University of Mississippi

eGrove

Honors Theses

Honors College (Sally McDonnell Barksdale Honors College)

Spring 5-10-2023

An Application of the PageRank Algorithm to NCAA Football Team Rankings

Morgan Majors
University of Mississippi

Follow this and additional works at: https://egrove.olemiss.edu/hon_thesis

Part of the Algebra Commons, and the Applied Statistics Commons

Recommended Citation
Majors, Morgan, "An Application of the PageRank Algorithm to NCAA Football Team Rankings" (2023). *Honors Theses*. 2999.
https://egrove.olemiss.edu/hon_thesis/2999

This Undergraduate Thesis is brought to you for free and open access by the Honors College (Sally McDonnell Barksdale Honors College) at eGrove. It has been accepted for inclusion in Honors Theses by an authorized administrator of eGrove. For more information, please contact egrove@olemiss.edu.

An Application of the Pagerank Algorithm to NCAA Football Team Rankings

Morgan Majors

A thesis submitted to the faculty of The University of Mississippi in partial fulfillment of the requirements of the Sally McDonnell Barksdale Honors College.

Oxford

April 2023

Approved by

....

Advisor: Talmage James Reid

....

Reader: Laura Sheppardson

....

Reader: Mark Wilder

An Application of the Pagerank Algorithm to NCAA Football Team Rankings

Morgan Majors

Abstract

We investigate the use of Google's PageRank algorithm to rank sports teams. The PageRank algorithm is used in web searches to return a list of the websites that are of most interest to the user. The structure of the NCAA FBS football schedule is used to construct a network with a similar structure to the world wide web. Parallels are drawn between pages that are linked in the world wide web with the results of a contest between two sports teams. The teams under consideration here are the members of the 2021 Football Bowl Subdivision. We achieve a total ordering of the 2021 FBS teams by applying the PageRank algorithm to the results of the regular and bowl seasons. A statistical method of correlation is used to compare the final AP rankings with PageRank models based on Margin of Victory and Total Points Scored.

Dedication

I dedicate my thesis work to my parents, Penn and Ashley Majors. Their constant support has propelled me to the finish line of this experience. They have stood by my side since day one, and without them, I would not be the person I am today. The immense love I feel from them daily is truly overwhelming.

A special feeling of gratitude to my father, whose love for mathematics and sports I inherited. Some of my favorite memories have been alongside my father at a multitude of Ole Miss sporting events. The long nights spent at the dining room table while in elementary school learning basic mathematics are still persistent memories. His encouragement throughout all levels of my education has undoubtedly helped me to pursue my ambitions.

Acknowledgements

I wish to thank my committee members for their generous time and expertise spent helping me make my thesis experience the best it could be. A special thanks to James Reid who was always patient, knowledgeable, and encouraging when I needed assistance and advice. This project could not have been completed without him. Thank you to Dr. Reid, Dr. Sheppardson, and Dr. Wilder for agreeing to serve on my committee.

I would also like to acknowledge my family for the immense love they have always provided me. My parents, brother, and grandparents have always shown sincere support toward my academic goals and me. A special thanks to my brother, Penn Earl Majors, for always pushing me to be a better version of myself. He has been a constant pillar of support in my life. I also would like to thank my grandfather, Penn Majors, who inspired me to pursue mathematics. As a former mathematics teacher, his wisdom always encouraged me to follow in his footsteps. My grandmothers, Trudy Bomer and Mary Majors, have been my ultimate encouragers throughout this journey.

I would like to thank everyone who has been involved in this process. All of my former teachers and professors have played a hand in where I am today. I am eternally grateful to the Sally McDonnell Barksdale Honors College for the opportunities they have provided.

Contents

Chapter 1

Introduction

College football is a profitable business with nineteen teams reporting at least 100 million dollars in Revenue in 2019 [8]. Hence it is of intense interest to the Universities participating in College Football to achieve an accurate ranking of their performance. This thesis considers an approach to ranking these teams that uses the Google PageRank algorithm for ranking webpages. The ranking of nodes in networks and sports teams by an eigenvalue approach is an active area of research (see, for example, [7, 9, 12, 18, 19, 20, 24, 27, 36, 38]). The use of these ranking methods here is motivated by the 2008 doctoral dissertation of Anjela Yuryevna Govan [19].

We next discuss the organization of the paper. In Chapter 1 we discuss the history of College Football ranking methods. We then discuss the PageRank algorithm of Sergey Brin and Larry Page that brought organization to the world wide web. We next discuss the widespread applications of the algorithm to areas of academic and scientific interest. We provide background on the NCAA 2021 Football Bowl Subdivision teams used in this study to conclude Chapter 1.

In Chapter 2 we give the mathematical background of the results presented here. We begin with the notion of the World Wide Web as a network or graph. We then discuss the matrix theory from mathematics used to produce ranking eigenvectors by our algorithm. Then the mathematical formulation of the PageRank algorithm is given. We discuss the modification of the PageRank algorithm which will be used to produce our results to conclude Chapter 2.

In Chapter 3, we use the modified PageRank algorithm to rank the results of teams in the 2021 FBS College Football season using two methods. The first method takes into account the Margin of Victory results of each team to compute its ranking. The second method is based on the Total Score of each team in each of its games to produce its ranking. We then compare the Margin of Victory and Total Score rankings with the final AP rankings for that season. Finally, we make some conclusions and observations about the efficiency of these ranking methods.

1.1 College Football Rankings

There are multiple ranking methods for College Football Teams that determine the order of teams from best to worst based on results. These ranking methods are important in determining which teams will play in the most prestigious bowls. The most notable ranking polls are the Associated Press, USA Today's Coaches, Bowl Championship Series, and College Football Playoff Polls. The ranking of college football teams began in 1936 with the release of the first Associated Press (AP) poll [29]. The first AP Poll included only the top twenty teams, with Minnesota ranking at number one [11]. College football gained national attention at the time of this first poll. Football was created in the northeast United States and was initially dominated by Ivy League universities. These schools not only dominated the world of football physically, but the rules committee was also made up of Ivy League alumni, coaches, and players. These schools were considered to have played a superior brand of football. However, the first AP Poll illustrated how teams from other areas of the county were now superior to the Ivy League teams. Minnesota, a Midwestern team, finished in the one spot, with schools from the Deep South, West, and Great Plains finishing in the top five places in the first poll [11].

The USA Today Coaches Poll that debuted in the 1950 season ranked the top 20 NCAA football and basketball teams. In 1990 the Coaches Poll conformed to the method of the other polls by ranking the top 25 teams. Until the 1973 season, the poll was released only in December, after the conclusion of the regular season but before the bowl games began. A scandal regarding controversial voting practices within the Coaches Poll arose in 2005, resulting in ESPN dropping its sponsorship of this poll [1]. The Coaches Poll became the USA Today Coaches poll in 2005 [1].

The Bowl Championship Series (BCS) Poll debuted in 1998 [29]. The BCS Poll was a revolutionary idea that propelled the popularity of college football. Prior to the BCS Poll, the idea of a true championship game between the top two teams was uncommon [32]. This was the reason for the creation of the BCS Poll. After naming two teams as Co-National Champions in back-to-back years in the 1990 and 1991 seasons, college football fans wanted clarity on a true champion. Some efforts were made through organizations such as the Coalition, the Alliance, and the Super Alliance to construct a number 1 versus number 2 National Championship matchup, but none of these organizations could get every conference on board. The year of the first BCS did exactly what it was created to do, create the matchup between the two best teams in the country. However, issues still arose such as when the BCS poll claimed LSU as the 2003 National Champions, but the AP claimed USC as the best team of the season. These issues led to the birth of the College Football Playoff. The BCS Poll bridged the gap between the poll era and the College Football Playoff Poll [32].

All of the polls differ in their ranking strategies, but the AP, Coaches, and BCS Polls have one thing in common: the method they use to rank the teams. All of these polls use a committee, ranging in size, to construct a ranking. This is done by each committee member casting votes for what team should be placed at 1, and so forth. The team that the committee member thinks should rank at 1 receives 25 points, the 2 spot receives 24 points, and so on until the team at 25 receives 1 point. Each poll releases a new ranking of teams 1-25 weekly. Today, the AP Poll committee is made of 63 members, the Coaches Poll is 62, BCS is 62, and CFP is 13. The CFP Poll differs from the rest because it does not construct a ranking until well into the season [31].

The first College Football Playoff (CFP) Poll was released in 2014 [29]. While other polls simply rank teams for the audience's engagement, the CFP Poll has a great deal of meaning. This poll single-handedly chooses the top four teams that will play for the National Championship. Another responsibility of the CFP Poll is to place teams that did not receive an invitation to the playoffs into the New Year's Bowls. Six games occur on New Year's Eve and New Year's Day. These games are the Cotton Bowl, Fiesta Bowl, Orange Bowl, Peach Bowl, Rose Bowl Game, and Sugar Bowl. The two Playoff Semifinal games rotate annually through two of these games [2].

The ranking of college football teams is important for the advancement of the sport and its nationwide recognition. The top 4 teams in the College Football Playoff Poll receive the chance to compete for a National Championship. The team ranked at 1 plays against the 4-ranked team, and the team ranked at 2 plays against the 3-ranked team. The winners of the two semifinal games face off against one another in the final Championship game. The 2022 National Championship was played on January 10, 2022, between the Georgia Bulldogs and the Alabama Crimson Tide. The first National Title by the Georgia Bulldogs since 1980 was watched by 24.5 million viewers. [10]. This was the most-watched non-NFL sporting event in two years, with a 19 percent increase in viewers since the 2021 National Championship [10]. Along with the national attention that a team receives from being selected into the College Football Playoffs, teams receive large payouts for their selection. The four teams selected into playoffs receive a 6 million dollar payout from the CFB Revenue Pool[4]. No additional payout occurs for being one of the last two remaining teams, but all expenses are covered [4]. Similarly, popular bowls have a large payout [4]. With the financial implications of rankings to college football teams, interest in their methodology continues to grow.

1.2 The PageRank Algorithm

The origins of the methods considered here can be found in the desire to find order in the World Wide Web. The need for such an order arose in the late 20th century. There was an explosion in the number of webpages that occurred during this time. For example, in the years 1995, 1996, and 1997 the number of webpages increased from $23,500$ to $250,601$ to $1,117,255$ (see `https://www.internetlivestats.com/total-number-of-websites` for a live count of the number of pages on the web, currently close to 2 billion). Sergey Brin and Larry Page approached the problem of organizing the web by developing a web search engine in 1995. They call their web search engine Google. With this Google web search engine, they sought to place links to webpages of the greatest interest at the top of a search query. Brin and Page were computer science doctoral students at Stanford University at the time (see [22, Chapter 3]). They used their dorm rooms as offices for their business. A tech report from 1998 [9] announced the development of the search engine. Google's parent company Alphabet is currently valued at over $1.4 trillion [13].

The basis for the Google web search engine is the PageRank algorithm. The World Wide Web is organized by hyperlinks. Selecting a hyperlink to webpage Y that appears on webpage X will send the reader to webpage Y. Each webpage typically contains links to several other webpages. A webpage will be considered important if a link to it appears on other important webpages. This reasoning appears to be circular because two webpages might be considered to be important if each says the other is important. However, we show in Chapter 2 that applying mathematical tools to this reasoning leads to a well-defined ranking of webpages and therefore of sports teams.

We may think of the web as a network of webpages with links between some of the pages. A user of the web can be thought of as "surfing" between these pages. The network evolves as the user at a particular page has certain probabilities of traveling to other linked pages. Markov's Theory is useful in describing the evolution of such a network. This theory can be traced back to 1906 with the work of the Russian mathematician A.A. Markov [17]. A Markov process describes a sequence of possible events where the probability of each event only depends on the state of the previous. In order to assign ranks to webpages with the ranks being real numbers versus complex numbers, Brin and Page needed a result from Linear Algebra of Perron [30] and Perron and Frobenius (see [22, p. 172]). It is important to assign a real number rank to a web page as the set of real numbers is ordered, e.g. the number 0.3 is larger than the number 0.2. The set of complex numbers is not ordered. For example, there is no way to know whether a complex number such as 1+3i is larger than another complex number, so it is important to obtain ranking numbers that are positive real numbers.

The Power Method [26], developed by von Mises and Pollaczek-Geiringer, was a key piece of the puzzle necessary to make the large matrix computations needed to rank webpages using PageRank. Brin and Page needed to compute eigenvectors of large matrices (see Section 2.2) to provide their rankings and adaptations of this method of updating ranks iteration by iteration stabilized on well-defined ranking eigenvectors in a computable number of steps [22, Chapter 9]. All of these tools were synthesized by Brin and Page in 1998 to create the revolutionary PageRank algorithm. At the heart of this algorithm is the beautiful equation given in 1.2.1.

$$\pi^T = \pi^T(\alpha S + (1-\alpha)E) \tag{1.2.1}$$

Some consider this equation to be as important as equations from the theory of relativity such as $E = mc^2$ of Einstein and the Energy Wave Equation $E = hf$ suggested by Planck. Graham Farmelo [16] edited a book of beautiful equations entitled "It Must Be Beautiful: Great Equations of Modern Science" that contains Einstein's and Planck's equations as well as Equation 1.2.1 of Brin and Page. We explain the derivation of this equation that underlies the organization of the world wide web in Section 2.3.

An important alternative algorithm to the PageRank algorithm is the Hypertext Induced Topic Search algorithm, abbreviated HITS. The HITS algorithm was founded in 1998 by Jon Kleinberg. [22]. Thus this algorithm is a contemporary of the PageRank algorithm. The HITS algorithm is a system for ranking webpages while considering a webpage's popularity. It is very similar to the PageRank algorithm except it relies on the query to rank webpages and produces two popularity scores for each web page whereas PageRank is not dependent on the query and only produces one popularity score [28]. Also, HITS operates by associating webpages with authorities and hubs. An authority is a webpage with many inlinks, and a hub is a webpage with many outlinks [22]. Authorities and hubs have a dependent relationship, as good hubs result in good authorities and good authorities result in good hubs [28].

Methods of ranking webpages remain of considerable interest. After the founding of PageRank and HITS, less popular ranking methods were developed. A few years after the success of PageRank and HITS, in 2000, a Stochastic Approach to Link Structure Analysis was developed [22]. This stochastic algorithm later became known as SALSA (Stochastic Approach for Link-Structure Analysis) [15]. This approach has similarities to both PageRank and HITS. It creates both hub and authority scores while being derived from Markov chains. While it combines some of the best features from both major algorithms, SALSA has one major setback [15]. The major dependence on queries from SALSA is what keeps this algorithm from being well-known. This is because the query-dependence results in a ranking are not unique [22].

A new wave of ranking methods was introduced around 2005 through meta-search engines [22]. TrafficRank attempted to rank webpages by merging the results from several different ranking algorithms [12]. It hoped to provide the most accurate and precise rankings through the use of a multi-link algorithm. TrafficRank encouraged the mindset that the World Wide Web does not consist of webpages connected by a single road, instead, it is billions of webpages connected by billions of connections [22].

Although the PageRank algorithm is primarily used in web-search engines, it is used in many different ways throughout society. The PageRank algorithm has been manipulated and modified to fit a number of applications. A history of the development of the PageRank algorithm and its uses are given in Table 1.1 (see [17]). We will discuss the applications of this algorithm in the next section of the thesis.

Table 1.1: PageRank History

1906	Markov	Markov theory
1907	Perron	Perron theorem
1912	Frobenius	Perron-Frobenius theorem
1929	von Mises and Pollaczek-Geiringer	Power method
1941	Leontief	Econometric model
1949	Seeley	Sociometric model
1952	Wei	Sport ranking model
1953	Katz	Sociometric model
1965	Hubbell	Sociometric model
1976	Pinksi and Narin	Bibliometric model
1998	Kleinberg	HITS
1998	Brin and Page	PageRank

1.3 Applications

The PageRank algorithm is primarily associated with the World Wide Web and the way web pages are linked. However, the PageRank algorithm can be applied to many different subjects. For example, the PageRank algorithm can be applied within the fields of chemistry, biology, literature, and social networks to organize the complex relationships between data.

In chemistry, PageRank is used to study molecules and their valences through the study of change within water molecules linked by hydrogen bonds. Ultimately, it is used to determine if the molecules have a potential hydrogen bond to a solute molecule [18]. Teleportation occurs to determine structural differences. The development of PageRank within chemistry has evolved into the ability to provide analytical derivatives and illustrations [39]. Through different experiments, it can be determined which solvent rearrangement has the most impact on the reaction pathway [39].

Common uses of the PageRank algorithm are in biology such as GeneRank [27], ProteinRank [36], and IsoRank[24]. All of these applications study the dynamics of network data in bioinformatics. They result in the sharing of localized information about the graph produced by the PageRank algorithm. The reactions between the genes, proteins, and isotopes are transformed into interactions between nodes in the network, therefore allowing the PageRank algorithm to be applied to rank objects in the network [18]. In a recent experiment, the measurement of the graph centrality of the network associated with the PageRank algorithm was used in the identification of cancer genes [14]. Experimental data is used to determine the accuracy of the centrality within biology [14].

A fascinating aspect of the PageRank algorithm is that it is used in the world of literature through BookRank [37]. Elliot Yates and Louise Dixon in [37] state that the BookRank algorithm can shed light on " ...the central questions of literature: What are the most important books? Which story is most likely to occur within a novel? And what should I read next?" BookRank is used to catalog books on the web. This is useful for the organization of the multitude of written works. The idea of the BookRank algorithm is similar to the use of PageRank in Google, but book titles are exchanged for webpages in this algorithm [18].

PageRank is used in social networking [5] through BuddyRank and TwitterRank. The people interacting on social media platforms act as the nodes, while the relationships formed on social networking sites act as the edges or links between nodes. It can make predictions about who an individual might connect with next and evaluate the centrality of estimating someone's social status. Also, it can estimate the potential influence of a person's opinion of the social network. Both of these ranking programs typically use a standard alpha of .85 to perform a reverse application of PageRank [18].

PageRank allows endless possibilities for its uses. Whether they are used in everyday life or as crucial research opportunities, it is undeniable that PageRank has millions of possible applications. Biology, chemistry, literature, and social networks are just the surface of what PageRank can do. Although it is typically used in web search engines and mathematics, PageRank can be used in any subject whose interactions can be modeled by a network.

1.4 NCAA 2021 Football Season

The 2021 NCAA Football Season began with a game between Illinois and Nebraska played on August 28, 2021, at 1:20 pm Eastern time and concluded with a game between Georgia and Alabama at 8:00 pm Eastern time on January 10, 2022 [3]. NCAA Division I Football is categorized into two divisions, the Football Bowl Subdivision (FBS) and the Football Championship Subdivision (FCS)[6]. Our research is centered around the 130 teams of the 2021 NCAA FBS season. There are an additional 128 NCAA FCS teams that participated in the 2021 season. Games are primarily played between a pair of FBS teams and a pair of FCS teams although a limited number of games are played between teams of different levels.

The 2021 FBS teams are organized into 11 conferences and 7 independent schools as follows. The Atlantic Coast Conference includes the Atlantic division of Wake Forrest, Clemson, North Carolina State, Louisville, Flordia State, Boston College, Syracuse, and the Coastal division of Pitt, Miami (FL), Virginia, Virginia Tech, North Carolina, Georgia

Tech, and Duke. The American Athletic Conference includes Cincinnati, Houston, UCF, East Carolina, Tulsa, SMU, Memphis, Navy, Temple, South Florida, and Tulane. The Big 12 Conference includes Baylor, Oklahoma State, Oklahoma, Iowa State, Kansas State, West Virginia, Texas Tech, Texas, Texas Christian, and Kansas. The Big Ten Conference includes the East division of Michigan, Ohio State, Michigan State, Penn State, Maryland, Rutgers, Indiana, and the West division of Iowa, Minnesota, Purdue, Wisconsin, Illinois, Nebraska, and Northwestern. The Conference USA includes the East division of Western Kentucky, Marshall, Old Dominion, Middle Tennessee State, Charlotte, Florida Atlantic, Florida International, and the West division of UTSA, UAB, North Texas, UTEP, Rice, Louisiana Tech, and Southern Mississippi. The Mid-American Conference includes the East division of Kent State, Miami (OH), Ohio, Bowling Green, Buffalo, Akron, and the West division of Northern Illinois, Central Michigan, Toldeo, Western Michigan, Eastern Michigan, and Ball State. The Mountain West Conference includes the Mountain division of Utah State, Air Force, Boise State, Wyoming, Colorado State, and New Mexico, and the West division of San Diego State, Fresno State, Nevada, Hawaii, San Jose State, and Nevada-Las Vegas. The Pac-12 Conference includes the North division of Oregon, Washington State, Oregon State, California, Washington, and Stanford, and the South division of Utah, UCLA, Arizona State, Colorado, USC, and Arizona. The Southeastern Conference includes the East division of Georgia, Kentucky, Tennessee, South Carolina, Missouri, Florida, and Vanderbilt, and the West division of Alabama, Ole Miss, Arkansas, Texas A&M, Mississippi State, Auburn, and LSU. The Sun Belt Conference includes the East division of Appalachian State, Coastal Carolina, Georgia State, Troy, and Georgia Southern, and the West division of Louisiana, Texas State, South Alabama, Louisiana-Monroe, and Arkansas State. The Independent schools that are not in conferences are Notre Dame, New Mexico State, Liberty, Massachusetts, Connecticut, Army, and BYU. [33]. Each team in a conference typically plays most of its games against other teams in the same conference. Each team in the FBS typically has one or two opponents in the FCS with the remaining games against other FBS teams. Teams that are successful during the season with a winning record typically play in a Bowl Game after the conclusion of the regular season. Since there are 130 FBS teams and most teams play between 11 and 13 games, most pairs of teams do not play each other. This observation is the reason that College Football Ranking systems are useful to compare pairs of teams that do not play each other and the relative rankings of paired teams can be in dispute.

Chapter 2

Mathematics of PageRank

We discuss the mathematical concepts used in this research in this chapter. The PageRank algorithm uses an interesting mix of tools from a variety of mathematical disciplines. The research uses models and concepts from Graph Theory, Matrix Algebra, and Statistics. An understanding of the basic principles of all of these fields is necessary to develop the simple premise of the PageRank Algorithm:

Axiom 2.1. The ranking of a team is proportional to the sum of the rankings of the teams that it defeats.

In order to examine the implications of Axiom 2.1 to team ranking we will begin with a discussion of Graph Theory.

2.1 A Graph Model of College Football

Graph Theory is an area of mathematical research that is concerned with modeling the interactions between objects of interest in a subject using a graph. The Graph Theory terminology used here can be found in (see [35]. A graph is a pair $G = (V, E)$, where V is a nonempty set called the vertex set of G and E is a set called the edge set. The elements of

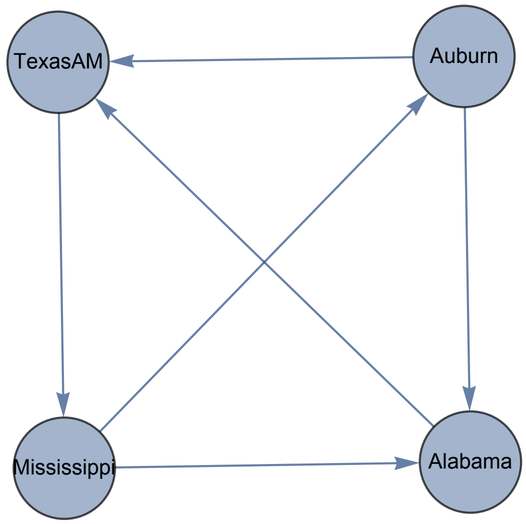

Figure 2.1: A four team example

E consist of pairs of vertices (the plural of vertex) of V. A graph is called directed when the two vertices of each edge are given an order or equivalently a direction. Consider the directed graph $G = (V, E)$ whose diagram is given in Figure 2.1. The vertex set of this graph consists of four elements labeled "Alabama", "Auburn", "Mississippi", and "TexasAM". There are edges between each distinct pair of vertices in G such as between the vertices labeled by Alabama and Auburn. Further, the edges have directions indicated by arrows such as the arrow that points from Auburn to Alabama. The graph G represents the results of all the games played between the University of Alabama, Auburn University, the University of Mississippi, and Texas A&M University in the NCAA 2021 football season. The directed edge from Auburn to Alabama in the graph G represents that Alabama defeated Auburn during this season. Thus this small graph represents that Alabama defeated Auburn and Mississippi, Auburn defeated Mississippi, Mississippi defeated Texas A&M, and Texas A&M defeated both Auburn and Alabama. One may think of this as vertex Alabama voting for vertex TexasAM, in the same way, a web page link "votes for" or "suggests" that the web user visits the linked page.

Assign a vertex to each of the 130 NCAA FBS football teams from the 2021 season, and

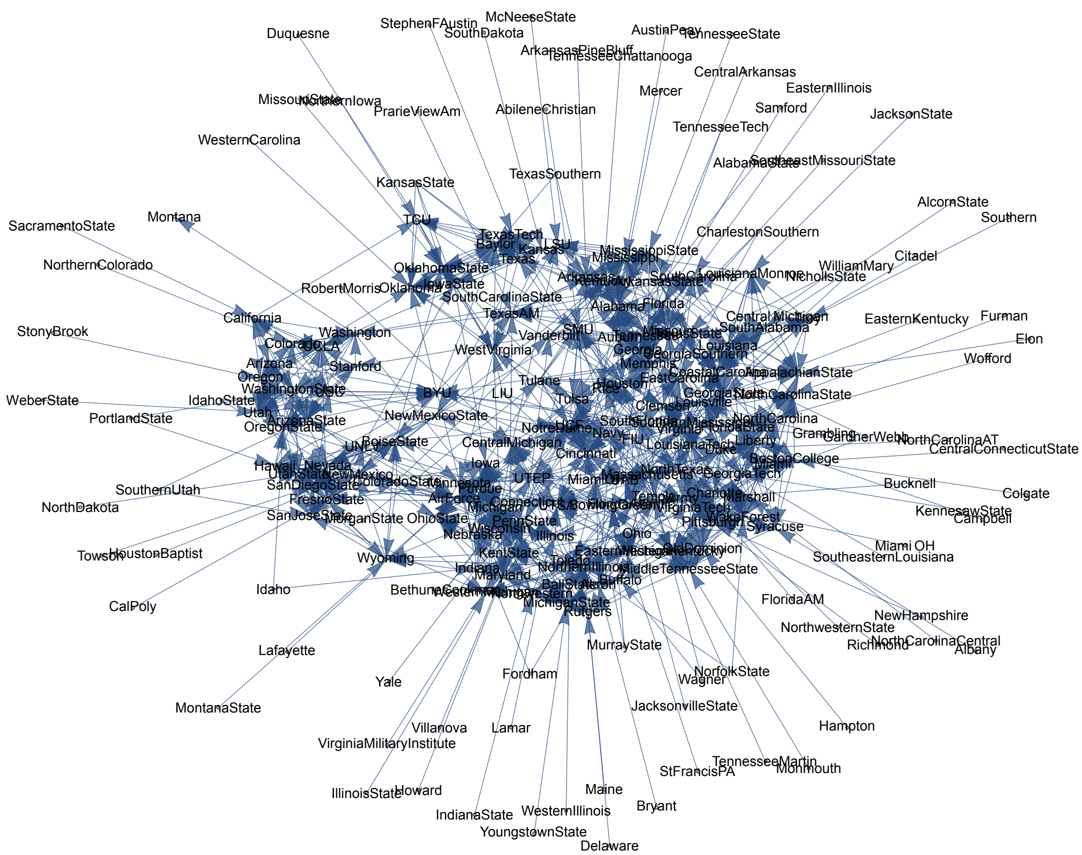

Figure 2.2: The 2021 NCAA Football Season as a Graph

to each of the FCS teams that played at least one FBS team during the season to form the vertex set of a graph such as in Figure 2.2. We draw an edge between two vertices when the corresponding teams played during either the regular or bowl season. This is the web-type network for which we seek an ordering of the vertex set. This graph was produced using the computer program Mathematica. One can note how the FCS teams appear in the periphery of the graph while teams in the interior of the graph are clustered mostly by conference and geographic location.

2.2 Matrix Algebra and Eigenvectors

We develop the notion of a ranking vector in this section of the thesis. The Linear Algebra notation used here follows the textbook Leon [23]. The equations given in the table of Figure 2.2.1 arise naturally from Axiom 2.1 and the graph given in Figure 2.1. Here the word "proportional" is modeled by the multiplication of the sums by the real number $\frac{1}{\lambda}$. For example, in the first row of the table in Equation 2.2.1 the rank of Alabama contains the proportionality constant $\frac{1}{\lambda}$ times the sum of the ranks of Auburn and Mississippi, the two teams in this list of four that Alabama defeated. This reasoning for computing rankings appears to be circular so we will examine the implications of such a computation from the perspective of Linear Algebra. Equation 2.2.1 can be written using matrix multiplication as in shown in Figure 2.4.

$$\begin{aligned}
r(Alabama) &= \frac{1}{\lambda}(0 \cdot r(Alabama) + 1 \cdot r(Auburn) + 1 \cdot r(Mississippi) + 0 \cdot r(TexasA\&M)) \\
r(Auburn) &= \frac{1}{\lambda}(0 \cdot r(Alabama) + 0 \cdot r(Auburn) + 1 \cdot r(Mississippi) + 0 \cdot r(TexasA\&M)) \\
r(Mississippi) &= \frac{1}{\lambda}(0 \cdot r(Alabama) + 0 \cdot r(Auburn) + 0 \cdot r(Mississippi) + 1 \cdot r(TexasA\&M)) \\
r(TexasA\&M) &= \frac{1}{\lambda}(1 \cdot r(Alabama) + 1 \cdot r(Auburn) + 0 \cdot r(Mississippi) + 0 \cdot r(TexasA\&M))
\end{aligned} \tag{2.2.1}$$

We let $\mathbf{R}$ be the ranking vector and $\mathbf{A}$ be the matrix given in Figure 2.3. Then the matrix equation $\mathbf{R} = \frac{1}{\lambda}\mathbf{A}\mathbf{R}$ is obtained. So $\mathbf{R}$ represents a vector with four rows and one column. The entries obtained in $\mathbf{R}$ will be the ranking score of Alabama, Auburn, Mississippi, and Texas A&M, respectively, read from top to bottom. The vector $\mathbf{R}$ is called an eigenvector of the matrix $\mathbf{A}$ while the number λ is called an eigenvalue. There is a problem with this method in finding the ranking vector $\mathbf{R}$ in that the entries may be complex numbers and not real numbers. The field of complex numbers is not ordered as is the field of real numbers.

This potential problem is addressed in the next section of the thesis. For example, the four eigenvalues obtained as solutions in λ to the equation $\lambda\mathbf{R} = \mathbf{AR}$ are 1.3953, $-.4604+1.1393i$, $-.4604-1.1393i$, and $-.4746$. There is a unique positive real eigenvalue, 1.3953. The eigenvector corresponding to the eigenvalue 1.3953 is $\langle .5516, .3213, .4484, .6256\rangle^T$. This would correspond to Alabama having the second highest ranking of .5516, Auburn having the lowest ranking of .3213, Mississippi having the second lowest ranking of .4484, and Texas A&M having the highest ranking of .6256. It is natural that Alabama and Texas A&M are ranked higher in this little four-team league as both of those teams had two wins. Texas A&M is ranked higher than Alabama as it defeated Alabama. Both Auburn and Mississippi had one win and Auburn defeated Mississippi, but Mississippi is ranked higher than Auburn. The explanation for this eigenvalue ranking is that Mississippi defeated the highest-ranked team Texas A&M. This small example ranking will be generalized to rank large complex networks such as that given in Figure 2.2. The eigenvalue ranking presented here is simpler than the PageRank algorithm presented in the next section of the thesis.

2.3 Modified PageRank Algorithm

We next develop the Google PageRank algorithm for ranking the relative importance of webpages. Consider the directed graph G given in Figure 2.5. Suppose this graph models the web links of six websites labeled by 1 through 6 (see [22, Chapter 4]). So WebPage 1 has links to WebPages 2 and 3, for example. Following Axiom 2.1, Brin and Page let $r(P_i)$ denote the rank of page P_i. Let B_i be the set of pages that point to page P_i for each i. So

$$R = \begin{bmatrix} r(Alabama) \\ r(Auburn) \\ r(Mississippi) \\ r(TexasA\&M) \end{bmatrix} \qquad \text{and} \qquad A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

Figure 2.3: The ranking vector

$B_2 = \{P_1, P_3\}$ and $B_5 = \{P_3, P_4\}$ for example. So they obtained the equation given in 2.3.1 where $|P_j|$ denotes the total number of outlinks from page P_j.

$$r(P_i) = \sum_{P_j \in B_i} \frac{r_k(P_j)}{|P_j|} \tag{2.3.1}$$

So consider the ranking of the page P_2 by the method of Equation 2.3.1. We obtain $r(P_2) = \frac{r(P_1)}{|P_1|} + \frac{r(P_3)}{|P_3|} = \frac{P_1}{2} + \frac{P_3}{3}$. In order to start this process we need ranks for P_1 and P_3. So we let each of the six vertices of G have rank $\frac{1}{6}$. This assignment has the advantage that each row of H will sum to 1. So the rank of P_2 becomes $\frac{1}{6}(\frac{1}{2} + \frac{1}{3})$. In terms of vectors we can write a row vector $\pi^T = \langle \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6} \rangle$ where the ranking of vertex i is found in the ith place of the vector.

Let $\pi^T{=}\pi(0)^T$ and $\pi^{(1)T}{=}\pi^{(0)T}H$. Then, we obtain a new ranking vector $\pi^{(1)}{=}\langle 1/18, 5/36, 1/12, 1/4, 5/36, 1/6 \rangle$. Repeating this process for $K = 1, 2, \ldots$, we obtain $\pi^{(K+1)T} = \pi^{(K)T}H$. Now, the surprising thing noticed by Brin and Page is that the ranking vectors converge to a unique vector of positive, real entries. The reason for this convergence is the Perron-Frobenius Theorem from 1908 [22, p. 172]. The description here is the Power Method for computing the ranking Eigenvector.

Equation 2.3.1 can cause a division by 0 if a page has no inlinks. This problem can be solved by adding non-zero entries to H as follows. Let $\mathbf{e}$ be the 6×1 vector of ones and S be the matrix $Z = \alpha S + (1-\alpha)\frac{1}{6}ee^T$. Then Z is called the *Google Matrix*. The number α is chosen between 0 and 1, typically $\alpha = 0.85$ is chosen. This parameter is called the *teleportation index*. So if $\alpha = 0.85$, and we surf the web of the graph G with 85% probability, we follow the

$$\begin{bmatrix} r(Alabama) \\ r(Auburn) \\ r(Mississippi) \\ r(TexasA\&M) \end{bmatrix} = \frac{1}{\lambda} \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} r(Alabama) \\ r(Auburn) \\ r(Mississippi) \\ r(TexasA\&M) \end{bmatrix}$$

Figure 2.4: Matrix form of Axiom 2.1

Figure 2.5: PageRank Example Graph G

hyperlink structure of H, and with 15% probability, we randomly teleport to a new vertex with probability $\frac{1}{6}$. So we multiply the ranking vector transpose on the left of G iteratively until the ranking vector for the pages is obtained. In general, we use n instead of 6 for a graph of n vertices. This ranking can be considered P_1 "voting" for P_2 with half-credit as P_1 has two out nodes and P_3 "voting" for P_3 with one-third credit as P_3 has three out nodes. Computing rankings for each vertex of the graph G of Figure 2.5 by Equation 2.3.1 we obtain a matrix H in Figure 2.6 where the rank of each vertex P_i is the sum of the entries in column i of the matrix. The matrix H is called the Hyperlink matrix of the Graph G. We continue to iterate the process of applying Equation 2.3.1 where the new rank of each P_i is computed by multiplying the transpose of the current ranking vector by H. These ranking vector for this example converges to $\pi = \langle 0.33550.02500.35020.10940.09300.0870 \rangle$. If $R(i)$ is the ranking of node i, then $R(1) = 0.3355$, $R(2) = 0.0250$, $R(3) = 0.3502$, $R(4) = 0.1094$, $R(5) = 0.0930$, and $R(6) = 0.0870$. In this framework, smaller ranked nodes represent websites that are more likely to be visited. So, the most important nodes in order would be 2, 6, 5, 4, 1, and 3. One can imagine that Page 2 is the most likely visited page as one surfs this six-page web as it has inlinks but not outlinks. On average, the nodes on the right side of the graph are higher ranked than the nodes on the left side of the graph due to the link from Page 3 to Page 5.

$$
\begin{array}{c} d \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}
\begin{array}{c}
\begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\
\begin{pmatrix}
0 & 1/2 & 1/2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\
0 & 0 & 0 & 0 & 1/2 & 1/2 \\
0 & 0 & 0 & 1/2 & 0 & 1/2 \\
0 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}
\end{array}
$$

Figure 2.6: Hyperlink matrix H

$$
\begin{pmatrix}
0 & 1/2 & 1/2 & 0 & 0 & 0 \\
1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\
1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\
0 & 0 & 0 & 0 & 1/2 & 1/2 \\
0 & 0 & 0 & 1/2 & 0 & 1/2 \\
0 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}
$$

Figure 2.7: Stochastic Matrix S, Row entries sum to 1

$$
\begin{pmatrix}
1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\
1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\
1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\
1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\
1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\
1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6
\end{pmatrix}
$$

Figure 2.8: The matrix $1/n\ \mathbf{ee}^T$

2.4 Kendall Rank Correlation

We discuss the Kendall Rank Correlation of two data sets in this section of the thesis. Kendall Rank Correlation is a non-parametric measure of relationships between columns of ranked data [25]. We use the Kendall Rank Correlation to test the similarities in the ordering of our data [25].

Let $x_1, x_2, \cdots, x_n$ and $y_1, y_2, \cdots, y_n$ be two ranked lists of observations. Then the Kenall Rank Correlation τ of the variables is given in Equation 2.4.1.

$$\tau = \frac{\text{number of concordant pairs} - \text{number of discordant pairs}}{\text{number of pairs}} \tag{2.4.1}$$

Here a pair of observations (x_i, y_i) and (x_j, y_j), $i < j$, is concordant if the sort order of x_i and x_j and the sort order of y_i and y_j agree, otherwise the pair is said to be discordant. The number of pairs is $\binom{n}{2}$. Note that if the rankings are the same, then $\tau = 1$ while if the rankings are reversed, then $\tau = -1$. The closer the rankings are to each other, the closer τ will be to one. A small example of computing this parameter for the rankings $1, 2, 3, 4$ and $1, 4, 2, 3$ is given next. Here $\binom{n}{2} = \binom{4}{2} = 6$. The concordant pairs are 12, 13, 14, and 23 while the discordant pairs are 24 and 34. Thus $\tau = \frac{4-2}{6} = \frac{1}{3}$. That τ is a positive measure that the lists agree in order more than they disagree.

We can see the results of the Kendall Rank in 3.2 and 3.3. The linear regression line can be given in the equation $Y = a + bX$, where X is the explanatory variable and Y is the dependent variable [34]. Plots above the linear regression line are more highly Y-ranked and plots below the linear regression line are more highly X-ranked. A graph with more plots clustered close to the line of regression gives a more consistent evaluation. The plots in 3.2 show a closer relationship to the final AP rankings than 3.3 because of the lesser distance between the linear regression line and the majority of plots.

Chapter 3

Results

We give a total ordering of all teams in the FBS based on the results of the 2021 season by two different methods in this chapter. The methods used are adapted from those used in the 2008 Ph.D. thesis of Anjela Govan at North Carolina State University [19]. Following Govan's work and a subsequent paper of Laurie Zack, Ron Lamb, and Sarah Ball [38] that appeared in the journal Involve we investigate the rankings produced using both a Margin of Victory approach and a Total Points scored in a contest approach. Our rankings are for the 130 teams that were in the FBS of College football in 2021. The ranking of Govan and Zack et. al. was for the 32 teams in the National Football League. It was a challenge to compile the data from the large pool of teams in the FBS.

3.1 Margin of Victory Rankings

We present a total ordering of the FBS football teams by using the computer program Matlab with code as given in Figure 3.1. Note that the line headings "Line 1" etc. were not part of the code used, they are added for reference's sake. The matrix **R** is at the heart of the ranking. This matrix had both rows and columns indexed by the 130 FBS teams in alphabetic order. For the matrix entry in the column labeled "Ole Miss" and the row labeled "LSU", we accounted for the score of the game between LSU and Ole Miss as follows. Ole Miss won the game 31 to 17 so a 14 was entered in this entry as that was the Margin of Victory for Ole Miss over LSU. So each team such as Ole Miss received a positive entry in its column at each corresponding row where the team won a game against that opponent. If a team such as Ole Miss either lost a game or did not play a game against the team labeling the corresponding row of its column, then a "0" was entered. If two teams played twice, then the scores were added for both games and a Margin of Victory positive number was entered for the team with the most total points in its column. On occasion, some teams played another team more than once. Rematches of this type occurred four times during the season. The most notable example of a rematch was Alabama versus Georgia in the SEC Championship and in the National Championship. In the SEC Championship, Alabama won by a score of 41-24 with a Margin of Victory of 17. In the National Championship, Georgia won by a score of 33-18 with a Margin of Victory of 15. Both Margins of Victory scores were recorded. So returning to Line 1 of Figure 3.1 the matrix R features prominently. This matrix is 130 rows by 130 columns. The matrix $\mathbf{C}$ is merely a square matrix of the same dimension as R with the number 0.1 in each entry. We add $\mathbf{C}$ to $\mathbf{R}$ so that we do not divide by 0 in the computations that follow as the matrix $\mathbf{R}$ has a 0 in the majority of its entries. Line 2 of the code is necessary to make the matrix called Hv row stochastic where the entries in each row sum to 1. Line 3 sets the teleportation index α equal to 0.85. This is the number between 0 and 1 commonly used in applications of the PageRank algorithm. We found that this α was a good choice based on experiments with choices ranging from 0.5 to 0.99. We note that in the matrix $\mathbf{R}$ we ignored the results of matches between FBS and FCS teams as these did not appreciably affect rankings according to our experiments. In Line 6 we create the Google matrix $\mathbf{Ga}$. If we imagine ourselves randomly traveling around the Margin of Victory network, then with probability α we follow nodes suggested by the numbers in $\mathbf{Hv}$

```
Line 1: S=sum(R+C,1);
Line 2: Hv=(R+C)./S;
Line 3: alpha = 0.85;
Line 4: e = ones(130,1);
Line 5: v = e/130;
Line 6: Ga = alpha*Hv+(1-alpha)*v*e';
Line 7: q = rand(130,1);  % random initial vector
Line 8: q = q/sum(q);   % normalise p to have unit sum
Line 9: for k=1:100
Line 10: q = Ga*q;
Line 11: end
```

Figure 3.1: Matlab Code for Finding a Pagerank Vector

where teams lose to other teams vote for them based on the Margin of Victory, while with probability $1-\alpha$ we randomly jump around the network based on the numbers in the square matrix $v * e'$ with each entry as 1/130. The initial vector q produced in Lines 7 and 8 is a random vector with entries between 0 and 1 that is normalized so that its entries summed to 1 which is our initial ranking. An amazing aspect of this algorithm is that it doesn't matter what is our initial ranking. In Line 10 of the algorithm, we iterate q by left multiplying it by the matrix **Ga** 100 times. Eventually, the numbers in q stabilize to a final ranking vector.

In Table 3.1 we list the top 25 teams obtained by the algorithm of Figure 3.1. The q value obtained for each team is listed there. The complete rankings for all 130 teams are given in the Appendix. In this table, a smaller q value produces a better ranking. Note that the algorithm could be adapted to a transposed matrix so that the larger q value would produce a better ranking. We also give the Final AP Top 25 in the last column for comparison's sake.

The top 25 in Table3.1 produced some expected results and some surprising results. Georgia and Alabama are highly ranked in both polls as expected. Teams such as Notre Dame and Coastal Carolina seem to be more highly ranked than would be expected. Note that the 2021 coach of Notre Dame, Brian Kelly, received a large increase in salary to join LSU

	Team	q-Value	AP Top 25
1	Notre Dame	0.00194679056358042	Georgia
2	Georgia	0.00196000393177937	Alabama
3	Alabama	0.00196797504119794	Michigan
4	Coastal Carolina	0.00196988141542532	Ohio State
5	Cincinnati	0.00198379033469467	Baylor
6	Oklahoma State	0.00200420842507026	Oklahoma State
7	Michigan	0.00202229517821375	Cincinnati
8	Oklahoma	0.00206023741216896	Notre Dame
9	Clemson	0.00207774962269084	Oklahoma
10	Baylor	0.0021016193527964	Michigan State
11	Ohio State	0.00220550673438769	Ole Miss
12	Air Force	0.00229666661999109	Arkansas
13	BYU	0.00229756915826271	Utah
14	Pittsburg	0.00230684479094631	Pitt
15	Michigan State	0.0023120444199844	NC State
16	Penn State	0.00241098386284439	Clemson
17	Ole Miss	0.00246464202733623	Minnesota
18	Arkansas	0.00248728348591432	Wisconsin
19	East Carolina	0.00250162674755332	Purdue
20	NC State	0.00250988269888323	Tennessee
21	Texas A&M	0.00252239514451533	Kentucky
22	Boise State	0.00253635971904086	Iowa
23	Wake Forest	0.00255176433875926	Wake Forrest
24	Kentucky	0.00255453226853235	Utah State
25	SMU	0.00260051664523215	San Diego State

Table 3.1: Margin of Victory

after the 2021 season. Similarly, the coach of Coastal Carolina, Jamey Chadwell, received a large increase in salary to join Liberty after the 2022 season. The quality of these two teams' seasons was noticed by the college football community. Now we discuss a statistical method for comparing the PageRank list with the AP List. Note that we compared all 130 teams in both lists.

3.2 Total Score Ranking

We constructed the Total Score Ranking very similarly to Margin of Victory. This method differs from Margin of Victory because it uses two weighted arrows to represent each team's score rather than one for the Margin of Victory. [38]. Using the same example in Chapter 3.1, Ole Miss defeated LSU 31 to 17. For the Total Score Ranking, a 31 would be placed in the matrix entry in the column labeled "Ole Miss" and the row labeled "LSU". Similarly, a 17 would be placed in the matrix entry in the column labeled "LSU" and the row labeled "Ole Miss". This way Ole Miss was positively recorded for scoring 31 points and defeating LSU and LSU was recorded for scoring 17 points. If a team such as Ole Miss did not play a game against the team labeling the corresponding row of its column, then a "0" was entered. If two teams played twice, then the total number of points scored by that team was added together and placed in the corresponding matrix entry of the row and column. The same code used in the Margin of Victory computations was used for Total Score Rankings, see below. The only change that was made was the matrix R. We imported a 130 row by 130 column matrix with the Total Score entries. Our code for the Matlab computation of the Margin of Victory Ranking used q. We used p to note the difference in the matrix imported. Eventually, the numbers in p stabilize to a final ranking vector.

In Table 3.2 we list the top 25 teams obtained by the algorithm of Figure 3.4. The p value obtained for each team is listed there. The complete rankings for all 130 teams are given

Calculation results

Kendall's τ	**0.64452263**
Z score	**10.87594765**
P-value (τ; H_0: τ ≤ 0)	**0**
95% Confidence Interval	**[0.5284 , 0.7607]**
95% Right-Sided Interval	**[0.5470 , +1]**
95% Left-Sided Interval	**[-1 , 0.7420]**
Number of (x,y) pairs	**130**
Least-Squares Regression Line	**y = 0.82237911 · x + 11.63416816**

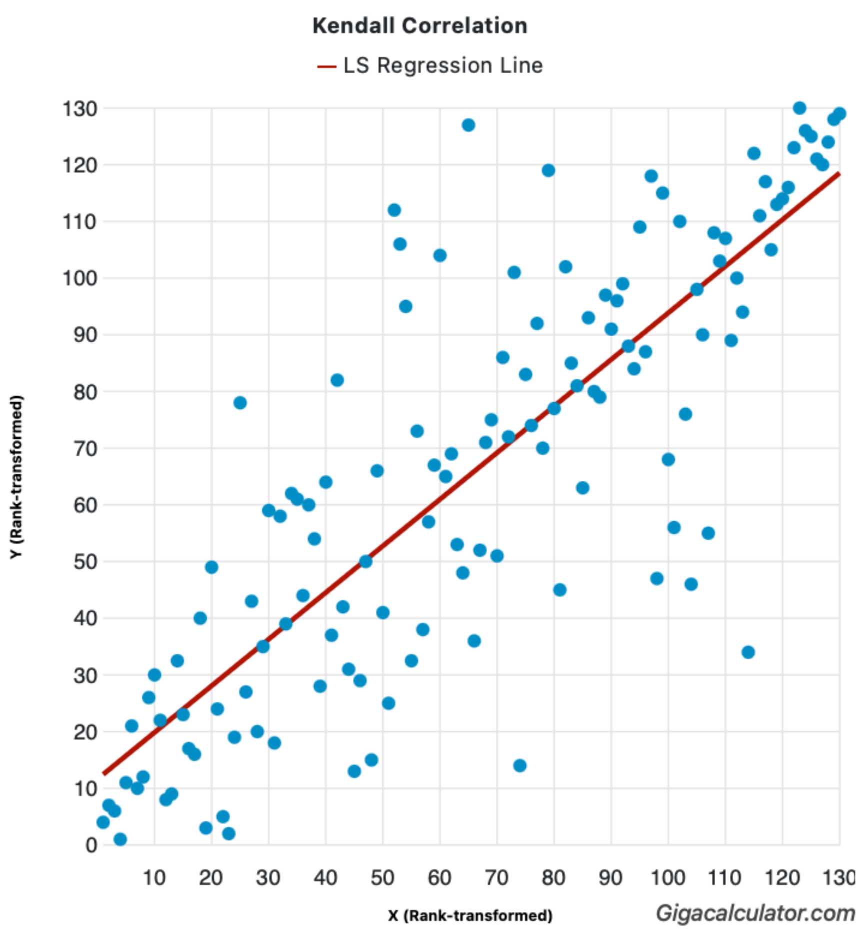

Figure 3.2: Margin of Victory Kendall's Tau

in the appendix. In this table, a smaller p value produces a better ranking. Note that the algorithm could be applied to a transposed matrix so that the larger p value would produce a better ranking. We also give the Final AP Top 25 in the last column for comparison's sake.

The top 25 in Table 3.2 produced some expected results and some surprising results. Georgia and Alabama are highly ranked in both polls as expected. Teams such as Liberty and Texas A&M seem to be more highly ranked than would be expected. Note that the head coach of the Liberty Flams, Hugh Freeze, received a large increase in salary to join the Auburn Tigers after the 2022 season. Jimbo Fisher, the head coach of the Texas A&M Aggies, received a four-year contract extension following the 2020 season that raised his annual salary by 1.5 million dollars [21]. The conclusions for Total Score differ from the results of Margin of Victory. This is because not all of the teams shown in 3.2 represent teams with a heavy win season. While many of the teams listed in 3.2 did win the majority of their games, having a winning schedule was not an exclusive factor, as it was in Margin of Victory. The teams listed in 3.2 scored many points, whether they were victorious in the game or not. Teams in the Margin of Victory table outscored their opponents the most, while teams in the Total Score table scored many points, regardless of a win or loss.

The top 25 in Table3.2 produced some expected results and some surprising results. Georgia and Alabama are highly ranked in both polls as expected. Teams such as Liberty were not as expected to be ranked highly. Note that the head coach of the Liberty Flames, Hugh Freeze, received a large salary increase by joining the Auburn Tigers after the 2022 season. Jimbo Fisher, the head coach of the Texas A&M Aggies, received a four-year contract extension following the 2020 season that raised his annual salary by 1.5 million dollars [21]. The conclusions for Total Score differ from the results of Margin of Victory. This is because not all of the teams shown in 3.2 represent teams with a heavy win season. While many of the teams listed in 3.2 did win the majority of their games, having a winning schedule was

Calculation results

Kendall's τ	**0.51627907**
Z score	**8.71236635**
P-value (τ; H_0: τ ≤ 0)	**0**
95% Confidence Interval	**[0.4001 , 0.6324]**
95% Right-Sided Interval	**[0.4188 , +1]**
95% Left-Sided Interval	**[-1 , 0.6138]**
Number of (x,y) pairs	**130**
Least-Squares Regression Line	**y = 0.71359434 · x + 18.75957066**

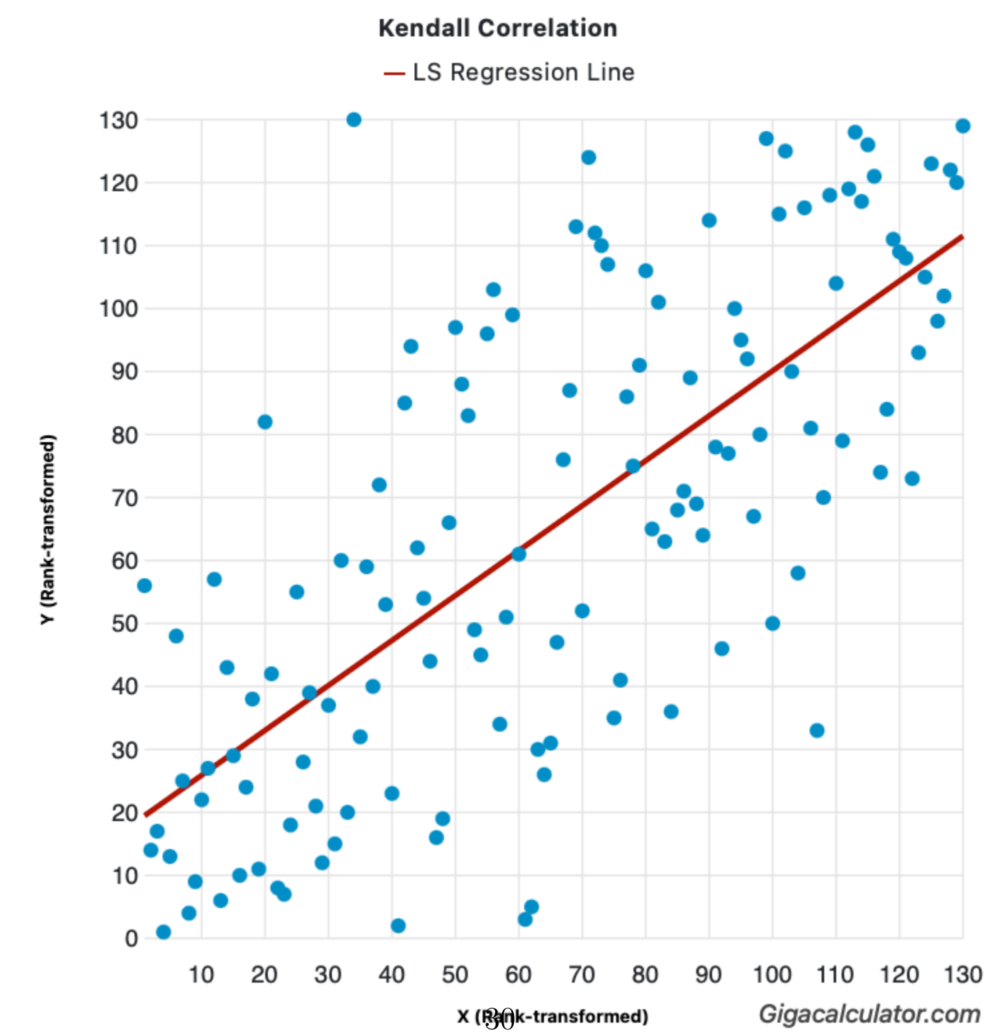

Figure 3.3: Total Score Kendall's Tau

```
Line 1: S=sum(H+C,1);
Line 2: Hv=(H+C)./S;
Line 3: alpha = 0.85;
Line 4: e = ones(130,1);
Line 5: v = e/130;
Line 6: Ga = alpha*Hv+(1-alpha)*v*e';
Line 7: p = rand(130,1);  % random initial vector
Line 8: p = p/sum(p);   % normalise p to have unit sum
Line 9: for k=1:100
Line 10: p = Ga*p;
Line 11: end
```

Figure 3.4: Matlab Code for Finding a Pagerank Vector

	Team	p-Value	AP Top 25
1	Liberty	0.00304746969912415	Georgia
2	Georgia	0.00381204199025146	Alabama
3	Wisconsin	0.00407542269923828	Michigan
4	Penn State	0.00417366519034118	Ohio State
5	Clemson	0.00419166990380779	Baylor
6	TexasA&M	0.0042417589958973	Oklahoma State
7	Nebraska	0.00473118334230752	Cincinnati
8	Oklahoma State	0.00486888996498965	Notre Dame
9	Michigan	0.00488610520420468	Oklahoma
10	NC State	0.00497641969495136	Michigan State
11	Alabama	0.00521606093022588	Ole Miss
12	Purdue	0.00530534261905736	Arkansas
13	Iowa	0.00533300262674794	Utah
14	Minnesota	0.00534577785603451	Pitt
15	Kansas State	0.00536983025416077	NC State
16	Iowa State	0.0054653187130986	Clemson
17	Auburn	0.00563143171890239	Minnesota
18	Syracuse	0.00566076793675136	Wisconsin
19	Illinois	0.00574882147723986	Purdue
20	Arkansas	0.00575824184598954	Tennessee
21	Florida State	0.00576672770807803	Kentucky
22	Ole Miss	0.00592753339050079	Iowa
23	Michigan State	0.0059897626199456	Wake Forrest
24	California	0.00601905226057784	Utah State
25	Florida	0.0061003127503729	San Diego State

Table 3.2: Total Score

not an exclusive factor, as it was in Margin of Victory. The teams listed in 3.2 scored many points, whether they were victorious in the game or not. Teams in the Margin of Victory table outscored their opponents the most, while teams in the Total Score table scored many points, regardless of a win or loss.

3.3 Conclusions

The Google PageRank algorithm proved to be an accurate and effective tool for providing rankings that correlated well with the Final AP rankings. The Margin of Victory method provided a better correlation with the AP rankings than did the Total Score rankings. Mathematica was an effective tool for illustrating the network involved in the algorithm. The combination of Matlab and the power method is a better tool for computing matrix eigenvalues of large matrices such as the 130 by 130 NCAA FBS season results. The FBS versus FCS matchups did not appreciably affect the rankings of the teams and were ignored. This observation was verified computationally and supported by previous research [20]. Future directions for research could be the effect of first downs, turnovers, yards gained, passing yards, running yards, and many other game parameters to see which factors contribute the most to team success.

The method of applying the PageRank algorithm to rank sports teams can be directed to many future research opportunities. This method can be used to produce and compute rankings for any sport. In a similar approach, this ideology could be applied to any factor regarding sports. For example, time of possession, turnovers, total yardage, yards gained by pass and run, points gained by field goals, time of possession, and many other factors could be applied to the PageRank method. Another way to further research the ranking of sports teams would be the changing of the value of α. Throughout our research, we used a standard α of 0.85 to evaluate the teleportation index. The value of α could range from 0 to 1. Changing the value of α allows for a different ranking vector to be obtained.

Chapter 4

Appendix

A complete list of the Margin of Victory PageRank rankings of the 2021 FBS football season is given next.

Air Force Falcons 0.002297 Akron Zips 0.061325 Alabama Crimson Tide 0.001968 Appalachian State Mountaineers 0.002951 Arizona Wildcats 0.015740 Arizona State Sun Devils 0.004497 Arkansas Razorbacks 0.002487 Arkansas State Red Wolves 0.015871 Army Black Knights 0.003912 Auburn Tigers 0.003000 Ball State Cardinals 0.008654 Baylor Bears 0.002102 Boise State Broncos 0.002536 Boston College Eagles 0.004283 Bowling Green Falcons 0.057965 Buffalo Bulls 0.028531 BYU Cougars 0.002298 California Golden Bears 0.014806 Central Michigan Chippewas 0.004611 Charlotte 49ers 0.013100 Cincinnati Bearcats 0.001984 Clemson Tigers 0.002078 Coastal Carolina Chanticleers 0.001970 Colorado Buffaloes 0.012612 Colorado State Rams 0.009721 Duke Blue Devils 0.012657 East Carolina Pirates 0.002502 Eastern Michigan Eagles 0.012743 FIU Panthers 0.030598 Florida Gators 0.003968 Florida Atlantic Owls 0.006102 Florida State Seminoles 0.002742 Fresno State Bulldogs 0.002873 Georgia Bulldogs 0.001960 Georgia Southern Eagles 0.009894 Georgia State Panthers 0.005741 Georgia Tech Yellow Jackets 0.004574 Hawaii Rainbow Warriors 0.009230 Houston Cougars 0.002779 Illinois Fighting Illini 0.004247 Indiana Hoosiers 0.007545 Iowa

Hawkeyes 0.002654 Iowa State Cyclones 0.005514 Kansas Jayhawks 0.014448 Kansas State Wildcats 0.002845 Kent State Golden Flashes 0.008974 Kentucky Wildcats 0.002555 Liberty Flames 0.004807 Louisiana Ragin' Cajuns 0.003103 Louisiana-Monroe Warhawks 0.018261 Louisiana Tech Bulldogs 0.013464 Louisville Cardinals 0.002998 LSU Tigers 0.003185 Marshall Thundering Herd 0.003596 Maryland Terrapins 0.004653 Memphis Tigers 0.004637 Miami (FL) Hurricanes 0.002730 Miami (OH) RedHawks 0.004825 Michigan Wolverines 0.002022 Michigan State Spartans 0.002312 Middle Tennessee Blue Raiders 0.006750 Minnesota Golden Gophers 0.009449 Mississippi State Bulldogs 0.004004 Missouri Tigers 0.003702 Navy Midshipmen 0.005613 Nebraska Cornhuskers 0.003237 Nevada Wolf Pack 0.003568 New Mexico Lobos 0.012356 New Mexico State Aggies 0.015929 North Carolina Tar Heels 0.006411 North Carolina State Wolfpack 0.002510 North Texas Mean Green 0.010420 Northern Illinois Huskies 0.004430 Northwestern Wildcats 0.011050 Notre Dame Fighting Irish 0.001947 Ohio Bobcats 0.031091 Ohio State Buckeyes 0.002206 Oklahoma Sooners 0.002060 Oklahoma State Cowboys 0.002004 Old Dominion Monarchs 0.004932 Ole Miss Rebels 0.002465 Oregon Ducks 0.004454 Oregon State Beavers 0.005858 Penn State Nittany Lions 0.002411 Pittsburgh Panthers 0.002307 Purdue Boilermakers 0.002760 Rice Owls 0.011668 Rutgers Scarlet Knights 0.007779 San Diego State Aztecs 0.003019 San Jose State Spartans 0.009861 SMU Mustangs 0.002601 South Alabama Jaguars 0.012184 South Carolina Gamecocks 0.003553 South Florida Bulls 0.006962 Southern Miss Golden Eagles 0.018827 Stanford Cardinal 0.010281 Syracuse Orange 0.004115 TCU Horned Frogs 0.005330 Temple Owls 0.018647 Tennessee Volunteers 0.003038 Texas Longhorns 0.009096 Texas A&M Aggies 0.002522 Texas State Bobcats 0.010770 Texas Tech Red Raiders 0.006355 Toledo Rockets 0.004997 Troy Trojans 0.011609 Tulane Green Wave 0.005181 Tulsa Golden Hurricane 0.002877 UAB Blazers 0.006338 UCF Knights 0.003262 UCLA Bruins 0.002863 UConn Huskies 0.026607 UMass Minutemen 0.022084 UNLV Rebels 0.008076 USC Trojans 0.008906 UTEP Miners 0.003932 UTSA Roadrunners 0.004133 Utah Utes 0.002646 Utah State Aggies

0.005562 Vanderbilt Commodores 0.010507 Virginia Cavaliers 0.004298 Virginia Tech Hokies 0.005800 Wake Forest Demon Deacons 0.002552 Washington Huskies 0.004219 Washington State Cougars 0.002647 West Virginia Mountaineers 0.003823 Western Kentucky Hilltoppers 0.006111 Western Michigan Broncos 0.005934 Wisconsin Badgers 0.002887 Wyoming Cowboys 0.012582

A complete list of the Points Scored PageRank rankings is given below.

Air Force Falcons 0.006603 Akron Zips 0.013724 Alabama Crimson Tide 0.005241 Appalachian State Mountaineers 0.007499 Arizona Wildcats 0.007898 Arizona State Sun Devils 0.006329 Arkansas Razorbacks 0.005779 Arkansas State Red Wolves 0.012095 Army Black Knights 0.007541 Auburn Tigers 0.005650 Ball State Cardinals 0.009814 Baylor Bears 0.006459 Boise State Broncos 0.006511 Boston College Eagles 0.006776 Bowling Green Falcons 0.010981 Buffalo Bulls 0.010814 BYU Cougars 0.007602 California Golden Bears 0.006032 Central Michigan Chippewas 0.012192 Charlotte 49ers 0.010518 Cincinnati Bearcats 0.006180 Clemson Tigers 0.004215 Coastal Carolina Chanticleers 0.009016 Colorado Buffaloes 0.006502 Colorado State Rams 0.008345 Duke Blue Devils 0.009839 East Carolina Pirates 0.006889 Eastern Michigan Eagles 0.009999 FIU Panthers 0.007864 Florida Gators 0.006119 Florida Atlantic Owls 0.006928 Florida State Seminoles 0.005791 Fresno State Bulldogs 0.007081 Georgia Bulldogs 0.003839 Georgia Southern Eagles 0.009574 Georgia State Panthers 0.009059 Georgia Tech Yellow Jackets 0.007608 Hawaii Rainbow Warriors 0.010284 Houston Cougars 0.008187 Illinois Fighting Illini 0.005771 Indiana Hoosiers 0.006566 Iowa Hawkeyes 0.005357 Iowa State Cyclones 0.005490 Kansas Jayhawks 0.009831 Kansas State Wildcats 0.005392 Kent State Golden Flashes 0.013347 Kentucky Wildcats 0.006186 Liberty Flames 0.003070 Louisiana Ragin' Cajuns 0.007135 Louisiana-Monroe Warhawks 0.009357 Louisiana Tech Bulldogs 0.009366 Louisville Cardinals 0.006815 LSU Tigers 0.006599 Marshall Thundering Herd 0.006855 Maryland Terrapins 0.007592 Memphis Tigers 0.009482 Miami (FL) Hurricanes 0.006310 Miami (OH) RedHawks 0.008865 Michigan Wolverines 0.004910 Michigan State Spartans 0.006014 Middle Tennessee Blue Raiders 0.007790 Minnesota Golden Gophers 0.005361 Mississippi State Bulldogs 0.006659 Missouri Tigers 0.008643 Navy Midshipmen 0.007442 Nebraska Cornhuskers 0.004760 Nevada Wolf Pack 0.008679 New Mexico Lobos 0.008127 New Mexico State Aggies 0.010140 North Carolina Tar Heels 0.007635 North Carolina State Wolfpack 0.005000 North Texas Mean Green 0.007913 Northern Illinois Huskies 0.011972 Northwestern Wildcats 0.006766 Notre Dame Fighting Irish 0.006568 Ohio Bobcats 0.010453 Ohio State Buckeyes 0.006285 Oklahoma Sooners 0.007265 Oklahoma State Cowboys 0.004892 Old Dominion Monarchs 0.008413 Ole Miss Rebels 0.005949 Oregon Ducks 0.008070 Oregon State Beavers 0.007943 Penn State Nittany Lions 0.004200 Pittsburgh Panthers 0.007238 Purdue Boilermakers 0.005332 Rice Owls 0.009068 Rutgers Scarlet Knights 0.006489 San Diego State Aztecs 0.007002 San Jose State Spartans 0.007742 SMU Mustangs 0.008127 South Alabama Jaguars 0.009226 South Carolina Gamecocks 0.006317 South Florida Bulls 0.008463 Southern Miss Golden Eagles 0.007797 Stanford Cardinal 0.008242 Syracuse Orange 0.005685 TCU Horned Frogs 0.007999 Temple Owls 0.011338 Tennessee Volunteers 0.007035 Texas Longhorns 0.008135 Texas A&M Aggies 0.004266 Texas State Bobcats 0.010629 Texas Tech Red Raiders 0.007797 Toledo Rockets 0.008604 Troy Trojans 0.008350 Tulane Green Wave 0.007789 Tulsa Golden Hurricane 0.008222 UAB Blazers 0.007027 UCF Knights 0.007679 UCLA Bruins 0.007636 UConn Huskies 0.009882 UMass Minutemen 0.010104 UNLV Rebels 0.008571 USC Trojans 0.008731 UTEP Miners 0.007895 UTSA Roadrunners 0.008369 Utah Utes 0.006886 Utah State Aggies 0.008797 Vanderbilt Commodores 0.009217 Virginia Cavaliers 0.007191 Virginia Tech Hokies 0.007171 Wake Forest Demon Deacons 0.007518 Washington Huskies 0.006219 Washington State Cougars 0.007747 West Virginia Mountaineers 0.006717 Western Kentucky Hilltoppers 0.009320 Western Michigan Broncos 0.009984 Wisconsin Badgers 0.004102 Wyoming Cowboys 0.009729

Bibliography

[1] *Coaches' poll*, https://americanfootballdatabase.fandom.com/wiki/Coaches%27_Poll.

[2] *College football playoff history*, https://collegefootballplayoff.com/sports/2016/9/30/overview, journal=College Football Playoff.

[3] *2021 year summary: College football at sports*, https://www.sports-reference.com/cfb/years/2021.html, journal=Reference.com, 2021.

[4] *College football playoff payouts 2022-2023*, https://businessofcollegesports.com/college-football-playoff-payouts/, January 2022.

[5] Bahman Bahmani, Abdur Chowdhury, and Ashish Goel, *Fast incremental and personalized pagerank*, Proceedings of the VLDB Endowment **4** (2010), no. 3, 173–184.

[6] Stan Becton, *Everything you need to know about the 2021 fcs championship semifinals*, https://www.ncaa.com/news/football/article/2021-12-14/2021-fcs-championship-everything-you-need-know-about-2021-fcs-championship-semifinals, December 2021.

[7] Joost Berkhout and Bernd F. Heidergott, *Ranking nodes in general networks: a Markov multi-chain approach*, Discrete Event Dyn. Syst. **28** (2018), no. 1, 3–33. MR 3772939

[8] Ryan Brewer, *College football rankings*, https://graphics.wsj.com/table/NCAA_2019.

[9] S. Brin, L. Page, R. Motwami, and T. Winograd, *The pagerank citation ranking: Bringing order to the web*, Tech. report, Stanford University, 1998.

[10] Amanda Brooks, *2022 college football playoff national championship nets 22.6 million viewers, cable's top telecast in two years*, https://espnpressroom.com/us/press-releases/2022/01/2022-college-football-playoff-national-championship-nets-22-6-million-viewers-cable January 2022.

[11] Matt Brown, *What we can learn from college football's very first ap poll, where minnesota was no. 1*, http://www.sbnation.com/college-football/2018/10/19/18001148/ap-top-25-poll-history, October 2018.

[12] Stephen K. Callaway, *Internet banking and performance*, American Journal of Business **26** (2011), no. 1, 12–25.

[13] Raju Choudhary, *Google net worth 2022: Top 5 google (alphabet) shareholderrs*, https://caknowledge.com/google-net-worth/, September 2022.

[14] Christopher Engström, *Pagerank in evolving networks and applications of graphs in natural language processing and biology*, Ph.D. thesis, Mälardalen University, 2016, p. 1–256.

[15] Ayman Farahat, Thomas LoFaro, Joel C. Miller, Gregory Rae, and Lesley A. Ward, *Authority rankings from hits, pagerank, and salsa: Existence, uniqueness, and effect of initialization*, SIAM Journal on Scientific Computing **27** (2006), no. 4, 1181–1201.

[16] Graham Farmelo (ed.), *It must be beautiful*, Granta Publications, London, 2003, Great equations of modern science. MR 1990427

[17] Massimo Franceschet, *Pagerank*, Communications of the ACM **54** (2011), no. 6, 92–101.

[18] David F. Gleich, *PageRank beyond the web*, SIAM Rev. **57** (2015), no. 3, 321–363. MR 3376760

[19] Anjela Yuryevna Govan, *Ranking theory with application to popular sports*, ProQuest LLC, Ann Arbor, MI, 2008, Thesis (Ph.D.)–North Carolina State University. MR 2712816

[20] Ilse C. F. Ipsen and Teresa M. Selee, *PageRank computation, with special attention to dangling nodes*, SIAM J. Matrix Anal. Appl. **29** (2007), no. 4, 1281–1296. MR 2369296

[21] Sam Kahn Jr, `https://theathletic.com/4201182/2021/09/01/texas-am-jimbo-fisher-agree-to-4-year-extension-through-2031-and-raise/`, August 2021.

[22] Amy N. Langville and Carl D. Meyer, *Google's PageRank and beyond: the science of search engine rankings*, Princeton University Press, Princeton, NJ, 2006. MR 2262054

[23] Steven J. Leon, *Linear algebra with applications*, Macmillan, Inc., New York; Collier Macmillan Publishers, London, 1980. MR 566769

[24] C.-S. Liao, K. Lu, M. Baym, R. Singh, and B. Berger, *Isorankn: Spectral methods for global alignment of multiple protein networks*, Bioinformatics **25** (2009), no. 12, i253–i258.

[25] Joseph Magiya, *Kendall rank correlation explained.*, `https://towardsdatascience.com/kendall-rank-correlation-explained-dee01d99c535`, June 2019.

[26] Richard von Mises and Hilda Pollaczek-Geiringer, *Praktische verfahren der gleichungsauflösung*, VDI-Verlag, 1929.

[27] Julie L Morrison, Rainer Breitling, Desmond J Higham, and David R Gilbert, *Generank: Using search engine technology for the analysis of microarray experiments*, BMC Bioinformatics **6** (2005), no. 1.

[28] Saeko Nomura, Satoshi Oyama, Tetsuo Hayamizu, and Toru Ishida, *Analysis and improvement of hits algorithm for detecting web communities*, Systems and Computers in Japan **35** (2004), no. 13, 32–42.

[29] Spencer Parlier, *College football history: Notable firsts and milestones*, `https://www.ncaa.com/news/ncaa/article/2020-01-31/college-football-history-notable-firsts-and-milestones`, June 2022.

[30] Oskar Perron, *Zur Theorie der Matrices*, Math. Ann. **64** (1907), no. 2, 248–263. MR 1511438

[31] Wayne Staats, *College football rankings: Every poll explained and how they work*, `https://www.ncaa.com/news/football/article/2019-07-08/college-football-rankings-every-poll-explained-how-they-work`, August 2022.

[32] Andy Staples, *The chaos and consequences of the bcs, 20 years after its inaugural season*, `https://www.si.com/college/2018/07/09/bcs-history-20th-anniversary-controversy-tennessee-florida-state`, July 2018.

[33] Richard Tovar, *Ncaa college football 2021: Lists of the fbs conferences for this season*, `https://bolavip.com/en/sports/ncaa-college-football-2021-lists-of-the-FBS-conferences-for-this-season-20210824-00`
`html`, 2021.

[34] Yale University, *Linear regression*, http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm.

[35] Douglas B. West, *Introduction to graph theory*, Prentice Hall, Inc., Upper Saddle River, NJ, 1996. MR 1367739

[36] Gang Wu, Ying Zhang, and Yimin Wei, *Accelerating the Arnoldi-type algorithm for the PageRank problem and the ProteinRank problem*, J. Sci. Comput. **57** (2013), no. 1, 74–104. MR 3095291

[37] Elliot J. Yates and Louise C. Dixon, *Pagerank as a method to rank biomedical literature by importance*, Source Code for Biology and Medicine **10** (2015), no. 1.

[38] Laurie Zack, Ron Lamb, and Sarah Ball, *An application of Google's PageRank to NFL rankings*, Involve **5** (2012), no. 4, 463–471. MR 3069048

[39] Tiecheng Zhou, Ernesto Martinez-Baez, Gregory Schenter, and Aurora E. Clark, *Pagerank as a collective variable to study complex chemical transformations and their energy landscapes*, The Journal of Chemical Physics **150** (2019), no. 13, 134102.