Pioneer academics

Tackling your Ranking

Tim Chartier

Department of Math & CS

DAVIDSON

WAIT?

Why?

Better rank wins

beats

2 rank

Two purposes of ranking are:

- to rank the teams for their play over an entire season, and
- to predict future play.

When life is linear

- We will use a linear model used by the Bowl Championship Series for college football.
- This model gives interdependence of the teams.

Picture credit: http://www.bcsfootball.org

Just solve it…

- The linear system:

$$2x + y = 5$$

$$3x - y = 10$$

has 2 equations and 2 unknowns.

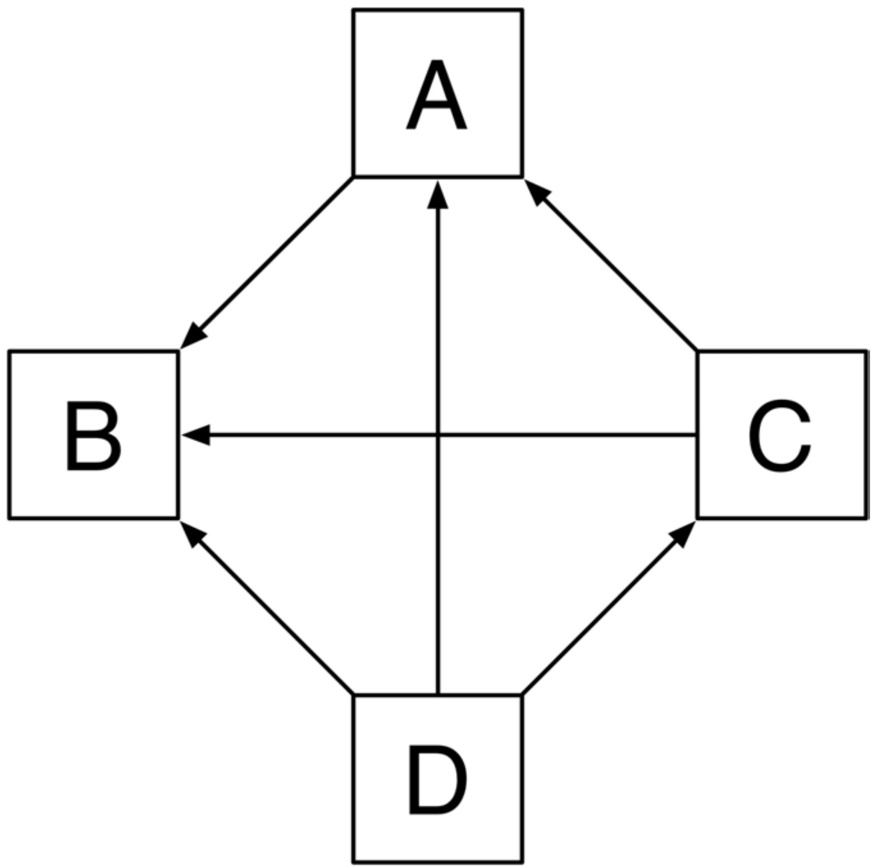
- Our system will simply be bigger – the number of unknowns equals the number of teams.

Getting graphic

- Suppose we only know who won and who lost each game.
- We denote this with an arrow (directed edge) from the winning to the losing team.

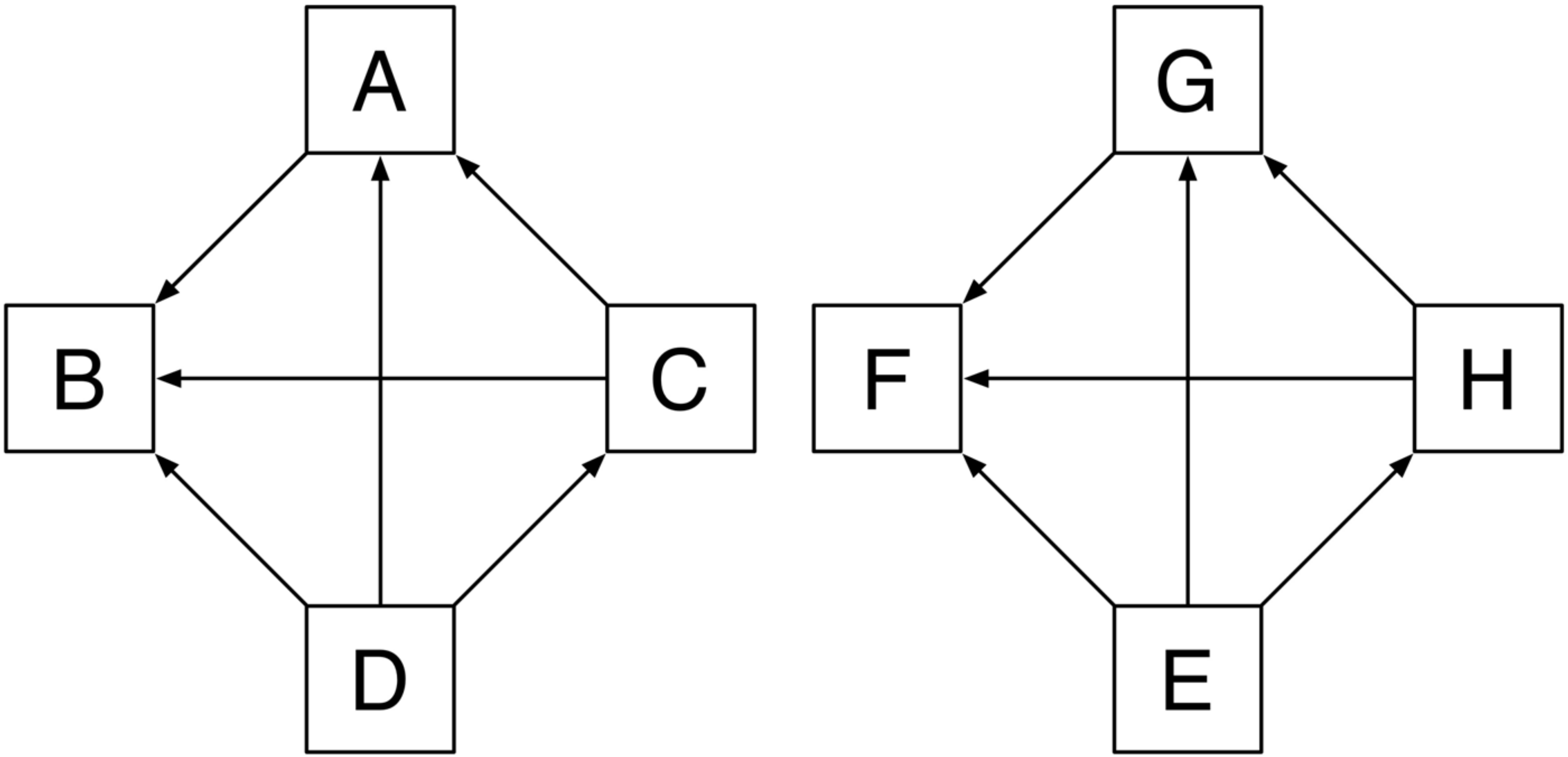
Ready to Rank

How would you rank the following teams?



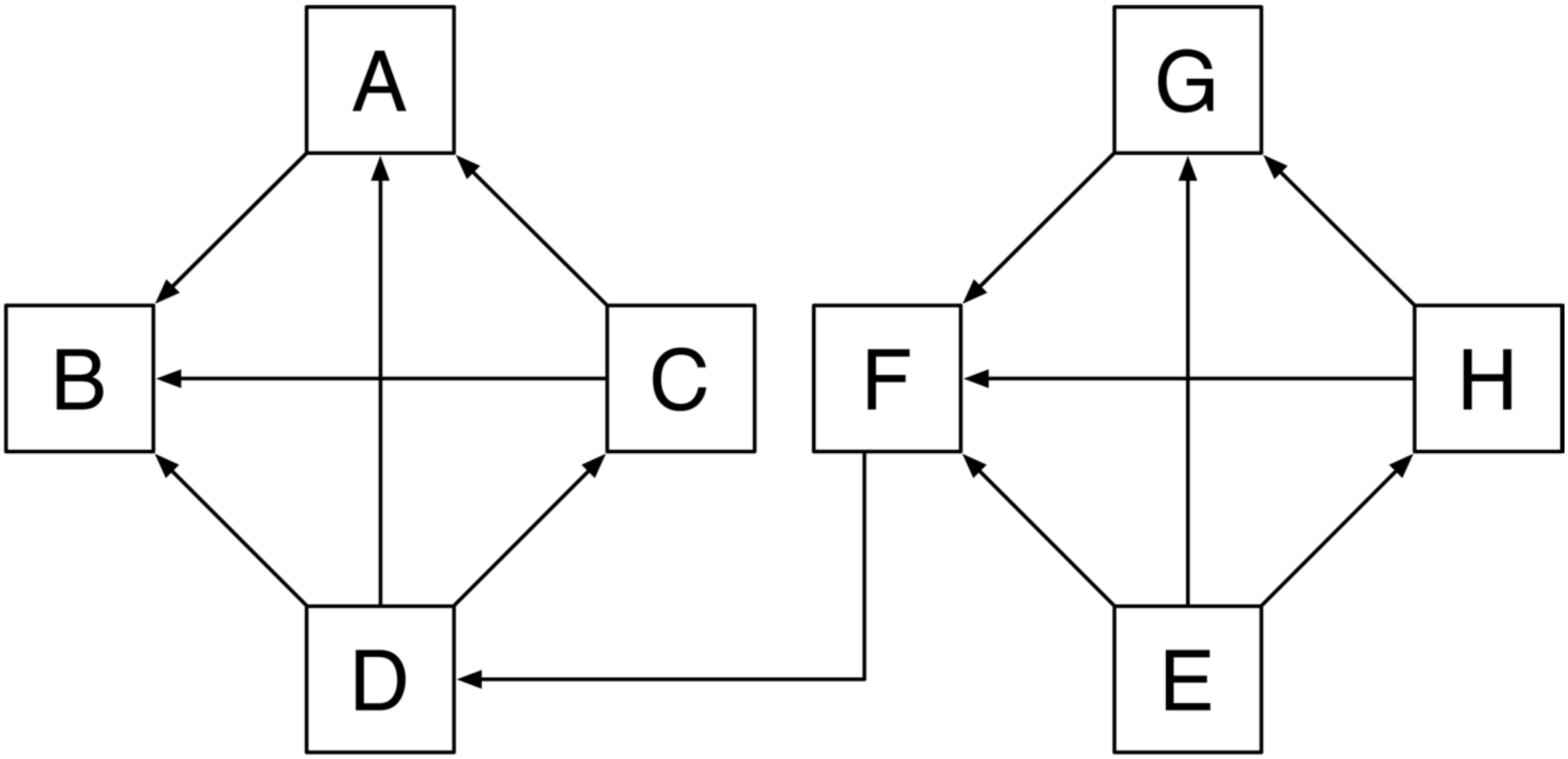
Two conferences

Suppose we have two conferences.



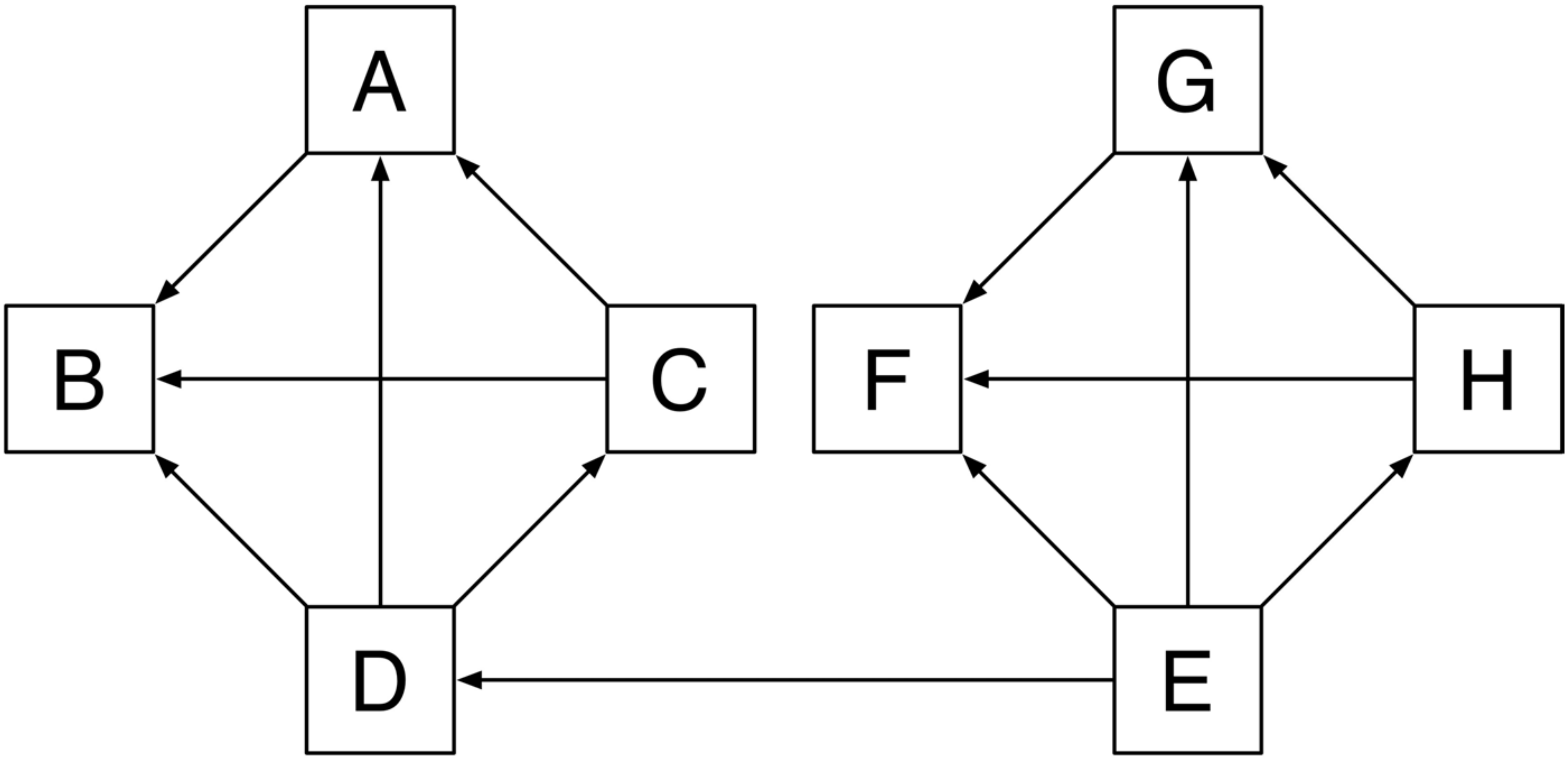
Rank again...

How would you rank now?



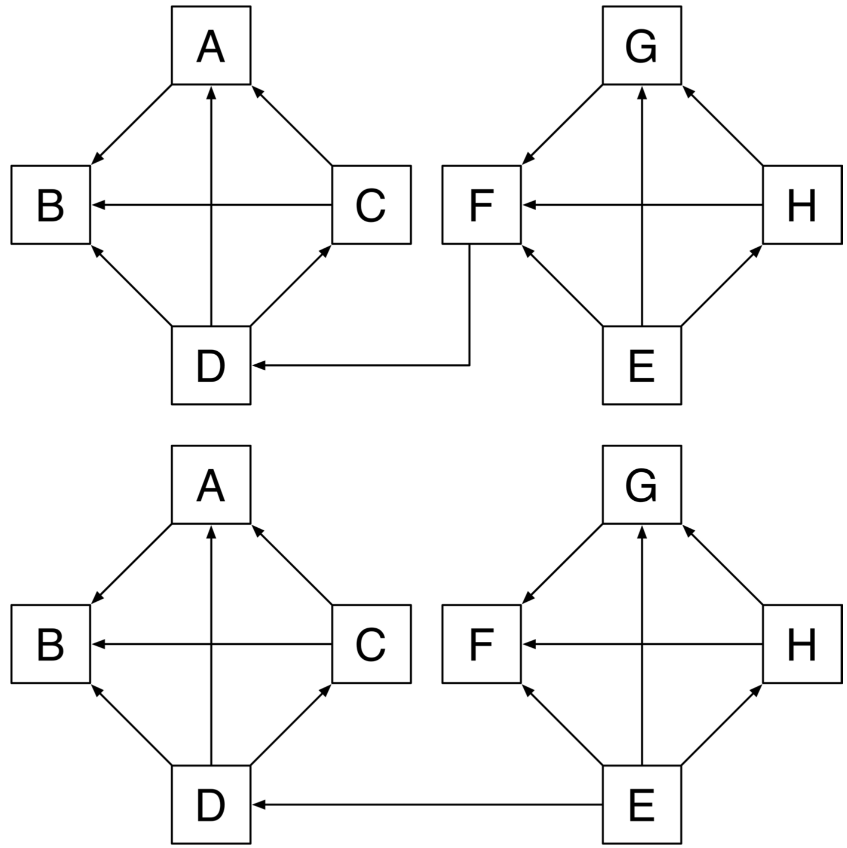
and again…

How would you rank now?



Dependable math

Our rankings will have a sense of “quality” of a win or loss.



METHOD

Why linear?

- Sports often rank by winning percentage.
- Applying this to the results from the 10th week of the 2008-2009 season.

8 wins, 2 losses
80%

7 wins, 3 losses
70%

6 wins, 4 losses
60%

5 wins, 5 losses
50%

Rate to Rank

- A percentage above is the *rating* of a team.
- Using these, we get the *ranking*.

8 wins, 2 losses
80%

7 wins, 3 losses
70%

6 wins, 4 losses
60%

5 wins, 5 losses
50%

Analysis

<u>Concern</u> Defeating the worst team improves one's rating just as much as defeating the best team.

A future win against the Detroit Lions would result in the same effect on the Panther's final rating as a win against the NY Giants.

0 wins, 10 losses   9 wins, 1 loss

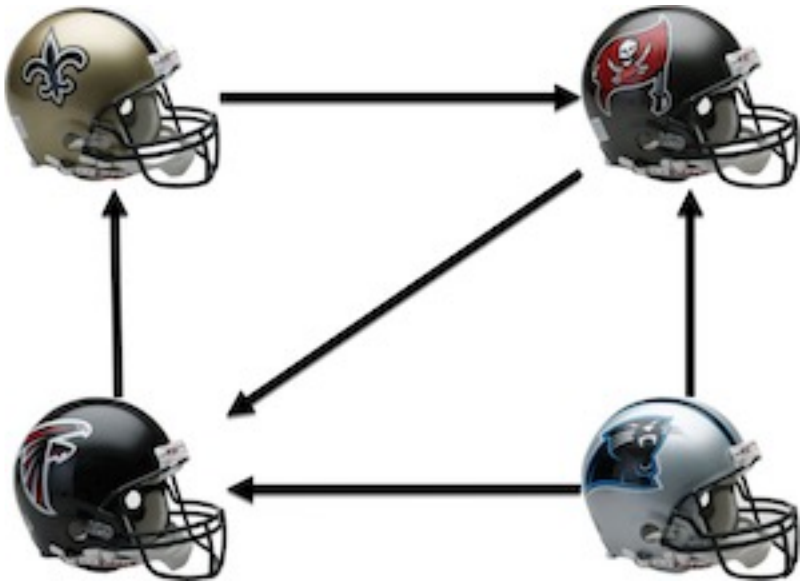
Nuts and bolts included

- Our method will take this into account.
- We will bypass the derivation of the method and focus on the mechanics of forming the linear system.
- More details on the math behind the method will be in a future homework reading.

Picture credit: http://www.boltdepot.com/images/Chrome-nuts-and-bolts.jpg

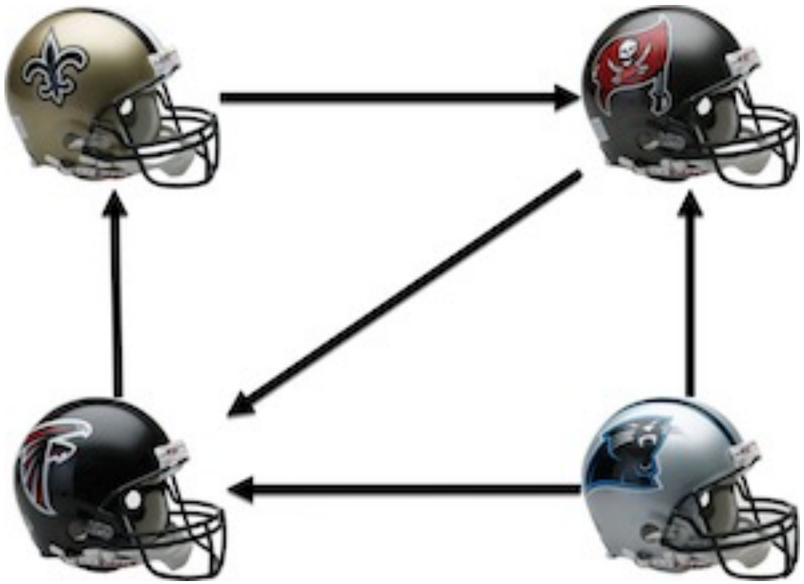
Fictional NFL

- Let's use a fictional series of games within the South Division of the NFL.
- We'll represent the records of the teams by a graph.



Go ahead and point

- Again, an arrow represents a game and points from the winning to the losing team.
- Below the Panthers beat both the Buccaneers and the Falcons but did not play the Saints.

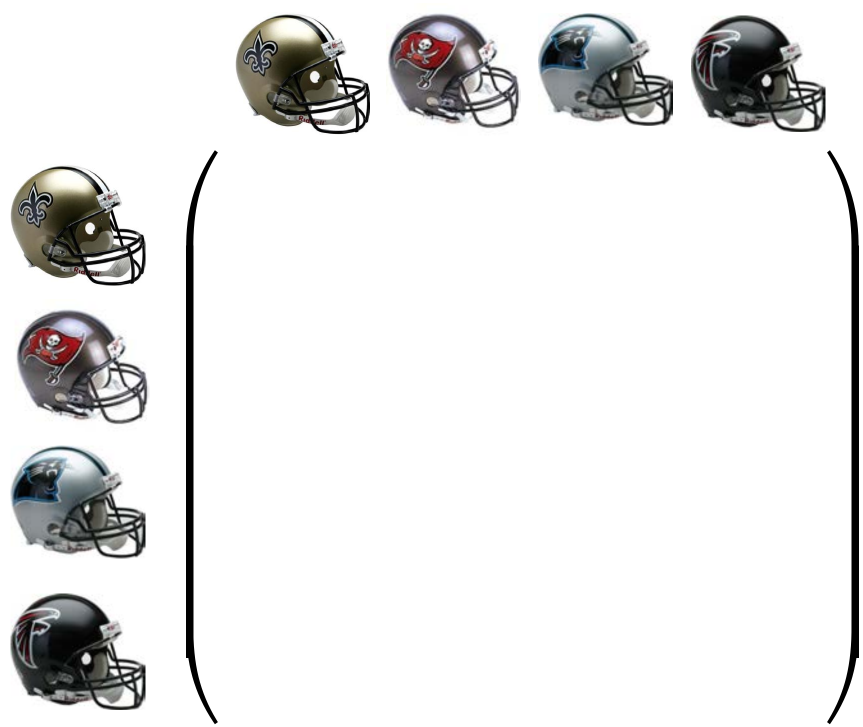


Row, row, row

- There are 4 teams so we need a 4 x 4 matrix.
- Each row of the matrix corresponds to a team.

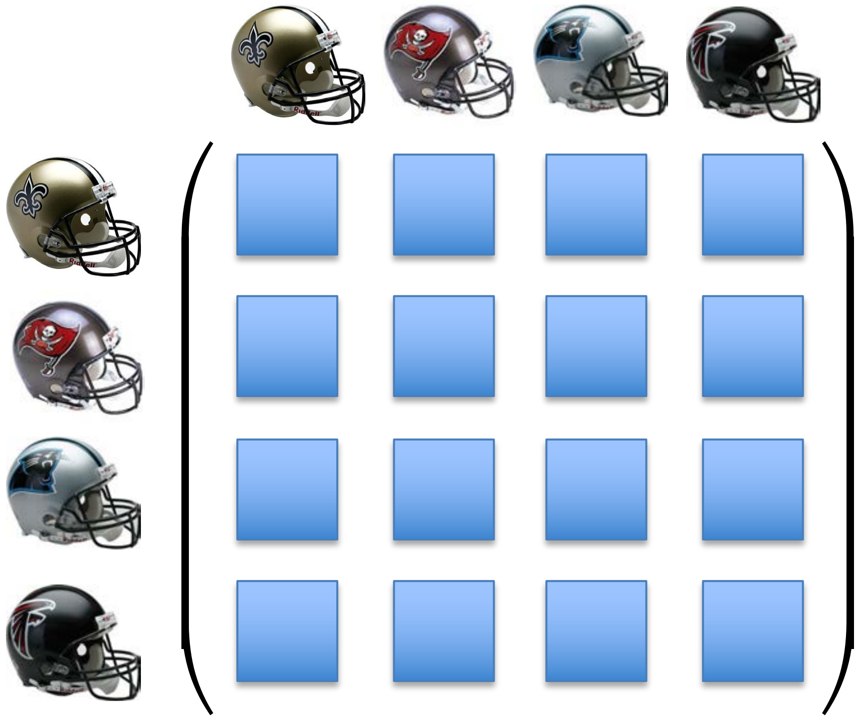
Naming your columns

The columns have the same ordering as the rows.



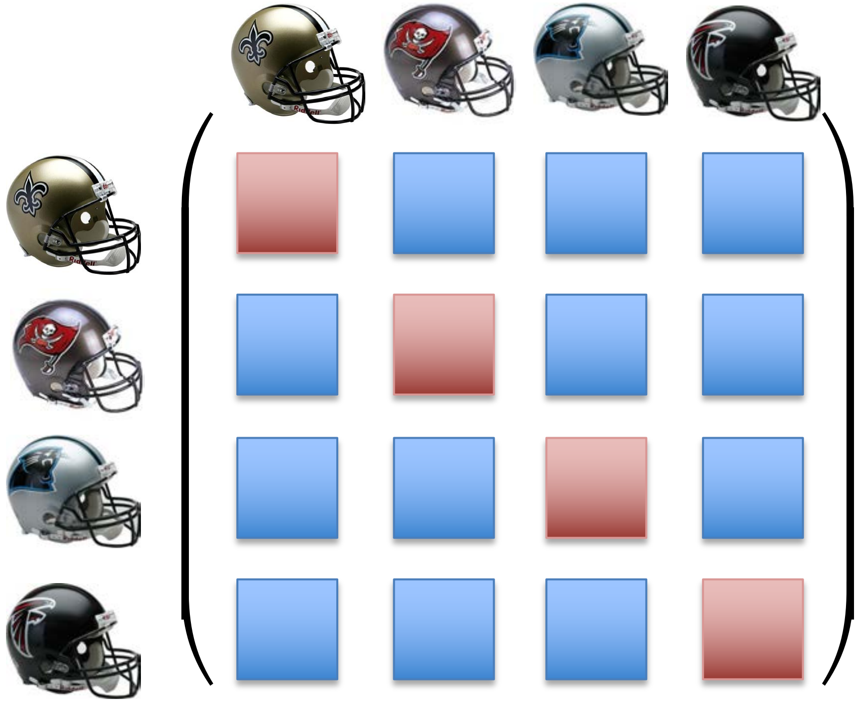
Fill in the blanks

Now, we need to know how to fill in the linear system.



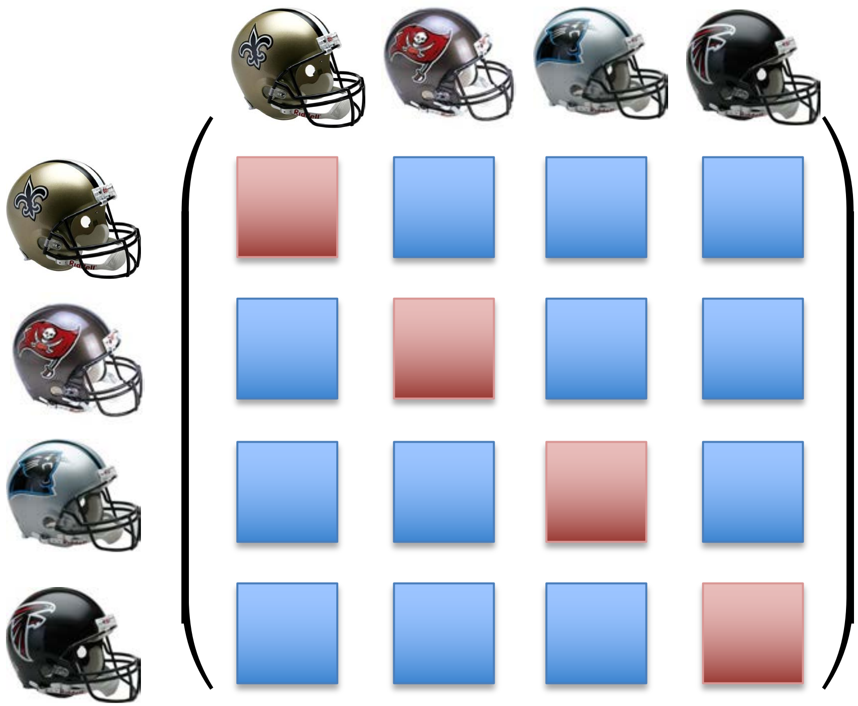
Corner to Corner

Each diagonal element equals 2 + t, where t equals the number of games the corresponding team played.



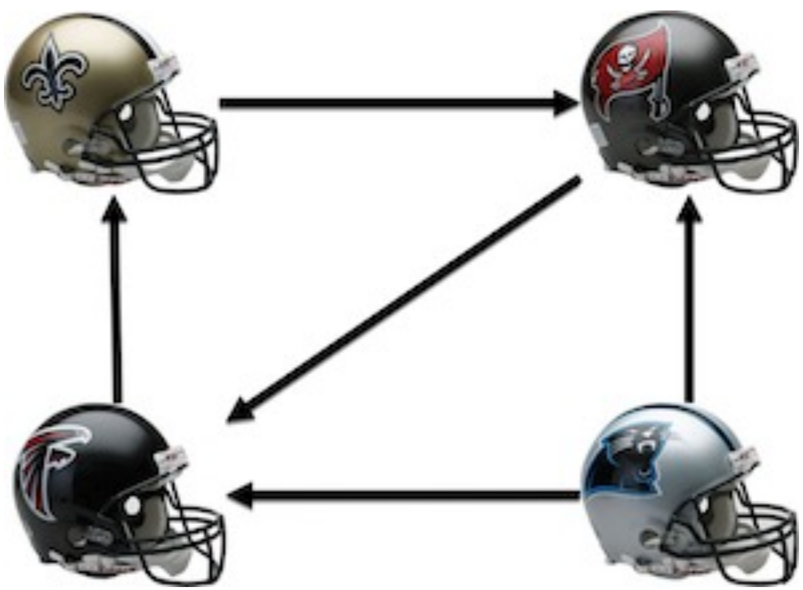
Team to team

Each off-diagonal element equals -*g* where *g* is the number of times the corresponding teams played.



‘Tis the season

Let’s construct the matrix for our season.



$$\begin{pmatrix} 4 & -1 & 0 & -1 \\ -1 & 5 & -1 & -1 \\ 0 & -1 & 4 & -1 \\ -1 & -1 & -1 & 5 \end{pmatrix}$$

Go to the right

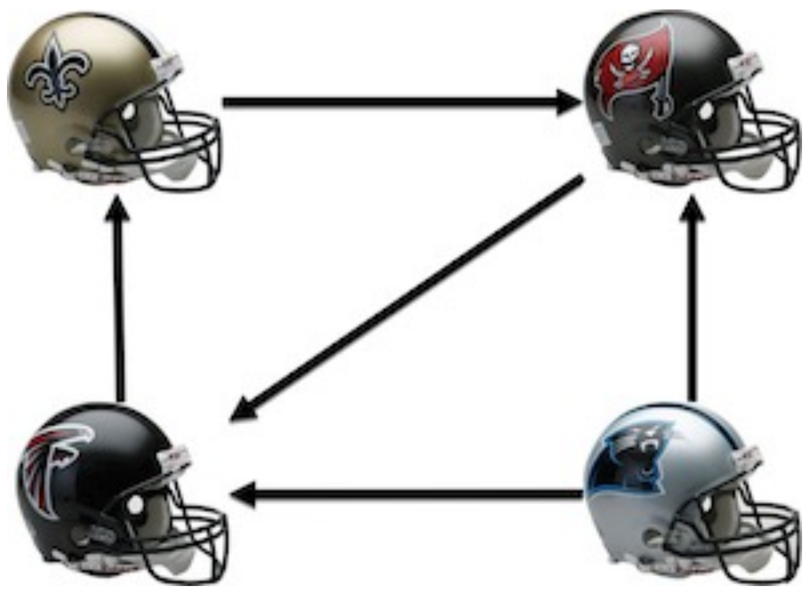
- We need to form the vector for the right-hand side of the linear equation.
- Each row in that vector corresponds to the same team that it did in the matrix.

A Saint

- Let W equal the number of wins by the Saints and L equal the number of losses.
- Then in the first row of the vector, we place the number $1 + 1/2(W - L)$.
- The Saints had 1 win and 1 loss so we place a 1 in that entry in the vector.

I’m right

For the season, the right-hand side vector is, (remember the formula: $1 + 1/2(W - L)$)



$$\begin{pmatrix} 1 \\ 1/2 \\ 2 \\ 1/2 \end{pmatrix}$$

Playing the system

The entire linear system is then:

$$\begin{pmatrix} 4 & -1 & 0 & -1 \\ -1 & 5 & -1 & -1 \\ 0 & -1 & 4 & -1 \\ -1 & -1 & -1 & 5 \end{pmatrix} \begin{pmatrix} S \\ B \\ P \\ F \end{pmatrix} = \begin{pmatrix} 1 \\ 1/2 \\ 2 \\ 1/2 \end{pmatrix}$$

Solving problems

WolframAlpha computational knowledge engine

LinearSolve[{{4,-1,0,-1},{-1,5,-1,-1},{0,-1,4,-1},{-1,-1,-1,5}},{{1},{1/2},{2},{.5}}]

Examples Random

Input:

$$\text{LinearSolve}\left[\begin{pmatrix} 4 & -1 & 0 & -1 \\ -1 & 5 & -1 & -1 \\ 0 & -1 & 4 & -1 \\ -1 & -1 & -1 & 5 \end{pmatrix}, \begin{pmatrix} 1 \\ \frac{1}{2} \\ 2 \\ 0.5 \end{pmatrix}\right]$$

Result:

{{0.458333}, {0.416667}, {0.708333}, {0.416667}}

Computed by Wolfram *Mathematica*

Download page

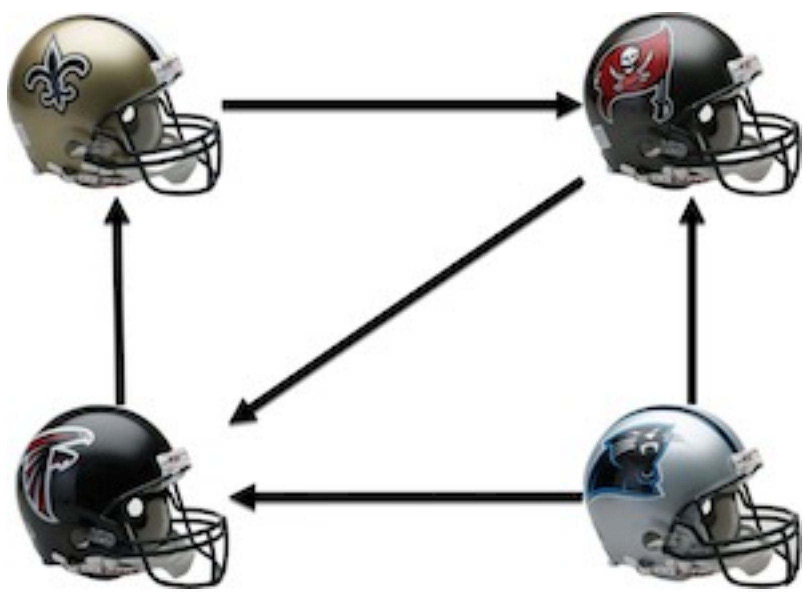
Go to the right

- This gives us the rating for the teams.
- The higher the rating the “better” the team.

$$\begin{pmatrix} 0.458 \\ 0.417 \\ 0.708 \\ 0.417 \end{pmatrix}$$

Who's #1?

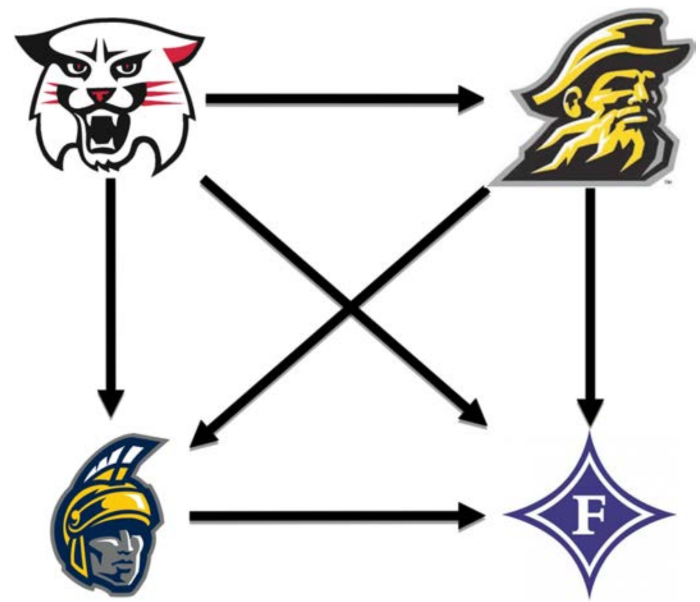
For this season, the teams are ranked (from best to worst) Panthers, Saints and for a tie for third between the Buccaneers and Falcons.



$$\begin{pmatrix} 0.458 \\ 0.417 \\ 0.708 \\ 0.417 \end{pmatrix}$$

Your Turn

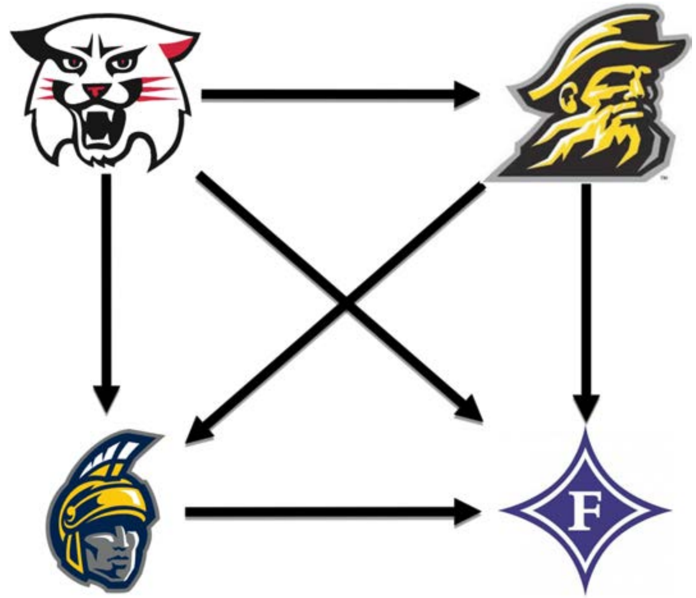
Form the linear system that will rate the teams for the following season:



$$\begin{pmatrix} & & & \\ & & & \\ & & & \\ & & & \end{pmatrix} \begin{pmatrix} D \\ A \\ F \\ U \end{pmatrix} = \begin{pmatrix} \\ \\ \\ \\ \end{pmatrix}.$$

Your Turn

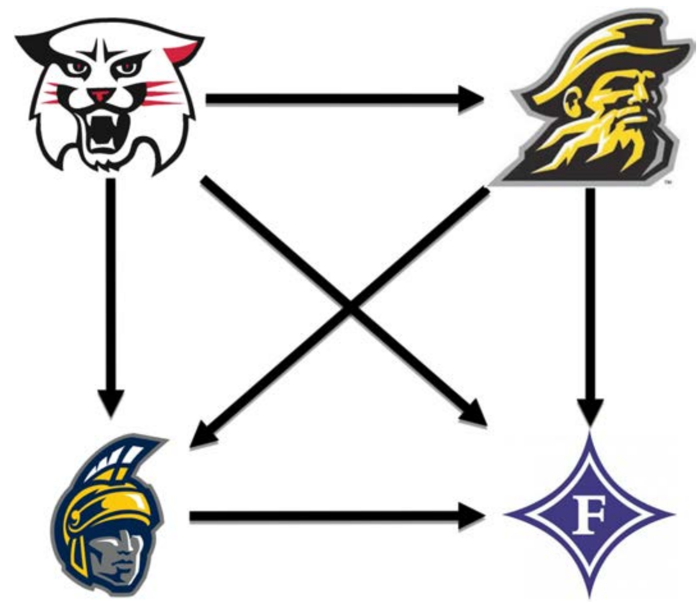
Form the linear system that will rate the teams for the following season:



$$\begin{pmatrix} 5 & -1 & -1 & -1 \\ -1 & 5 & -1 & -1 \\ -1 & -1 & 5 & -1 \\ -1 & -1 & -1 & 5 \end{pmatrix} \begin{pmatrix} D \\ A \\ F \\ U \end{pmatrix} = \begin{pmatrix} 5/2 \\ 3/2 \\ -1/2 \\ 1/2 \end{pmatrix}.$$

Solution

Form the linear system that will rate the teams for the following season. Then find the ratings for these teams.



$$\begin{pmatrix} D \\ A \\ F \\ U \end{pmatrix} = \begin{pmatrix} 0.75 \\ 0.58 \\ 0.25 \\ 0.42 \end{pmatrix}$$

After math processing

For our earlier season results, solving this linear system yields our revised ratings.

0.58

0.58

0.42

0.42

After math processing

- Let's take the earlier case that the Panthers then beat the Falcons.
- Note how the ratings adjusted for this change in the rating of the Panthers and the Falcons.

0.625

0.583

0.375

0.417

Code it!

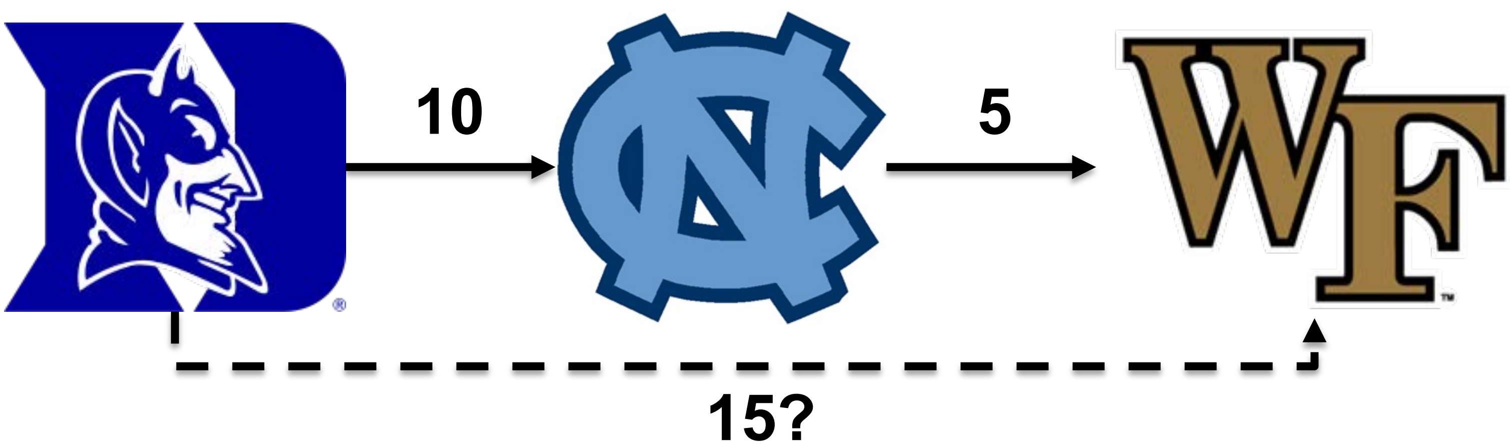
Let's see this in python code.

colleyRankingPython.ipynb

METHOD

Transitivity?

- Suppose Duke beat UNC-Chapel Hill by 10 points and UNC-Chapel Hill beat Wake Forest by 5 points.
- Could we predict Duke would beat Wake Forest by 15 points?



Crystal ball not included

Can it be true approximately?

Picture credit: http://www.photo-dictionary.com/phrase/5923/crystal-ball.html

Rated r

- Let r_1, r_2, and r_3 be the ratings for Duke, UNC-Chapel Hill and Wake Forest.
- This method will take:

$$r_1 - r_2 = 10, \text{ and}$$

$$r_2 - r_3 = 5.$$

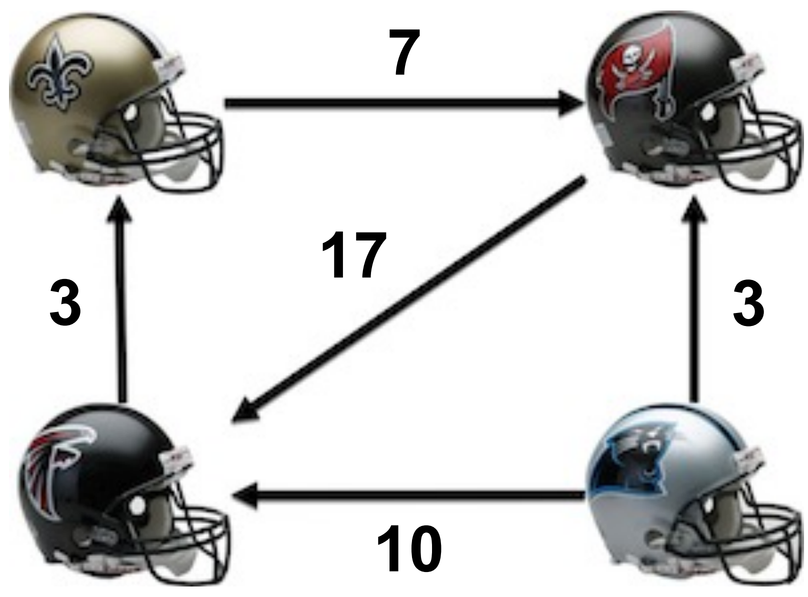
Rated r

- Again, a future reading will discuss the math behind the method.
- For now, we will get the nuts and bolts of forming the linear system.

Picture credit: http://www.itracing.co.uk/suspension-nuts-and-bolts

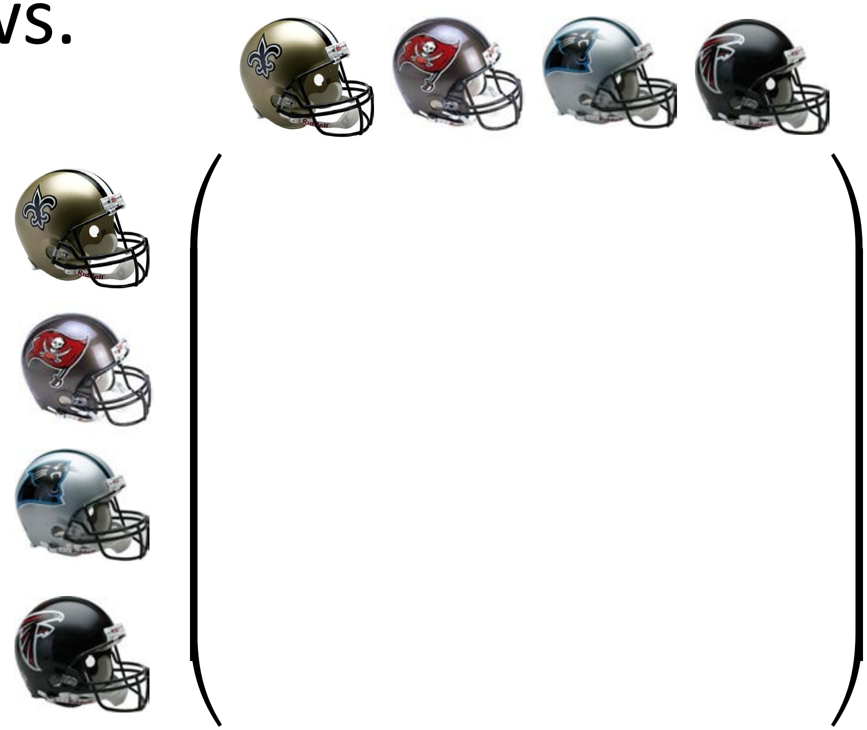
Fictional NFL

- Let's return to our fictional series of games within the South Division of the NFL.
- An arrow points from the winning to the losing team with the weight equaling the point differential of the final scores.



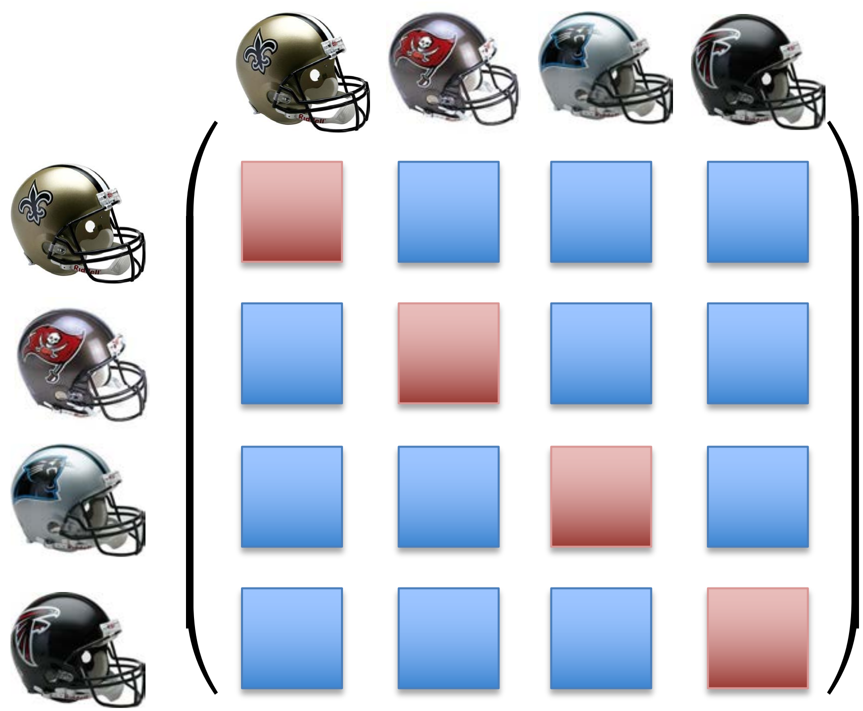
Line ‘em up

- There are 4 teams so we need a 4 x 4 matrix.
- Each row of the matrix corresponds to a team and the columns have the same ordering as the rows.



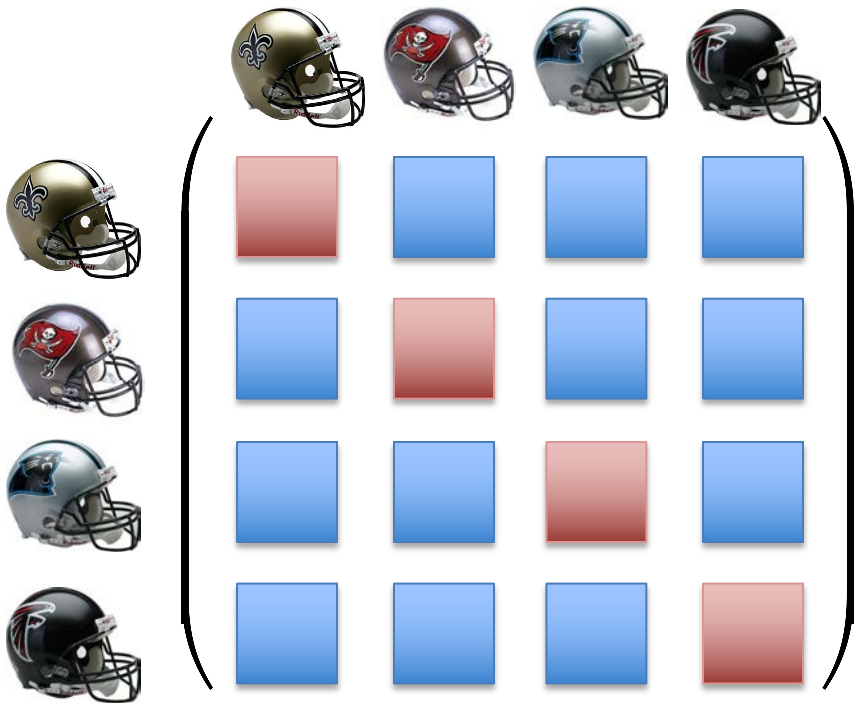
Fill in the blanks

Each diagonal element equals the number of games played by the corresponding team.



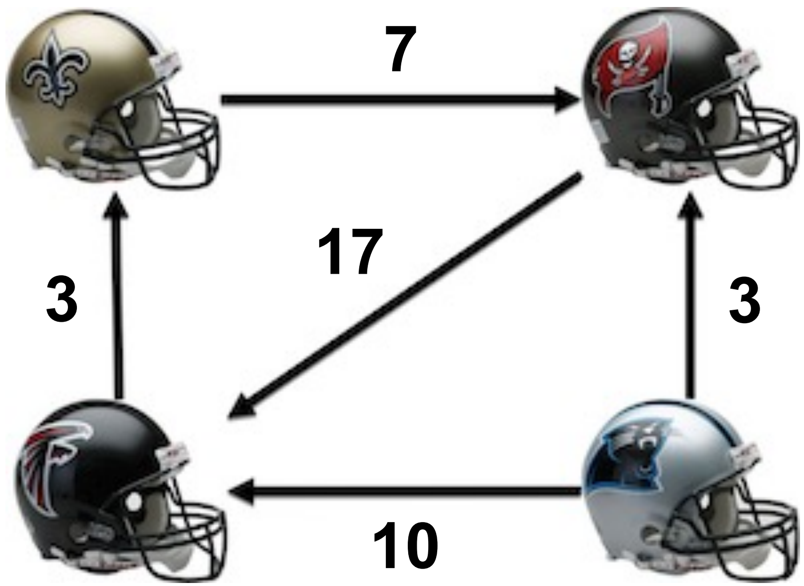
Team to team

Each off-diagonal element equals -*g* where *g* is the number of times the corresponding teams played.



‘Tis the season

Let’s construct the matrix for our season.



$$\begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 2 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix}$$

Go to the right

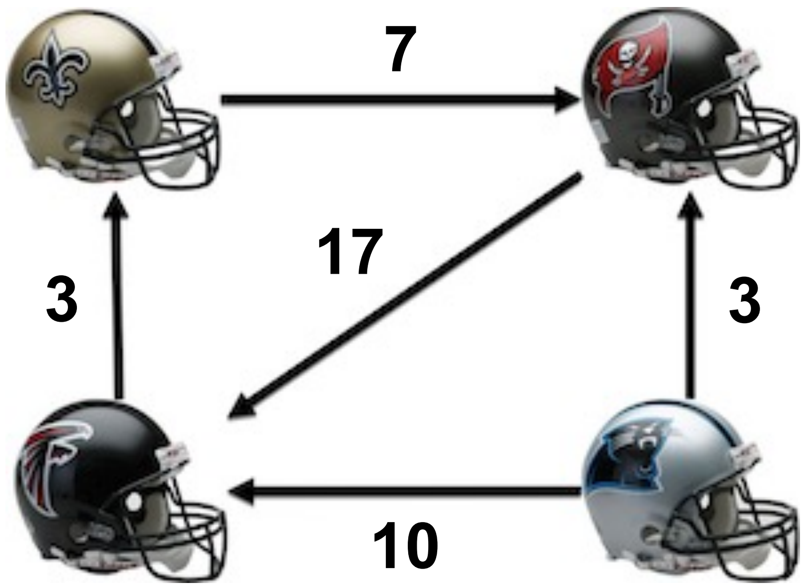
- We need to form the vector for the right-hand side of the linear equation.
- Each row in that vector corresponds to the same team that it did in the matrix.

A Saint

- The RHS equals the sum of the point differentials where a win gives a positive differential and a loss, a negative one.
- The Saints won by 7 and lost by 3. So, the RHS associated with this team would be 7-3 = 4.

I’m right

For the season, the right-hand side vector is,



$$\begin{pmatrix} 4 \\ 7 \\ 13 \\ -24 \end{pmatrix}$$

Playing the system

The entire linear system is then:

$$\begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 2 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix} \begin{pmatrix} S \\ B \\ P \\ F \end{pmatrix} = \begin{pmatrix} 4 \\ 7 \\ 13 \\ -24 \end{pmatrix}$$

Note: There is one more step.

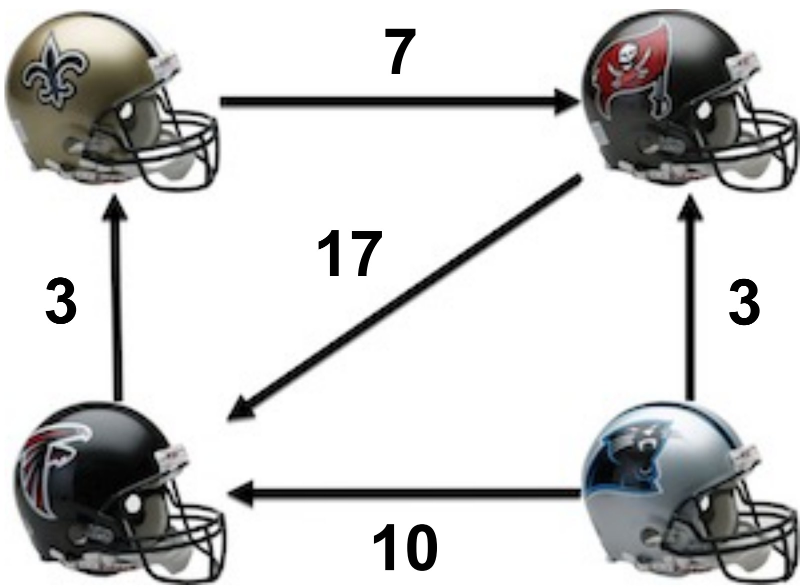
Playing the system

Replace any row by 1's in the matrix and 0 on the RHS.

$$\begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 2 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} S \\ B \\ P \\ F \end{pmatrix} = \begin{pmatrix} 4 \\ 7 \\ 13 \\ 0 \end{pmatrix}$$

Who’s #1?

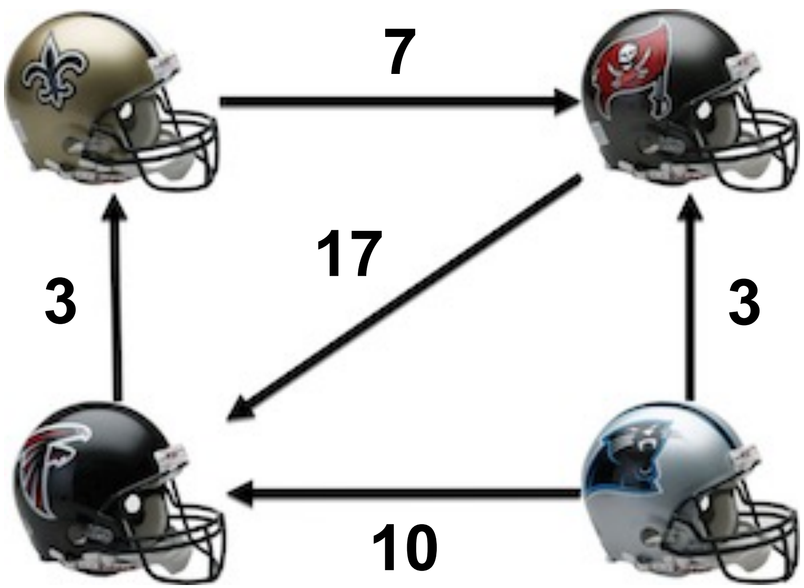
For this season, the teams are ranked (from best to worst) Panthers, Buccaneers, Saints and Falcons.



$$\begin{pmatrix} -0.125 \\ 1.75 \\ 4.375 \\ -6 \end{pmatrix}$$

Colley vs Massey

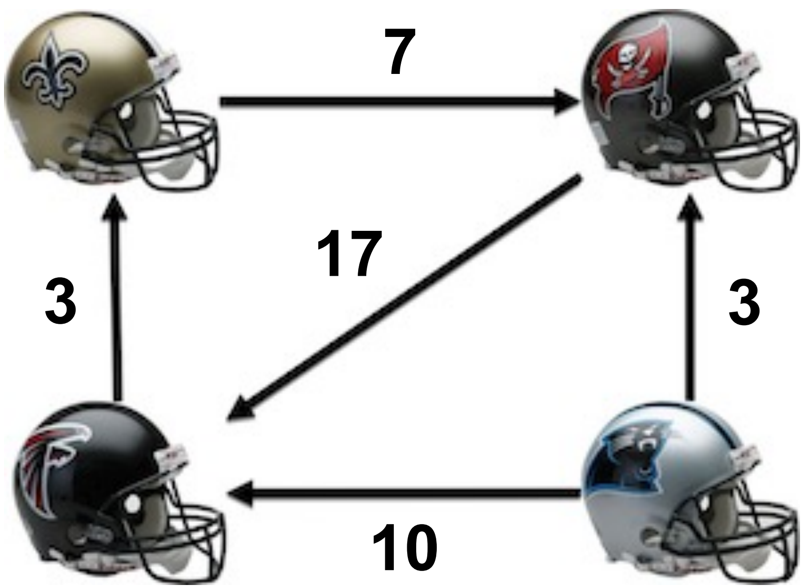
- Note that the Buccaneers and Saints switched places versus their ranking with Colley.
- Can you see why?



$$\begin{pmatrix} -0.125 \\ 1.75 \\ 4.375 \\ -6 \end{pmatrix}$$

Prediction

So, if the Panthers played the Saints, this method predicts the Panthers will win by 4.375 + 0.125 = 4.5 points.



$$\begin{pmatrix} -0.125 \\ 1.75 \\ 4.375 \\ -6 \end{pmatrix}$$

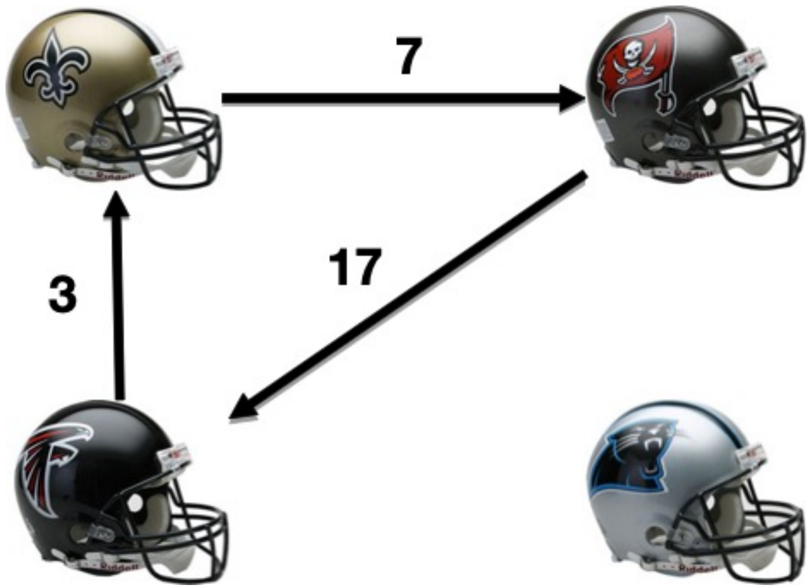
An honorable start

The Massey method started as an honors math project by Ken Massey while he was undergraduate at Bluefield College.

VIRGINIA'S
Bluefield College
Live the Challenge

Important!!!

Massey must have a connected network.



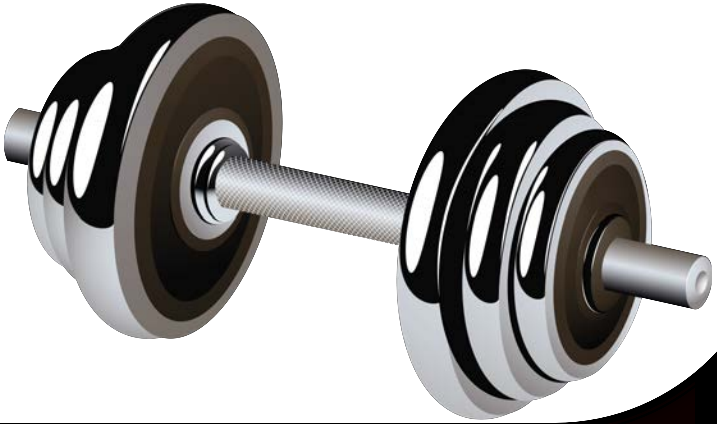
Code it!

Let’s see this in python code.

masseyRankingPython.ipynb

Individualized rankings

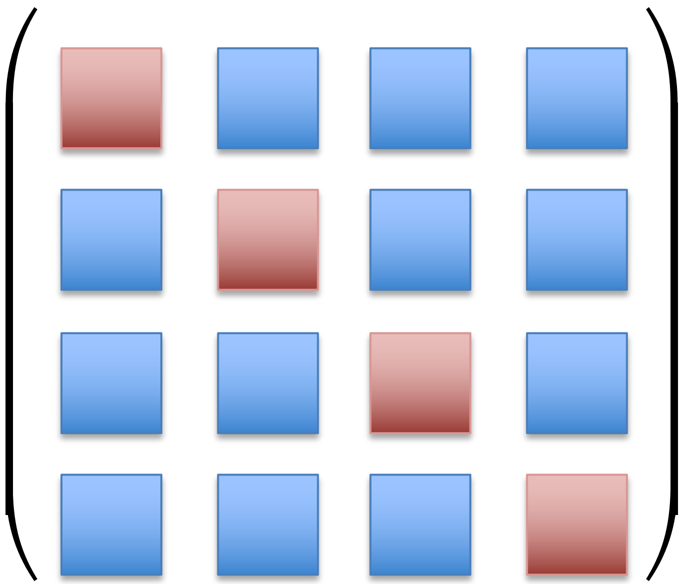
- Let's individualize Colley and Massey.
- We add different "weights" to the games.
- **Question**: How do you weight different games?
- This idea can be adapted to PageRank and possibly other ranking methods.



COLLEY

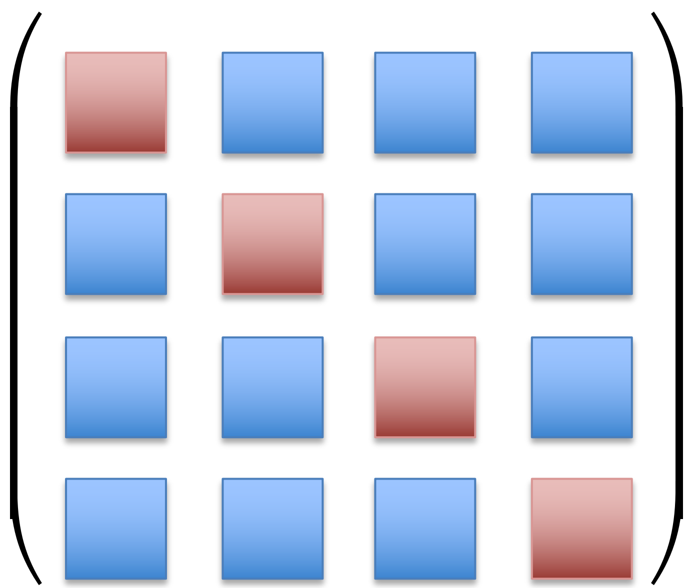
Corner to Corner

Each diagonal element equals 2 + t, where t equals the number of weighted games the corresponding team played.



Team to team

Each off-diagonal element equals -*g* where *g* is the number of weighted games the corresponding teams played.



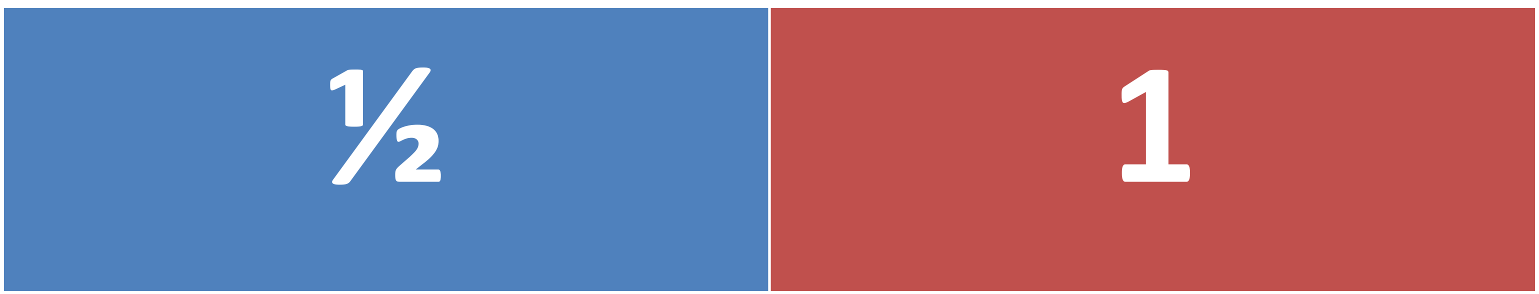
Go to the right

Each element in the right-hand side vector equals $1 + 1/2(W - L)$, where again W and L are the number of weighted wins and losses for the corresponding team.

$$\begin{pmatrix} \\ \\ \\ \\ \end{pmatrix}$$

Weighting time

Let’s break the season in half and count games in the first half as ½ and the second half as 1.

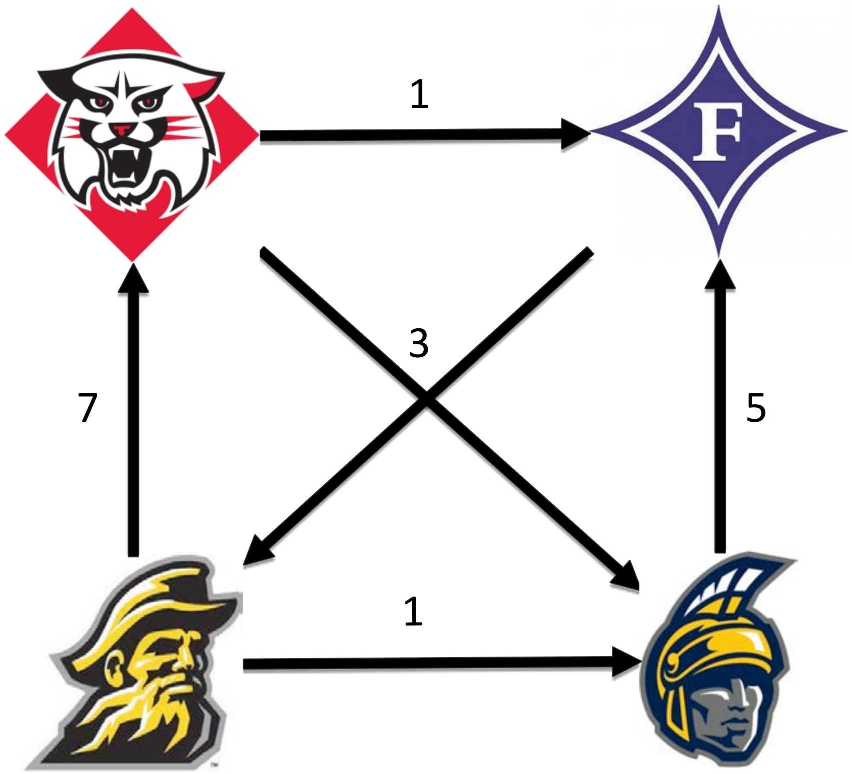


Davidson games

Assume half the season is at 7/2 = 3.5.

Davidson

- won 1/2 + 1/2 = 1
- lost 1



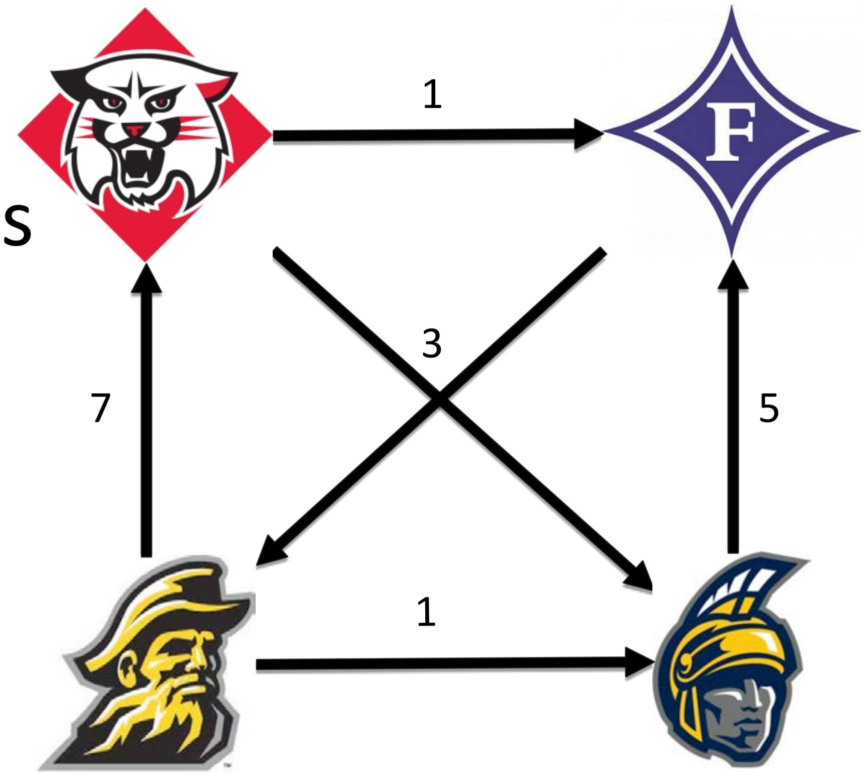
Weighted record

Davidson – 1 win, 1 loss

Furman – 1/2 win, 3/2 loss

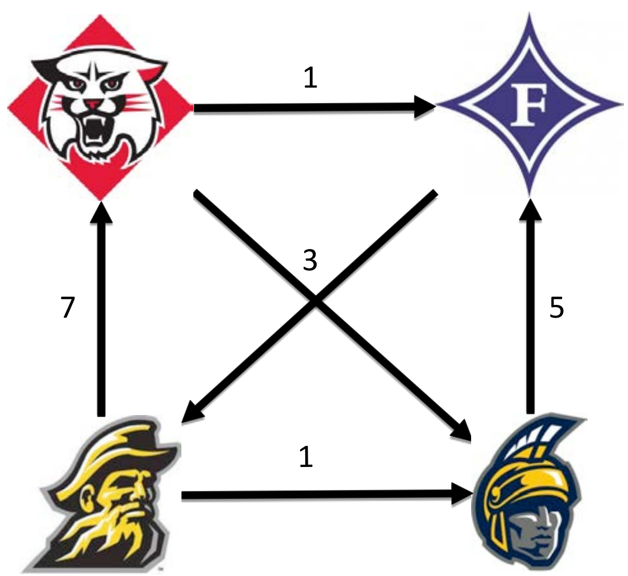
UNCG – 1 win, 1 loss

App – 3/2 win, 1/2 loss



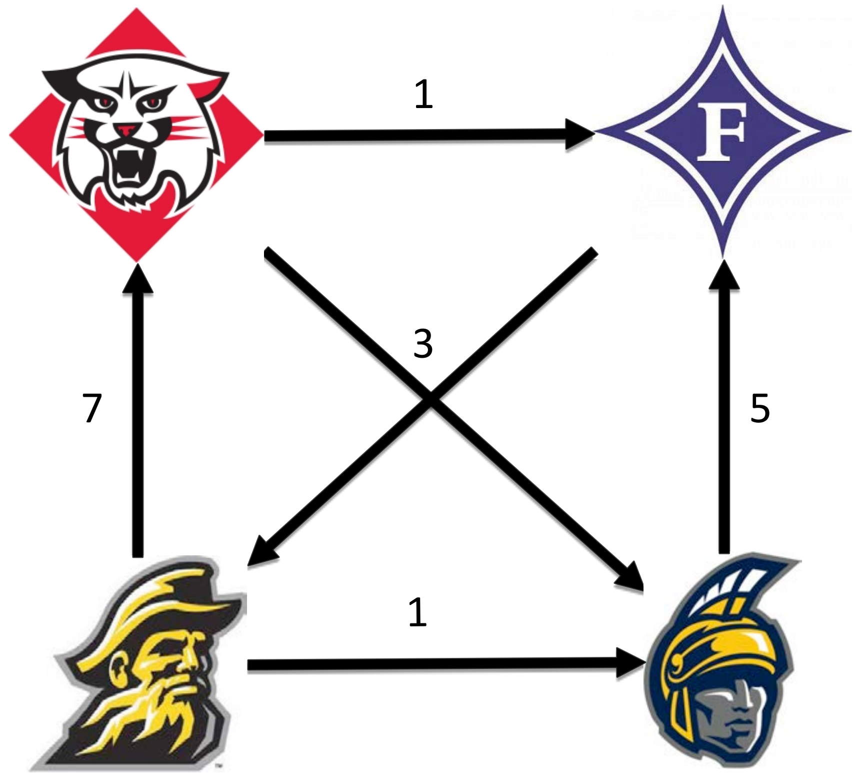
Davidson games

$$\begin{pmatrix} 4 & -\frac{1}{2} & -\frac{1}{2} & -1 \\ -\frac{1}{2} & 4 & -1 & -\frac{1}{2} \\ -\frac{1}{2} & -1 & 4 & -\frac{1}{2} \\ -1 & -\frac{1}{2} & -\frac{1}{2} & 4 \end{pmatrix} \begin{pmatrix} D \\ F \\ U \\ A \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{1}{2} \\ 1 \\ \frac{3}{2} \end{pmatrix}$$



Ranking

$$\begin{pmatrix} D \\ F \\ U \\ A \end{pmatrix} = \begin{pmatrix} 0.5125 \\ 0.3875 \\ 0.4875 \\ 0.6125 \end{pmatrix}$$



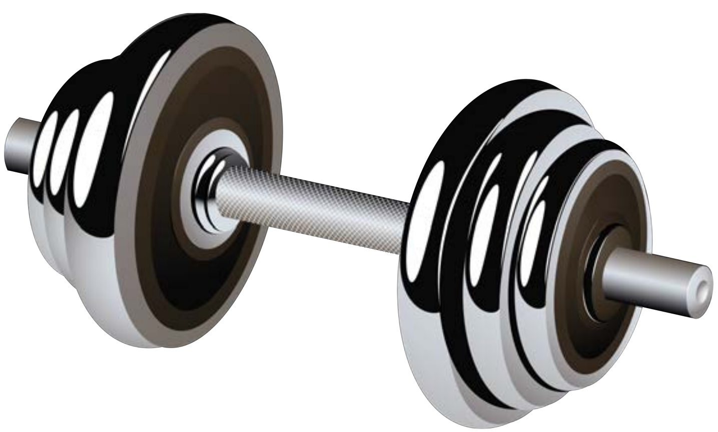
Ranking

1. App State
2. Davidson
3. UNCG
4. Furman

MASSEY

Massey adds weight

- We use weighted least squares for Massey.
- Again, how much will different games count?



Weighted Massey

- Recall the Massey method is:

$$M^T M\, \mathbf{r} = M^T\, \mathbf{b}$$

- Now, we take:

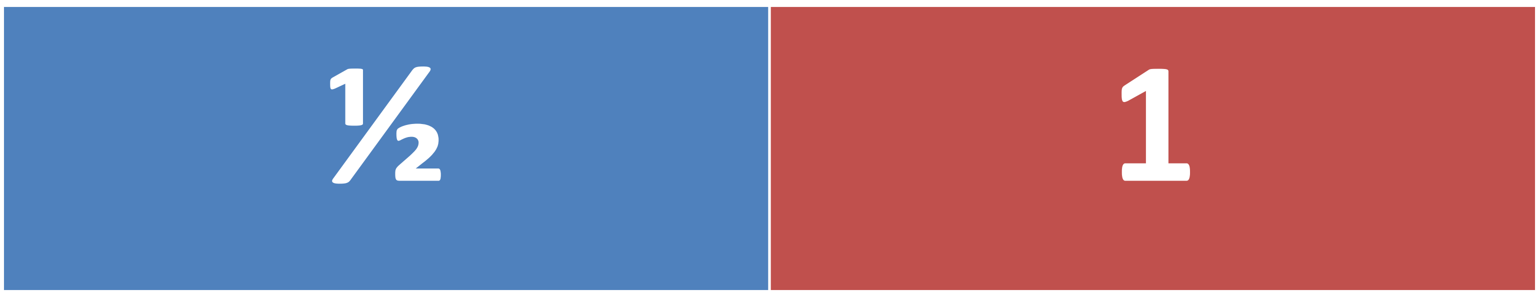
$$M^T W M\, \mathbf{r} = M^T\, W\mathbf{b},$$

where W is a diagonal matrix and w_{jj} is the weight of the jth game.

- Then, we again remove a row, replace it with 1's and 0 on the RHS.

Weighting time

Again, let's break the season in half and count first half games as ½ and second half as 1.



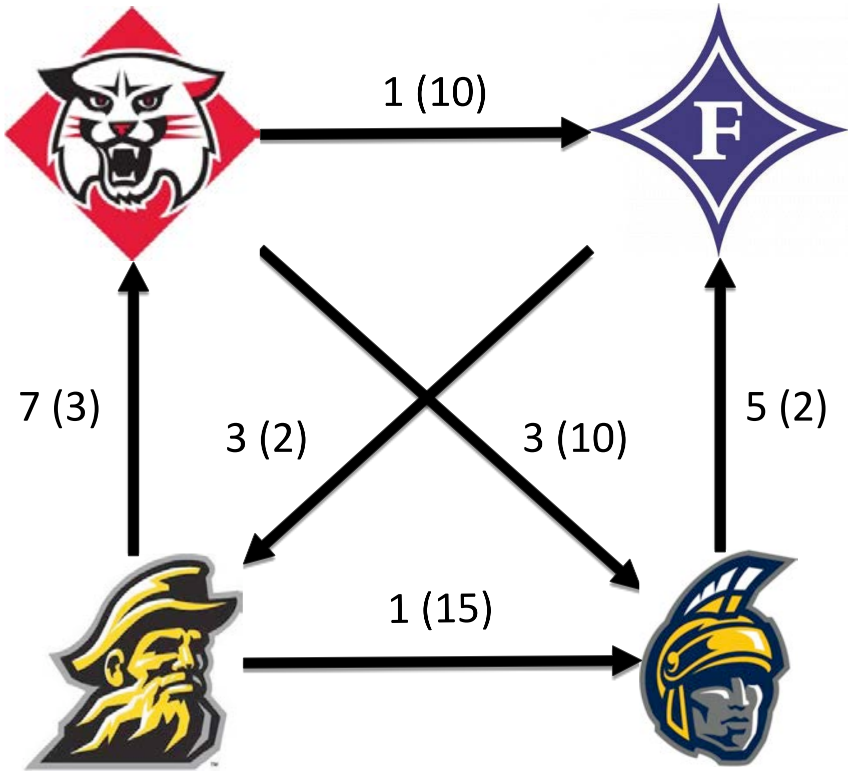
Davidson games

So, we have:

$r_1 - r_2 = 10$ $\quad$ $r_3 - r_2 = 2$

$r_1 - r_3 = 10$ $\quad$ $r_2 - r_4 = 2$

$r_4 - r_1 = 3$ $\quad$ $r_4 - r_3 = 15$



Least Square Matrices

So, the matrices for the least square linear system are:

$$M = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 10 \\ 10 \\ 3 \\ 2 \\ 2 \\ 15 \end{pmatrix}$$

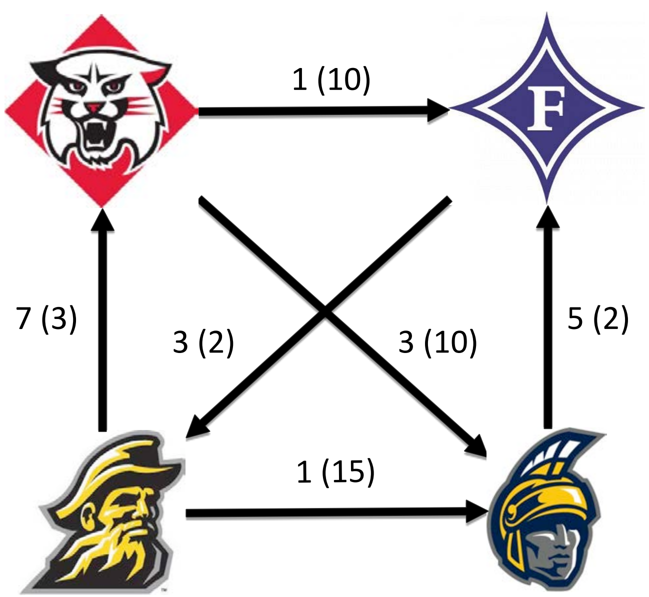
Adding weight

Adding weight, we get:

$$W = \begin{pmatrix} 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 \end{pmatrix}$$

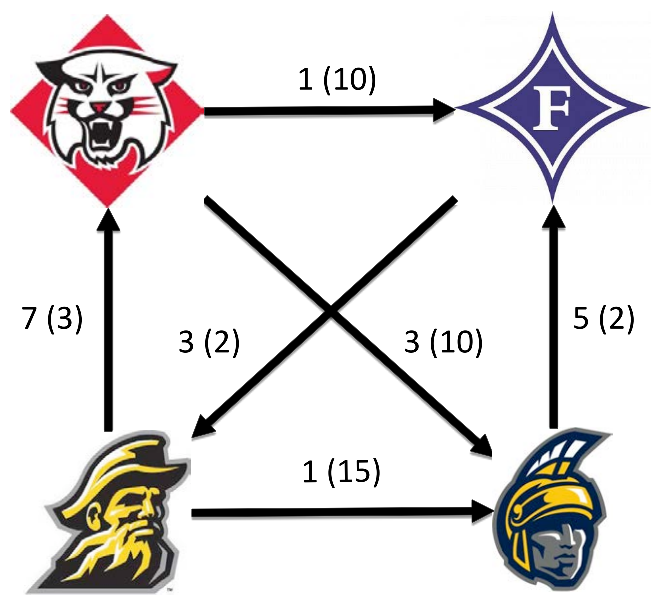
Davidson games

$$\begin{pmatrix} 2 & -0.5 & -0.5 & -1 \\ -0.5 & 2 & -1 & -0.5 \\ -0.5 & -1 & 2 & -0.5 \\ -1 & -0.5 & -0.5 & 2 \end{pmatrix} \begin{pmatrix} D \\ F \\ U \\ A \end{pmatrix} = \begin{pmatrix} 7 \\ -6 \\ -10.5 \\ 9.5 \end{pmatrix}$$



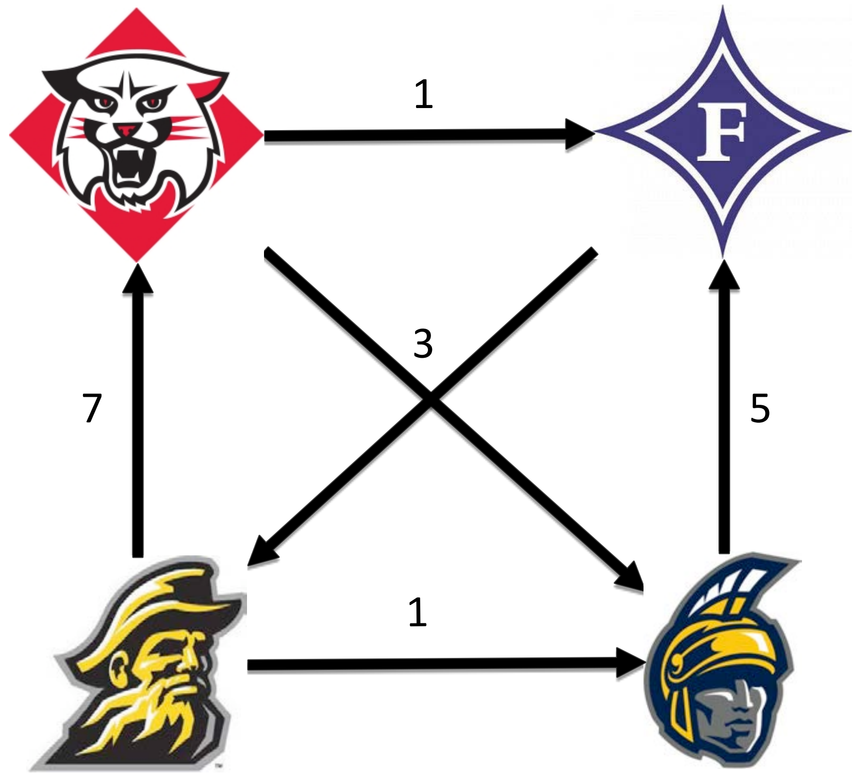
2 sum 2 zero

$$\begin{pmatrix} 2 & -0.5 & -0.5 & -1 \\ -0.5 & 2 & -1 & -0.5 \\ -0.5 & -1 & 2 & -0.5 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} D \\ F \\ U \\ A \end{pmatrix} = \begin{pmatrix} 7 \\ -6 \\ -10.5 \\ 0 \end{pmatrix}$$



Ranking

$$\begin{pmatrix} D \\ F \\ U \\ A \end{pmatrix} = \begin{pmatrix} 3.7083 \\ -3.3750 \\ -4.8750 \\ 4.5417 \end{pmatrix}$$



Ranking

1. App State
2. Davidson
3. UNCG
4. Furman

How do we do?

- We've entered our brackets into the ESPN Tournament Challenge and competed against more than 4 million brackets!
- You get 10 points for each correct choice in the first round and each successive round doubles in the points allotted to a correct prediction.

2009

Method	Percent
Colley – no weight	62
Massey – no weight	88
Colley – bi-weekly	97
Massey – bi-weekly	79
Obama	80
Mike Greenberg (sports analyst)	70
Mike Golic (sports analyst)	43

10 and 11

- In 2010, a student in my math modeling class beat 99% of the close to 5 million brackets.
- In 2011, the best bracket beat most of the sports experts in the ESPN Challenge.

weight in python

Let’s see how we will control weights in our python code.

```
# Set weights for home, away and neutral wins
weightHomeWin = 1
weightAwayWin = 1
weightNeutralWin = 1
segmentWeighting = [1/2,2]

# Will you use weighting?
useWeighting = True
```

Code it!

Let’s see this in python code.

colleyWeightedRanking.ipynb

Code it!

Let’s see this in python code.

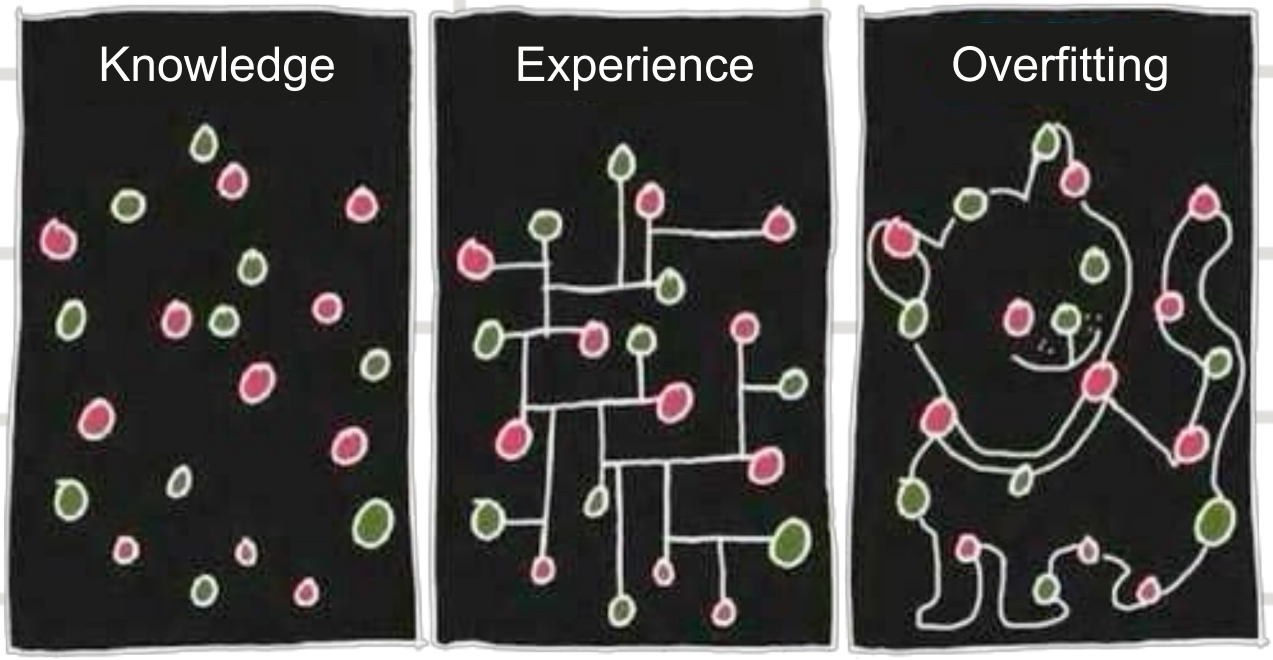
masseyWeightedRanking.ipynb

Be a bracketologist

- You’re ready to create a bracket with math or partially with math.
- Got ideas? You will get a chance to try them and see how you fare this year!

WARNING

beware overfitting



MATHness

http://marchmathness.davidson.edu/

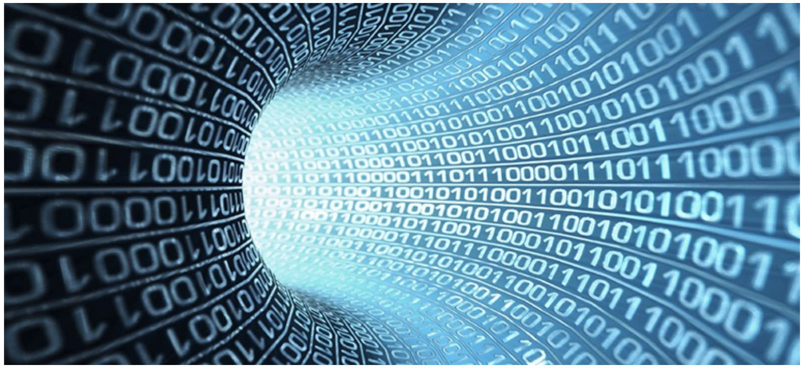
WHICH IS

Homework – 1

- Watch *The Elo Rating System for Chess and Beyond* on YouTube.
- In 2-3 paragraphs describe what you found most interesting about the Elo method and how it compares to the Massey and Colley methods.
- We will briefly cover this ranking method in a later session.

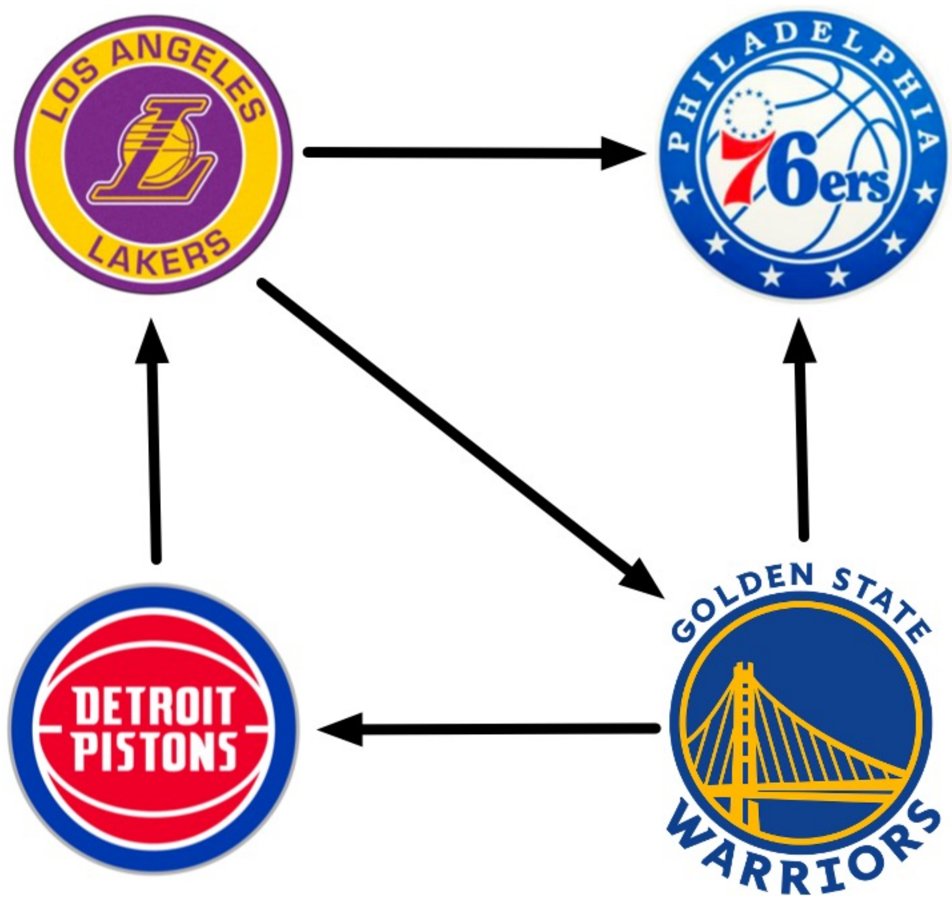
Homework – 2

- Read *If Your Data Is Bad, Your Machine Learning Tools Are Useless*
- In 2-3 paragraphs describe what you found most interesting about this article. If you have thoughts about how it relates to your research, please share them.
- In the next session, we will look at scraping data.



Homework – 3

In this problem, you rank with the Colley method.



Homework – 4

- Run Colley and Massey on 2020-21 regular season NBA data.
- The rankings will differ. So, compare the rankings. You could use:
  - exposition
  - a data visualization
  - a metric to give numerical insight

- Note, this can but does not need to involve any discussion of basketball or the NBA teams.

Homework – 5

Again, share your thoughts on your research.