

RANKING NFL TEAMS AWESOME!

Sean J. Taylor / @seanjtaylor

INTRODUCTION

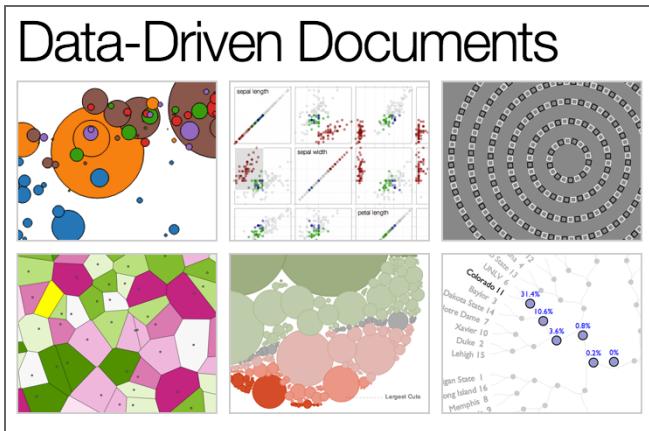
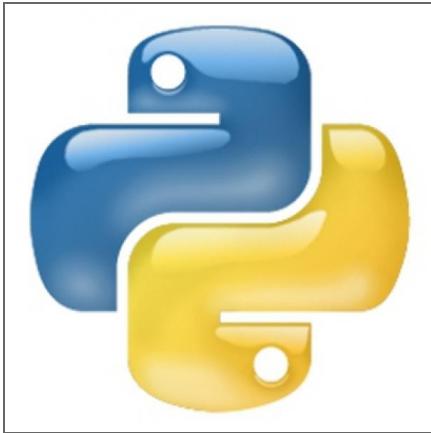


5th year PhD candidate at NYU Stern

INTERESTS

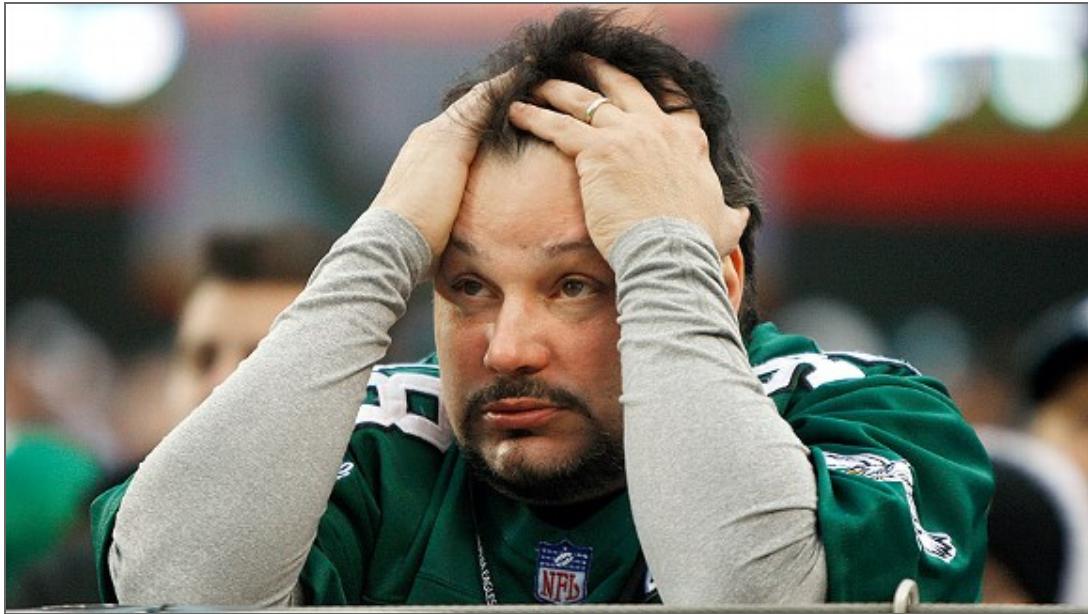
- Social influence
- Causal Inference and Web Experiments
- Bayesian modeling
- Prediction

TOOLBOX





LIFELONG EAGLES FAN

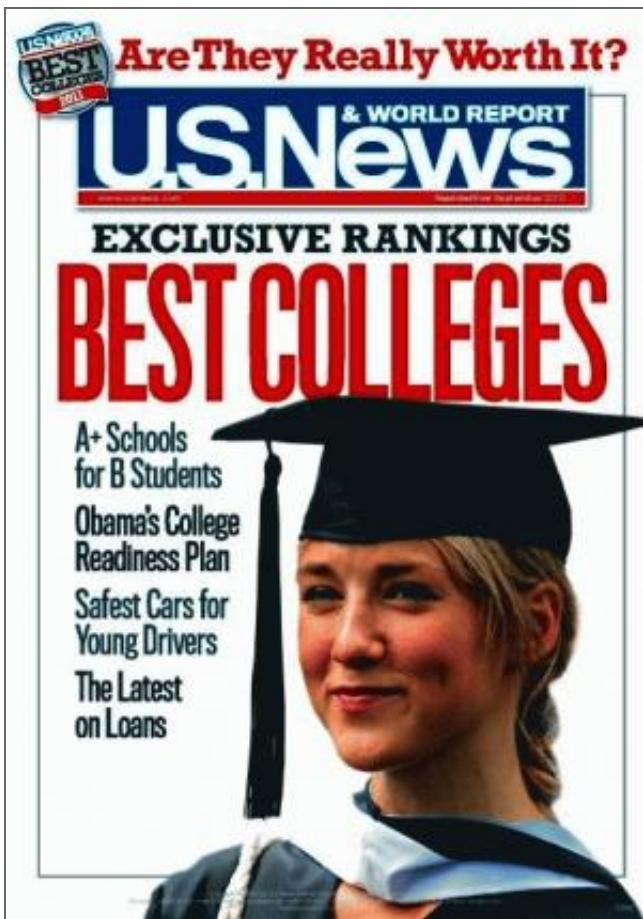


2002 Power Rankings: Week 18				
RANK	TEAM / RECORD	TRENDING	COMMENTS	
1	 Eagles 12-4	-- Last Week: NR	Eight of the last 12 NFC teams with home-field advantage have made the Super Bowl.	
2	 Raiders 11-5	-- Last Week: NR	Raiders' offense has been much more balanced in recent weeks. It cost Rich Gannon a chance to break Marino's record, but it could help in the playoffs.	
3	 Buccaneers 12-4	-- Last Week: NR	Win over the Bears was big. Bucs get first-round bye and break their streak of 20 straight losses when it's under 40 degrees.	
4	 Packers 12-4	-- Last Week: NR	Ouch -- Packers go from a chance at home-field throughout the playoffs to having to play a first-round game.	
4	 Titans 11-5	-- Last Week: NR	At 1-4 and coming off a 31-14 loss to the Redskins, there couldn't have been too many people predicting a first-round bye in the playoffs for Tennessee.	
6	 Steelers 10-5-1	-- Last Week: NR	The saying goes that it's difficult to beat a team three times. Steelers fans better hope that's not true.	
7	 Jets 9-7	-- Last Week: NR	As important as Chad Pennington's emergence has been, the return to health and prominence of Curtis Martin is as significant.	

OUTLINE

- What is a ranking?
- Evaluating rankings
- Six methods
- Empirical evaluation
- Ranking other things

RANKING == DECISION TOOL



where to go to college?



bars

union square

Suggestions for bars

BEST NEARBY

Expand to filter your results



SideBAR

1 9.1 118 E 15th St. (at Irving Pl.)
Bar • 2 here now



For a weekday binge, start at this sports bar. From noon to 7pm, hoist \$3 Bud Light drafts and two-for-one Absolut cocktails, Cruzan mojitos and bottles of Bud.
- Time Out New York

Bar connoisseurs tend to come here

Save Like



Old Town Bar

2 9.3 45 E 18th St. (btwn Broadway & Park Ave. S)
Bar • \$\$\$\$ • See menu



In a city known for hot dogs, the hefty all-beef Sabrett at Old Town Bar reigns supreme. It's one of our #100best dishes and drinks of 2011. - Time Out New York

Bar connoisseurs tend to come here

Save Like



Lillie's Union Square

3 9.3 13 E 17th St (btwn Broadway & 5th Ave)
Bar • 6 here now • \$\$\$\$ • See menu



Drink their generously portioned drinks - E H.

Save Like



Pete's Tavern

4 9.0 129 E 18th St. (at Irving Pl.)
Pub • 3 here now • \$\$\$\$ • See menu



where to get a drink after the talk?

PBR belt buckle 22 items

Sort by: Relevancy ▾

Ships to United States ✕



Vintage Pabst Blue Ribbon ...
heyltjeroose \$36.00 usd



PBR Pabst Blue Ribbon Mod...
BUCKLELO... \$24.00 usd



Belt Buckle Recycled Beer C...
pureGARAGE \$18.00 usd



Bottle Opener Buckle, "PBR ...
metallogical \$40.00 usd



Pabst Blue Ribbon PRB Belt ...
americanga... \$38.00 usd



Bronco Rodeo Hand Engrav...
thegigglesg... \$65.00 usd



Bull Rider Silver and Green ...
VintageDiehls \$20.00 usd



PBR Pabst Blue Ribbon Mod...
BUCKLELO... \$24.00 usd

which product to check out?

RANKING == DESCRIPTION

Table 3 A Revealed Preference Ranking of Colleges		
rank	College Name	Elo pts
1	Harvard	2800
2	Yale	2738
3	Stanford	2694
4	Cal Tech	2632
5	MIT	2624
6	Princeton	2608
7	Brown	2433
8	Columbia	2392
9	Amherst	2363
10	Dartmouth	2357
11	Wellesley	2346
12	U Penn	2325
13	U Notre Dame	2279
14	Swarthmore	2270
15	Cornell	2236
16	Georgetown	2218
17	Rice	2214
18	Williams	2213
19	Duke	2209
20	U Virginia	2197
21	Northwestern	2136

which colleges are being chosen over others?

"A Revealed Preference Ranking of U.S. Colleges and Universities" (Avery et al. 2004)



which college football team is the best?

Emirates ATP RANKINGS

DEFINING EXCELLENCE SINCE | 1973

14.01.2013

Top 100

All Countries

Go

Rank, Name & Nationality	Points	Week Change	Tourn Played
1 Djokovic, Novak (SRB)	12,920	0	18
2 Federer, Roger (SUI)	10,265	0	20
3 Murray, Andy (GBR)	8,000	0	20
4 Nadal, Rafael (ESP)	6,600	0	17
5 Ferrer, David (ESP)	6,505	0	26
6 Berdych, Tomas (CZE)	4,680	0	25
7 Del Potro, Juan Martin (ARG)	4,480	0	22
8 Tsonga, Jo-Wilfried (FRA)	3,375	0	25
9 Tipsarevic, Janko (SRB)	3,090	0	28
10 Gasquet, Richard (FRA)	2,720	0	23
11 Almagro, Nicolas (ESP)	2,515	0	25
12 Monaco, Juan (ARG)	2,430	0	24
13 Isner, John (USA)	2,215	0	26
14 Cilic, Marin (CRO)	2,210	0	24
15 Raonic, Milos (CAN)	2,175	0	24

which tennis player is the best?

RANKING == TOTAL ORDERING

- set of items: X
- binary relation: \geq
- transitivity: $a \geq b \wedge b \geq c \Rightarrow a \geq c$
- totality: $a \geq b \vee b \geq a \forall a, b \in X$
- anti-symmetry: $a \geq b \wedge b \geq a \Leftrightarrow a = b$

RANKING == CLASSIFICATION

Ranking problem can be transformed into a two-class classification problem via the **pairwise transform**.

Goal: Learn $f : X \times X \rightarrow \{-1, 1\}$

RANKING == REGRESSION

All we have to do is use: (\mathbb{R}, \geq)

- Obviously satisfies total ordering criteria.
- But has a bit more: we have/need a metric.
- Shows ranking can be a regression problem.

Goal: Learn scoring function $q : X \rightarrow \mathbb{R}$

GETTING IT WRONG: LOSS FUNCTIONS

- Descriptive loss
- Predictive loss
- Ranking agreements

ZERO-ONE LOSS

Count the games where the ranking-predicted winner is upset.

MISRANK LOSS

If the ranking-predicted winner loses, add the difference in ranks to the loss.

WIN MATRIX

Season: 2012

Week: 17

RANKING AGREEMENT

Kendall tau distance is a metric that counts the number of pairwise disagreements between two rankings.

Two rankings R_1, R_2 disagree on pair i, j if:

$$R_1(i) > R_1(j) \wedge R_2(i) < R_2(j)$$

Useful if we want to match some existing ranking, say ESPN.com Power Rankings.

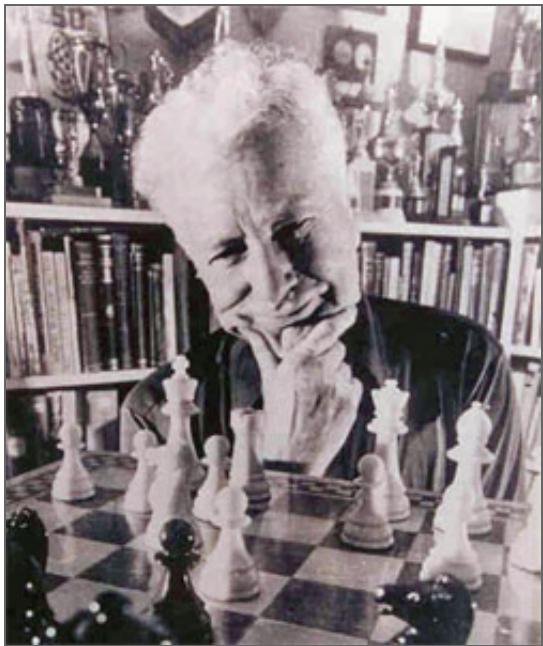
METHODS

- Elo
- Pythagorean wins
- Eigenvector
- Bradley-Terry-Luce
- Feature engineering
- Optimal Rankings

ELO



ELO



ELO RANKING

Ratings of players A and B: R_A, R_B

Expected score of A while playing B:

$$E(R_A, R_B) = \frac{1}{1 + 10^{(R_B - R_A)/400}}$$

Note that $E(R_A, R_B) = 1 - E(R_B, R_A)$

ELO SCORE UPDATES

Outcome when A plays B is $S_{AB} \in \{0, 0.5, 1\}$

$$R'_A \leftarrow R_A + K(E(R_A, R_B) - S_{AB})$$

$$R'_B \leftarrow R_B + K(E(R_B, R_A) - S_{BA})$$

$$R'_B \leftarrow R_B - K(E(R_A, R_A) - S_{AB})$$

PROBLEMS

- How to set parameters?
- How often to update?

PYTHAGOREAN WINS

“From 1988 through 2004, 11 of 16 Super Bowls were won by the team that led the NFL in Pythagorean wins, while only seven were won by the team with the most actual victories. Super Bowl champions that led the league in Pythagorean wins but not actual wins include the 2004 Patriots, 2000 Ravens, 1999 Rams and 1997 Broncos.”

— Football Outsiders Almanac (2011)

See also **this post.**

FORMULA

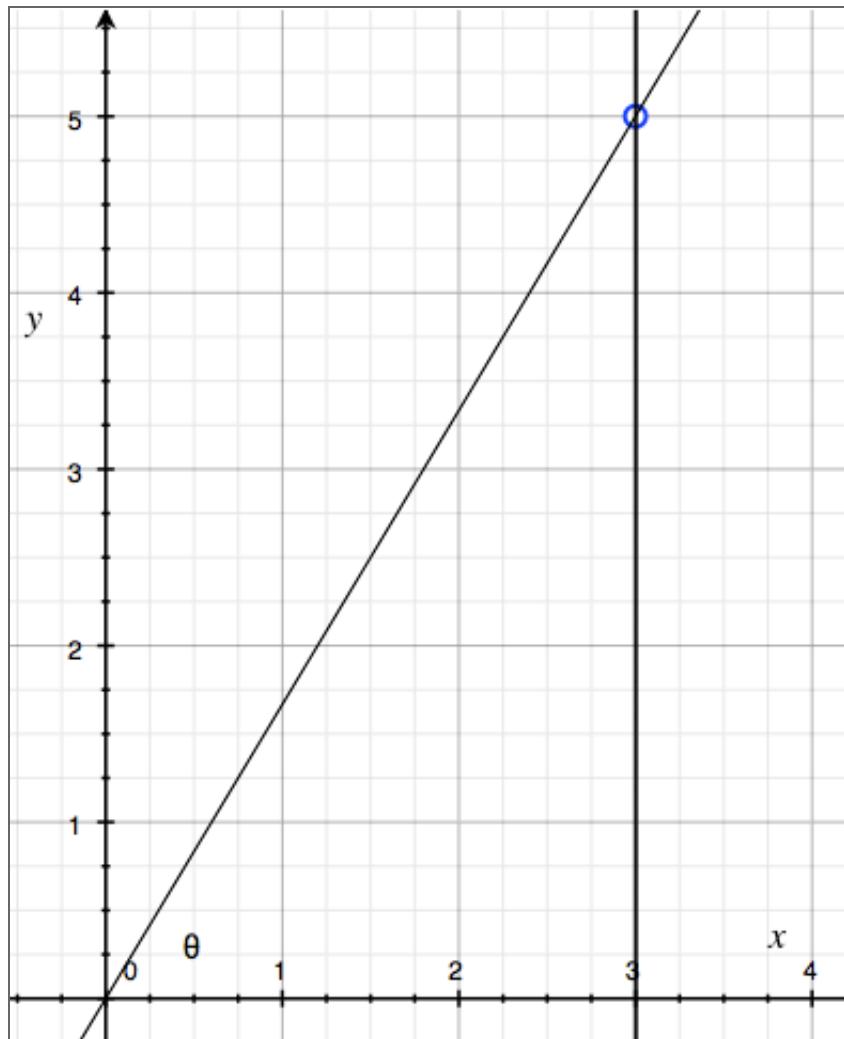
Points for = y Points against = x

$$\text{Win Rate} = \frac{y^\beta}{x^\beta + y^\beta}$$

Set β empirically, different for each sport.

β can be thought of as a shrinkage parameter.

VISUALIZATION



PYTHAGOREAN WINS

1		9		17		25	
2		10		18		26	
3		11		19		27	
4		12		20		28	
5		13		21		29	
6		14		22		30	
7		15		23		31	
8		16		24		32	

SCORE



0 - 7

EIGENVECTOR METHOD

Idea: iteratively adjust for strength of schedule. Give more credit to victories over teams which are "good"

r_i is the score of team i . a_{ij} is 1 if i beat j .

$$\lambda r_i = \frac{1}{n_i} = \sum_{j=1}^N a_{ij} r_j$$

$$\mathbf{A}\mathbf{r} = \lambda\mathbf{r}$$

PERRON-FROBENIUS

If A is nonnegative and irreducible, it has a strictly positive eigenvector corresponding to its largest eigenvalue.

INTUITION

$$\begin{matrix} DAL \\ N Y G \\ P H I \\ W A S \end{matrix} \begin{pmatrix} 0 & 1 & 2 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 2 & 1 & 2 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ 1 \\ 5 \end{pmatrix}$$

SECOND ORDER WINS

$$\begin{pmatrix} 0 & 1 & 2 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 2 & 1 & 2 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 2 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 2 & 1 & 2 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 9 \\ 3 \\ 11 \end{pmatrix}$$

Must normalize this output vector by its norm.

K-TH ORDER RANK

$$r_k = \frac{A^k r_0}{\|A^k r_0\|}$$

The limit of this is just the first eigenvector!

POWER ITERATION METHOD

```
def power_iteration(m, iters=10000):
    x0 = np.ones(m.shape[0])

    for i in range(iters):
        x0 = np.dot(m,x0)
        x0 /= np.linalg.norm(x0,1)

    return x0
```

ASIDE: PAGERANK

Redefine the preference matrix A :

$$A' = (1 - \alpha)A + \alpha \frac{1}{N} M$$

where M is a matrix of all ones.

EIGENVECTOR METHOD

1		9		17		25	
2		10		18		26	
3		11		19		27	
4		12		20		28	
5		13		21		29	
6		14		22		30	
7		15		23		31	
8		16		24		32	

SCORE



0 — 14

BRADLEY-TERRY-LUCE MODEL

$$P(i > j) = \frac{1}{1 + e^{-(q_i - q_j)}} = \text{logit}^{-1}(q_i - q_j)$$

(Bradley & Terry, 1952; Luce, 1959)

This is basically the same as a **Rasch model** from education.

BTL ESTIMATION

```
dummies = {}
for team in teams:
    dummies[team] = (df['H'] == team).astype(np.int) - (df['V'] ==
= team).astype(np.int)
df2 = pd.DataFrame(dummies)
df2['win'] = (df['PTSH'] > df['PTSV']).astype(np.int)
y, X = dmatrices('win ~ 0 + %s' % ' + '.join(teams), df2)
m = linear_model.LogisticRegression(C=c, penalty='l2', tol=1e-6,
                                    fit_intercept=False)
m.fit(X, y)
quality_scores = clf.coef_[0]
```

Could easily use Bayesian estimation here.

BTL EVOLVED

- Add a parameter for home-field advantage.
- Weight observations so that older games count less toward the likelihood.
- Regularization.
- Add control variables for random events that affect outcomes but are NOT part of team quality.
- e.g. fumbles lost, defensive touchdowns, missed field goals.

BTL METHOD

1		9		17		25	
2		10		18		26	
3		11		19		27	
4		12		20		28	
5		13		21		29	
6		14		22		30	
7		15		23		31	
8		16		24		32	

SCORE



0 — 21

FEATURE ENGINEERING

- Idea: don't score a team, score the inputs to success.
- We can estimate fewer parameters and get some intuition for rankings.
- Approach taken by Brian Burke at AdvancedNFLStats.com
- Use logistic regression: Generic WP = $\text{logit}^{-1}(\beta' \mathbf{x}_i)$

FEATURES

Should be correlated with success AND stable over time

- Offensive Net Yards Per Attempt
- Defensive Net Yards Per Attempt
- Offensive run success rate
- Defensive run success rate
- Interception rate
- Defensive interception rate
- Fumble rate
- Penalty rate

Team	netYPA	Run SR%	Int%	Fum%	D netYPA	D Run SR%	D Int%	Pen Rate
SF	7.1	48	1.7	2.5	5.6	52	2.3	0.48
BLT	6.6	39	1.6	1.4	6.0	56	2.4	0.51
Avg	6.2	41	2.5	2.1	6.3	57	2.4	0.41

WIN PROBABILITY



38% – 62%

OPTIMAL RANKINGS

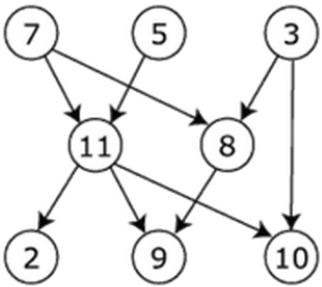
What's the absolute best we can do in *describing* the data with a ranking?

GAMES AS A GRAPH

Season: 2011

Week: 17

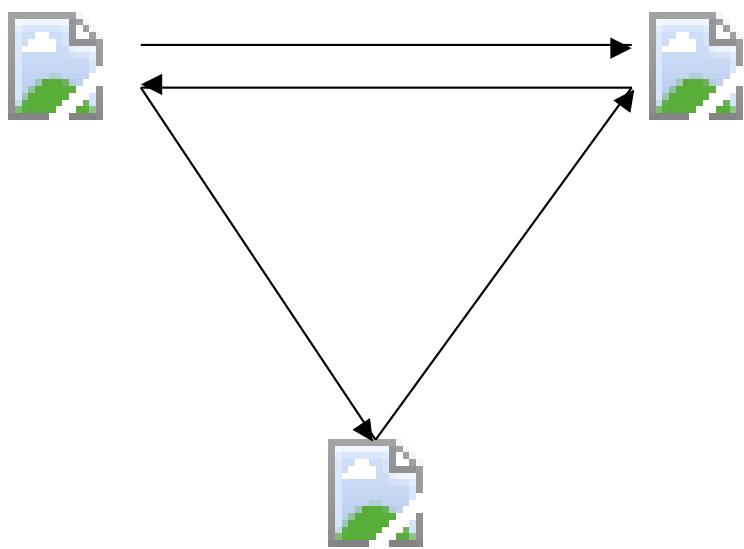
TOPOLOGICAL SORT



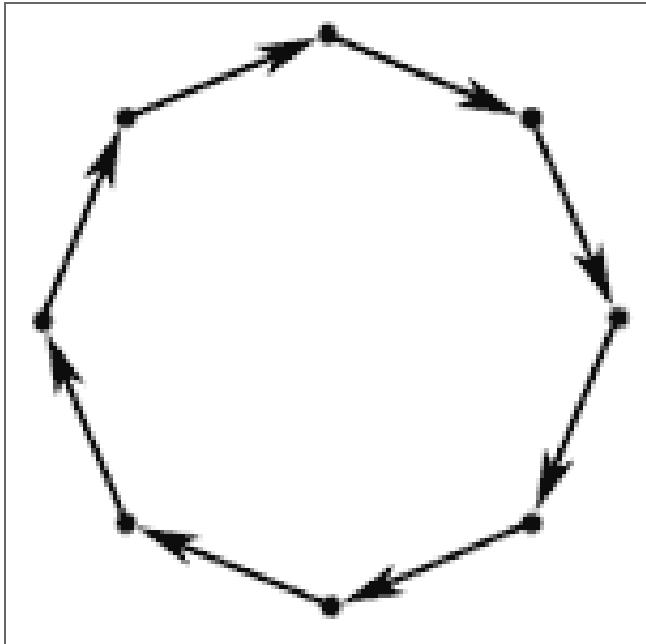
The graph shown to the left has many valid topological sorts, including:

- 7, 5, 3, 11, 8, 2, 9, 10 (visual left-to-right, top-to-bottom)
- 3, 5, 7, 8, 11, 2, 9, 10 (smallest-numbered available vertex first)
- 3, 7, 8, 5, 11, 10, 2, 9
- 5, 7, 3, 8, 11, 10, 9, 2 (fewest edges first)
- 7, 5, 11, 3, 10, 8, 9, 2 (largest-numbered available vertex first)
- 7, 5, 11, 2, 3, 8, 9, 10

CYCLES



LONG CYCLES



MOVING THE CYCLES

A *feedback arc set* is a set of edges whose removal makes the graph acyclic.

```
# g = igraph.Graph()
feedback_arcs = g.feedback_arc_set(method='exact')
g.delete_edges(feedback_arcs)
```

NP-Complete!

OPTIMAL RANK

1		9		17		25	
2		10		18		26	
3		11		19		27	
4		12		20		28	
5		13		21		29	
6		14		22		30	
7		15		23		31	
8		16		24		32	

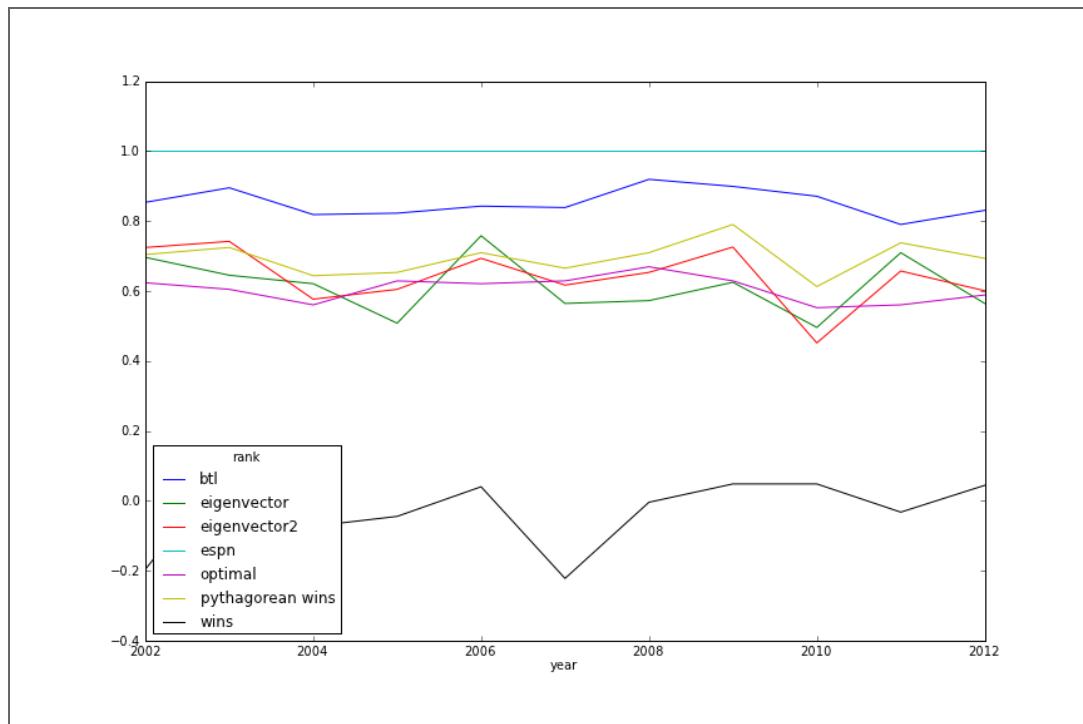
SCORE



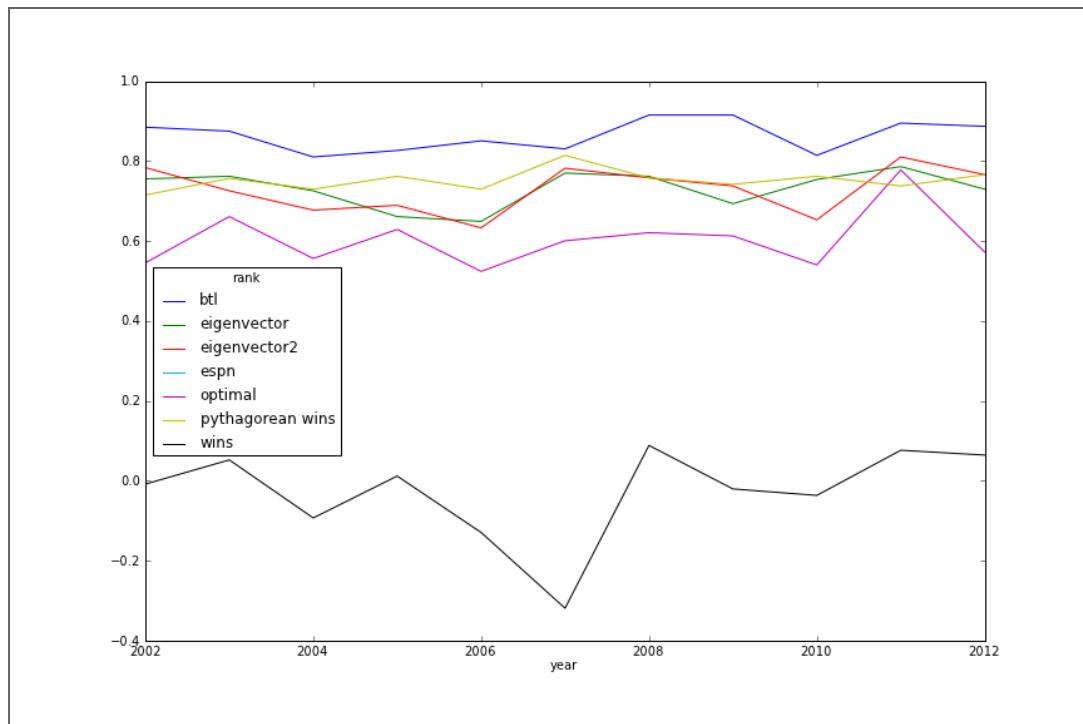
0 – 28

EVALUATION

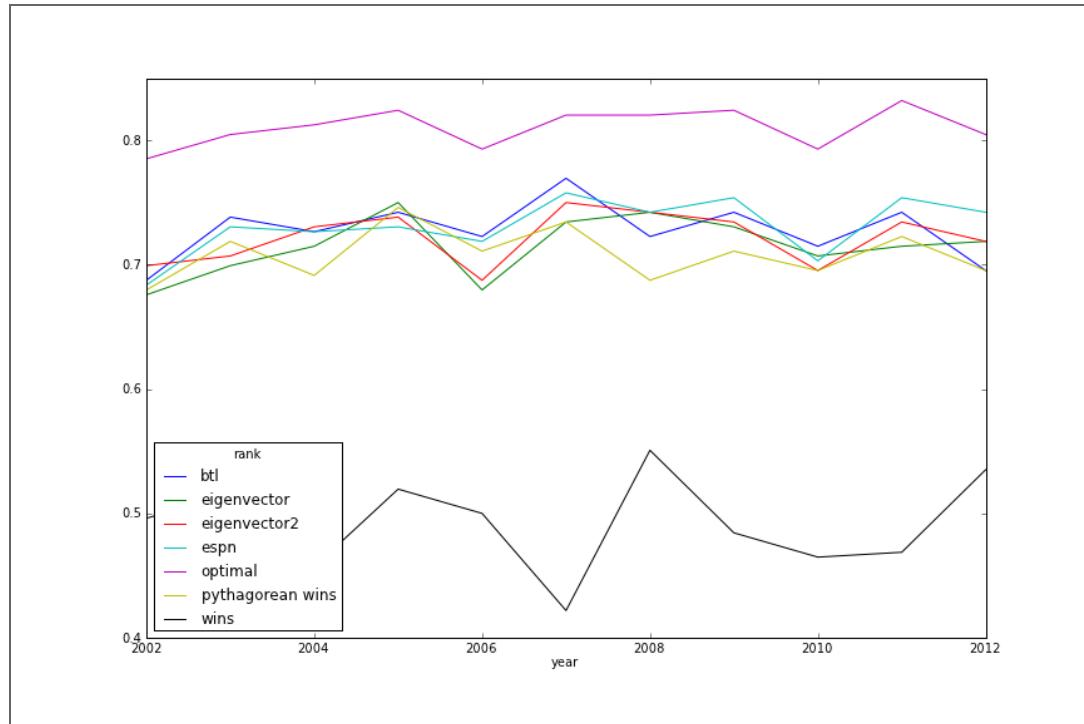
WEEK 8 POWER RANKINGS SIMILARITY



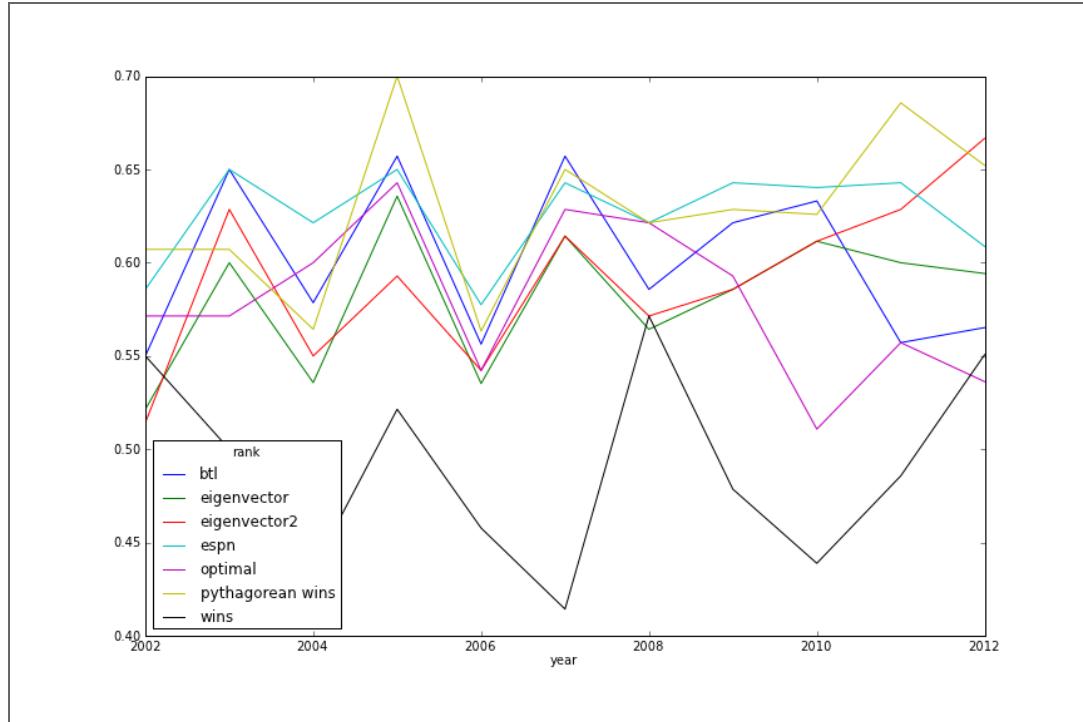
WEEK 17 POWER RANKINGS SIMILARITY



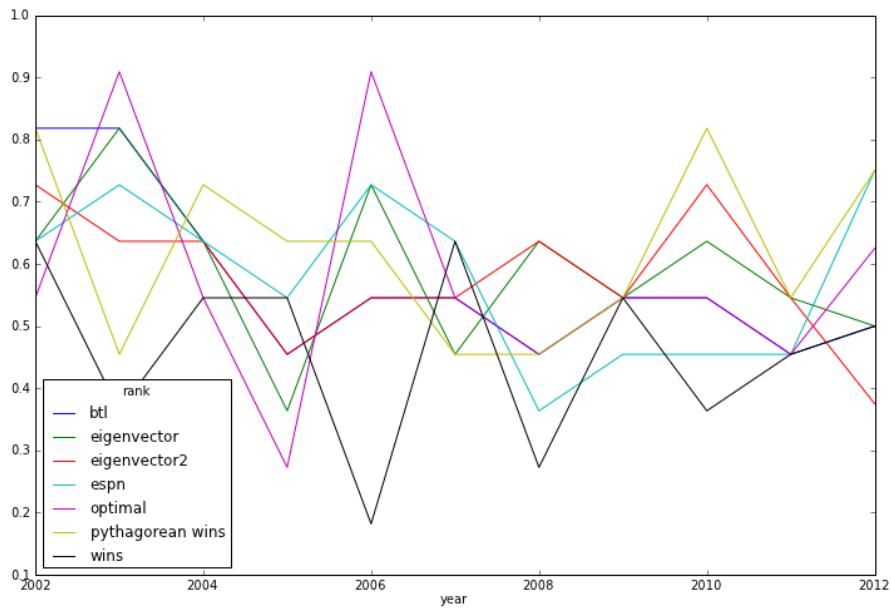
DESCRIPTIVE



IN-SEASON PREDICTION



POST-SEASON PREDICTION

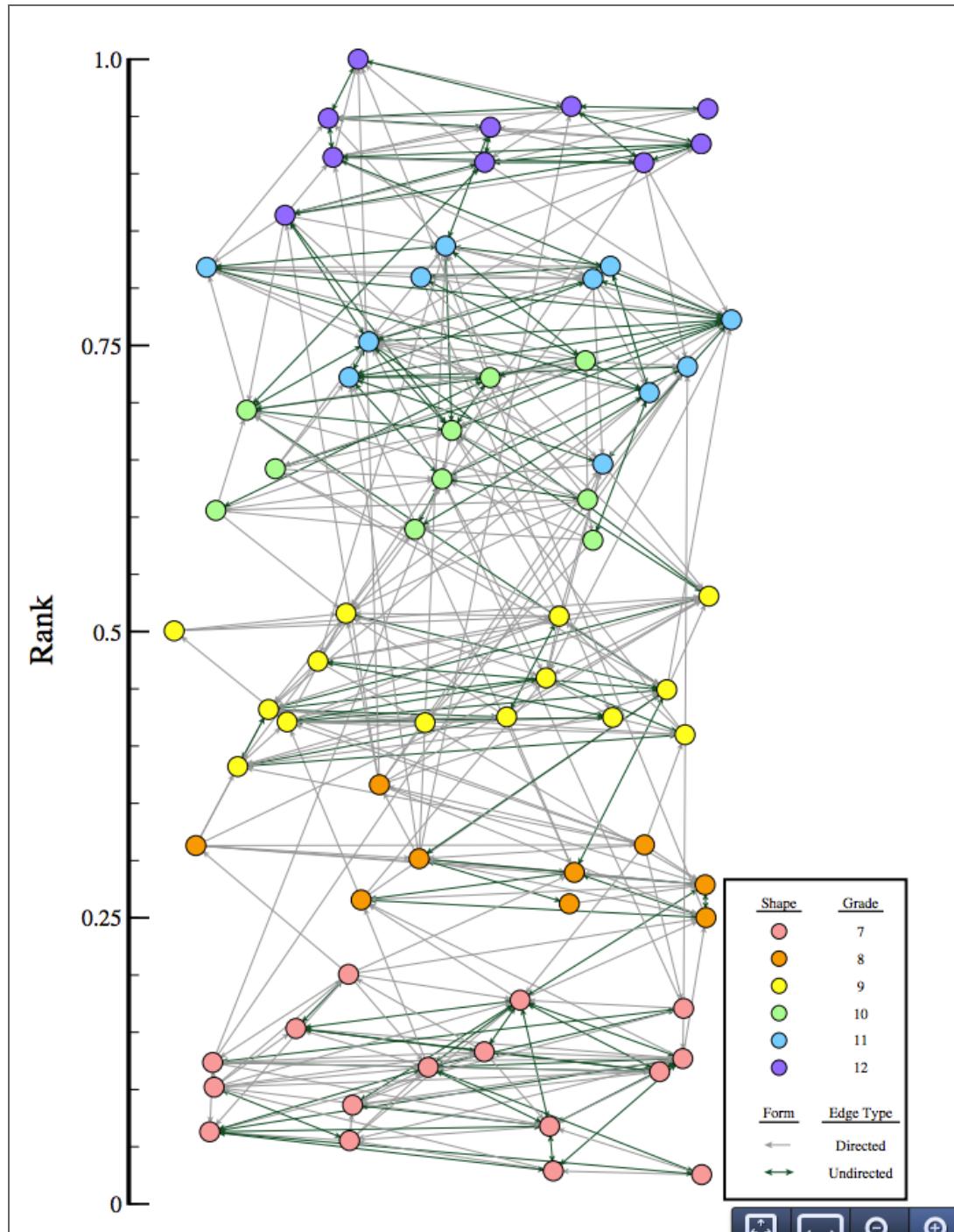


SCORE



24 – 28

RANKING PEOPLE



TURN IT INTO A PAIR-WISE COMPARISON

- Twitter reputation: reciprocity of follow-relationship
- Influence: which sharer of same link gets more clicks?
- Online dating: who spurns messages from whom?

**MAKE YOUR SUPER BOWL
PREDICTIONS**

HTTP://CREDS.ME



(shameless plug)

THANKS! ANY QUESTIONS?



SEAN J. TAYLOR

@SEANJTAYLOR

HTTP://SEANJTAYLOR.COM

<https://github.com/seanjtaylor/NFLRanking>