix} = \alpha Px + v$$

Importance to transmit
α less than 1.

Average rank given to all pages
(n pages total)

STANFORD
UNIVERSITY

We can solve iteratively

- Initialize with, for example

$$x^{(0)} = \begin{bmatrix} 1/n \\ \vdots \\ 1/n \end{bmatrix}$$

- Repeat until convergence:

$$x^{(k+1)} = \alpha P x^{(k)} + v$$

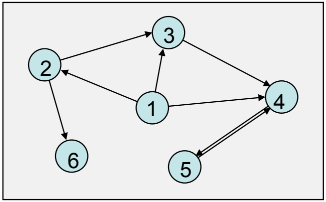
This is a Jacobi iteration: oldie but goodie

STANFORD UNIVERSITY

Applied to our example with α=0.85

$$x^{(0)} = \begin{pmatrix} .17 \\ .17 \\ .17 \\ .17 \\ .17 \\ .17 \end{pmatrix} \quad x^{(1)} = \begin{pmatrix} .17 \\ .21 \\ .28 \\ .50 \\ .31 \\ .24 \end{pmatrix}$$

$$x^{(2)} = \begin{pmatrix} .17 \\ .21 \\ .30 \\ .72 \\ .58 \\ .26 \end{pmatrix} \quad \dots \; x^{(\infty)} = \begin{pmatrix} .17 \\ .21 \\ .30 \\ 2.2 \\ 2.0 \\ .27 \end{pmatrix}$$



This was just a simple fix

In reality it is done a little differently

Convergence of PageRank

In the next HW you'll prove that PageRank converges. Keep in mind these properties of P.

- non-negative: $P_{ij} \geq 0$
- left (column) stochastic: $\vec{\mathbf{1}}^T P = \vec{\mathbf{1}}^T$
  or $\sum_{i=1}^{n} P_{ij} = 1$ for all j
- zero diagonal: $P_{ii} = 0$

PageRank

Jacobi & Gauss-Seidel

Iterative Methods

Recall for (stationary point) iterative methods we split $A = M - N$
And then solve $Mx^{(k+1)} = Nx^{(k)} + b$ where

- Jacobi: $M = D, N = -(L + U)$
- Gauss-Seidel: $M = D + L, N = -U$

Jacobi

For Jacobi iteration

$$\vec{x}^{(k+1)} = D^{-1}(\vec{b} - (L+U)\vec{x}^{(k)})$$

And we know the inverse of a diagonal matrix so ...

$$\vec{x}_i^{(k+1)} = \frac{1}{a_{ii}}(\vec{b}_i - \sum_{j \neq i}^{n} a_{ij}\vec{x}_j^{(k)})$$

The update for the i^{th} coordinate of $\vec{x}^{(k+1)}$ doesn't depend on any other coordnate and we can do these updates in parallel. But what if we wanted to update each coordinate in order and use the update as soon as we computed it?

Gauss-Seidel I

Jacobi updates as (same as last slide)

$$\vec{x}_i^{(k+1)} = \frac{1}{a_{ii}} \left(\vec{b}_i - \sum_{j \neq i}^{n} a_{ij} \vec{x}_j^{(k)} \right)$$

but as we update coordinate i, the updates for all previous coordinates are available so we might change this to...

$$\vec{x}_i^{(k+1)} = \frac{1}{a_{ii}} \left(\vec{b}_i - \sum_{j<i} a_{ij} \vec{x}_j^{(k+1)} - \sum_{j>i} a_{ij} \vec{x}_j^{(k)} \right)$$

Gauss-Seidel II

If we write this new update equation (same as last slide)

$$\vec{x}_i^{(k+1)} = \frac{1}{a_{ii}} \left(\vec{b}_i - \sum_{j<i} a_{ij} \vec{x}_j^{(k+1)} - \sum_{j>i} a_{ij} \vec{x}_j^{(k)} \right)$$

in matrix form we get

$$\vec{x}^{(k+1)} = D^{-1} \left(\vec{b} - L\vec{x}^{(k+1)} - U\vec{x}^{(k)} \right)$$

or

$$(D + L)\vec{x}^{(k+1)} = \vec{b} - U\vec{x}^{(k)}$$

so we see this is a split where $M = D + L$ and $N = -U$ which is the Gauss-Seidel update.