# Ranking National Football League Teams Using Google's PageRank.

A. Govan     C. Meyer     R. Albright

North Carolina State University,
SAS Institute Inc.

INFORMS Annual Meeting, November 2007

# Outline of PageRank

- $\mathbf{G} = \alpha[\mathbf{H} + (1/n)\mathbf{a}\mathbf{e}^T] + (1 - \alpha)\mathbf{e}\mathbf{v}^T$
  - $\mathbf{H}$ is a hyperlink matrix,
  - $0 < \alpha < 1$, $\mathbf{e}$ is a vector of 1's
  - $\mathbf{a}_i = 0$ if $\mathbf{H}_i^T$ is nonzero and $\mathbf{a}_i = 1$ otherwise,
  - $\mathbf{v} > 0$ is a probability distribution vector.
- 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.

# Generalizing PageRank, GeM

- $\mathbf{G} = \alpha_0 \mathbf{S}_0 + \alpha_1 \mathbf{S}_1 + ... + \alpha_p \mathbf{S}_p$
  - Each $\mathbf{S}_i$ is stochastic (derived from sports statistics),
  - $0 \leq \alpha_i \leq 1$ and $\Sigma \alpha_i = 1$.
- The vector $\boldsymbol{\pi}$ contains the rating scores of each team, such that

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

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

## **Gem1** - Basic model

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

- $0 < \alpha < 1$,
- $\mathbf{H}$ is based on game scores,
- $\mathbf{a}_i = 0$ if row $i$ of $\mathbf{H}$ is nonzero and $\mathbf{a}_i = 1$ otherwise,
- $\mathbf{u}$ is a probability distribution.

## **Gem2** - Feature vectors model

$$\mathbf{G} = \alpha_0[\mathbf{H} + \mathbf{a}\mathbf{u}^T] + \alpha_1\mathbf{e}\mathbf{v}_1^T + ... + \alpha_p\mathbf{e}\mathbf{v}_p^T$$

- $0 \leq \alpha_i \leq 1$, $\Sigma\alpha_i = 1$.
- $\mathbf{H}$ is derived using game scores,
- $\mathbf{a}_i = 0$ if row $i$ of $\mathbf{H}$ is nonzero and $\mathbf{a}_i = 1$ otherwise,
- $\mathbf{u}$ and $\mathbf{v}_i$ are probability distribution vectors.
- Compute $\mathbf{v}_i$ using nonnegative matrix factorization of a matrix containing all of the statistics.

## **Gem3** - Offense-defense model

$$\mathbf{G} = \alpha_0[\mathbf{H} + \mathbf{a}\mathbf{u}^T] + \alpha_1\mathbf{e}\mathbf{o}^T + \alpha_2\mathbf{e}\mathbf{d}^T$$

- $0 \leq \alpha_i \leq 1$, $\Sigma\alpha_i = 1$,
- $\mathbf{H}$ is derived using game scores.
- $\mathbf{a}_i = 0$ if row $i$ of $\mathbf{H}$ is nonzero and $\mathbf{a}_i = 1$ otherwise,
- $\mathbf{u}$ is a probability distribution vector.
- Compute offense vector $\mathbf{o}$ and defense vector $\mathbf{d}$ using modified HITS.

## **Gem4** - Feature matrices model

$$\mathbf{G} = \alpha_0[\mathbf{H}_0 + \mathbf{a}_0\mathbf{u}_0^T] + ... + \alpha_p[\mathbf{H}_p + \mathbf{a}_p\mathbf{u}_p^T]$$

- $0 \leq \alpha \leq 1$, $\Sigma\alpha_i = 1$,
- $\mathbf{H}_j$ is derived using statistic $j$ (e.g. scores, yards, etc.),
- $\mathbf{a}_{i_j} = 0$ if row $i$ of $\mathbf{H}_j$ is nonzero and $\mathbf{a}_{i_j} = 1$ otherwise,
- $\mathbf{u}_j$ is a probability distribution vector.

# Game Prediction Results



**Results of NFL game predictions.**