

# College Coaching Legends

## 2014 MCM

### Problem B

Lorena Maxwell  
Adam Michael  
Yunhui Shi

September 2014

## 1 Problem

College sports are a hotly debated topic. Every years, millions of people speculate, analyze and forecast thousands of games involving thousands of athletes. One major factor in a collegiate sports team's success is the guidance of a head coach. Analyzing the impact of a coach is more difficult than that of a player because almost all of his effects are intangible, however there is no doubt that these effects can make or break a team. For this and several other reasons, comparing coaches is very difficult. Our team was tasked with constructing and analyzing a model to determine the best all-time college coaches.

## 2 Summary

Comparing collegiate coaches requires the consideration of many factors, several of which are highly subjective. As such, no one approach to this problem can accurately determine the best coach. Instead, we approached the problem from several different angles and looked for patterns and similarities among the result sets. Our goal is to design a method to rank the top coaches in any collegiate sport. For the purpose of proving the model, we will analyze three collegiate sports and determine the top 5 coaches from each. The three sports, Men's Football, Men's Basketball and Men's Baseball, were selected due to large amount of publicly available data for each.

Our total approach to this problem is as follows:

- Collect as much data on all collegiate coaches for each of the selected sports as is publicly available
- Construct a general model that can be used to directly compare any two coaches in a given collegiate sport
- Apply the constructed direct model to our data sets
- Construct a general model to address subjectivity in rankings that allows coaches to be ranked without needing to directly compare every pair of coaches
- Apply the constructed indirect model to our data sets
- Identify common high-ranking coaches across multiple models

### 3 Data Collection

Data was collected on the following sports:

- basketball
- football.

Sports Reference<sup>1</sup> was used to collect statistics about documented coaches. The data included numbers for the number of years having coach, number of wins and losses, percentage of wins, and number of championships (in their respective sport) having won.

### 4 Direct Method

The simplest model we created was to directly compare coaches using a scaled weight for their win percentages. This naive approach will later be compared to more complicated models. We hope to find whether an involved model is necessary, or if win percentages are a good enough estimator.

The simple model is as follows:

$$w_{wins} = \frac{c_{wins}}{W_{wins}}$$
$$I = w_{wins} \cdot p_{wins},$$

where  $c_{wins}$  is the number of games won by the coach,  $W_{wins}$  is the total number of games won by all coaches,  $p_{wins}$  is a coach's win percentage.  $w_{wins}$  is the scaled weight for each coach's win percentage and this is multiplied by the win percentage to create the index  $I$ . The coaches were then ordered from highest to lowest, with the coach having the largest index representing the best coach for the model.

This method favors coaches who have won a lot of games throughout their career. The weighted approach was needed because when directly comparing win percentages, coaches who only played a few games could seem better than coaches who had played a great amount more. For example, a coach who played three games and won all of them appeared superior than a coach who had played eight hundred games, but had only won eighty percent of them.

### 5 Indirect Method

Comparing coaches is inherently subjective and often times it is difficult to directly compare two coaches. However, sometimes the superior coach is clear. In our indirect model, we identify clear coaching victories but defer the comparison of two similar coaches to the model. We construct a graph of clear victories and use this graph to make the murkier comparisons. Specifically, we use Google's PageRank algorithm to identify the most significant vertex in the graph according to a metric of eigenvector centrality.

A  $n$  by  $n$  matrix was used, where  $n$  was the total number of coaches, to hold the outputs of different functions when information about each coach was fed into it. Detailed descriptions of the functions are given below.

#### 5.1 Comparison Functions

Our ranking model relies on some kind of comparison function. This function determines if one coach is definitively better than another or if it is too close to call. Clearly, any definition of this function is subjective, so to accurately portray the status of each coach we will consider several different comparison functions inside of our analysis framework.

---

<sup>1</sup><http://www.sports-reference.com/>

### 5.1.1 Championship-based

With this comparison function, we prioritize the number of championships and championship appearances. Specifically, if one coach has several more championship victories then that coach is the clear superior. If they have a similar number of championships, but one has several more championship appearances, then that coach is the clear superior. If they have a similar number of championships and championship appearances, but one had a significantly shorter career, then he is the clear winner. If none of these criteria are true, then there is no definite winner and the comparison is inconclusive.

Rank	Football	Basketball	Baseball
1	Bear Bryant	John Wooden	Coach A
2	Bobby Bowden	Adolph Rupp	Coach A
3	Tom Osborne	Phog Allen	Coach A
4	Joe Paterno	Jim Boeheim	Coach A
5	Mack Brown	Dean Smith	Coach A

Table 1: Best college coaches according to the championship-based indirect method

### 5.1.2 Publicity-based

With this comparison function, we consider the public opinion of the coaches.

Rank	Football	Basketball	Baseball
1	Bobby Bowden	Jim Boeheim	-
2	Woody Hayes	Bob Knight	-
3	Joe Paterno	Dean Smith	-
4	Bear Bryant	Denny Crum	-
5	Lou Holtz	Lute Olson	-

Table 2: Best college coaches according to the publicity-based indirect method

### 5.1.3 Extrapolation

This comparison method accounts for differences in lengths of coaching careers. When comparing two coaches of different career lengths, we use a least squares regression to predict the next several years of the less-tenured coaches career and then compare the number of wins of the coaches.

Rank	Football	Basketball	Baseball
1	Walter Camp	Brad Underwood	-
2	John Hartwell	Brad Stevens	-
3	Samuel Thorne	Shaka Smart	-
4	Malcolm McBride	Craig Neal	-
5	Frank Trutico	Gary Cunningham	-

Table 3: Best college coaches according to the extrapolation indirect method

This comparison function produced less than ideal results. The model is dominated by coaches with extremely short careers who happened to win an unusually high percentage of games in their few years in the league.

#### 5.1.4 Linear Combination

Rank	Football	Basketball	Baseball
1	Bobby	A	-
2	Bear	A	-
3	Tom	A	-
4	Joe	A	-
5	Mack	A	-

Table 4: Best college coaches according to the linear combination indirect method

## 6 Results

## 7 Conclusion

## 8 Bibliography