GENERALIZING GOOGLE'S PAGERANK TO RANK NATIONAL FOOTBALL LEAGUE TEAMS

Anjela Govan, Carl D. Meyer, and Russell Albright

North Carolina State University,
SAS Institute Inc.

March 2008

Definition

A *ranking* is a technique of assigning a positive integer (rank) to each object in a finite set.

Ranks are often based on a rating score denoting the degree of importance of each object in the set.

PageRank Model

Use link structure of the web to rank web pages.

Random web surfer:

- Choose a link in the current web page at random
- The proportion of time spent at web page i is the rating of web page i
- Random surfer will spend the larger proportion of time on important web sites

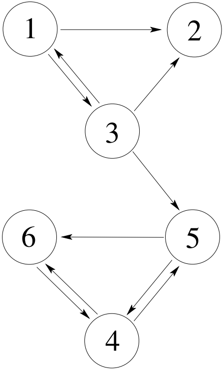
A web page is important if it is pointed to by other important web page(s).

Problem with random web surfer model: "rank sinks"

PageRank Algorithm, hyperlink matrix

1. Web with n web pages $\rightarrow$ directed graph, with n nodes.
2. Form hyperlink matrix $\mathbf{H}$.

$$H_{ij} = \begin{cases} 1/deg^{-}(P_i) & P_i \rightarrow P_j \\ 0 & P_i \nrightarrow P_j \end{cases}$$



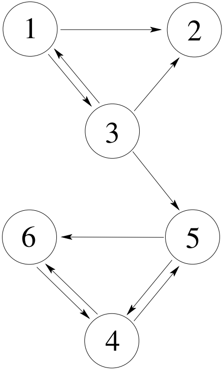
$$\mathbf{H} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

PageRank Algorithm, stochastic matrix

3. Form stochastic matrix $\mathbf{S}$ (fix dangling nodes)

$$\mathbf{S} = \mathbf{H} + \mathbf{a}\mathbf{u}^T \qquad a_i = \begin{cases} 1 & \text{if } \mathbf{H}_i = 0 \\ 0 & \text{otherwise} \end{cases}$$

$\mathbf{u}$ is a probability distribution vector.



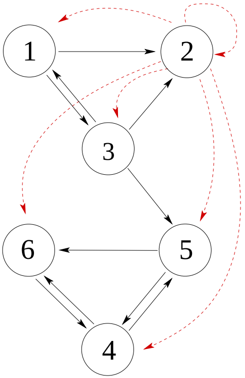
$$\mathbf{H} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

PageRank Algorithm, stochastic matrix

3. Form stochastic matrix $\mathbf{S}$ (fix dangling nodes)

$$\mathbf{S} = \mathbf{H} + \mathbf{a}\mathbf{u}^T \qquad a_i = \begin{cases} 1 & \text{if } \mathbf{H}_i = 0 \\ 0 & \text{otherwise} \end{cases}$$

$\mathbf{u}$ is a probability distribution vector.



$$\mathbf{S} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

PageRank Outline, Google matrix

4. Form Google matrix $\mathbf{G}$

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

$0 < \alpha < 1$, and $\mathbf{v} > 0$ is a probability distribution vector.

5. The vector containing the ratings of each web page is $\boldsymbol{\pi}$ such that

$$\boldsymbol{\pi}^T = \boldsymbol{\pi}^T \mathbf{G}$$

Use rating scores in $\boldsymbol{\pi}$ to rank web pages.

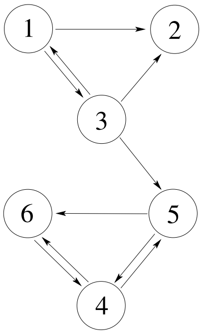
Small Example, Google matrix

$$\mathbf{G} = 0.85\mathbf{S} + (1-0.85)(1/6)\mathbf{e}\mathbf{e}^T =$$

$$\begin{pmatrix} \frac{1}{40} & \frac{9}{20} & \frac{9}{20} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{37}{120} & \frac{37}{120} & \frac{1}{40} & \frac{1}{40} & \frac{37}{120} & \frac{1}{40} \\ \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{9}{20} & \frac{9}{20} \\ \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{9}{20} & \frac{1}{40} & \frac{9}{20} \\ \frac{1}{40} & \frac{1}{40} & \frac{1}{40} & \frac{7}{8} & \frac{1}{40} & \frac{1}{40} \end{pmatrix}$$

Small Example, rating vector π

$$\boldsymbol{\pi}^T \approx \begin{pmatrix} 0.05170 & 0.0737 & 0.0574 & 0.3487 & 0.1999 & 0.2686 \end{pmatrix}$$



The list of the web pages in the order of rating scores (highest to lowest) is

$$4 \quad 6 \quad 5 \quad 2 \quad 3 \quad 1$$

GeM (Generalized Markov Method)

1. Sport season with n teams $\rightarrow$ weighted directed graph with n nodes.

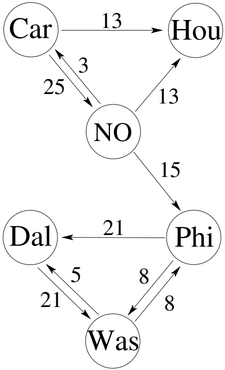
   Game between i and j $\rightarrow$ edge from loser to winner with weight equal to the positive difference of the game statistic (e.g. scores, yards, fumbles, etc.).

   If i lost to team j more than once during a season w_{ij} is the sum of the positive differences of the statistic of the games team i lost to j.

GeM Method, Gamelink matrix

2. Form matrix $\mathbf{H}$.

$$H_{ij} = \begin{cases} 1/\sum_{k=1}^{n} w_{ik} & \text{team } i \text{ lost to } j \\ 0 & \text{otherwise} \end{cases}$$



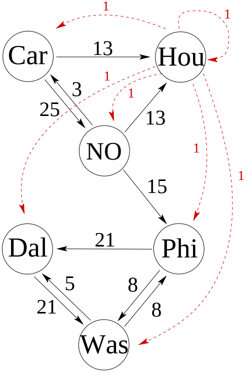
	Car	Dal	Hou	NO	Phi	Was
Car	0	0	$\frac{13}{38}$	$\frac{25}{38}$	0	0
Dal	0	0	0	0	0	1
Hou	0	0	0	0	0	0
NO	$\frac{3}{31}$	0	$\frac{13}{31}$	0	$\frac{15}{31}$	0
Phi	0	$\frac{21}{29}$	0	0	0	$\frac{8}{29}$
Was	0	$\frac{5}{13}$	0	0	$\frac{8}{13}$	0

GeM Method, Stochastic matrix

3. Form matrix $\mathbf{S}$.

$$\mathbf{S} = \mathbf{H} + \mathbf{a}\mathbf{u}^T, \qquad a_i = \begin{cases} 1 & \text{if } \mathbf{H}_i = 0 \\ 0 & \text{otherwise} \end{cases}$$

$\mathbf{u}$ is a probability distribution vector.



	Car	Dal	Hou	NO	Phi	Was
Car	0	0	$\frac{13}{38}$	$\frac{25}{38}$	0	0
Dal	0	0	0	0	0	1
Hou	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
NO	$\frac{3}{31}$	0	$\frac{13}{31}$	0	$\frac{15}{31}$	0
Phi	0	$\frac{21}{29}$	0	0	0	$\frac{8}{29}$
Was	0	$\frac{5}{13}$	0	0	$\frac{8}{13}$	0

GeM, GeM matrix

4. Form GeM matrix $\mathbf{G}$.

$$\mathbf{G} = \alpha_0 \mathbf{S}_0 + \alpha_1 \mathbf{S}_1 + ... + \alpha_p \mathbf{S}_p$$

   where $0 \leq \alpha_i < 1$, $\sum \alpha_i = 1$ and $\mathbf{S}_i$ is a stochastic matrix $0 \leq i \leq p$, we will call it an ith feature matrix.

5. The vector containing the rating scores of each team is $\boldsymbol{\pi}$ such that

$$\boldsymbol{\pi}^T = \boldsymbol{\pi}^T \mathbf{G}$$

   Use rating scores in $\boldsymbol{\pi}$ to rank teams.

Feature Matrices GeM (FMGeM).

$$\mathbf{G} = \alpha_0\mathbf{S}_0 + \alpha_1\mathbf{S}_1 + ... + \alpha_p\mathbf{S}_p$$

where $0 \leq \alpha_i < 1$, $\sum \alpha_i = 1$ and $\mathbf{S}_i$ is the ith feature matrix $0 \leq i \leq p$.

$\mathbf{S}_i = \mathbf{H}_i + \mathbf{a}_i\mathbf{u}_i^T$, where $\mathbf{H}_i$ is a gamelink matrix formed using ith statistic (e.g. game scores, total yards, rushing yards)

Rank One update GeM (ROGeM).

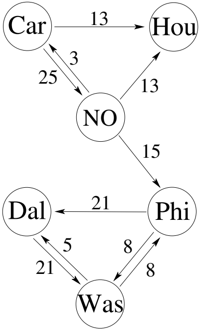
$$\mathbf{G} = \alpha[\mathbf{H} + \mathbf{a}\mathbf{u}^T] + (1-\alpha)\mathbf{e}\mathbf{v}^T$$

where $0 < \alpha < 1$, $\mathbf{v} > 0$ and $\mathbf{u}$ are probability distribution vectors and

$$a_i = \begin{cases} 1 & \text{if } \mathbf{H}_i = 0 \\ 0 & \text{otherwise} \end{cases} .$$

Small Example, rating vector π

$$\pi^T \approx \begin{pmatrix} 0.0389 & 0.2824 & 0.0656 & 0.056 & 0.2289 & 0.3281 \end{pmatrix}$$



The list of the teams in the order of rating scores (highest to lowest) is

Was Dal Phi Hou NO Car