The Best There Ever Was: Mathematical Models for Ranking NCAA Coaches Across Sports and Time

February 11, 2014

Abstract

The passion of college sports fans extends not just to their team's players but also to the coaches. Which coach really was the best college coach in history? In this study, we develop mathematical models for ranking NCAA coaches with the purpose of definitively identifying the best in history.

For the purposes of this study, we consider data across several sports ranging from 1913 to 2013, using individual game outcome data wherever possible.

The task of modeling is divided in to two distinct sub-problems: ranking teams from the input data and ranking coaches from the team ranks. Each sub-problem is analyzed using several different models, and the system is generalized to facilitate comparisons between sports.

The teams are ranked primarily by minimizing a graph theoretic metric over the graph of the outcomes of all games in each season. This problem is shown to be NP-Hard, and three different algorithms for the problem are implemented and investigated: a randomized heuristic algorithm, the famous PageRank algorithm, and a bounded approximation algorithm.

Once team ranks are established, coach ranks are determined using two different methods. First we measure each coach's skill by the absolute ranking of their teams. Then, we measure each coach's skill by the year-over-year improvement in their teams rank, using a first-principle difference equation. Every observation of coaching skill is treated as a sample from a normal distribution of actual skill, and corresponding confidence bounds are established. Both coach ranking approaches are shown to have merit, and their differences are investigated.

The modeling techniques are assessed in their limiting assumptions, strengths and weaknesses, self-validation, and parameter sensitivity.

Finally, the results of formulations across graph theory and statistics are used to investigate larger trends in college athletics across different sports, different time periods, and different genders. Our models, derived solely from mathematical formulations and raw data, was able to very effectively analyze the system. Our models indicate that within the datasets analyzed, football legend Tom Osborne is definitively the greatest coach in NCAA history.

Contents

1 Introduction

The National College Athletic Association (NCAA) is the unifying body for college sports in the United States. The organization represents a very large set of players and sporting events, including more than 440,000 student athletes on over 18,000 teams in 2011 [1].

Coaches in the NCAA take on great challenges in leading their teams to victory. They must use skills in both strategy and leadership in a constant competition to outdo their opponents. Which coach, in the history of the NCAA across every sport, was the best?

Behind this question lies many mathematical nuances and computational challenges. In this study, we use mathematical models to provide a rigorous, definitive method for ranking the best NCAA coaches of all time.

We begin by identifying data for analysis. Then, we give several models for ranking teams given the raw data. Once the teams have been ranked, we use two different measures of the coaches skill across the teams they coached. Finally, we use these models to present several different kinds of analysis on the data and the system.

2 Problem Statement

This study seeks to identify the **best overall coaches in the history of college athletics** in their respective sports. The search will span all sports, genders, and years.

2.1 Problem Assumptions

In order to make the problem tractable, several key assumptions are made.

- **Coaching skill is measured only by team performace**
  In the public eye, the best coaches are not just the ones who win the most games. They may have inspired the fanbase, taken the ethical high ground, or overcome situational adversity. In these models, we measure coaches solely by their team's (or teams') in-game performances.

- **A coach's skill is constant**
  A more advanced model could assume that a coach's skill rises and fall through his or her career. Here we assume that a coach has a single, constant skill level throughout his or her career (though we do take statstical samples from this value). This assumption is further relevant since the we are seeking the best coach over time, not the best coach for a season.

- **A team's skill is constant in each season**
  We assume that a team is exactly as good at the beginning of a season as it is at the end. This significant assumption, because a good coach likely improves their team constantly, not just between seasons. Nonetheless, it greatly simplifies the task of fitting a coach's skill to data.

3 Data Sources

The process of designing models must begin with considering the data available.

The first challenge to overcome in the process of ranking coaches is the acquisition of raw data. While the results of athletic events are consistently recorded by the NCAA, it is difficult to find unified, accessible datasets, especially from seasons before the internet era.

This study seeks to analyze leagues with varying levels of data availability to assess the models' performance in the absence of complete data. To that end, we study FBS Football, Division I Men's Basketball, and Division I Women's Soccer.

- **FBS Football**
  The NCAA Division I Football Bowl Series (FBS) is by far the most well documented sport in NCAA history. We utilize a complete dataset spanning the past 100 years of football from 1913-2013. The dataset includes the outcome of every FBS football game as well as a nearly complete list of coaches. The data is gathered from Sports-Reference.com, and is already santized to use consistent naming schemes and other conventions [2]. The largest season in the dataset contains roughly 120 teams and 1300 games.

- **Division I Men's Basketball**
  The next most widely documented sport is Division I Men's Basketball. From the 1980-1981 season onward, full data from every game played is available. Before 1980 the dataset does not contain game outcomes but only net win/loss data. A nearly complete list of coachest is given across all years. The data is gathered from Sports-Reference.com, and is already santized to use consistent naming schemes and other conventions [3]. The largest season in the dataset contains roughly 345 teams and 3200 games.

- **Division I Women's Soccer**
  Finding consistent, openly accessibile data for sports other than football and men's basketball is extremely challenging. To develop the models on an interesting sport with less data we study Division I Women's Soccer. Our dataset extends from 2002-2013 and includes only net win/loss outcomes. Coach data is only sporadically available, and is correllated manually from a list of top coaches by win ratio. Data is gathered from NCAA website reports [4, 5]. The largest dataset contains roughly 330 games.

These different datasets fall across the spectrum of data availability. On one side FBS football has 100 years of complete data, while on the other side Women's soccer contains just 11 years of incomplete data. Men's Basketball falls in between with data that is consistent but only includes game outcomes for a portion of the time period.

When deriving and discussing performance of our models, we focus most intensely on football because it represents the most complete modeling task. It is assumed that if these models were used in a more extensive setting, complete data could be obtained from the NCAA. Nonetheless, it is still important to analyze Women's Soccer to assess model performance on incomplete data.

4 Modeling Methodology

We consider the modeling challenge to consist of two main components: first generating a meaningful ranking of team perforamnce, then assessing each coach based on all the teams they led. Since these components are largely disconnected, we consider several different models for each, which are futher elaborated in the subsequent sections.

The general outline of the system is as follows, displayed graphically in Figure 1.

- **Team Skill**
  - **Victory Graph: Edge Inversion Minimization**
    We consider the outcomes of every game in a season to form a graph, and define a metric for how well a ranking corresponds to that graph. We then seek to find the ranking that minimizes that metric (which turns out to be an NP-Hard problem).
    * **Stochastic Heuristic Algorithm**
      We design a simple randomized algorithm which attempts to solve the problem by making optimal decisions whenever they are obvious and random decisions otherwise. The algorithm is run many times and the best result is returned as the ranking.
    * **PageRank Algorithm**
      The famous PageRank Algorithm, designed for ranking webpages, is applied to the graph to generate a ranking.
    * **Approximation Algorithm**
      We demonstrate that this minimization corresponds to the Minimum Feedback Arc Set problem in Operations Research. Since this problem is known to be NP-Hard, a bounded approximation algorithm from the literature is implemented to generate a ranking.
  - **Win/Loss Records**
    Assess each team in a season by its win/loss ratio for the season, and rank them accordingly. This is clearly an inferior method, but can be applied to datasets where individual game data is unavailable.
- **Coach Skill**
  - **Absolute Performance Analysis**
    Assess each coach's skill by the ranking their teams achieved each season. Each season ranking is treated as sample from a true coaching skill distribution for the coach.
  - **Performance Improvement Analysis**
    Assume that a coach's skill is directly expressed in their teams' change in ranking. Assess each coach's skill by their ability to move teams to the top of the rankings and keep them there. Each of these values is treated as sample from a true coaching skill distribution for the coach.

5 Modeling Team Rankings

5.1 Victory Graph Edge Inversion Minimization

When full game-by-game data is available, we model each season as a directed graph $G = (V, E)$ where the vertices V are teams and the directed edges E represent a victory pointing towards the winning team (Figure 2). In the main portion of our study, an unweighted graph with $w_i = 1$ is used. In the context of this model we seek a ranking $R : V \to \mathbb{N}$, which is a one-to-one correspondence labelling the teams in the graph with integer rankings. The best team has $R(t) = 1$.

For a given R, we define an *edge inversion* to occur when for two team $a, b \in V$ there is an edge $e = (a, b)$ but $R(a) > R(b)$ under R. Colloquially, edge inversions are upsets, games where a lower ranked team beats a higher ranked team.

We argue that the ranking of the teams in V should be defined as the ranking R^* that incurs the fewest edge inversions on the victory graph.

$$R^* = \arg\min_R \sum_{\substack{R(v_i)<R(v_j)\\(v_i,v_j)\in E}} w(v_i, v_j) \tag{1}$$

The true ranking will not be unique; in general there can be a set of rankings all of which incur a minmal inversion count. Minimizing this result corresponds to a maximum-likelihood estimate (MLE) of the ranking most likely to generate the observed game results.

Now, we seek to algorithmically find the optimal ranking for these graphs. As is discussed below, this problem is equivalent to the Minimum Feedback Arc Set (FAS) problem. Unfortunately, this problem is both NP-Hard and APX-Hard [7]. Because it is NP-Hard, we cannot solve the program in polynomial time (barring a major advance in Computer Science which is beyond the scope of this study). Because it is APX-Hard, approximate solutions cannot be generated in polynomial time that are guaranteed to lie within some constant factor of the optimal solution. As such, we consider various schemes to give acceptable alternative approximate solutions to the problem.

5.1.1 Stochastic Heuristic Algorithm

Our first approach is to design our own heuristic algorithm to find minimizing rankings. We recognize that because our graphs are relatively small in a computing sense ($|V| < 300$), a randomized algorithm will be able to explore a considerable subset of the solution space.

We implement a simple stochastic algorithm which iteratively ranks the teams from best to worst by applying the following policy:

Find the team $v_i \in V$ with the minimal $|V_w|$, $V_w \subset V$, $\forall v_j \in V_w \quad (v_i, v_j) \in E$, then rank v_i and remove it and all edges connecting to it from G. (*chose a team with the fewest losses to lower ranked teams*). Select a v_i at random if multiple meet this criteria.

This algorithm is restarted 1000 times, and the ranking R_h which incurs the fewest minimzation is reported.

5.1.2 PageRank

The PageRank algorithm rose to fame after being developed by Google's Larry Page and Sergey Brin for ranking webpages [8]. PageRank simulates a user navigating the internet, going from one webpage to another. In this context, webpages are vertices and hyperlinks are edges. At every instant in time, the user can go

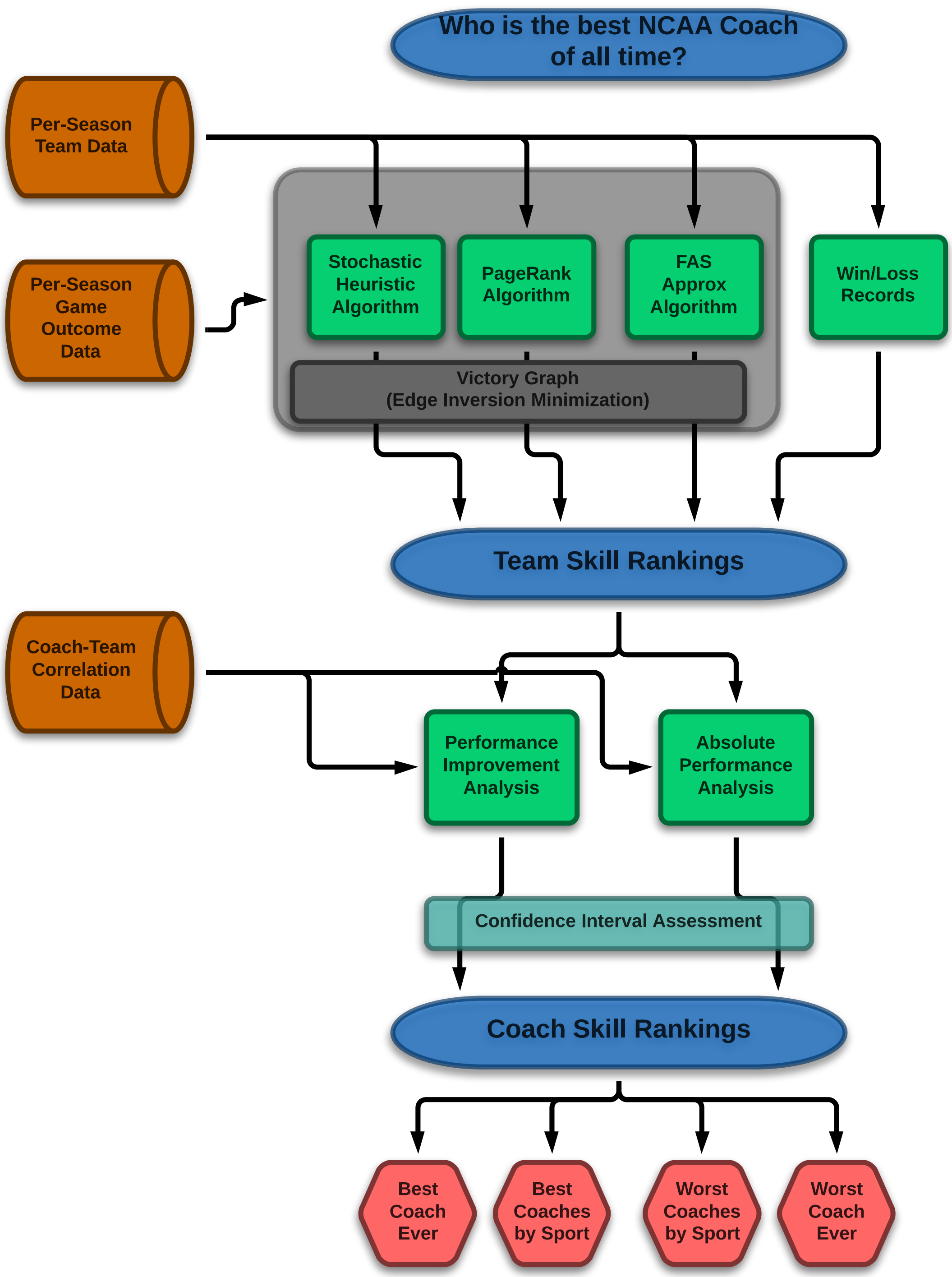
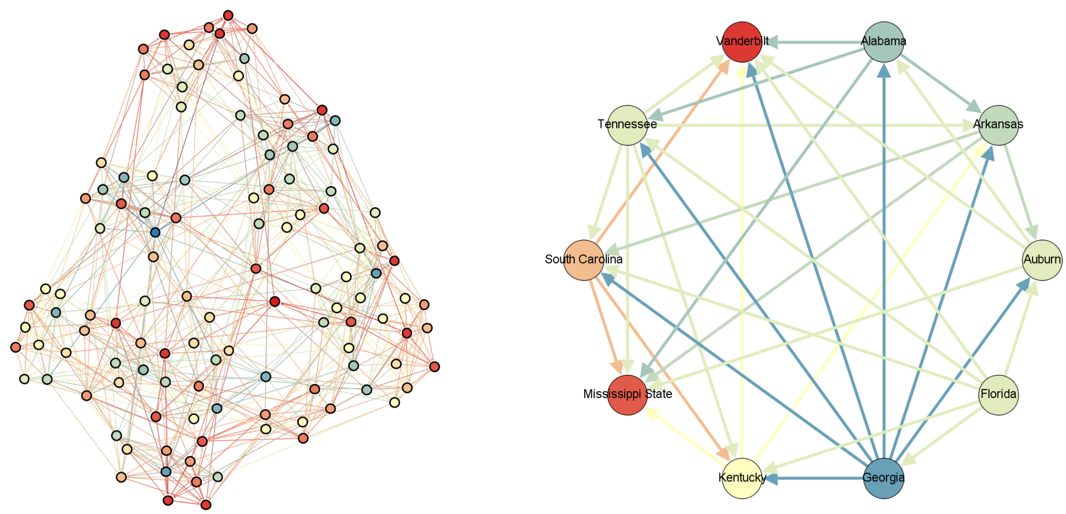


Figure 1: The general structure of the coach ranking model. Any one of the team ranking methods could be mixed with any one of the coach skill generation methods.



(a) All NCAA Division 1 FBS teams. The clusters correspond to the league conferences.

(b) The SEC conference teams and in-conference games.

Figure 2: Victory graphs for the 2002 NCAA FBS Football season. Each directed edge represents a victory in a game. Teams are colored corresponding to their overall FBS record for the season, blue being the best record and red being the worst. Visualizations created with Gephi [6].

from its current page to one of the neighboring pages (via a link), or can go to any page in the set of all pages.

A PageRank implementation is applied to the graph, and a ranking R_p is derived from the weights that PageRank assigns each vertex [9]. The inversions for this ranking are then measured to asses validity.

5.1.3 FAS Appoximation Algorithm

Under further examination and research, it was discovered that this problem formulation is mathematically identical to the Minimum Feedback Arc Set (FAS) problem in Operations Research. Current research efforts are only recently reporting analyses of the problem, but there has been some effort in to provide practical methods for approximating solutions.

Demetrscu et. al. state a polynomial time approximation algorithm (Algorithm 1) for the problem [10]. Because the problem is APX-Hard, no constant bound can be placed on the ratio of approximation, but it is known that the ratio scales asymptotically with the length of the longest simple cycle in the graph.

This algorithm is implemented and applied to the datasets to generate a ranking R_a. In the analysis below we show that this algorithm consistently has the best performance of our implemented methods, and it **is selected as team ranking method for generating coach ranks**.

5.2 Team Ranking by Win/Loss Ratio

The most basic quantitative measure of team skill is calculated directly by analyzing the ratio of a team's wins to its losses and ranking teams according to this ratio. This method is clearly inferior to graph-based ranking, as it has no concept of "strength of schedule", but can be applied to teams for which game-by-game data is unavailable.

Algorithm 1 Approximating a minimal feedback arc set of a weighted directed graph. Described in [10].

procedure FAS($G = (V, E); w : E \to R^+$)
  $F \leftarrow \emptyset$ ▷ Phase 1
  while $(V, E \setminus F)$ is not acyclic **do**
    Let C be a simple cycle in $(V, E \setminus F)$
    Let (x, y) be a minimum weight edge in C and let ϵ be its weight
    for each $(v, w) \in C$ **do**
      $w(v, w) \leftarrow w(v, w) - \epsilon$
      if $w(v, w) = 0$ **then**
        $F \leftarrow F \cup \{(v, w)\}$
      end if
    end for
  end while
  for each $(v, w) \in F$ **do** ▷ Phase 2
    if $(V, (E \setminus F) \cup \{(v, w)\})$ is acyclic **then**
      $F \leftarrow F \setminus (v, w)$
    end if
  end for
return F ▷ F is the feedback arc set
end procedure

The win/loss ratio ranking can be calculated for any season in our dataset. For the seasons which also have game outcome data, an inversion count R_r is calculated and compared the graph-based ranking methods.

5.3 Method Comparisons

All of these methods are applied each of the datasets (where possible), and the results are reported in the inversions table for comparisons.

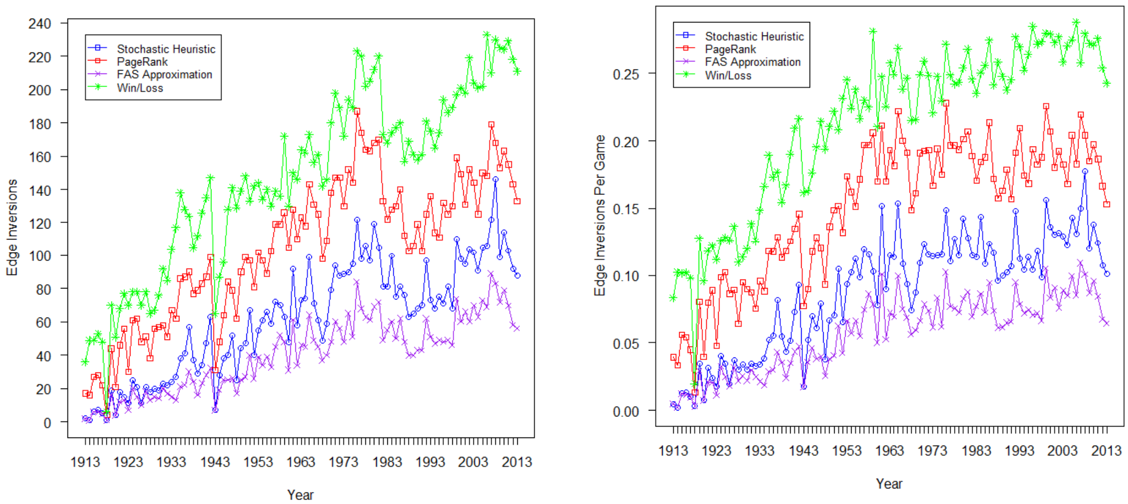
Over both the football and basketball datasets, it was observed that the Demetrescu approximation algorithm produces rankings with the fewest inversions, with the randomized algorithm close behind and PageRank a distant third.

This result is of interest because several recent studies have applied PageRank to sports data with the assumption that it naturally extends from websites to sports [11, 12]. We believe that this method underperforms in the inversion metric because hyperlinks when ranking web pages are not directly equivalent to victories when ranking sports teams.

	Randomize Heuristic R_h	Pagerank R_p	Approx R_a	WinLoss R_r
FBS Football	65.63	104.95	40.47	147.83
Men's Basketball	585.29	681.74	579.74	833.53
Womens's Soccer	N/A	N/A	N/A	N/A

6 Modeling Coach Rankings

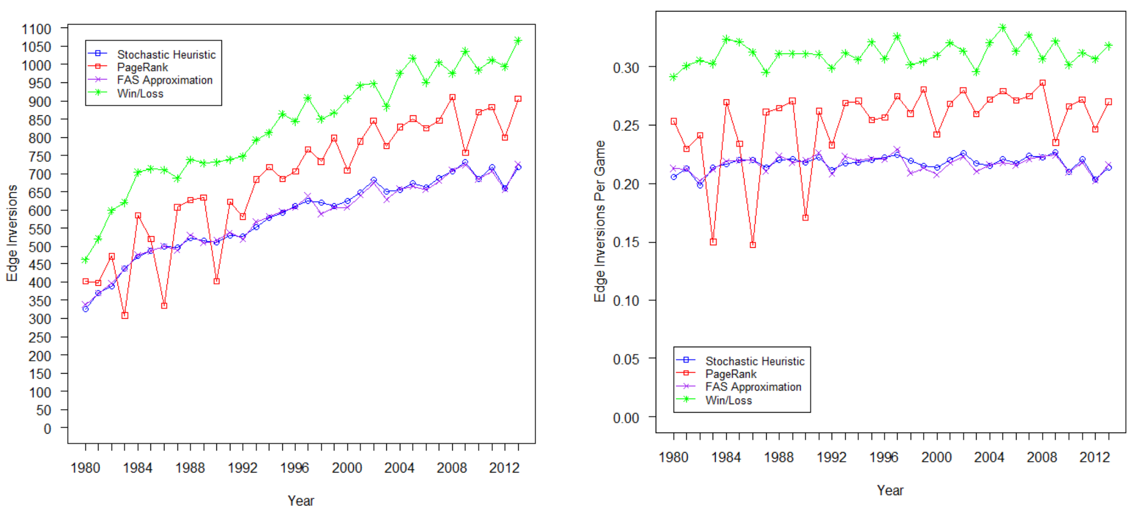
Given rankings for each team in each season, we now must assess the performances of coaches across their careers. Rankings within each season are normalized into the interval $[0, 1]$, where 1 is the best according to



(a) The absolute number of edge inversion.

(b) The edge inversions per game played in the league, useful for comparing across years.

Figure 3: Edge inversions for the rankings found from each method in the FBS Football data.



(a) The absolute number of edge inversion.

(b) The edge inversions per game played in the league, useful for comparing across years.

Figure 4: Edge inversions for the rankings found from each method in the Men's Basketball data. Analysis is only possible after 1980 when game outcome data is available.

$$r_{norm} = \frac{N_{teams} - r_{raw}}{N_{teams} - 1} \tag{2}$$

to enable comparison across sports and seasons with different numbers of teams.

Two different methods are considered for generating coach ranks from team ranks. Each method represents a different formulation of what coaching skill means. Both methods are implemented on our datasets and used to generate all-time coach rankings.

6.1 Ranking Coaches by Absolute Performance

The most direct method to rank each coach is to examine the performance of each of his or her teams.

> This method corresponds to the most straightforward definition of coaching skill, that a coach's skill is measured by the performance of his or her teams.

We assume that each coach has a skill C, constant throughout his or her lifetime. **Each season is then a sample from a normal distribution about the coach's true skill.** We further assume that all coaching distributions have the same deviation.

Given the set of team performances associated with each coach, we use the pooled variance statistic (from all the coaches) to fit a normal distrbution for the coach. **The reported coach skill is then the lower bound of the 99% confidence interval for the coach's distribution.** This reflects the idea that if a coach is active for just a single season, and has an excellent season, it is more likely that he or she got lucky than that he or she is the best coach of all time.

6.2 Ranking Coaches by Performance Improvements

We additionally the consider the idea that the skill of a coach could be reflected by a measure more complex than just the absolute rank of their teams.

> This method represents the idea that a coach's skill is reflected not in the performance of their teams, but in the change in performance of their teams. The best coaches move their teams to the top and keep them there.

Mathematically, this is accomplished by modeling each teams year-to-year improvement with a difference equation. We consider a first-order model with two components to the change in rank. First, the coach has a constant increasing or decreasing effect on the teams rank. Second, there is a linearly scaling tendency for teams at the top to move downward and teams at the bottom to move up. We justify modeling this trend as linear because the n'th ranked team has n teams below it competing for its spot. This tendency is assigned an known magnitude f, which we expect to be negative. The difference equation then becomes:

$$r_i - r_{i-1} = C + fr_{i-1} \tag{3}$$

where r_i is the normalized rank of the team in a given year, C is the skill of the coach of the team in that year, and f is constant. The difference equation involves some significant assumptions which are elaborated below.

We consider f be a single constant for each sport, which in a qualitative sense describes how strongly a team at the top of the ranking is getting forced downward, relative to a team in the middle.

We fit f for each sport with a simple linear regresion, made possible by the assumption that C is independent of r. This regression is shown in Figure 5.

Once f has been calculated for each sport, all terms but C are known in Equation 3 and we calculate C for each team-season. These values are then given the exact same stastical treament as in the previous section on Absolute Performance, where each is viewed as a sample from a normal distribution and the lower bound of a 99% confidence interval is taken to find final results.

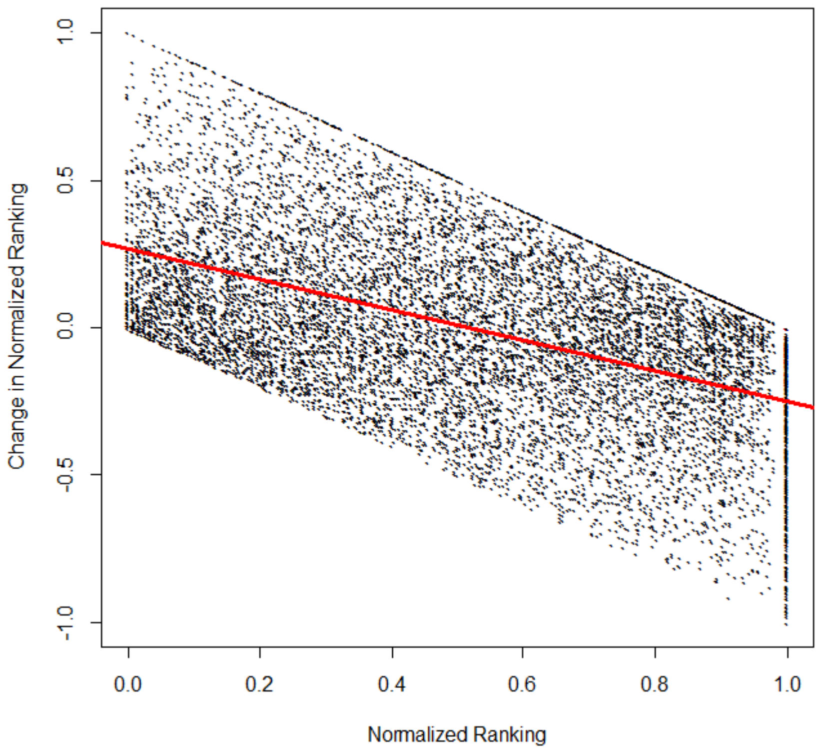


Figure 5: The change in rank of FBS football teams by their rank the previous season. Data spans the past 100 years of FBS football. The fitted line has slope $f = -0.52027$ with $R^2 = 0.2601$, which inidicates significant trend. The slope f gives the magnitude of the tendency for bad teams to get better and good teams to get worse. The visible lines of points along the boundaries of the data correspond to motion of the top ranked teams which all have $r = 1$.

It should be noted that the C values calculated here are dimensionally different from those calculated in the absolute analysis and not directly comparable. The C values from absolute anaylsis have the dimension of *rank*, whereas these values have dimension *change of rank*.

7 Model Analysis

7.1 Model Data Limitations

Certain limitations must be made to apply the model to our datasets. Some of these assumptions are immaterial but important to note, others represent weak points in the model. All are properties of our analysis, not the system, and could be improved upon with further effort.

- **Analyze only "game-style" sports**
  This model only applies to sports that consist of a series of two-team matchups. Some NCAA sports, such as Swimming or Wrestling, are played in other formats such as many-team competitions and individual-athelete competitions.

- **Discard ties**
  When treating game results as a victory graph, only wins and losses are relevant. Games with a tie reported as the result are discarded.

- **Discard inter-division games**
  When analyzing a given division we consider only games played within the division. In example, if a Division I plays a Division II team, that game is discarded when modeling either division. This could be resolved by modeling the entire sport rather than each division separately.

7.2 Model Analysis Assumptions

In order to leave tractable mathematical problems, we must limit the system with simplifying assumptions, some of which are more signifcant than others.

- **Assume same deviation for all coach skill distributions**
  In reality some coaches are likely more consistent then others, and season samples from their skill distribution would have a tigher deviation. Instead, we assume that the deviation of skill observed each season is constant for all coaches. This make the problem of fitting a distribution to a coach with a single season well-defined, as we only need to fit a mean.

- **Relative performance difference equation - first-order**
  The relative performance measure is quantified by Equation 3, which involves several significant assumptions. First, it assumes the relationship is first-order, though it likely has higher-order effects.

- **Relative performance difference equation - constant C**
  Equation 3, which describes the change in performance for each team, has a constant term for C, the coaching skill. This implies that coaching skill is independent of r, when in reality we can clearly state that better teams have better coaches. This assumption is needed here to simplify fitting f, but likely introduces significant bias in to the model and should be elimiated in future work.

7.3 Model Weaknesses

- **Difficulty balancing confidence intervals**
  If too wide of a confidence bound is used, the model will over-emphasize coaches that have been coaching for a long time, whereas narrow bounds favor coaches who had a single season of greatness. This is bound is a subjective choice by the modeler and should be elimiated if possible.

- **Interesting results require interesting data**
  While it's very likely that, given simple win-loss data, the model will generate some obvious top picks, the more interesting results that are less obvious will likely be missed. This is due to the reliance upon minimizing inversions, the model loses much of its value without game outcome data.

	95%		99%	
Rank	Coach	Combined Rank	Coach	Combined Rank
1	Urban Meyer	5	Tom Osborne	2
2	Bob Stoops	7	Bear Bryant	8
3	Les Miles	11	Joe Paterno	9
4	Nick Saban	12	Frank Leahy	10
5	Pete Carroll	13	Steve Spurrier	11

Table 1: Comparison between results with different confidence intervals

7.4 Model Strengths

– **Sport independent**
One of the beauties of the model is that coaches from any sport can be compared as long as the sport is one in which matches are played between two teams.

– **Flexible with data qualiity**
We were seamlessly able to incorporate rich datasets alongside sparsely populated game data. While the results may suffer in accuracy due to poor data, the model remains consistent.

– **Unreliant upon subjective measures of quality**
Most rankings involve some form of human interaction, whereas the proposed ranking system is based purely upon objective data.

8 Validation and Verification

8.1 Sensitivity analysis

Due to the complex nature of the sports datasets, we do not attempt to modify the these data points to observe the effect on the model. While this could yeild significant insights in to the system and model, it is an extremely complex process and judged beyond the scope of this study.

However, we can perform a sensitivity analysis on the confidence level for coach skill used.

We see that if too narrow of a confidence bound is used, coaches who had a single good season are ranked disproportionally high. If too wide of a bound is used, the model simply ranks coaches who coached the most seasons. This issue most be carefully considered in selecting a confidence found.

9 Coach Analyses

9.1 FBS Football

With the richest dataset, FBS Football yields our most accurate result. We were able to obtain every team's game-by-game record from 1913-2013 as well as the head coach of the team at the time. This allowed us to get consistent statistics and rankings of each team throughout the entire time range.

9.2 Division I Men's Basketball

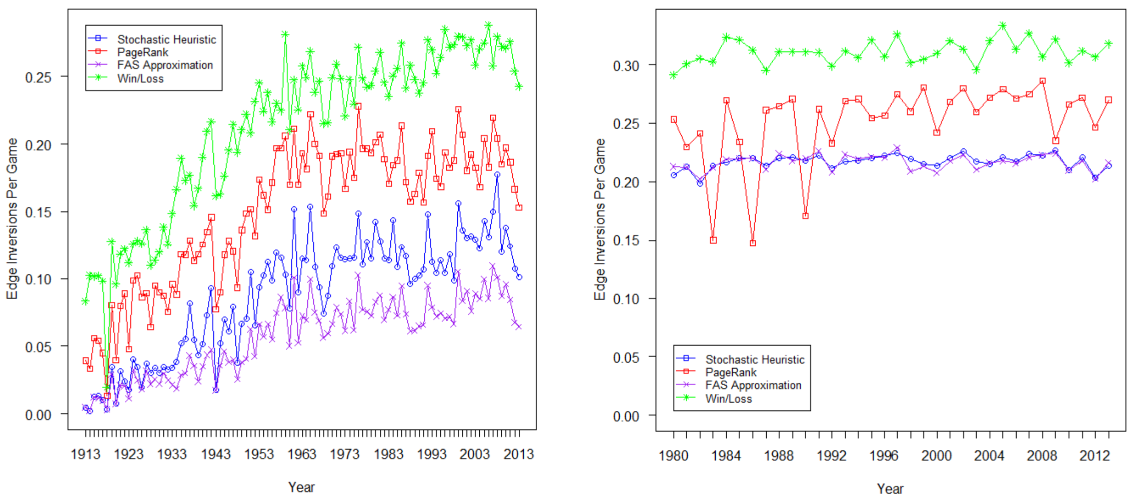
Men's Basketball only yielded rich data from 1980-2013, so we were forced to improvise and obtain simple win-loss records of teams from 1913-1980. Due to the nature of the data, it is not possible to perform inversion counts on years where we don't have game-specific data. Despite this, the results still seemed quite accurate.

9.3 Division I Women's Soccer

Women's Soccer had the largest data deficiency. With only 12 years of simple win-loss data and mostly missing coach records, we weren't extremely hopeful about getting accurate or interesting results. That said, the coaches that made the top 5 were all extremely notable and talented.

9.4 Change in Coaching Over Time

The edge inversion analysis described above yields a subtle yet insightful method for understand how coaching has changed over time. The plot of the rate of edge inversions per game is reproduced in Figure 6, which gives a measure of how "volatile" the league is, a higher number indicates more upsets.



(a) The edge inversions per game played in FBS Football over all 100 years.

(b) The edge inversions per game played in Men's Division I Basketball over the years data is available.

Figure 6: Edge inversions per game played, which allows comparison over time. As time progresses, college football has become significantly more volatile while basketball has not. This means there are more upsets, and coaches much coach a higher fraction of "big games".

There is a clear trend where the football has an inreasing edge inversion rate over time, while the basketball edge inversion rate stays relatively stable, across all methods. This is evidence that **FBS football has become a more competitive sport over the available dataset, while Men's Basketball has not changed.**

This has a significant effect on the nature of coaching because a coach in a more volatile league has to play more close games, as opposed to a coach in a league where most games strongly favor one team or the other.

9.5 The best coach of all time

By comparing each coach, independent of sport, by the sum of their Absolute Rank and Performance Improvement Rank, we can give an un-biased answer to the question, "Who is the best coach of all time"?

From the results posted above/below?, we can see that **Tom Osborne has the highest cumulative raw score and rank. By our model he is definitively the best coach of all time in the analyzed datasets.**

Rank	Coach	Absolute Rank (raw)	Performance Improvement Rank (raw)	Years as a Coach
1	Tom Osborne	1 (0.8059)	1 (0.3625)	25
2	Bear Bryant	6 (0.7475)	2 (0.3584)	38
3	Joe Paterno	4 (0.7504)	5 (0.3484)	45
4	Frank Leahy	3 (0.7591)	7 (0.3415)	13
5	Steve Spurrier	8 (0.7348)	3 (0.3510)	24

Table 2: Top 5 FBS Football Coaches

Rank	Coach	Absolute Rank (raw)	Performance Improvement Rank (raw)	Years as a Coach
1	Mark Few	1 (0.8344)	3 (0.3521)	15
2	Adolph Rupp	2 (0.7932)	2 (0.3571)	41
3	Jerry Tarkanian	3 (0.7746)	4 (0.3429)	27
4	Bill Self	6 (0.7414)	6 (0.3374)	17
5	Denny Crum	8 (0.7374)	5 (0.3387)	30

Table 3: Top 5 Division I Men's Basketball Coaches

Rank	Coach	Absolute Rank (raw)	Performance Improvement Rank (raw)	Years as a Coach
1	Anson Dorrance	1 (0.8309)	2 (0.2027)	34
2	Randy Waldrum	5 (0.8060)	4 (0.1891)	15
3	Chris Petrucelli	9 (0.7847)	1 (0.2142)	23
4	G. Guerrieri	7 (0.7926)	3 (0.1960)	22
5	John Daly	6 (0.8005)	5 (0.1871)	26

Table 4: Top 5 Division I Women's Soccer Coaches

10 Conclusions

Through a survey of three sports: one with dense amounts of data, one with mixed data, and one with sparse data, we investigated various models for their effectiveness in the distinct tasks of ranking teams within a season as well as ranking coaches within their career.

Unsurprisingly, we found that the ability of a model to describe the system was heavily dependent on the data available. The most effective method for assigning team ranks, victory graph edge inversion minimization, is only possible when the outcome of each game within a season is known.

The approximation of a minimal feedback arc set has proven to be a very effective method of ranking teams, outperforming PageRank as well as other naive approaches in mimizing edge inversions. Although the formal problem is NP-Hard and APX-Hard, the approximate algorithm generated solutions entirely sufficient for continued analysis and innovative investigation.

We consider two different methods for assessing coach performance, and demonstrate that both have value in measuring different kinds of coaching effectiveness. The absolute ranking method gives the most obvious measure of how well the coach's teams performed. While the relative ranking difference equation measures the more subtle skill of improving team performance, it is significantly sensitive to the confidence level imposed. By creating a composite model from both of these metrics, we are able to rank coaches in a way that captures many of the nuances of their performance.

All of these models together permit considerable analysis of NCAA atheletics. The most direct result of our model is definitely ranking Tom Osborne as the greatest NCAA coach of all time. We are also able to draw deeper conclusions about the nature of NCAA athletics overtime, most significantly the fact that FBS Football has become more competitive over our dataset, while Men's Division I Basketball has not.

Even with a small subset of historical NCAA data, were are able to present a novel analysis of college sports using a variety of mathematical techniques. The methods presented here hold great value in the future modeling of college athletics over increased datasets.

References

[1] NCAA.com. Ncaa participation rates going up. Web Article, November 2011.

[2] Justin Kubatko. College football at sports-reference.com. Online, February 2014.

[3] Justin Kubatko. College basketball at sports-reference.com - college basketball statistics and history. Online, February 2014.

[4] NCAA. Women's soccer coaching records. Online, Februrary 2014.

[5] NCAA. Division i women's soccer statistics. Online, Februrary 2014.

[6] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: An open source software for exploring and manipulating networks, 2009.

[7] Claire Kenyon-Mathieu and Warren Schudy. How to rank with few errors. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 95–103. ACM, 2007.

[8] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: bringing order to the web. 1999.

[9] Joshua O'Madadhain, Danyel Fisher, Padhraic Smyth, Scott White, and Yan-Biao Boey. Analysis and visualization of network data using jung. *Journal of Statistical Software*, 10(2):1–35, 2005.

[10] Camil Demetrescu and Irene Finocchi. Combinatorial algorithms for feedback problems in directed graphs. *Information Processing Letters*, 86(3):129–136, 2003.

[11] Anjela Govan, Carl D Meyer, and Russell Albright. Generalizing google's pagerank to rank national football league teams. 2008.

[12] Dorit S Hochbaum. Ranking sports teams and the inverse equal paths problem. In *Internet and Network Economics*, pages 307–318. Springer, 2006.