WASP:

An Algorithm for Ranking College Football Teams

WASP:
AN ALGORITHM FOR RANKING COLLEGE FOOTBALL TEAMS

BY

JONATHAN EARL, B.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Science.

Master of Science (2016)	McMaster University
(Mathematics & Statistics)	Hamilton, Ontario, Canada

TITLE:	WASP: An Algorithm for Ranking College Football Teams
AUTHOR:	Jonathan Earl B.Sc. (Mathematics) Redeemer University College, Ancaster, Canada
SUPERVISOR:	Dr. Paul D. McNicholas
NUMBER OF PAGES:	x, 69

Lay Abstract

The problem of ranking is a ubiquitous problem, appearing everywhere from Google to ballot boxes. One of the more notable areas where this problem arises is in awarding the championship in American college football. This paper explains why this problem exists in American college football, and presents a bias-free mathematical solution that is compared against how American college football awards their championship.

Abstract

Arrow's Impossibility Theorem outlines the flaws that effect any voting system that attempts to order a set of objects. For its entire history, American college football has been determining its champion based on a voting system. Much of the literature has dealt with why the voting system used is problematic, but there does not appear to be a large collection of work done to create a better, mathematical process. More generally, the inadequacies of ranking in football are a manifestation of the problem of ranking a set of objects. Herein, principal component analysis is used as a tool to provide a solution for the problem, in the context of American college football. To show its value, rankings based on principal component analysis are compared against the rankings used in American college football.

Acknowledgements

I'm thankful for my supervisor Dr. Paul D. McNicholas for providing scholarly guidance and financial support for my research. I also must thank my parents and sister Joanna for their love and support.

Notation and abbreviations

PCA Principal Component Analysis.

NCAA National Collegiate Athletic Association.

FBS Football Bowl Subdivision.

FCS Football Championship Subdivision.

AP Associated Press.

SOS Strength of Schedule.

SEC Southeastern Conference.

ACC Atlantic Coast Conference.

Pac 10 Pacific 10 Conference.

AQ Automatically Qualifying.

BCS Bowl Championship Series.

CFP College Football Playoff.

FDW First Degree Wins.

FDL First Degree Losses.

SDW Second Degree Wins.

SDL Second Degree Losses.

CDF Cumulative Distribution Function.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The problem of ranking is a problem that arises in many different areas. One of the most well known ranking problems and solutions is the Google PageRank algorithm [17]. In general, a concise description of the problem of ranking is "given a set of n observations $X = \{\vec{x}_1, \vec{x}_2, ..., \vec{x}_n\}$, where each observation $\vec{x}_i$ has j characteristics, $\vec{x}_{i,1}, \vec{x}_{i,2}, ..., \vec{x}_{i,j}$, produce an ordering of the observations based upon their characteristics". While it is good to have a *ranking* of the objects, having a *rating* for the objects is more informative. A *ranking* simply assigns each object a positive integer from $1, 2, ..., n$. In contrast, the *rating* of an object can take on any numeric value, although the domain could be restricted to $\mathbb{R}^+$ or $[0, 1]$ to ease interpretation. One obvious issue that arises when attempting to order objects is that there are $n!$ possible orderings and determining a "good" ordering is a difficult task. This means that an objective standard which can be used for validation of the proposed ordering does not generally exist. Therefore, intuition is usually required to differentiate a "good" ordering from a "bad" ordering.

One approach is to consider the observations in terms of *games* between the observations with definitions for *wins* and *losses*. For the purpose of this thesis, terms such as *win* and *loss* will have the conventional, everyday meaning, except where it is explicitly stated that they are being given an alternative definition. When considering the number of wins and losses that an object has, they will be expressed as $w - l$ to indicate that the object has w wins and l losses. In addition to Google ranking web pages, another context in which ranking is very important is determining the top teams in American college football. As will be seen later, the top teams are determined by polls. These polls are voting methods where the voters are media members and football coaches. Arrow's Impossibility Theorem [1] illustrates the inherent problem of ranking via a voting method. The theorem assumes that a pool of voters are tasked with producing a ballot ranking all n objects, and then the ballots are aggregated together to produce a final ranking for the group. The theorem that considers the existence of a ranking system that meets four criteria:

- *unrestricted domain*: a ballot is allowed to be any of the $n!$ orderings.
- *independence of irrelevant alternatives*: when ranking some subset of the n objects, the ranking of the subset is preserved when considering all n objects. I.e., if a voter compares objects A and B and ranks A ahead of B, then when considering all objects, A will be ranked ahead of B.
- *non-dictatorship*: no voter is able to determine the outcome of the final ranking.
- *unanimity* if all voters rank object A ahead of object B, then the final ranking ranks A ahead of B.

The very surprising conclusion of the theorem is that it is impossible to construct a ranking system that contains all four of these criteria.

Using mathematics for ranking football teams (see [6], [10], [12], [13], and [18]) is one way to combat the impediment present in voting methods. An additional benefit of using a mathematical approach is that mathematics allows for abstraction to then apply these methods is other contexts. For example, Park and Newman [18] describe an abstraction of *game*, *win*, and *loss* and apply it to a study of bison in Montana.

As an overview of this thesis, a brief description of principal component analysis (PCA) concludes this chapter. That is followed by an in-depth history and overview of how ranking is done in the context of American college football (Chapter 2). Chapter 3 describes the data used in this thesis for ranking. Chapter 4 introduces the Weighted And Schedule-utilizing PCA (WASP) ranking algorithm and illustrates WASP, along with an example of how the process works. Chapter 5 investigates various manifestations of WASP applied to the 2008 season of American college football and compares the results against the polls, as well as one of these manifestations applied to twenty years worth of games. Lastly, Chapter 6 provides some concluding remarks.

1.1 Principal Component Analysis

PCA is a data reduction technique that is used to find where the most variation occurs within multivariate data. Suppose that a p-dimensional random vector $\vec{x}$ is given with mean vector $\vec{\mu}$ and covariance matrix $\mathbf{\Sigma}$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p$ be the (ordered) eigenvalues of $\mathbf{\Sigma}$, and let $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_p$ be the corresponding eigenvalues of $\mathbf{\Sigma}$. Now, the i^{th} principal component is given by $W_i = \vec{v}_i^T(\vec{x} - \vec{\mu})$, for $1 \leq i \leq p$. The W_i are linear combinations of the original variables, and for any $i \neq j$, W_i and W_j are uncorrelated and orthogonal with each other. Lastly, if given that $W = \vec{v}_i^T(\vec{x} - \vec{\mu})$ and $\vec{v}^T\vec{v} = 1$, $\mathbb{V}\text{ar}[W]$ is maximized if W is the first principal component W_1. Subsequently, if W is uncorrelated with the first $k < p$ principal components, then $\mathbb{V}\text{ar}[W]$ is maximized if W is the $(k+1)^{th}$ principal component W_{k+1}. Lastly, for some i such that for $1 \leq i \leq p$, the proportion of variance explained by the i^{th} component is given by $\dfrac{\lambda_i}{\text{tr}(\boldsymbol{\Sigma})}$. Because PCA focuses on finding the direction of most variation in the data, PCA can be a useful tool for finding the most unique data points. As far as ranking is concerned, the extreme points (the "best" data points on one extreme and "worst" data points on the other) will be the points that have the most variation; hence, PCA should be able to produce reasonable rankings. In the literature, PCA does not appear to have been applied to the problem of ranking.

Chapter 2

Crowning a Champion

2.1 Conventional Methods for Awarding Championship

In general, the champion of some sports league is determined either through a round-robin type of schedule (where every team plays every other team), or some kind of playoff, where entry and position in the playoff is determined by positioning in the regular season. While a playoff is the most common method of determining a champion in North America, round-robin scheduling or scheduling that ensures all teams play each other multiple times is also used for awarding championships. Generally, double round-robin scheduling is used, because it ensures that each team plays every other team at home and also on the road. Double round-robin scheduling is used by all major European soccer associations such as the English Premier League, German Bundesliga, Italian Serie A, and Spanish La Liga. It is also routinely used for awarding cricket championships and determining the winner of chess tournaments [11].

As an illustration of playoffs, consider the four major sports leagues in North

America – the National Football League (NFL), Major League Baseball (MLB), National Hockey League (NHL), and National Basketball Association (NBA). Of the four leagues, the NFL is the one that has schedules that is the least similar to a round-robin schedule. The NFL divides thirty-two teams into two sixteen team conferences, the American Football Conference (AFC) and National Football Conference (NFC). Then each of the sixteen teams in each conference are separated into North, South, East, and West divisions, each containing four teams.

Each team will play its three divisional rivals at home and on the road (six games), six games against other teams in the same conference but in a different division, and four games against teams in the other conference for a total of sixteen games. At the conclusion of the regular season, six AFC teams and six NFC teams compete to win their respective conferences. In each division, the team with the best win-loss record is the division champion and enters the playoffs. Additionally, of the teams that were not division winners, the two teams with the best win-loss records are given "wild card" entries into the playoffs for their conference. Following the conference playoffs, the AFC champion and NFC champion meet in the Super Bowl to crown the NFL champion. The 2015 season standings are provided as an example in Table 2.1, where the win-loss record is included in parentheses; bold teams were the teams that reached the playoffs. As in the NFL, the teams in each of the MLB, NHL, and NBA are separated into two conferences (East and West in the NHL and NBA, and American League and National League in the MLB). But unlike the NFL, MLB scheduling ensures that each team plays every team in its conference at least once; the NHL guarantees that every team plays every other team; and the NBA requires every team to play every other team at least twice. As an example of this partial round-robin scheduling, consider the Toronto Raptors in the Eastern Conference of

Table 2.1: Summary of the 2015 NFL season.

American Football Conference			
East	West	North	South
Patriots (12-4)	**Broncos** (12-4)	**Bengals** (12-4)	**Texans** (9-7)
Jets (10-6)	**Chiefs** (11-5)	**Steelers** (10-6)	Colts (8-8)
Bills (9-7)	Raiders (7-9)	Ravens (5-11)	Jaguars (5-11)
Dolphins (6-10)	Chargers (4-12)	Browns (3-13)	Titans (3-13)
National Football Conference			
East	West	North	South
Redskins (9-7)	**Cardinals** (13-3)	**Vikings** (11-5)	**Panthers** (15-1)
Eagles (7-9)	**Seahawks** (10-6)	**Packers** (10-6)	Falcons (8-8)
Giants (6-10)	Rams (7-9)	Lions (7-9)	Saints (7-9)
Cowboys (4-12)	49ers (5-11)	Bears (6-10)	Buccaneers (6-10)

the NBA. Each of the Eastern and Western Conferences contain fifteen teams which are split into three divisions of five teams each, and each team plays an eighty-two game schedule. The Raptors play four games against each of the other four teams in their division and two games against each of the fifteen teams in the Western conference. Of the remaining ten teams in the Eastern Conference, the Raptors play four games against six of these teams, and three games against the remaining four teams, to bring their total to eighty-two. The schedules in MLB and NHL are somewhat similar to the NBA schedule. In addition to having schedules that see each team play, at a minimum, a majority of the opposing teams, all of the MLB, NHL, and NBA use a playoff system to determine their champion. Both the NBA and NHL have a sixteen team playoff (eight from each conference), while MLB has a ten team playoff (five from each conference), culminating in the conference champions playing for the league championship.

2.2 Ranking in American College Football

The problem of ranking and determining a champion is especially visible at the highest level of American college football. Division I, the highest level of competition in the National Collegiate Athletic Association (NCAA), is broken into two subdivisions for football. Formerly known as Division I-A, the Football Bowl Subdivision (FBS) is the highest level of football competition in the NCAA, while the Football Championship Subdivision (FCS, formerly Division I-AA), constitutes the other subdivision of Division I football. For simplicity, FBS football will subsequently be referred to as NCAA football. NCAA football is very distinct in how its championship is awarded, because its champion is determined by neither a round-robin schedule nor a playoff, in the same sense as described above. In contrast to the FBS, the FCS champion is determined through a playoff. Currently, there are more than 125 universities in the FBS that compete for the national championship in a given year, and each team plays twelve, or possibly thirteen, regular season games. Also, unlike the four major sports leagues that have only two conferences, NCAA football is currently broken into ten conferences of varying size.

A majority of each team's regular season games, typically eight or nine, are played within its conference. Each conference that has at least twelve members is separated into two divisions and is permitted, by NCAA legislature, to have a conference championship game between the winners of the two divisions. At the conclusion of the regular season and conference championship games, many teams are invited to play in "bowl" games. Although these bowl games come after the regular season, they share nothing in common with the playoffs of the NFL, MLB, NHL or NBA. The participants in a given bowl game are usually fixed by conference affiliation. As a consequence of this structure, not only will teams play schedules with vastly different degrees of difficulty, but there are also very few direct games between different conferences to allow for comparisons.

Further complicating matters is the fact that most FBS teams tend to schedule at least one game and occasionally two games against FCS opponents each year. The relationship between FBS and FCS teams playing against each other is further considered in Section 3.2. Another consequence is that the bowl games do not naturally lead to any kind of a championship game or playoff. Furthermore, in all of men's and women's hockey, basketball, soccer, and tennis; men's baseball; women's softball; and football, all of which are played at all three divisions of the NCAA, the only one that did not utilize a playoff of any kind to determine the champion prior to 2014 was FBS football. An additional quality of the FBS conference structure is that these conferences are very fluid, unlike the divisions in the four major professional leagues in North America which are more or less fixed. The conference alignments are able to change frequently because the individual universities are permitted to change their conference affiliation. This freedom lead to wide scale changes in conference alignments from 2011 to 2014. Table 2.2 gives an overview of the conference structure for the 2008 season. Also unlike the professional leagues, collegiate schedules are not completely determined by the league or conference. Each conference sets its conference schedule, but pairs of schools will mutually agree to schedule an out of conference game. Instead of allowing results of games to purely dictate how the championship is awarded, NCAA football has utilized polls to rank teams. Since the 1930s, the Associated Press (AP) has been publishing an NCAA football poll. Select sports writers fill out a ballot of the top 25 teams, in that writer's personal opinion. Teams receive twenty-five points for a first place vote, twenty-four points for a second place, down

Table 2.2: NCAA Conferences from 2008 season.

Conference Name	Teams	Referenced Teams
Southeastern Conf.	12	Florida, Auburn, Louisiana St., Alabama
Pacific 10 Conf.	10	S. California, California, Washington St.
Big 12 Conf.	12	Oklahoma, Texas, Baylor
Big 10 Conf.	11	Michigan, Ohio State
Atlantic Coast Conf.	12	Florida St.
Big East Conf.	8	Cincinnati
Independent	3	Notre Dame, Naval Academy
Mountain West Conf.	9	Utah, Texas Christian
Western Athletic Conf.	9	Boise State
Conference USA	9	
Mid-American Conf.	13	Miami (Ohio)
Sun Belt Conf.	9	

to one point for a twenty-fifth place vote. Then, the teams are ordered based on the number of points awarded from the ballots. A preseason poll is performed before any games are played, and a subsequent poll is performed following each week of play. At the conclusion of the bowl games, the team that finishes first in the final AP poll would be declared the Associated Press National Champion. But the AP Poll is not the only major poll used to award a championship. There is also the Coaches Poll which works along the same guidelines as the AP Poll, but the members are a portion of the head coaches of the various teams. Similarly, the Coaches Poll awards the American Football Coaches Association National Championship to whichever team finishes first in the final Coaches Poll. Further adding to the confusion, from its inception in 1950 to 1973, the final Coaches Poll occurred at the conclusion of the regular season and before the bowl games, making the results of the bowl games irrelevant in regards to the Coaches Poll. The lack of a post-bowl poll was especially problematic in the 1971 and 1973 seasons when Texas and Alabama, the #1 teams in the respective Coaches Polls, lost their bowl games. Because there are two distinct polling groups that award championships, there is also the potential to have a "split championship" where the polls award their championship to different teams. This embarrassment has occurred eleven times: in the 1954, 1957, 1965, 1970, 1973, 1974, 1978, 1990, 1991, 1997, and 2003 seasons.

Even the use of polls has become a source of controversy, due to accusations of bias. By implication, if bias is present in the polls, then they would be unsuitable to be used at all. Based on studying the results from AP polls for the 2007 season, Coleman et al. [5] conclude that the media members that constitute the voter pool for the AP Poll were biased in favour of schools that are from the same state as the writer; schools that are in the same conference as schools in his state; schools that are in some of the prestigious "power" conferences; and schools that frequently play in marquee games on national television networks, such as ABC, NBC, or ESPN. The six bold faced conferences from Table 2.2 constitute the six power conferences of NCAA football. Of these six, Coleman et al. deduced that the SEC, Pac 10, and Big 12 conferences were the beneficiaries of voter bias, while the non-power conferences ("mid-major" conferences) were given preference that was equal to or better than the ACC, Big 10 and Big East conferences. The authors also provide an interesting commentary on why this may be the case. Of the six power conferences, the Big East and ACC have historically been better at basketball than football. Subsequently, these two conferences would be viewed as basketball-focused conferences as opposed to football-focused conferences. They also propose that while Big 10 schools such as Michigan and Ohio State had been premier teams in previous decades, the Big 10 was not as dominant in 2007 as it had been in years past, which led to the voters to be biased against the Big 10 conference. Interestingly, Goff [8] was also interested in bias in the polls, and conducted analysis on the polls from 1980–1989 and gave evidence to suggest that Ohio State and Michigan were among the teams that were given the most preferential treatment by voters during the 1980s. This period of voter bias in their favour coincided with when they were consistently upper echelon teams. As another example of voter bias, Goff also indicates that teams that were highly ranked in preseason polls were given higher rankings in the final polls than they deserved.

2.3 Historical Overview of the NCAA Football Championship

Throughout the history of NCAA football, there have been three different "eras" that describe how the champion has been determined: the Poll Era (1930s–1997), the Bowl Championship Series (1998–2013), and the College Football Playoff (2014–present). As alluded to, the Poll Era of NCAA football determined a champion exclusively through the use of polls. The possibility existed for the #1 and #2 teams of a poll to face each other in a bowl game, but this event did not occur frequently. Between 1960 and 1997, the #1 and #2 teams in the AP Poll played each other in a bowl game only eleven times. Following the bowl games that concluded the season, the polls would award their respective championships.

The 1997 season was the final straw that necessitated the creation of a championship game between the top two teams. Entering the bowl part of the schedule Michigan held the #1 position in both the AP Poll and Coaches Poll. However, Michigan (winner of the Big 10) faced #8 Washington State (winner of the Pac 10) in the Rose Bowl and won 21-16. Meanwhile, Nebraska went to the Orange Bowl as the #2 team in both the AP and Coaches Poll faced Tennessee (#3 in the AP poll) in the Orange Bowl. Nebraska went on to win 42-17. Following the bowl games, Michigan scored 1731.5 points in the AP Poll to the 1698.5 points that Nebraska was given, which kept Michigan at #1 in the AP Poll [3]. It is worth noting that the pollsters viewed Michigan as a less overwhelming #1 team after their bowl game than before. Michigan only received fifty-eight of seventy first place votes after receiving sixty-nine first place votes prior to their bowl game [26]. Unlike the AP Poll, the Coaches Poll saw Nebraska win 1520 points (including thirty-two of sixty-two first place votes) in the final poll to 1516 points for Michigan (including the remaining thirty first place votes) to narrowly give Nebraska the #1 position and the national championship [3].

BCS Era

The fallout of the 1997 season lead to the creation of the Bowl Championship Series (BCS) in 1998. The BCS was not actually "under" the NCAA, but rather an outside organization given the authority to determine the champion. This transition led to the #1 and #2 teams in the newly formed BCS poll playing a national championship game. The BCS poll combined the results of the AP and Coaches Polls along with various computer polls designed to rank NCAA football teams. Of the six computer polls used, only the Colley Method [6] is fully reproducible, while the other five are black box methods. In addition to the championship game, the BCS was also responsible for determining the matchups for four of the most prestigious bowl games: the Rose Bowl, Sugar Bowl, Orange Bowl, and Fiesta Bowl. For the first eight years of the BCS, these four bowls would feature the national championship game, rotating

Table 2.3: BCS Bowl Allignments (2007-2013).

National Championship	BCS #1 v. BCS #2
Rose Bowl	Pac 10 #1 v. Big 10 #1
Sugar Bowl	SEC #1 v. Wild Card
Orange Bowl	ACC #1 v. Wild Card
Fiesta Bowl	Big 12 #1 v. Wild Card

through a four year cycle. Historically, these four games had operated along the guidelines outlined in Table 2.3, and the BCS attempted to maintain these relationships, except where the BCS #1 and #2 teams were concerned. For example, if the Pac 10 winner was also the BCS #1 or #2, and the national championship was to be the Orange Bowl, then the Pac 10 winner would play in the Orange Bowl, and the Pac 10 spot in the Rose Bowl would become a wild card entry.

In 2007, the BCS was tweaked to make the national championship game a separate entity from these four bowls — making five BCS bowls where there were previously four. Similar to the prior system, the historical alignments from Table 2.3 were guaranteed, unless the BCS #1 or #2 teams were involved. Implicit in the BCS was the favouring of the six BCS power conferences (the bold faced conferences from Table 2.2) as well as The University of Notre Dame. Each of the winners of these six conferences (known as Automatically Qualifying conferences, or AQ) were guaranteed to participate in a BCS bowl. Of the six AQ conferences, only the Big East does not appear in Table 2.3, because the Big East was not tied to any one bowl. Instead, the Big East champion would fill in either one of the Wild Card positions or one of the AQ positions for a team that went to the national championship game. Additionally, Notre Dame, an independent, was given an automatic qualification to a BCS bowl if it finished at #8 or better in the BCS poll [7]. After the automatic qualifications were distributed, there would be a maximum of four remaining slots (two in the years before 2007) for a mid-major school to get into BCS games as a wild card entrant.

All hope was not lost for the mid-major schools: they were also given a chance to automatically qualify for a BCS bowl. Of the five champions from the non-AQ conferences from Table 2.2, the one that had the highest ranking in the BCS poll would automatically qualify if it had a BCS ranking of at least #6. In 2004, this standard was generously relaxed so that the highest ranked mid-major conference champion automatically qualified if it was ranked no lower than #12, or if it was ranked no lower than #16 and was ranked higher than one of the AQ conference champions. Also working against the mid-major schools is the fact that, historically, they lack the prestige that many of the power schools have, which further removes the incentive for a mid-major to be selected for a BCS bowl as a wild card. Just like the BCS, each of the BCS bowls are run by organizations that are separate from both the BCS and the NCAA. As such, the built-in disadvantages that mid-major schools tend to face limit their attractiveness to the BCS bowls. After all of rules for automatic qualification have been handled, the organizations behind the BCS bowls are free to offer invitations to whomever they choose to fill any unfilled, wild card positions for their bowl game.

The 2009 NCAA football season provides a clear example of the up-hill battle that a mid-major football team encountered. Alabama (SEC champion), Texas (Big 12 champion), Cincinnati (Big East Champion), Texas Christian (TCU, Mountain West Champion), and Boise State (Western Athletic Champion) all finished the 2009 regular season with undefeated records. Ultimately, these teams finished #1, #2, #3, #4, and #6, respectively, in the BCS poll. This finish placed Alabama and Texas in the national championship, while Cincinnati automatically qualified for a BCS bowl by virtue of winning the Big East. Based on being the highest ranked champion of a non-AQ conference, TCU also automatically qualified for a BCS bowl by finishing #4, while Boise State became the first and only team from a non-AQ conference (aside from Notre Dame) to earn a wild card entry to a BCS bowl. Instead of scheduling TCU and Boise State to face power conference opponents in separate bowls, they were selected to face each other in the Fiesta Bowl. This controversial decision further illustrated the inequity of the system for the mid-major schools and has derisively been called the "Separate But Equal Bowl" by *Sports Illustrated* author Andy Staples, while Utah Congressman Jim Matheson called the game the "Kids' Table Bowl" [22]. Clearly, the move from using polls to determine a champion to the BCS resulted in the championship being determined by a championship game as opposed to votes. However, the BCS moved the point of contention from determining the champion to determining admission to the championship game, something that became no less controversial of a task than its predecessor. In many cases, the BCS generated more controversy than before.

Despite worthwhile intentions, the BCS generally became synonymous with controversy, questions of fairness, and United States Department of Justice inquiries. Although the BCS was formed with the goal of ensuring a consensus champion, this was not always the case. The 2003 season featured six teams with one loss: Southern California (USC), Louisiana State (LSU), and Auburn from power conferences, as well as Miami (Ohio), Boise State, and TCU from mid-major conferences. At the conclusion of the regular season, both the AP and Coaches Polls had USC at #1; LSU at #2; and Oklahoma at #3, with the other three teams at #14, #18, and #19, respectively. However, the results from the computer polls pushed the BCS poll to put Oklahoma at #1, while LSU narrowly beat out USC for the #2 position. LSU proceeded to defeat Oklahoma in the national championship game, but USC also defeated BCS #4 Michigan in the Rose Bowl. The final Coaches Poll had LSU as the #1 team; however, three coaches dissented and kept USC as #1, despite a contractual obligation to place LSU as the unanimous #1 team in the final poll. Additionally, USC remained the #1 team in the final AP Poll, winning the national championship of the AP Poll. The BCS resulted in the very thing that it was created to prevent: a split national championship.

A result similar to the 2003 season occurred in 2004: USC, Oklahoma, Auburn, Utah, and Boise State ended the regular season unbeaten. Except this time USC and Oklahoma were the #1 and #2 teams in the AP, Coaches, and BCS polls. In addition to USC defeating Oklahoma in the national championship game, Auburn and Utah won their bowl games. Despite three teams finishing undefeated, USC was regarded as the undisputed champion. Unlike the 2003 season, the controversy of the 2004 season was not over which teams ought to be #1 or #2, but rather which team should be #4. As the Pac 10 champion, USC would have been guaranteed a berth in the Rose Bowl. However, because USC was #1 in the BCS poll, they would play in the national championship, which would in turn change the automatic entry in the Rose Bowl into a wild card. The BCS rules at the time stipulated that a team that did not win their conference but finished at least #4 in the BCS poll was guaranteed entry into one of the four BCS bowls from Table 2.3. Entering the final week of the regular season, California held a narrow lead on Texas for the #4 position in the BCS poll, but California had a game left to play, while Texas did not.

Because California would be Pac 10 runner-up to USC, and Texas would be Big 12 runner-up to Oklahoma, the race for #4 would ensure not only the ability to play in the prestigious Rose Bowl, but also the guaranteed $14 million payout for participating in the Rose Bowl. In the human polls before the final week of the regular season, California led Texas by eighty-five points in the AP Poll and forty-eight points in the Coaches Poll. Following Texas's final game, the head coach of Texas began campaigning for Texas to be receive more fourth place votes. California won their game by a rather unconvincing 26-16 margin over a mediocre opponent. Although California had an opportunity to run up the score with a late touchdown, they chose to take the path of sportsmanship and settled for the ten point victory. In the subsequent polls, California saw their lead over Texas fall to sixty-two points and five points in the AP and Coaches Polls, respectively. The drastic change in the Coaches Poll included four coaches dropping California to #7, along with two coaches that placed California at #8 in their ballots, despite no one ranking California worse than #6 the week prior. These results, combined with the computer polls favouring Texas over California, were enough to push Texas into the #4 position and give them the wild card entry to the Rose Bowl.

The 2008 season provides a final example of the issues that plagued the BCS. In the 2003 and 2004 seasons, mid-major schools such as Utah and Boise State were undefeated teams that were passed over for entry into the national championship game in favour of undefeated teams from power conferences. Unlike the 2003 and 2004 seasons, only Utah and Boise State finished the regular season undefeated. As in prior years, both Utah (#6) and Boise State (#9) failed to be in the top two. Although Utah would automatically qualify for a BCS bowl game, based on the revised conditions from 2006, both teams were passed over for the championship game in favour of Oklahoma and Florida, who were both 12-1. While Boise State lost their bowl game, Utah responded by defeating #4 Alabama in the Sugar Bowl to finish the season as the only undefeated team in the FBS, but not recognized as the national champion. In Section 5.2, these seasons will be investigated using the WASP algorithm .

The BCS and the U.S. Government

In addition to years of on-field controversy, the BCS era gained scrutiny from the U.S. Department of Justice over possible violations of American antitrust laws. After

Utah failed to gain entry to the BCS national championship game in 2008, U.S. Senator Orrin Hatch (Utah) took issue with how the BCS was perceived to be limiting competition among all of the FBS institutions, which would violate the Sherman anti-trust act. Sen. Hatch proceeded to call for the Department of Justice to investigate the legality of the BCS, with the desire to re-categorize the BCS as a "cartel". In 2011, the Department of Justice inquired as to why NCAA football was the only sport in the NCAA that lacked a playoff. Also in 2011, Utah Attorney General Mark Shurtleff proposed filing an anti-trust lawsuit against the BCS. In 2012, the decision was made to abolish the BCS in favour of a playoff, and no lawsuit was ever filed. Utah was not the only state with politicians that raised questions about the legality of the BCS. Rep. Joe Barton (Texas) proposed the *College Football Playoff Act of 2009*. This Act was proposed "To prohibit, as an unfair and deceptive act or practice, the promotion, marketing, and advertising of any post-season NCAA Division I football game as a national championship game unless such game is the culmination of a fair and equitable playoff system" [23]. If approved, the Act would consider the BCS to be involved in false advertising. This designation would outlaw the promotion of any game (or sale of any merchandise in relation to any game) that was deemed to be a "national championship game" but was not the result of a playoff in which all FBS teams were eligible for entrance into the playoff. The Act was passed by a subcommittee of the House Energy and Commerce Committee before ultimately dying in the U.S. House of Representatives in December 2009.

CFP Era

In 2014, after years of unrest over the lack of a playoff system, the NCAA finally adopted the College Football Playoff (CFP): a four team, single-elimination playoff that replaced the BCS. This was a drastic change to the way admission to the championship was determined, because the polls, while still conducted, had no effect on the ranking of teams. Instead the CFP created a thirteen-person Selection Committee that had the job of producing a Top 25 ranking. At the conclusion of the regular season and conference championships, the top four teams would play a #1 versus #4 and #2 versus #3 semi-final, with the two winners playing for the national championship. Naturally, in the first year of a four team playoff, there were six teams that had a case for entry into the playoff. Entering the final week of the 2014 regular season, the Selection Committee had a top six of Alabama, Oregon, TCU, Florida St., Ohio State, and Baylor. All six had one game remaining. Alabama, Oregon, Florida State and Ohio State would all play for their respective conference championships; while TCU (which had moved to the Big 12 in 2012) faced lowly 2-9 Iowa State; and Baylor drew #9 Kansas State. After all six teams won their final games — including TCU defeating Iowa State by a resounding 55-3 margin — the final CFP ranking ordered these six teams Alabama, Oregon, Florida State, Ohio State, Baylor, and TCU.

2.4 Summary

This chapter has described the two usual methods for determining champions with examples of where the round-robin and playoff formats are used. The usage of these two formats is contrasted with the poll-based ranking in NCAA football. Following this comparison, a summary of the history of NCAA football with a specific look at a few years of controversy shows the inadequacy of the poll-based system used to rank the NCAA football teams. A natural inclination would be to conclude that wherever an injustice occurs in the football rankings, those associated with the football team—whether players, coaches, or fans—are the only ones who feel the impact of the injustice. However, research has shown that the university may also be adversely affected. Pope and Pope [19] look at how success in basketball and football relates to student applications for all Division I schools from 1983–2002. They conclude that a school that wins a football or basketball national championship could expect to see a 7% or 8% increase in enrolment applications. Additionally, they also state that this increase is not merely a result of an influx of high school students with low SAT scores. Rather, the extra applications that a school receives would include a combination of both low and high SAT scores. Lastly, they show that if a private university, such as Stanford, Notre Dame or Duke, has the same athletic success as a public institution, the private university could expect to see an application percentage increase of at least double what the public school receives. Interestingly, they also provide evidence that private institutions increase tuition following success in basketball but not in football.

Chapter 3

Details of Data

3.1 Introduction

The most basic approach to ranking teams is to simply use wins and losses, so that whoever wins the most or has the best win-loss ratio is the top team. As mentioned in Section 2.2, this method simply will not work in the context of NCAA football, due to the disparity in the difficulty of the individual schedules. With this in mind, the minimum amount of information needed to reasonably rank teams are the wins and losses of each team, as well as a strength of schedule measure (explained in Section 3.2), including, for a given team i, which teams i defeated and which teams defeated i. These three qualities are hardly an exhaustive set of criterion that could be included in a ranking system. Other possible candidates include the points that a team scores and allows; in addition to other qualities such as the number of turnovers each team commits. The point difference is probably the most common addition to any other algorithm, but it is a contentious property. After the 2001 season, the authors of the computer algorithms used as a part of the BCS poll were ordered, by the BCS committee, to remove margin of victory from their methods. Disregarding margin of victory has two principal benefits: first, teams are unable to "hack" their ratings by running up the score. A second benefit, and a corollary of the first, is the promotion of sportsmanship. Another criticism of margin of victory is that it has the potential to be very deceptive. In football, where a touchdown is worth six points with a conversion attempt for a seventh point, it is very easy to make close games appear more like blowouts and non-competitive games appear closer than they actually were. Because a blowout and a close game with late scoring that widens the point margin would look the same to a computer algorithm, it does not seem profitable to include point margins. Nevertheless, Chapter 5 includes comparisons involving point margins, for completeness.

3.2 Strength of Schedule

In Section 2.2, it was mentioned that one of the points of contention that the use of polls encountered was that the top mid-major teams were consistently ranked below their power conference counterparts. These results were justified on the belief that the best mid-major teams always play an inferior set of opponents compared to the best power conference teams; hence, a mid-major team that went undefeated did not carry the same impact as a power conference team that went undefeated, or in some cases, suffered a defeat. Due to the very diverse schedules that the various teams face, it is necessary to attempt to quantify the difficulty of the schedule of each team. The simplest way to represent the entire schedule of the league is as a directed graph. Let $\mathcal{S}$ be the digraph representing the schedule with vertex set $\mathcal{V} = \{1, 2, ..., n\}$, where n is the number of teams, and arc set $\mathcal{A} = \{(a, b)|a \text{ lost a game to } b\}$. It is also possible for teams to play each other multiple times. In that case, the arc (a, b) could appear twice in $\mathcal{A}$, or (a, b) as well as (b, a) could be in $\mathcal{A}$. For example, Figure 3.1 illustrates

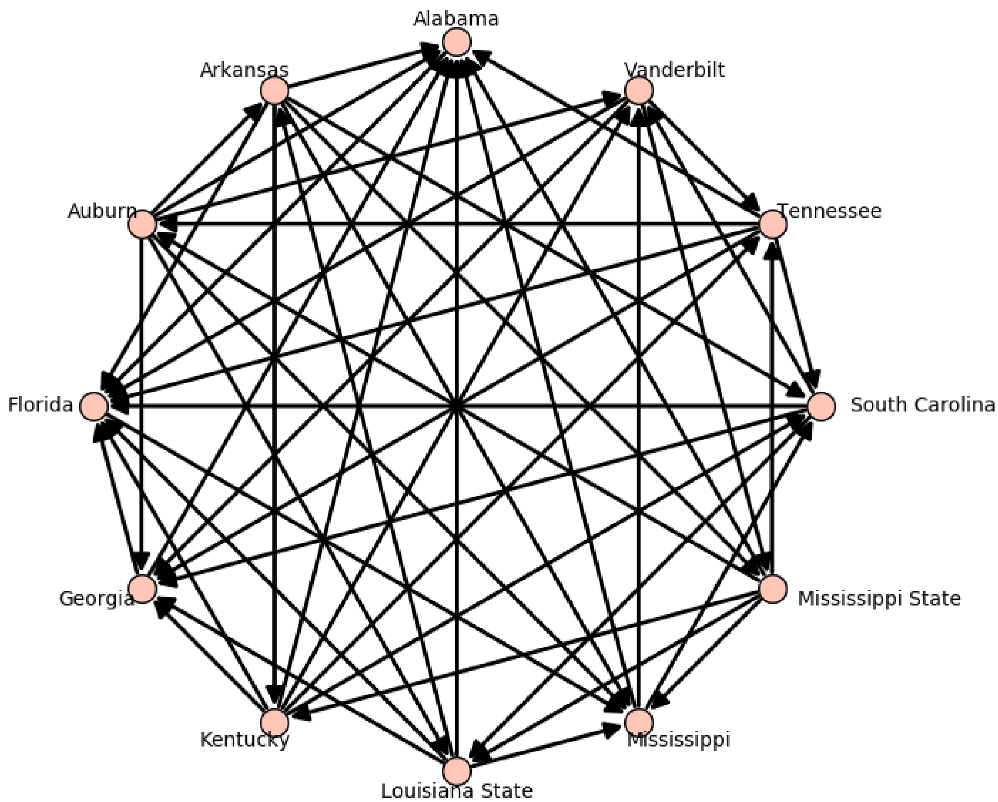

Figure 3.1: Schedule for 2008 SEC.

the Southeastern Conference (SEC) schedule for the 2008 season. Because Florida defeated Alabama, there is an arc going from Alabama to Florida. Likewise, Florida lost to Mississippi, so there is an arc from Florida to Mississippi. The simplest measure of strength of schedule (SOS) is to take the cumulative winning percentage of all opponents (first degree winning percentage). However, this discounts how those opponents accumulated their wins and losses; i.e., how difficult their schedules were. Thus, it seems reasonable to consider the cumulative winning percentage of all opponents of the opponents (second degree winning percentage). It is also worth noting that many games do not deserve to be considered in first and second degree winning percentage, because they will not provide any information about the SOS. Figure 3.2 shows the all of the games played by the 2008 Florida Gators, as well as all games

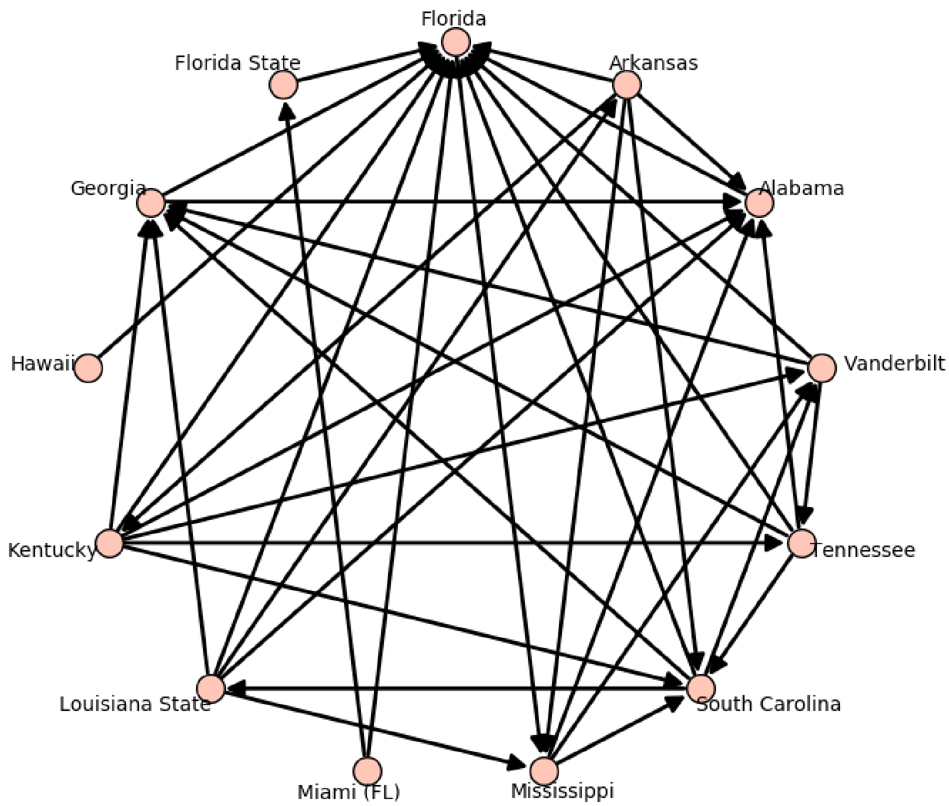

Figure 3.2: Graph of 2008 Florida and opponents.

played between opponents of Florida. For example, Miami (FL) played Florida State, so that game is included, because both Miami (FL) and Florida State played Florida. But Miami (FL) also played North Carolina, and because North Carolina did not play Florida, that game is not included in Figure 3.2. Now, when considering the first degree winning percentage for Florida, whenever two opponents of Florida have a game, say Miami (FL) and Florida State, the outcome does not change the first degree winning percentage. Whether Miami (FL) wins or Florida State wins, Florida will receive one win and one loss in the first degree winning percentage. As a result, every game between a pair of opponents can be discarded when calculating first degree winning percentage. Similarly, when looking at the second degree winning percentage, many of these games can also be ignored for similar reasons. Lastly, in Section 2.2, it was mentioned that teams in the FBS (the subdivision of interest) regularly play against teams in the FCS. The most thorough approach would be to include all of the FCS teams into the rankings as well. However, this does not solve the problem, because FCS teams will schedule games against Division II opponents; likewise, Division II opponents will schedule games against Division III opponents. Even if one endeavoured to produce a ranking of all FBS, FCS, Division II, and Division III teams, a ranking compiled from all these teams would be based on relationships between teams that are completely meaningless, and this requires accumulating excessive amounts of data that would be required to relate FBS teams to Division III teams.

A simpler approach would be to completely ignore all games between FBS and FCS teams. This is not an unreasonable approach, because the assumption is that an FBS team should always defeat an FCS team, and any FBS team that loses to an FCS team probably is not any good anyway. Thus, disregarding losses by weak FBS teams against FCS teams is unlikely to have a drastic impact upon the ratings, and these games can be ignored. Unfortunately, this premise is false. There have been three instances of an FCS team defeating an FBS team that was in the top 25 of the preseason poll: Appalachian State defeated #5 Michigan in 2007; James Madison beat #13 Virginia Tech in 2010; and #25 Oregon State lost to Eastern Washington in 2013. More generally, the Eastern Washington victory was one of sixteen wins by FCS teams against FBS opposition in 2013. In regards to the Virginia Tech loss, their defeat against James Madison was one of their two losses in the 2010 regular season, so ignoring that result would substantially boost their ranking.

The ineffectiveness of the prior approaches suggests that a third option would be to count all of the FCS opponents as one collective opponent. This third option is approximately the approach used when encountering a game involving an FCS opponent. As far as an FBS team is concerned, they do not ever get credit for a win in the event that they defeat an FCS team. This is reasonable, because it prevents teams from padding their win totals against what are effectively minor league opponents. However, in the event that an FBS team loses to an FCS team, as Michigan and Virginia Tech did, then that team gets credited with a loss. Because there are roughly 100 games played between FBS and FCS teams, the FCS "team" would carry substantial weight in the SOS calculation. So, to resolve this problem, the FCS "team" does not factor into the SOS ratings. This seems to be a reasonable trade off, because teams do not benefit from the additional win, but are also not penalized with a reduced SOS rating that would result from playing an FCS opponent. Finally, the FCS team is included in the rankings; however, the *win* and *loss* totals are scaled so that they sum to twelve, as a result of all other teams playing roughly twelve games.

Lastly, after removing many of the first (and second) degree games from consideration, the number of results that each team has will vary from as low as about thirty games up to 100 games. Teams in a conference, such as the Pac 10, where the conference schedule is a round-robin schedule, will be on the low end of the spectrum, while Independents such as Notre Dame or the Naval Academy will be at the upper end of the spectrum because their opponents are not primarily members of the same conference. This means that a team could have a very large first degree winning percentage based on a small number of results. To prevent this, the number of first degree results that each team has is normalized to some value by adding 1 first degree win (FDW) and 1 first degree loss (FDL) until the sum of FDW and FDL is within 1 result of the chosen value. When dealing with NCAA football seasons, the tenth most games have been used, because the Independent teams and their opponents tend to have higher amounts of first degree results. For example, if the tenth most number of first degree results is 88, and a team has 24 FDW and 6 FDL, that team would have a first degree winning percentage of 80%. After normalizing, said team would have 53 FDW and 35 FDL for a first degree winning percentage of 60.2%, which is a much more reasonable value than the 80% from before. This approach is also applied to normalize the number of second degree results as well to ensure that teams do not gain unfair advantages.

3.3 Flexibility of Inputs and Home Field Advantage

While the qualities used for inputs for ranking can vary substantially, the input values themselves are also not necessarily constants. The individual wins and losses do not have to be counted equally, but rather they are able to carry different weights. Table 3.1 gives the schedule of the 2008 Florida Gators, including the week that the game was played and where the game was played. If "Location" is blank, then Florida was at home; an "@" indicates that Florida was the road team; and "N" is used if the game was played at a neutral stadium. Two alternative methods for weighting games are based on when the game occurred and where the game was played. The reason to weight games is based on the view that some games are more significant than others. For example, the final game from Table 3.1 was the SEC championship game between undefeated Alabama and Florida, and the winner (Florida) would go to the BCS national championship game. Clearly, this is an immensely important game, and is substantially more important than Florida's opening game against a mediocre Hawaii team that finished the year 7-7. More generally, the importance of games tends to increase as the season nears the end of the schedule, and this could be encoded into the algorithm. There are a wide variety of schemes to weight games. A simple suggestion would be a binary weight that makes the results of games in the

Table 3.1: Weekly schedule of 2008 Florida Gators.

Date	Week		Location	Opponent	FLA Pts	Opp Pts
Aug 30	1	Florida		Hawaii	56	10
Sept 6	2	Florida		Miami (FL)	26	3
Sept 20	4	Florida	@	Tennessee	30	6
Sept 27	5	Florida		Mississippi	30	31
Oct 4	6	Florida	@	Arkansas	38	7
Oct 11	7	Florida		LSU	51	21
Oct 25	9	Florida		Kentucky	63	5
Nov 1	10	Florida	N	Georgia	49	10
Nov 8	11	Florida	@	Vanderbilt	42	14
Nov 15	12	Florida		S. Carolina	56	6
Nov 22	13	Florida		Citadel	70	19
Nov 29	14	Florida	@	Florida State	45	15
Dec 6	15	Florida	N	Alabama	31	20

second half of the season (say, week 9 and on, since there are 15 weeks) worth twice the result of games in the first half of the season. Using this method, Florida's loss to Mississippi would have much less impact than if Florida had (hypothetically) lost to South Carolina. Alternatively, each week could be given more weight than the previous week with a weighting function such as

$$w_i = 1 + \frac{i-1}{14},$$

where i is the week number. As in the first case, a win in the final week is worth twice the amount as a win in the first week, but now the impact is more evenly dispersed. The benefit of weighting games is not only that more significance is assigned to the seemingly more important games at the end of the year, but also that there is a focus on the performance of teams at the end of the year. This means that teams that finish the year strongly will place higher than teams that start strong but cannot maintain that level of play throughout the season. However, the same benefits that arise from weighting games also pose a noticeable problem.

Many NCAA football seasons feature marquee out of conference match ups between top teams in the first week of the season. For example, the 2010 season began with #3 Boise State beating #10 Virginia Tech (rankings based on pre-season polls); #3 LSU started their 2011 season with a victory over #4 Oregon while #5 Boise State beat #19 Georgia; #2 Alabama defeated #8 Michigan to start the 2012 season; and #8 Clemson claimed a win over #5 Georgia at the beginning of the 2013 season. All of these are games that are very significant for the winners because they are wins over other top teams. But by decreasing the weight of early games, these important games become relatively meaningless. Similarly, this is not overcome by giving additional weight to the first week of the schedule, while making the second week relatively insignificant. Due to the nature of scheduling, some teams could see the most difficult (i.e., most rewarding) part of their schedule be early in the year, which would put them at a substantial disadvantage when compared to a team that plays harder competition near the end of the year, when the games are more meaningful.

The second proposal for weighting games is based on the location of the game. Home field advantage has been a source of debate and also academic interest in the hope of settling with home field advantage is real or imagined. Different components of supposedly important factors of home field advantage such as crowd size and support, travel [16], the ability of the crowd to bias the officials in favour of the home team [14], and even the temperature on the day of the game [2] have been the subject of academic study. By far the most contentious issue is whether or not crowd support plays a role in home field advantage. Nevill et al. [15] support the belief that crowds influence players to perform at a higher level, while Strauss [20] rejects this hypothesis.

Going a step further, Thirer and Rampey [21] investigate how the home and visiting teams react to different crowd behaviour. They conclude that when the crowd is acting "politely" (i.e., booing), the home team is at a marked advantage. However, when the crowd is acting "rudely" (i.e., cursing), the performance of the visiting team is largely unaffected, while the home team performs *worse* than the visiting team. Assuming that there is a home field advantage and observing that home teams have a better winning percentage than the visiting team, a home win (and also a road loss) should not count for as much as a road win (and also a home loss). This is a perfectly reasonable assumption; however, the issue becomes determining how much of a difference home field advantage makes, and also how this would vary between pairs of teams. For example, a team from Ohio that is playing in Gainesville, Florida, in August in front of 90,000 people is intuitively at a much larger disadvantage than if that same team were playing in front of 20,000 people in Buffalo, New York, in August. But assigning numeric values to account for disadvantage in a consistent manner is not straightforward. Despite this difficulty, Chapter 5 includes a proposed method that attempts to account for the impact of location of games and reward the result accordingly.

Chapter 4

Methodology

4.1 Overview

The goal of this section is to describe the suitability of the WASP algorithm and show that it attains the following goals:

1. WASP contains no initial biases.
2. WASP is concerned with only results from the year or schedule in question.
3. WASP accounts for strength of schedule.
4. WASP accounts for teams defeated and teams lost to.
5. WASP adheres to Occam's Razor.
6. WASP provides final results that mesh with intuition.

Based on the description of PCA from Section 1.1, it is a useful tool for ranking based on the uniqueness of the data. Given a loading matrix

$$\boldsymbol{V} = \left[\begin{array}{c|c|c|c} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_j \end{array} \right]$$

that is combined with covariance matrix $\boldsymbol{\Sigma}$ produced by PCA, a meaningful ordering can be produced with a minimal ammount of assumptions. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_j \geq 0$ be the (ordered) eigenvalues of $\boldsymbol{\Sigma}$. Let $\vec{x}_i$ be the centred and scaled observations, and set f_i to be the full proportion of variance weighted value

$$f_i = \sum_{k=1}^{j} \frac{\lambda_k}{\text{tr}\left(\boldsymbol{\Sigma}\right)} \left(\vec{x}_i \cdot \vec{v}_k\right).$$

Similar to how components from PCA that explain very little variation are ignored, f_i can be reduced to r_i, defined by

$$r_i = \sum_{k=1}^{m} \frac{\lambda_k}{\text{tr}\left(\boldsymbol{\Sigma}\right)} \left(\vec{x}_i \cdot \vec{v}_k\right), \tag{4.1}$$

where m is the number of components selected. The process is more fully illustrated in the following (contrived) example.

Example 4.1.1 *Suppose that there is a group of twelve teams: the Arrows, Beavers, Camels, Defenestrators, Eagles, Flames, Giants, Hitmen, Icemen, Jazz, Knights, and Lumberjacks. Furthermore, the results of all the games in the schedule is outlined in Figure 4.1, where an arc going from vertex i to vertex j means that j defeated i.*

A basic investigation of the final standings in Table 4.1 would suggest that the Beavers should be the top team, because they were the only team that did not lose. Clearly, this overlooks the fact that the five wins of the Beavers came against weak competition in the form of the Eagles, Hitmen, Jazz, and Knights. Intuitively, this

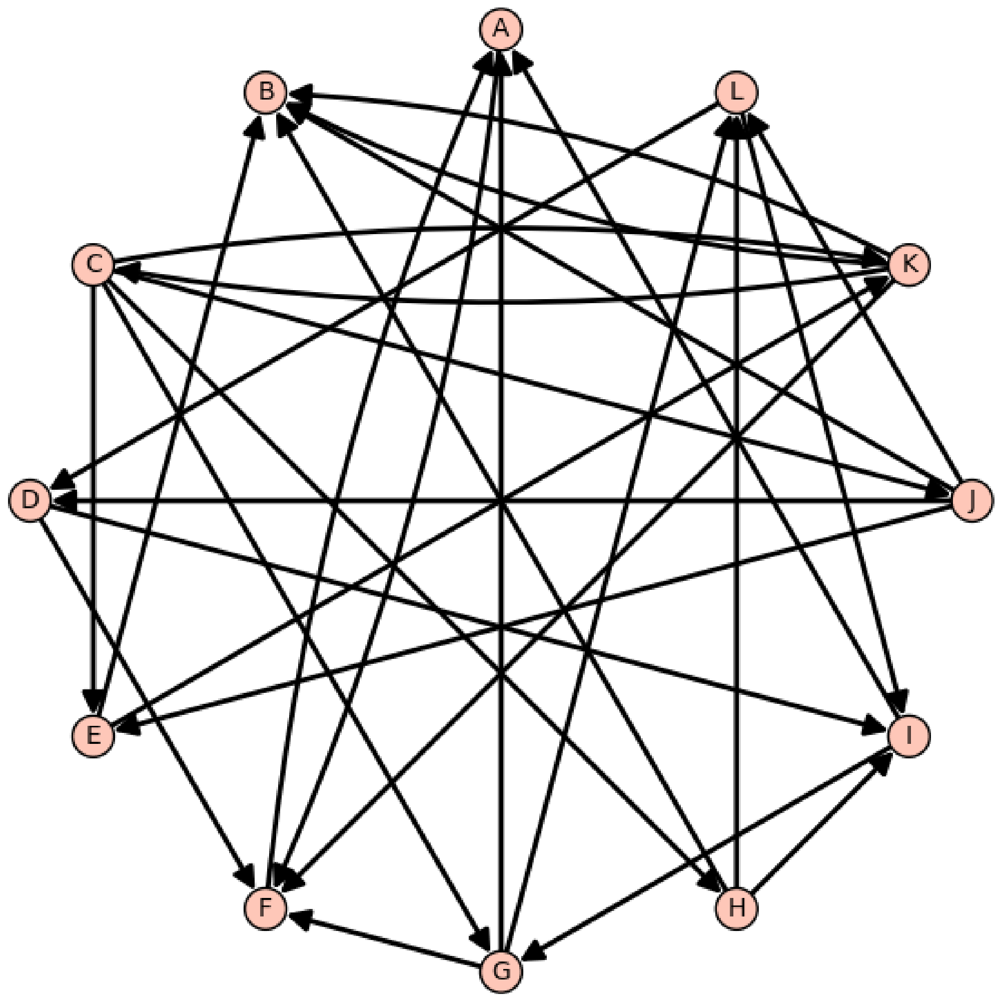

Figure 4.1: A digraph representation of a fictional schedule.

revelation might suggest that either the Arrows or Flames are more deserving of being the the #1 team, as a result of the Arrows having the best SOS, while the Flames have the second most wins and second best SOS. As sketched in Section 3.2, the SOS rating for the Arrows is based on their their opponents: Icemen, Giants, and Flames (twice). In games not against the Arrows or any opponent of the Arrows, the Flames had a record of 2-0. Because the Flames and Arrows played twice, this effect is counted twice. Likewise, the Giants had a record of 1-1, while the Icemen had a 3-0 record.

This gives the Arrows eight FDW and one FDL. In a similar manner, the FDW and FDL of all teams could be produced to show that the Camels have the largest number of first degree games with 17. Unlike in the NCAA football case, the normalizing will be done using the maximum number of games. This results in the Arrows having normalized FDW and FDL of 12 and 5, respectively. At the second degree level for

Table 4.1: Standings from the fictional schedule.

Team	Wins	Losses	SOS
Arrows	3	1	0.6265
Beavers	5	0	0.3725
Camels	1	5	0.3813
Defenestrators	2	2	0.5360
Eagles	2	2	0.3542
Flames	4	1	0.5677
Giants	2	3	0.5035
Hitmen	1	3	0.4345
Icemen	3	2	0.5425
Jazz	1	4	0.4112
Knights	2	4	0.4251
Lumberjacks	3	2	0.4328

the Arrows, the wins and losses of the opponents of the Giants, Icemen, and Flames (counted twice) need to be considered. When looking at the opponents of the Giants (for the Arrows), all results against the Giants, Arrows, Camels, Flames, Icemen, and Lumberjacks are ignored. The record for the Camels is 1-4. The Flames (this is only counted once, because the Giants and Flames only faced each other once) have a record of 2-0. The Icemen have a 2-0 record, while the record of the Lumberjacks is 2-1. So, in total the Giants are responsible for 7 and 5 second degree wins ("SDW") and losses ("SDL"), respectively, for the Arrows. Repeating this process for the Icemen and Flames shows that the Icemen account for 4 SDW and 3 SDL. Doubling the 6 SDW and 5 SDL from the Flames, shows that the Arrows have a total of 23 SDW and

18 SDL. In could also be shown that the Eagles have the most second degree games, at 54 games. After normalizing, the Arrows have 29 SDW and 24 SDL. Subsequently, the SOS of the Arrows is given by

$$\frac{\frac{12}{12+5} + \frac{29}{29+24}}{2} \approx 0.6265.$$

Setting $\boldsymbol{M}$ to be the scaled win, loss, and SOS matrix based on Table 4.1, and applying PCA to $\boldsymbol{M}$ produces the loading matrix $\boldsymbol{V}$ with variances $\vec{\lambda}$.

$$\boldsymbol{M} = \begin{array}{c} \begin{matrix} \text{W} & \text{L} & \text{SOS} \end{matrix} \\ \begin{pmatrix} 0.4704 & -0.9815 & 1.8471 \\ 2.0831 & -1.6743 & -1.0688 \\ -1.1424 & 1.7898 & -0.9688 \\ -0.3360 & -0.2887 & 0.8074 \\ -0.3360 & -0.2887 & -1.2799 \\ 1.2768 & -0.9815 & 1.1717 \\ -0.3360 & 0.4041 & 0.4343 \\ -1.1424 & 0.4041 & -0.3574 \\ 0.4704 & -0.2887 & 0.8822 \\ -1.1424 & 1.0970 & -0.6249 \\ -0.3360 & 1.0970 & -0.4657 \\ 0.4704 & -0.2887 & -0.3773 \end{pmatrix} \end{array},$$

$$\boldsymbol{V} = \begin{array}{c} \\ \text{W} \\ \text{L} \\ \text{SOS} \end{array} \!\!\! \begin{array}{c} \begin{array}{ccc} \text{PC1} & \text{PC2} & \text{PC3} \end{array} \\ \begin{pmatrix} 0.6398 & -0.3521 & 0.6831 \\ -0.6638 & 0.1947 & 0.7221 \\ 0.3872 & 0.9155 & 0.1091 \end{pmatrix} \end{array}, \quad \vec{\lambda} = \begin{pmatrix} 2.0528 \\ 0.8241 \\ 0.1231 \end{pmatrix}.$$

The loading matrix $\boldsymbol{V}$ shows that component 1 loads heavily on winning and "not losing", with some weight given to SOS. On the other hand, component 2 loads almost exclusively on SOS. Lastly, component 3 is heavily loaded on both wins and losses. This suggests that every team would have a similar value for component 3, because most teams play the same number of games. Based on the proportion of variances from λ_1, λ_2, and λ_3, PC1 account for about 68.43% of the variance, while PC2 and PC3 account for 27.47% and 4.10% of the variance, respectively. Because PC3 explains only a marginal amount of the total variance, the values from PC3 are ignored. For example, to calculate the rating for a team, say the Arrows, refer back to Equation 4.1, and use $m = 2$, because only two components are needed to adequately explain the variation among the teams. So, the rating of the Arrows is given by

$$0.684 \times \begin{bmatrix} 0.470 \\ -0.982 \\ 1.847 \end{bmatrix}^T \cdot \begin{bmatrix} 0.640 \\ -0.664 \\ 0.387 \end{bmatrix} + 0.275 \times \begin{bmatrix} 0.470 \\ -0.982 \\ 1.847 \end{bmatrix}^T \cdot \begin{bmatrix} -0.352 \\ 0.195 \\ 0.916 \end{bmatrix} \approx 1.508$$

Similarly, all of the ratings can easily be calculated based on Equation 4.1, with $m = 2$. Table 4.2 augments Table 4.1 by including the WASP ratings for each team. This shows that, despite having both lost games, the high degree of difficulty in the schedules of the Arrows and Flames was sufficient to push them ahead of the undefeated Beavers.

For interpretive and mathematical ease, it is best to convert the ratings from $\mathbb{R}$

to $[0, 1]$. While Table 4.2 says that the Arrows are better than the Beavers, it is not clear how much better they are. If the Arrows have a mapped rating near 1, and the Beavers have a mapped rating near 0.75, the relationship becomes much clearer and easier to interpret. From a mathematical perspective, it is awkward to deal with both positive and negative ratings. Because all of the ratings come from centred and scaled PCA, the ratings have a mean of 0 and a variance of approximately 1. So, it is reasonable to use $\Phi(x)$, the cumulative density function (CDF) of a standard normal random variable, to map the ratings from $\mathbb{R}$ to $[0, 1]$.

Table 4.2: Initial WASP ratings from the fictional schedule.

Team	Wins	Losses	SOS	Rating
Arrows	3	1	0.6265	1.5077
Flames	4	1	0.5677	1.4340
Beavers	5	0	0.3725	0.8295
Icemen	3	2	0.5425	0.7318
Defenestrators	2	2	0.5360	0.4181
Lumberjacks	3	2	0.4328	0.0813
Giants	2	3	0.5035	-0.0523
Eagles	2	2	0.3542	-0.6599
Hitmen	1	3	0.4345	-0.7362
Knights	2	4	0.4251	-0.7947
Jazz	1	4	0.4112	-1.1520
Camels	1	5	0.3813	-1.6072

4.2 Accounting for Value of Wins and Losses

Of the six goals given at the beginning of this chapter, five of the goals have been met. The one goal that is not currently attained is the fourth goal: given the teams that each team i plays, account for which ones i defeats and which ones defeat i. Returning to Example 4.1.1, we see that the Giants defeated the Icemen; the Icemen defeated the Lumberjacks; and the Lumberjacks defeated the Giants. In the event that each of these three results were reversed (i.e., the Icemen defeated the Giants, etc.), neither the SOS values, nor win-loss records would change, and as a result the final ratings would remain the same. Certainly, this is a problem, but it is a problem that is easily overcome. For a game where team i defeats team j, i ought to gain some value dependant upon the value of the two teams, while j ought to be penalized for losing. This suggests the use of functions $W(r_w, r_l)$ and $L(r_w, r_l)$, where the first argument in both functions is the rating of the winner.

Coinciding with proposing a $W(\cdot)$ and $L(\cdot)$ is *win probability.* Given teams i and j with respective ratings r_i and r_j in $[0, 1]$, then the probability that i defeats j is given by

$$p_{ij} = \frac{r_i}{r_i + r_j}.$$

This intuitive win probability frequently appears in the literature that pertains to ranking methods, e.g., [10], [12], and [13]. Therefore, the results of games where a tie is impossible can be broken into three general cases: the winner was highly likely to win, the winner had approximately the same probability to win as the loser, and the winner was unlikely to win. $W(\cdot)$ should be constructed so that it is most rewarding to a winning team that was unlikely to win, and provides a negligible value for a winner that was heavily favoured to win. Likewise, $L(\cdot)$ should be very penalizing to a loser that had a high probability of winning, while being forgiving to a loser that was likely to lose. As a result, three requirements follow immediately from this discussion, which can be imposed on a prospective $W(r_w, r_l)$ (with corresponding requirements for $L(r_w, r_l)$). First, $W(r_w, r_l)$ must be a positive function. Under no circumstance should a team be penalized for winning a game. Although a win may not be very meaningful, that win should still award some small positive value. Secondly, $W(r_w, r_l)$

must be an increasing function in r_l. This follows from the statement that the better the rating of the losing team, the more rewarding the win ought to be for a winner with fixed rating r_w. Lastly, $W(r_w, r_l)$ must be a decreasing function in r_w. As the rating of the winning team increases, a win against a loser with fixed rating r_l is less meaningful; hence, $W(r_w, r_l)$ ought to yield a smaller reward.

While this remains true, $W(\cdot)$ and $L(\cdot)$ also need to be based on more than just the win probability. While

$$W(r_w, r_l) = L(r_w, r_l) = \frac{r_l}{r_w + r_l}$$

would have the desired effects, they would be very problematic functions. Suppose that team x had the worst rating from WASP, and that rating was arbitrarily close to 0. Furthermore, suppose that x had an 1-11 win-loss record, and that the one win came against a team y that was lowly rated but was much better than x, with, say, a rating of $r_y = 0.1$. As defined, if z defeated x, then $L(r_z, r_x)$ will always be close to 0, because r_x is close to 0. Similarly, $W(r_x, r_y)$ will be close to 1. As a result, x will likely be ranked as an above average team, instead of a last place team. Defining $W(\cdot)$ and $L(\cdot)$ by

$$\begin{aligned} W(r_w, r_l) &= \left(\frac{r_l}{r_w + r_l}\right)^2 + r_l^2, \\ L(r_w, r_l) &= \sqrt{\frac{r_l}{r_w + r_l}} + \sqrt{1 - r_w} \end{aligned} \tag{4.2}$$

resolves this problem. Now, both $W(\cdot)$ and $L(\cdot)$ are functions of the win probability, as well as the rating of the opponent. $W(\cdot)$ is constructed to pick off good wins and ignore irrelevant wins, while $L(\cdot)$ harshly penalizes bad losses and is more generous towards "acceptable" losses. With $W(\cdot)$ and $L(\cdot)$ established, the wins and losses of team i can be accounted for by summing $W(r_i, r_j)$ for each j that i defeated and then subtracting the sum of $L(r_j, r_i)$ for each j that defeated i. As is customary, the Arrows will be used as an example of how this works. First, $\Phi(1.5077) \approx 0.9342$, which is the mapped rating for the Arrows. Combining the results of each of the games that the Arrows played with $W(\cdot)$ and $L(\cdot)$ produces the following values:

$$\begin{aligned}
W(r_A, r_F) &= \left(\frac{0.9242}{0.9242 + 0.9342}\right)^2 + 0.9242^2 && \approx 1.1015 \\
W(r_A, r_G) &= \left(\frac{0.4792}{0.4792 + 0.9342}\right)^2 + 0.4792^2 && \approx 0.3445 \\
W(r_A, r_I) &= \left(\frac{0.7678}{0.7678 + 0.9342}\right)^2 + 0.7678^2 && \approx 0.7931 \\
L(r_F, r_A) &= \sqrt{\frac{0.9342}{0.9242 + 0.9342}} + \sqrt{1 - 0.9242} && \approx 0.9843
\end{aligned}$$

So the Arrows will end up with a final rating of

$$1.1015 + 0.3445 + 0.7931 - 0.9843 \approx 1.2548.$$

When this is done for all teams, the resultant final rankings, which are produced in Table 4.3, are not actually any different from the initial ranking. There is one very crucial point to note about what is how this method works, which will be illustrated in Section 5.2 where WASP will be used to rank various NCAA football seasons. The final allocation of points based on the strength of the opponents does not actually focus on *who* the opponent is, but rather upon *how strong* the opponent is. By defeating the Arrows, the Flames were rewarded for getting a victory against a team with a rating of 0.9342, if there were multiple teams with a rating of 0.9342, they would get the same benefit.

Table 4.3: Final WASP ratings from the fictional schedule.

Team	Wins	Losses	SOS	Init. Rating	Final Rating
Arrows	3	1	0.6265	1.5077	1.2548
Flames	4	1	0.5677	1.4340	1.2030
Beavers	5	0	0.3725	0.8295	0.4415
Icemen	3	2	0.5425	0.7318	-1.2241
Defenestrators	2	2	0.5360	0.4181	-1.5608
Lumberjacks	3	2	0.4328	0.0813	-1.7204
Giants	2	3	0.5035	-0.0523	-2.0886
Eagles	2	2	0.3542	-0.6599	-2.4106
Hitmen	1	3	0.4345	-0.7362	-3.0824
Knights	2	4	0.4251	-0.7947	-3.9911
Jazz	1	4	0.4112	-1.1520	-4.2601
Camels	1	5	0.3813	-1.6072	-5.7733

4.3 A Minor Patch for WASP

In example 4.1.1, the PCA loading matrix was of the form

$$\boldsymbol{V} = \begin{array}{c} \\ \text{W} \\ \text{L} \\ \text{SOS} \end{array} \begin{array}{c} \begin{array}{ccc} \text{PC1} & \text{PC2} & \text{PC3} \end{array} \\ \begin{pmatrix} + & - & + \\ - & + & + \\ + & + & + \end{pmatrix} \end{array},$$

where the $+$ entries are positive, and the $-$ entries are negative. Depending on the data, it would be possible for $\boldsymbol{V}$ to be of the form

$$\boldsymbol{V} = \begin{array}{c} \\ \text{W} \\ \text{L} \\ \text{SOS} \end{array} \begin{array}{c} \begin{array}{ccc} \text{PC1} & \text{PC2} & \text{PC3} \end{array} \\ \begin{pmatrix} + & + & + \\ - & - & + \\ + & - & + \end{pmatrix} \end{array}.$$

This poses a minor annoyance because PC2 is now the negative of what it was previously. Furthermore, using this $\boldsymbol{V}$ to perform the rating in Equation 4.1 would give teams with weak SOS value a better rating than a team with a strong SOS value based on PC2. Now, the columns of $\boldsymbol{V}$ are orthogonal to each other, which means that PC1 is also orthogonal to any scalar multiple of PC2, including −PC2. So, $\boldsymbol{V}$ is constructed so that *wins* in PC1, *SOS* in PC2, and *wins* in PC3 are all positive. This patch will ensure that the teams with the best rating from WASP are the teams that win the most frequently and also have higher SOS values.

4.4 WASP Pseudocode

To concisely tie up the full WASP ranking algorithm, the pseudocode for WASP is described to conclude the chapter. This will just describe the basic WASP where only wins, losses, and SOS are used for ranking. Other qualities such as extra variables or weighting the value of wins and losses can be used, and the pseudocode would extend analogously.

Algorithm 1 WASP algorithm

procedure WASP(s) ▷ Compute rankings of all teams based on schedule s
    $T \leftarrow \{\}$ ▷ Set of teams
    for each distinct team $t \in s$ **do**
        $w_t \leftarrow 0$ ▷ wins by t
        $l_t \leftarrow 0$ ▷ losses by t
        $W_t \leftarrow \{\}$ ▷ Set of teams t beats
        $L_t \leftarrow \{\}$ ▷ Set of teams that beat t
        $T_t \leftarrow \{\}$ ▷ Set of teams t ties
        $R_{i,t} \leftarrow 0$ ▷ Initial rating of t

$R_{f,t} \leftarrow 0$ ▷ Final rating of t
$S_t \leftarrow 0$ ▷ SOS rating of t
$T \leftarrow T \cup \{t\}$
end for
for each game $g \in s$ **do** ▷ gives road and home teams, and points scored
$R \leftarrow$ road team
$H \leftarrow$ home team
$p_r \leftarrow$ road points
$p_h \leftarrow$ home points
if $p_r > p_h$ **then** ▷ Road team wins
$W_R \leftarrow W_R \cup \{H\}$
$L_H \leftarrow L_H \cup \{R\}$
$w_R \leftarrow w_R + 1$
$l_H \leftarrow l_H + 1$
else if $p_r < p_h$ **then** ▷ Home team wins
$L_R \leftarrow L_R \cup \{H\}$
$W_H \leftarrow W_H \cup \{R\}$
$l_R \leftarrow l_R + 1$
$l_H \leftarrow w_H + 1$
else ▷ Tie game
$T_R \leftarrow T_R \cup \{H\}$
$T_H \leftarrow T_H \cup \{R\}$
$w_R \leftarrow w_R + \frac{1}{2}$
$l_R \leftarrow l_R + \frac{1}{2}$
$w_H \leftarrow w_H + \frac{1}{2}$

$l_H \leftarrow l_H + \frac{1}{2}$
end if
end for
for each team $t \in T$ **do**
calculate S_t ▷ Based on section3.2
end for
$D \leftarrow [[w_t, l_t, S_t]$ **for each** team $t]$ ▷ D is the data matrix
$D \leftarrow \text{scale}(D)$ ▷ Centre and scale D
$V, \vec{\lambda} \leftarrow \text{PCA}(D)$ ▷ V is PCA loading matrix, $\vec{\lambda}$ is proportions of variance
for each team $t \in T$ **do**
$R_{i,t} \leftarrow \Phi\left(\lambda_1 (D \times V)_{t,1} + \lambda_2 (D \times V)_{t,2}\right)$ ▷ $\Phi(\cdot)$ is $N(0,1)$ CDF
end for
for each team $t \in T$ **do**
for each team $u \in W_t$ **do**
$R_{f,t} \leftarrow R_{f,t} + W(R_{i,t}, R_{i,u})$ ▷ $W(\cdot)$ as defined in equation 4.2
end for
for each team $u \in L_t$ **do**
$R_{f,t} \leftarrow R_{f,t} - L(R_{i,u}, R_{i,t})$ ▷ $L(\cdot)$ as defined in equation 4.2
end for
for each team $u \in T_t$ **do**
$R_{f,t} \leftarrow R_{f,t} + \frac{W(R_{i,t}, R_{i,u}) - L(R_{i,u}, R_{i,t})}{2}$
end for
end for
return T sorted by $R_{f,\cdot}$ in descending order ▷ Final Ranking
end procedure

Chapter 5

Results

This chapter will take the WASP algorithm described in Chapter 4 and apply it to various NCAA football seasons. The first half of the chapter will investigate how variations of WASP conform to the "expert" opinions of the polls for the 2008 season. The second half of the chapter will look at who WASP determines should be the champion or play for the championship over about twenty years of data. All schedules and boxscores from games where both teams are FBS teams come from `sports-reference.com` [24]. Unfortunately, `sports-reference.com` does not include the yards gained or turnovers committed, which are needed for the expanded variable set in Method 3, when one of the teams is an FCS team. Fortunately, ESPN [25] does provide this information, so ESPN is used in this case to gather the yards gained and turnovers committed.

5.1 Ranking for the 2008 Season

In Section 2.2, the 2008 season was one example given of controversy in the BCS. Here, WASP is used in multiple variations to rank teams and compared against the

AP, Coaches, and BCS polls. To provide this comparison, a variety of versions of WASP will be included. The first version will use WASP exactly as described in Example 4.1.1, using only wins, losses, and SOS as data entries for PCA. The second method will be similar to the first except the value of games will be weighted based on the location of the game and at what point in the season the game occurred. The third method will take WASP and modify it slightly to add other information about the teams to the data. The final method will be similar to 4.1.1, except it will redefine "win" and "loss" to create different data entries than before. For unknown reasons, ESPN did not have any game statistics available for three games: Weber State Hawaii, Western Kentucky Eastern Kentucky, and Murray State Western Kentucky. For these games, it is assumed that the yards and turnovers for both teams are 0. At the conclusion of this section, the result of these four methods, before and after looping over the schedule, are compared against the BCS poll.

Method 1: Regular WASP

Using the exact same routine as in the hypothetical league of Example 4.1.1 produces a PCA matrix and eigenvalues that are fairly similar to the ones from before. After accumulating all of the wins and losses, and calculating the strength of schedule for each team, centred and standard deviation-scaled PCA yields loading matrix V and a vector of variance proportions $\vec{\lambda}$ given by

$$\begin{array}{cc} & \begin{array}{ccc} \text{PC1} & \text{PC2} & \text{PC3} \end{array} \\ \boldsymbol{V} = \begin{array}{c} \text{W} \\ \text{L} \\ \text{SOS} \end{array} & \begin{pmatrix} 0.6689 & -0.2478 & 0.7009 \\ -0.6773 & 0.1854 & 0.7120 \\ 0.3064 & 0.9509 & 0.0438 \end{pmatrix} \end{array}, \quad \vec{\lambda} = \begin{pmatrix} 2.0881 \\ 0.8891 \\ 0.0228 \end{pmatrix}.$$

As before, PC1 loads on wins against losses, while PC2 is dominated by SOS, and PC3 loads on wins and losses. Similarly, PC1 accounts for nearly 70% of the variation; PC2 accounts for just under 30% of the variation; and PC3 is responsible for a negligible amount of the variation. Also of interest are the teams that maximize PC1 and maximize PC2. Figure 5.1 shows the PC1 scores against the PC2 scores, including labels for the ten teams with highest PC1 value and the three teams with the highest PC2 value. Many of the teams with PC2 values that are greater than 1 also have negative PC1 values. This is to be expected, because those teams play the most difficult schedules; hence, they would not be expected to win as frequently as a team that played a weak schedule. Similarly, many of the teams with large PC1 values have negative PC2 values. Again, this is expected, because teams that play poor opponents should win a majority of their games. But then there are the exceptional teams, such as Oklahoma, Florida, and Texas, in the top right corner that have large PC1 and PC2 values. Not only do they face difficult opponents, but they also win a almost all of their games. It should come as no surprise that Oklahoma, Florida, and Texas take the top 3 positions in the WASP ranking.

Method 2: WASP with Variable Win Values

In Section 3.3, the impact of home field advantage was considered. This method will vary the value of each wins or loss based on the location of the game and the week in which the game was played. Jamieson [9] cites various studies on home field advantage which indicate that the home team wins between 55% and 60% of all games. As opposed to attempting to quantify the different home field effects that each team benefits from, a flat value of 1.3 wins for a road win (1.3 losses for a home loss) and 0.7 wins for a home win (0.7 losses for a road loss) will be used. In the event of a game

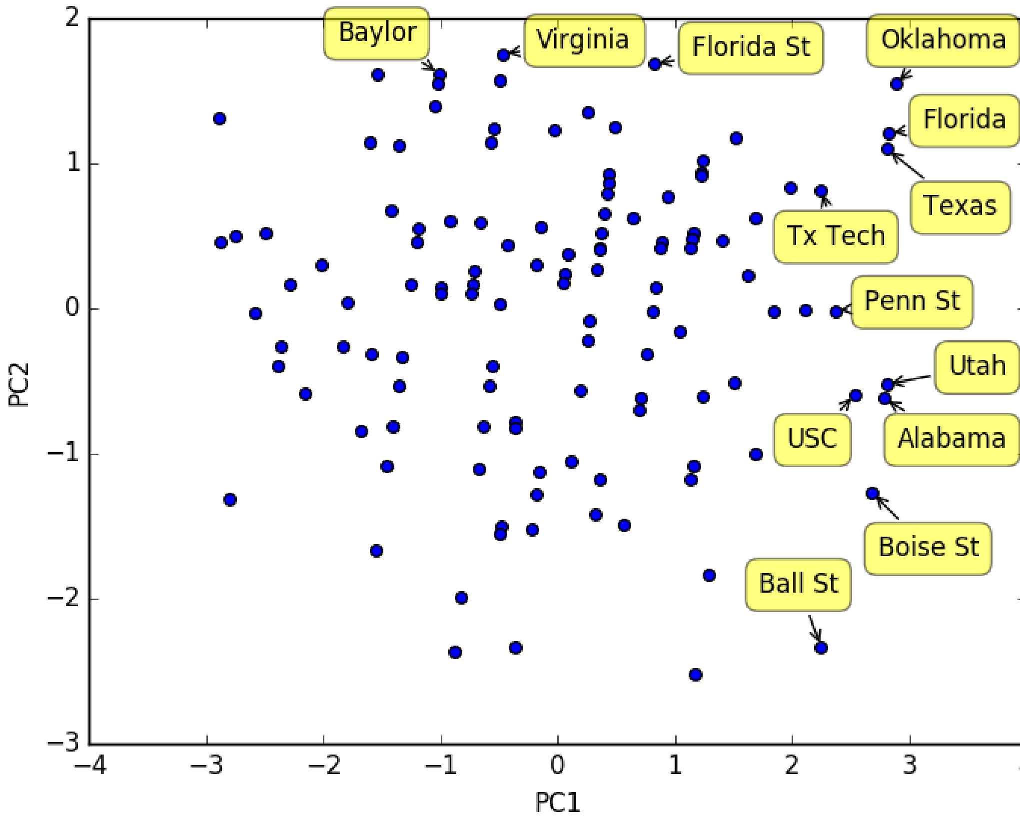

Figure 5.1: A plot of PC1 scores versus PC2 scores.

played at a neutral site, a win or loss will be worth the standard 1 unit. Additionally, the weekly game weights can be multiplied to the home field advantage modifier. To avoid extreme complexity, the linear weighting

$$w_i = 1 + \frac{i-1}{14}$$

will be used. Under these two modifications, the minimum value for a win is a home win in the first week, which will award 0.7 wins. While the maximum win value is 2.6, which is awarded by winning a road game in the fifteenth and final week of the regular season. After adding these effects, it is apparent that the columns of the loading matrix $\boldsymbol{V}$ are more extreme in this case than before. Now, PC1 is just wins against losses with an SOS value that is almost 0. As a result, the win and loss values for PC2 are much closer to 0 than anywhere else, while the SOS value is essentially 1. Lastly, PC3 loads on wins and losses, and SOS is negligible.

$$\boldsymbol{V} = \begin{array}{c} \\ \text{W} \\ \text{L} \\ \text{SOS} \end{array} \begin{array}{c} \begin{array}{ccc} \text{PC1} & \text{PC2} & \text{PC3} \end{array} \\ \begin{pmatrix} 0.7042 & -0.0594 & 0.7075 \\ -0.7032 & 0.0801 & 0.7065 \\ 0.0986 & 0.9950 & -0.0146 \end{pmatrix} \end{array}, \quad \vec{\lambda} = \begin{pmatrix} 1.9355 \\ 0.9910 \\ 0.0735 \end{pmatrix}.$$

Method 3: Expanded PCA variables

The intention of this third method is to test whether or not the assumption of simplicity is sufficient, or if more information about teams is required to obtain a sensible ranking. In order to do this, the *point differential* (the difference between *points for* and *points against*); *yard differential* (the difference between *yards for* and *yards against*), and *turnover differential* (the difference between *turnovers forced* and *turnovers committed*) will be additional statistics about the teams that will be used by WASP. Naturally, factoring in these three additional variables will create a very different result than the result based on just the three variables from Method 1.

$$\boldsymbol{V} = \begin{array}{c} \\ \text{W} \\ \text{L} \\ \text{SOS} \\ \text{PTS} \\ \text{YDS} \\ \text{TOs} \end{array} \begin{array}{c} \begin{array}{cccccc} \text{PC1} & \text{PC2} & \text{PC3} & \text{PC4} & \text{PC5} & \text{PC6} \end{array} \\ \begin{pmatrix} 0.495 & -0.027 & -0.103 & 0.502 & -0.302 & 0.633 \\ -0.503 & 0.017 & 0.036 & -0.410 & 0.068 & 0.757 \\ 0.167 & 0.560 & 0.808 & -0.046 & -0.048 & 0.039 \\ 0.500 & -0.033 & -0.050 & -0.200 & 0.827 & 0.152 \\ 0.474 & -0.111 & -0.090 & -0.732 & -0.467 & -0.033 \\ -0.003 & 0.820 & -0.570 & -0.050 & -0.008 & -0.020 \end{pmatrix} \end{array},$$

$$\vec{\lambda} = \begin{pmatrix} 3.7543 & 1.0739 & 0.8545 & 0.2259 & 0.0711 & 0.0202 \end{pmatrix}^T.$$

The loading matrix shows that PC1 loads wins, point differential, and yard differential against losses. PC2 loads on turnover differential with a lesser impact from SOS, while PC3 weighs SOS against turnover differential. PC4 loads wins against yard differential and losses. PC5 loads on point differential with lesser opposition from wins and yard differential. Lastly, PC6 loads on wins and losses. Additionally, the large drop offs from λ_1 to λ_2 and from λ_3 to λ_4 suggest that either one component be used or three components be used. To explain a larger amount of the variation, the rankings at the conclusion of this section will use three components.

Method 4: Alternate Win-Loss Definitions

In [12], Massey mentions an alternative view of a *win*. Instead of using the conventional definition for *win* and *loss*, *win* can be redefined as "scoring a point"; likewise, *loss* would be redefined as "surrendering a point." For example, when Florida played Miami (FL), Florida "won" 26-3. Accordingly, Florida would get credit for 26 wins and 3 losses against Miami (FL). Following these redefinitions, WASP proceeds analogously to the outline from Chapter 4, with minor alterations where necessary. First, SOS is calculated under the new definitions for *win* and *loss*. For example, Florida has over 3000 combined first degree wins and losses and over 37000 second degree wins and losses. These numbers are substantially larger than in Method 1, where there would be about 70 and 700 wins and losses, respectively. In this circumstance, the loading matrix and eigenvalues from PCA are quite different:

$$\boldsymbol{V} = \begin{array}{c} \\ \text{W} \\ \text{L} \\ \text{SOS} \end{array} \begin{array}{c} \begin{array}{ccc} \text{PC1} & \text{PC2} & \text{PC3} \end{array} \\ \begin{pmatrix} 0.5838 & -0.5595 & -0.5883 \\ -0.6750 & 0.0681 & -0.7357 \\ 0.4512 & 0.8260 & -0.3379 \end{pmatrix} \end{array}, \quad \vec{\lambda} = \begin{pmatrix} 1.3247 \\ 0.9489 \\ 0.7264 \end{pmatrix}.$$

Now, the first two components only account for 75% of the variation, so the third component is arguably worth considering. Also, SOS has much more impact on PC1 than before, because PC1 loads wins and SOS against losses. Also, PC2 is no longer just loading on SOS, but rather is loading SOS against wins. Lastly, PC3 is a weighted average of the three variables.

Summary of the Four Variations

As might be expected for the differing qualities that these four WASP variations are based on, the initial and final rankings from Tables 5.1 and 5.2 tell very different stories. The standard version of WASP (Method 1), described in Example 4.1.1, as well as WASP with the expanded variable set produced initial and final rankings that are very similar to the BCS poll. Of interest is the how highly USC finished in the initial rankings, only to fall back down in the final rankings of the expanded variable

Table 5.1: Comparison of WASP rankings with BCS poll for 2008 – pre-looping.

#	BCS	WASP-1	WASP-2	WASP-3	WASP-4
1	Oklahoma	Oklahoma	Oklahoma	Oklahoma	Florida
2	Florida	Florida	Florida	Cincinnati	Ohio St.
3	Texas	Texas	USC	Florida	Oklahoma
4	Alabama	Texas Tech	Texas	Texas	Penn St.
5	USC	Utah	Penn St.	Boise St.	USC
6	Utah	Alabama	Boise St.	Pittsburgh	Texas
7	Texas Tech	Ohio St.	Texas Tech	Utah	Florida St.
8	Penn St.	Penn St.	Alabama	Texas Tech	Georgia
9	Boise St.	USC	Utah	USC	Texas Tech
10	Ohio St.	Cincinnati	TCU	Ohio St.	Mississippi
11	TCU	Georgia	Ohio St.	Penn St.	TCU
12	Cincinnati	Boise St.	Georgia	Virginia Tech	Clemson
13	Oklahoma St.	Pittsburgh	Ball St.	Florida St.	South Carolina
14	Georgia Tech	TCU	Missouri	Georgia Tech	Oregon St.
15	Georgia	Michigan St.	Oklahoma St.	TCU	Virginia

set. USC received a high initial ranking because they scored almost five points for every point allowed, and gained more than two yards for every yard allowed, but then they fall back because they did not have very strong wins. Because both of these rankings are similar, the first method would be preferable because it requires fewer inputs and is simpler to implement.

As the third method showed, weighting games heavily penalized teams that lost games at the end of the year. Alabama lost only one regular season game, but it was the SEC championship game against Florida. This lone loss became so damaging that Alabama was ranked #17 in the initial ranking and #10 in the final ranking. Conversely, Cincinnati had two loses: at Oklahoma in the second week of the year, and at Connecticut in the ninth week of the year. Because they were both road games and they occurred fairly early in the schedule, the negative impact is mitigated, and Cincinnati jumped to third in the final rankings. Also of interest is the difference between the two undefeated mid-major schools: Boise State and Utah. Utah concluded

Table 5.2: Comparison of WASP rankings with BCS poll for 2008 – post-looping.

#	BCS	WASP-1	WASP-2	WASP-3	WASP-4
1	Oklahoma	Oklahoma	Oklahoma	Florida	USC
2	Florida	Florida	Florida	Oklahoma	Florida
3	Texas	Texas	Texas	Cincinnati	Penn St.
4	Alabama	Utah	Utah	Boise St.	Alabama
5	USC	Texas Tech	Texas Tech	Utah	Texas
6	Utah	Alabama	Alabama	Texas	Oklahoma
7	Texas Tech	Penn St.	USC	Penn St.	Iowa
8	Penn St.	USC	Penn St.	Texas Tech	Ohio St.
9	Boise St.	Boise St.	Boise St.	USC	TCU
10	Ohio St.	Ohio St.	Ohio St.	Alabama	Georgia Tech
11	TCU	Cincinnati	Cincinnati	Pittsburgh	Mississippi
12	Cincinnati	Georgia Tech	Pittsburgh	Georgia Tech	Northwestern
13	Oklahoma St.	Ball St.	TCU	Virginia Tech	Wake Forest
14	Georgia Tech	TCU	Georgia Tech	Ball St.	Clemson
15	Georgia	Pittsburgh	Ball St.	Ohio St.	Arizona

the season with TCU and Brigham Young, two of their toughest opponents, near the end of the year, but both of those games were played at Utah. On the other hand, Boise State had very few notable opponents, but they played their toughest opponents on the road. The toughest opponent that Boise State faced was Oregon, in the fourth week, but because it was a road game, that game became more beneficial to Boise State than any game that Utah played. More important than the Oregon game was a late season win by Boise State at Nevada. Although Nevada was a slightly above average opponent, the combination of the game being near the end of the season and on the road made that game more valuable than even the Oregon game. Based on these results, it is clear why weighting the value of games pushed Boise State ahead of Utah.

Lastly, the alternative definitions for *win* and *loss* produced wildly different results from the BCS poll. Iowa, Mississippi, Wake Forest, Clemson, and Arizona all finished in the top 15 under these new definitions, yet none of these teams finished in the top 25 of the BCS poll. The reason that these teams were highly ranked here, but not in the BCS poll, is because they lost games, but they were fairly close losses. For example, Arizona lost five games, but they lost those five games by a combined twenty-eight points. In contrast, they won seven regular season games by a combined 217 points. Similarly, Iowa lost four games by a combined twelve points, but won eight games by a combined 216 points. This illustrates the problem mentioned in Chapter 3; i.e., teams are able to use blowout wins to overcome close losses and boost their ranking.

5.2 Comprehensive Historical Rankings

Based on the discussion from the previous section, the version of WASP used for Example 4.1.1 provided very good results and also used a minimal number of variables. As a result, this section will analyze twenty years of WASP rankings, and then provide some commentary on years of controversy, years where WASP diverges from consensus opinions, and also the example seasons cited in Section 2.2. In the appendix, Table A.1 contains the top 5 from the AP Poll, Coaches Poll, BCS or CFP Poll (where it exists), and from WASP for every season since 1997, as well as the 1989 season. As an aside, the penultimate Coaches Poll results do not appear to be available from 1998-2001, so there is no Coaches Poll given in those years. For the 1989 and 1997 seasons, the results will include all games played, while the seasons from 1998-2015 will exclude the bowl games. As can be seen in Table A.1, WASP provides very sensible rankings, that are generally in agreement with the polls. For a given year, bold text entries represent what teams would play in the playoff or championship game, or be declared the champion. Of the forty-two bold text entries, WASP agrees on thirty-five of the participants in the playoff or championship game or the declared champion, for an agreement percentage of 83%. One minor bookkeeping note is that in 2012, Ohio State was ineligible for the championship due to NCAA sanctions. As a result, Ohio State would appear in neither the BCS nor Coaches Polls. For the sake of interest, Oregon placed #6 in WASP in 2012.

In 2008, Utah finished the season as one of the two undefeated teams, but was ranked below a pair of teams with one loss. Utah does fare better in WASP, going from #6 in the BCS poll to #4 in WASP, but that would still not be good enough to get them into the championship game. In 2004, Texas was trailing California entering the final week of the season when the head coach of Texas campaigned for the voters to rank Texas ahead of California. Ultimately, the combination of some voters moving Texas ahead of California and the computer polls favouring Texas pushed Texas to #4 and dropped California to #5. While WASP agrees that Texas should have been ranked ahead of California due to Texas having better wins than California, the vote pandering is a clear flaw that does not ensnare mathematics.

The 2003 season had the controversy of the AP #1 team finishing #3 in the BCS poll; hence, out of the national championship game. The second controversy was three one-loss mid-major schools finishing at best #14 in either human poll and out of the lucrative BCS games. All three of the mid-major schools under consideration do much better when ranked by WASP. Miami (OH) finished #9 (#11 in the BCS poll); TCU finished #13 (#18 in the BCS poll); and Boise State finished #14 (#17 in the BCS poll).

An example of where WASP differs significantly from the polls is the 1989 season. The reason for the difference is likely the result of the point raised at the end of Section 4.2. Miami (FL) captured the #1 ranking in both the AP and Coaches polls after defeating then #1 Notre Dame in the final week of the regular season, followed by a victory over #7 Alabama in the Sugar Bowl. After losing to Miami (FL), Notre Dame fell to #4 in the AP poll before defeating new #1 Colorado in the Orange Bowl. After their Orange Bowl victory, Notre Dame finished #2 in the AP Poll and #3 in the Coaches Poll. Although this victory over Notre Dame is considered, Miami (FL) is still unable to surpass Notre Dame in the WASP ranking. The reason why Miami (FL) did not fare well in the WASP rankings was because they played a very weak schedule: only their victories against Michigan State, Pittsburgh, Notre Dame, and Alabama resulted in Miami (FL) gaining more than 0.4 points from $W(\cdot)$. And while Miami (FL) had victories over Michigan State and Pittsburgh, both of those teams also lost to Notre Dame. Additionally, Notre Dame had significant wins over Virginia, Michigan, USC, Penn State, and Colorado, all of which gave Notre Dame over 1 point from $W(\cdot)$. Coleman [4] discusses why minimizing the number of occurrences where i defeats j, but j is ranked ahead of i (a violation of the ranking) is an important consideration for ranking systems. In general, this is a desirable quality for a ranking system to possess, but in this case, it is clear that Notre Dame faced much better competition than Miami (FL) over the course of the year, and that Notre Dame taking the #1 position in the WASP rankings while Miami (FL) finished #1 should not necessarily be viewed as a problem.

Chapter 6

Conclusion

This thesis has presented the WASP method which is a very concise application of PCA to the problem of ranking. The WASP method attains six natural qualities for a ranking method:

1. WASP contains no initial biases.
2. WASP is concerned with only results from the year or schedule in question.
3. WASP accounts for strength of schedule.
4. WASP accounts for teams defeated and teams lost to.
5. WASP adheres to Occam's Razor.
6. WASP provides final results that mesh with intuition.

As indicated, WASP reasonably agrees with the "expert" opinion of the national polls. Lastly, the mathematical power present in abstraction allows for WASP to be of use in any situation in which a natural definition of "win" and "game" appear for ranking purposes. The problems associated with the polls of NCAA football provide insight into the ways in which rankings can become biased. One of the most important features of WASP is that it is an unbiased method for ranking, unlike the polls. Because WASP is unbiased, it provides an equal opportunity for every team to achieve a high ranking, as well as the rewards associated with a high ranking.

Appendix A

Table of WASP Rankings

Here, a table of the top 5 teams in the AP, Coaches, and BCS/CFP (where it exists) polls are compared against the rankings from the main version of WASPfrom the 1989 and 1997–2015 seasons. The high level of agreement between WASP and the polls indicates that WASP has value as a tool for ranking college football teams

Table A.1: Comparison of WASP rankings with polls for 1989 and 1997–2015.

2015	AP Poll	Coaches Poll	CFP Poll	WASP
1	**Clemson**	**Clemson**	**Clemson**	**Alabama**
2	**Alabama**	**Alabama**	**Alabama**	**Michigan St.**
3	**Michigan St.**	**Oklahoma**	**Michigan St.**	**Clemson**
4	**Oklahoma**	**Michigan St.**	**Oklahoma**	**Iowa**
5	Stanford	Ohio St.	Iowa	Stanford

2014	AP Poll	Coaches Poll	CFP Poll	WASP
1	**Alabama**	**Alabama**	**Alabama**	**Florida St.**
2	**Florida St.**	**Florida St.**	**Oregon**	**Alabama**
3	**Oregon**	**Oregon**	**Florida St.**	**Ohio St.**
4	**Baylor**	**Ohio St.**	**Ohio St.**	**Oregon**
5	Ohio St.	Baylor	Baylor	TCU

2013	AP Poll	Coaches Poll	BCS Poll	WASP
1	**Florida St.**	**Florida St.**	**Florida St.**	**Auburn**
2	**Auburn**	**Auburn**	**Auburn**	**Florida St.**
3	Alabama	Alabama	Alabama	Stanford
4	Michigan St.	Michigan St.	Michigan St.	Alabama
5	Stanford	Baylor	Stanford	Michigan St.

2012	AP Poll	Coaches Poll	BCS Poll	WASP
1	**Notre Dame**	**Notre Dame**	**Notre Dame**	**Notre Dame**
2	**Alabama**	**Alabama**	**Alabama**	**Florida**
3	Ohio St.	Oregon	Florida	**Ohio St.**
4	Florida	Florida	Oregon	Alabama
5	Oregon	Georgia	Kansas St.	Stanford

2011	AP Poll	Coaches Poll	BCS Poll	WASP
1	**LSU**	**LSU**	**LSU**	**LSU**
2	**Alabama**	**Alabama**	**Alabama**	**Oklahoma St.**
3	Oklahoma St.	Oklahoma St.	Oklahoma St.	Alabama
4	Stanford	Stanford	Stanford	Kansas St.
5	USC	Oregon	Oregon	Stanford

2010	AP Poll	Coaches Poll	BCS Poll	WASP
1	**Auburn**	**Oregon**	**Auburn**	**Auburn**
2	**Oregon**	**Auburn**	**Oregon**	**Oklahoma**
3	TCU	TCU	TCU	Oregon
4	Wisconsin	Wisconsin	Stanford	TCU
5	Stanford	Stanford	Wisconsin	Michigan St.

2009	AP Poll	Coaches Poll	BCS Poll	WASP
1	**Alabama**	**Alabama**	**Alabama**	**Alabama**
2	**Texas**	**Texas**	**Texas**	**Texas**
3	TCU	TCU	Cincinnati	Florida
4	Cincinnati	Cincinnati	TCU	Cincinnati
5	Florida	Florida	Florida	TCU

2008	AP Poll	Coaches Poll	BCS Poll	WASP
1	**Florida**	**Oklahoma**	**Oklahoma**	**Oklahoma**
2	**Oklahoma**	**Florida**	**Florida**	**Florida**
3	Texas	Texas	Texas	Texas
4	Alabama	Alabama	Alabama	Utah
5	USC	USC	USC	Texas Tech

2007	AP Poll	Coaches Poll	BCS Poll	WASP
1	**Ohio St.**	**Ohio St.**	**Ohio St.**	**LSU**
2	**LSU**	**LSU**	**LSU**	**Virginia Tech**
3	Oklahoma	Oklahoma	Virginia Tech	Georgia
4	Georgia	Georgia	Oklahoma	Missouri
5	Virginia Tech	Virginia Tech	Georgia	Ohio St.

2006	AP Poll	Coaches Poll	BCS Poll	WASP
1	**Ohio St.**	**Ohio St.**	**Ohio St.**	**Florida**
2	**Florida**	**Florida**	**Florida**	**Ohio St.**
3	Michigan	Michigan	Michigan	USC
4	LSU	LSU	LSU	Michigan
5	Louisville	Wisconsin	USC	Louisville

2005	AP Poll	Coaches Poll	BCS Poll	WASP
1	**USC**	**USC**	**USC**	**Texas**
2	**Texas**	**Texas**	**Texas**	**USC**
3	Penn St.	Penn St.	Penn St.	Penn St.
4	Ohio St.	Ohio St.	Ohio St.	Ohio St.
5	Notre Dame	Oregon	Oregon	Virginia Tech

2004	AP Poll	Coaches Poll	BCS Poll	WASP
1	**USC**	**USC**	**USC**	**USC**
2	**Oklahoma**	**Oklahoma**	**Oklahoma**	**Oklahoma**
3	Auburn	Auburn	Auburn	Auburn
4	California	California	Texas	Texas
5	Utah	Texas	California	Utah

2003	AP Poll	Coaches Poll	BCS Poll	WASP
1	**USC**	**Oklahoma**	**Oklahoma**	**Oklahoma**
2	**LSU**	**USC**	**LSU**	**LSU**
3	Oklahoma	LSU	USC	USC
4	Michigan	Michigan	Michigan	Ohio St.
5	Texas	Georgia	Ohio St.	Michigan

2002	AP Poll	Coaches Poll	BCS Poll	WASP
1	**Miami (FL)**	**Miami (FL)**	**Miami (FL)**	**Ohio St.**
2	**Ohio St.**	**Ohio St.**	**Ohio St.**	**Miami (FL)**
3	Iowa	Iowa	Georgia	Georgia
4	Georgia	Georgia	USC	USC
5	USC	USC	Iowa	Oklahoma

2001	AP Poll	Coaches Poll	BCS Poll	WASP
1	**Miami (FL)**		**Miami (FL)**	**Miami (FL)**
2	**Oregon**		**Nebraska**	**Nebraska**
3	Colorado		Colorado	Tennessee
4	Nebraska		Oregon	Colorado
5	Florida		Florida	Oregon

2000	AP Poll	Coaches Poll	BCS Poll	WASP
1	**Oklahoma**		**Oklahoma**	**Oklahoma**
2	**Miami (FL)**		**Florida St.**	**Florida St.**
3	Florida St.		Miami (FL)	Washington
4	Washington		Washington	Miami (FL)
5	Oregon St.		Virginia Tech	Virginia Tech

1999	AP Poll	Coaches Poll	BCS Poll	WASP
1	**Florida St.**		**Florida St.**	**Florida St.**
2	**Virginia Tech**		**Virginia Tech**	**Alabama**
3	Nebraska		Nebraska	Nebraska
4	Wisconsin		Alabama	Virginia Tech
5	Alabama		Tennessee	Michigan St.

1998	AP Poll	Coaches Poll	BCS Poll	WASP
1	**Tennessee**		**Tennessee**	**Tennessee**
2	**Florida St.**		**Florida St.**	**Florida St.**
3	Ohio St.		Kansas St.	UCLA
4	Kansas St.		Ohio St.	Texas A&M
5	Arizona		UCLA	Kansas St.
1997	AP Poll	Coaches Poll		WASP
1	**Michigan**	**Nebraska**		**Nebraska**
2	Nebraska	Michigan		Tennessee
3	Florida St.	Florida St.		Michigan
4	Florida	North Carolina		Florida
5	UCLA	UCLA		Florida St.
1989	AP Poll	Coaches Poll		WASP
1	**Miami (FL)**	**Miami (FL)**		**Notre Dame**
2	Notre Dame	Florida St.		Tennessee
3	Florida St.	Notre Dame		Miami (FL)
4	Colorado	Colorado		Colorado
5	Tennessee	Tennessee		Florida St.

Bibliography

[1] Arrow, Kenneth J. “A difficulty in the concept of social welfare.” The Journal of Political Economy 58 (1950): 328-346.

[2] Borghesi, Richard. “The home team weather advantage and biases in the NFL betting market.” Journal of Economics and Business 59.4 (2007): 340-354.

[3] Boyles, Bob, and Paul Guido. The USA Today College Football Encyclopedia. Skyhorse Publishing Inc., 2008.

[4] Coleman, B. Jay. “Minimizing game score violations in college football rankings.” Interfaces 35.6 (2005): 483-496.

[5] Coleman, B. Jay, et al. “Voter bias in the associated press college football poll.” Journal of Sports Economics 11.4 (2010): 397-417.

[6] Colley, Wesley N. “Colley’s bias free college football ranking method: The Colley matrix explained.” Princeton University, Princeton (2002).

[7] Frye, Sean. “Why Notre Dame Is the Only Team Sad to See the BCS Go”. Retrieved on July 31, 2016 from <http://bleacherreport.com/articles/1902645-why-notre-dame-is-the-only-team-sad-to-see-the-bcs-go>

[8] Goff, Brian. "An assessment of path dependence in collective decisions: evidence from football polls." Applied Economics 28.3 (1996): 291-297.

[9] Jamieson, Jeremy P. "The home field advantage in athletics: A meta-analysis." Journal of Applied Social Psychology 40.7 (2010): 1819-1848.

[10] Keener, James P. "The Perron-Frobenius theorem and the ranking of football teams." SIAM review 35.1 (1993): 80-93.

[11] Larson, Jeffrey, Mikael Johansson, and Mats Carlsson. "An integrated constraint programming approach to scheduling sports leagues with divisional and round-robin tournaments." *International Conference on AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*. Springer International Publishing, 2014.

[12] Massey, Kenneth. "Statistical Models Applied to the Rating of Sports Teams." Honors project in mathematics, Bluefield College, Bluefield, VA., (1997).

[13] Mease, David. "A penalized maximum likelihood approach for the ranking of college football teams independent of victory margins." The American Statistician 57.4 (2003): 241-248.

[14] Nevill, Alan M., Nigel J. Balmer, and A. Mark Williams. "The influence of crowd noise and experience upon refereeing decisions in football." Psychology of Sport and Exercise 3.4 (2002): 261-272.

[15] Nevill, Alan M., Sue M. Newell, and Sally Gale. "Factors associated with home advantage in English and Scottish soccer matches." Journal of Sports Sciences 14.2 (1996): 181-186.

[16] Pace, Andrew, and Albert V. Carron. "Travel and the home advantage." Canadian Journal of Sport Sciences (1992).

[17] Page, Lawrence, et al. "The pagerank citation ranking: Bringing order to the web". Technical report, Stanford Digital Library Technologies Project, 1998.

[18] Park, Juyong, and Mark EJ Newman. "A network-based ranking system for US college football." Journal of Statistical Mechanics: Theory and Experiment 2005.10 (2005): P10014.

[19] Pope, Devin G., and Jaren C. Pope. "The impact of college sports success on the quantity and quality of student applications." Southern Economic Journal, Forthcoming (2008).

[20] Strauss, B. E. R. N. D. "The impact of supportive spectator behavior on performance in team sports." International Journal of Sport Psychology 33.4 (2002): 372-390.

[21] Thirer, Joel, and Mark S. Rampey. "Effects of abusive spectators' behavior on performance of home and visiting intercollegiate basketball teams." Perceptual and Motor Skills 48.3 suppl (1979): 1047-1053.

[22] Wetzel, Dan, Josh Peter, and Jeff Passan. Death to the BCS: The definitive case against the Bowl Championship Series. Penguin, 2010.

[23] College Football Playoff Act of 2009, H.R. 390, 111th Cong. (2009).

[24] <http://www.sports-reference.com/cfb/>. Retrieved on April 28, 2016

[25] <http://espn.go.com/college-football/>. Retrieved on July 18, 2016

[26] databasefootball.com. "12/1/1997 A.P. Poll". Retrieved on July 31, 2016 from <http://www.databasefootball.com/College/polls/appoll.htm?PollID=403>