Ranking National Football League teams using Google's PageRank.

A. Govan C. Meyer

Department of Mathematics
North Carolina State University

A.A. Markov Anniversary Meeting, 2006

Outline

Google's ranking algorithm.

Ranking NFL.

Summary

Google's ranking.

- Think of the internet as a graph.
  - Webpages are nodes of the graph, *n* nodes.
  - Hyperlinks are directed edges.
- Basic Idea: $r(P) = \sum_{Q \in B_P} \frac{r(Q)}{|Q|}$
  where $r(P)$ is the rank of a webpage P, B_P is the set of webpages pointing to P, and $|Q|$ is the outdegree of a webpage Q.

Google Matrix.

- Hyperlink Matrix $\mathbf{H}$
  - $\mathbf{H}(i,j) = \begin{cases} 1/|i| & \text{there is a link from } i \text{ to } j \\ 0 & \text{otherwise} \end{cases}$
- Stochastic matrix $\mathbf{S}$
  - Obtained by modifying matrix $\mathbf{H}$.
  - Replace the zero rows of $\mathbf{H}$ with $(1/n)\mathbf{e}^T$, where $\mathbf{e}$ is a column vector of ones.
- Google Matrix $\mathbf{G}$.
  - Convex combination: $\mathbf{G} = \alpha\mathbf{S} + (1-\alpha)\mathbf{e}\mathbf{v}^T$, $\alpha \in (0,1)$ and $\mathbf{v}^T > 0$
  - Personalization vector $\mathbf{v}$.

PageRank vector π.

- $\mathbf{G}$ is the transition probability matrix.
- $\mathbf{G}$ is irreducible (and aperiodic).
- π is the stationary probability distribution vector.
- π is unique (up to a scalar multiple).

NFL web.

- Each NFL team is a node in a graph.

NFL web.

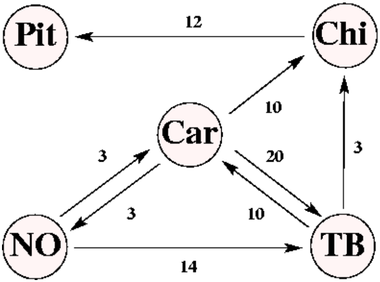
- Each NFL team is a node in a graph.
- A regular season game results in a directed edge from the loser to the winner.

NFL web.

- Each NFL team is a node in a graph.
- A regular season game results in a directed edge from the loser to the winner.
- The edges are weighted, the weight is the score difference of the corresponding game.

NFL web.

- Each NFL team is a node in a graph.
- A regular season game results in a directed edge from the loser to the winner.
- The edges are weighted, the weight is the score difference of the corresponding game.



Variations on PageRank.

- $\mathbf{H}(i,j) = \sum_t w_{ij}^t / (\sum_j (\sum_t w_{ij}^t))$
  where w_{ij} is the weight on the edge from team i to team j during week t.
- Dealing with the ith zero row (undefeated team i)

Variations on PageRank.

- $\mathbf{H}(i,j) = \sum_t w_{ij}^t / (\sum_j (\sum_t w_{ij}^t))$
  where w_{ij} is the weight on the edge from team i to team j during week t.
- Dealing with the ith zero row (undefeated team i)
  - $(1/32)\mathbf{e}^T$, equally likely to lose to any other team.

Variations on PageRank.

- $\mathbf{H}(i,j) = \sum_t w_{ij}^t / (\sum_j (\sum_t w_{ij}^t))$
  where w_{ij} is the weight on the edge from team i to team j during week t.

- Dealing with the ith zero row (undefeated team i)
  - $(1/32)\mathbf{e}^T$, equally likely to lose to any other team.
  - $\mathbf{e}_i^T$, where $\mathbf{e}_i$ is the standard basis vector.

Variations on PageRank.

- $\mathbf{H}(i,j) = \sum_t w_{ij}^t / (\sum_j (\sum_t w_{ij}^t))$
  where w_{ij} is the weight on the edge from team i to team j during week t.

- Dealing with the ith zero row (undefeated team i)
  - $(1/32)\mathbf{e}^T$, equally likely to lose to any other team.
  - $\mathbf{e}_i^T$, where $\mathbf{e}_i$ is the standard basis vector.
  - $\boldsymbol{\pi}_{t-1}^T$, using the ranking vector from previous week.

Ranking NFL with Keener (SIAM Review,1993)

- Nonnegative matrix $\mathbf{A}$
  - $\mathbf{A}(i,j) = h\left(\frac{S_{ij}+1}{S_{ij}+S_{ji}+2}\right)$, *Laplace's rule of succession*

    where S_{ij} is the amount of points scored by team i against team j.
  - $h(x) = \frac{1}{2} + \frac{1}{2}\text{sgn}(x - \frac{1}{2})\sqrt{|2x-1|}$
- Rank vector $\mathbf{r}$ is the Perron vector of $\mathbf{A}$.

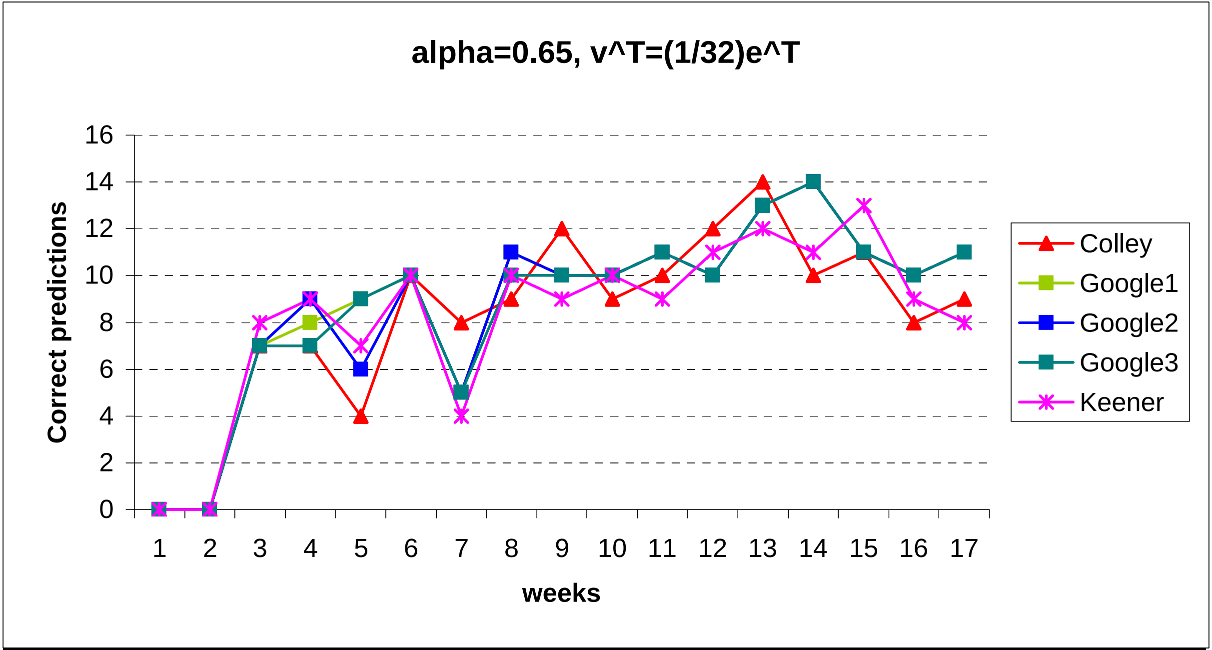
Ranking NFL with Colley (Colley's Bias Free Matrix Rankings)

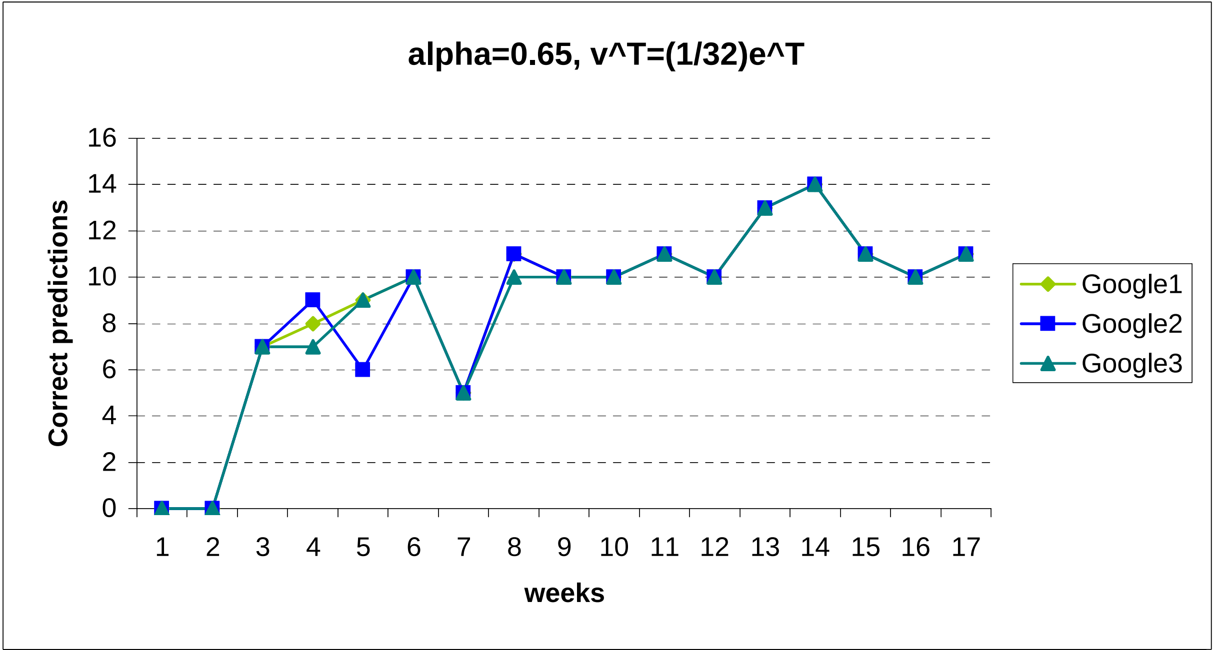
- Colley matrix $\mathbf{C}$
  - $\mathbf{C}(i,j) = \begin{cases} 2 + n_{tot,i} & i = j \\ -n_{j,i} & i \neq j \end{cases}$ *Laplace's rule of succession*

  where $n_{tot,i}$ is the total number of games played by team i, and $n_{j,i}$ is the number of times team i played team j.
- Ranking vector $\mathbf{r}$ is the solution to the linear system

$$\mathbf{C}\mathbf{r} = \mathbf{b}$$

where $\mathbf{b}(i,1) = 1 + (n_{w,i} - n_{l,i})/2$, given that $n_{w,i}$ is the number of games lost by team i and $n_{l,i}$ is the number of games won by team i.


alpha=0.65, v^T=(1/32)e^T
Correct predictions
0
2
4
6
8
10
12
14
16
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
weeks
Colley
Google1
Google2
Google3
Keener


alpha=0.65, v^T=(1/32)e^T
Correct predictions
0
2
4
6
8
10
12
14
16
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
weeks
Google1
Google2
Google3

Regular season 2005, α=0.65, $\mathbf{v}^T = (1/32)\mathbf{e}^T$											
	Colley		Google1		Google2		Google3		Keener		
	Correct	Spread	Correct	Spread	Correct	Spread	Correct	Spread	Correct	Spread	games
week 1	0	0	0	0	0	0	0	0	0	0	
week 2	0	0	0	0	0	0	0	0	0	0	
week 3	7	202	7	215	7	182	7	215	8	560	14
week 4	7	178	8	180	9	150	7	175	9	178	14
week 5	4	198	9	270	6	176	9	274	7	231	14
week 6	10	125	10	137	10	132	10	138	10	140	14
week 7	8	144	5	201	5	172	5	202	4	197	14
week 8	9	196	11	244	11	150	10	249	10	152	14
week 9	12	107	10	165	10	176	10	164	9	145	14
week 10	9	134	10	120	10	140	10	121	10	139	14
week 11	10	196	11	198	11	198	11	197	9	190	16
week 12	12	108	10	155	10	103	10	157	11	153	16
week 13	14	133	13	181	13	167	13	183	12	147	16
week 14	10	171	14	179	14	175	14	176	11	194	16
week 15	11	227	11	217	11	223	11	219	13	206	16
week 16	8	174	10	178	10	178	10	178	9	176	16
week 17	9	221	11	214	11	214	11	214	8	213	16
Total	140	2514	150	2854	148	2536	148	2862	140	3021	224

Results.

PageRank based ranking method:

- depends on α and the personalization vector $\mathbf{v}^T$.
- does not appear to depend on the method of adjusting the zero rows.

Summary

- Based on preliminary simulations PageRank variation method outperforms Keener's and Colley method.

- Future
  - Personalization vector(s) using season statistics data.
  - Automate the selection of the best α for a specified $\mathbf{v}^T$.
  - Extend the convex combination to include more then one personalization vector.