A Markov Method for Ranking College Football Conferences

R. Bruce Mattingly and Amber J. Murphy
SUNY Cortland

1 Introduction

The use of mathematical methods to develop rankings of sports teams is certainly not a new idea. See for instance, Ford [7]. Ranking methods are particularly important in Division I college football, because unlike nearly any other sport, the champion is not determined through a playoff system. Instead, the Bowl Championship Series (BCS) Standings are used to select the two teams that will compete for the national championship. The BCS Standings are determined through a combination of two human polls and six computer ranking methods [2]. These methods take a variety of factors into account, including a team's individual win-loss record, its strength of schedule, and so on. Most college teams are affiliated with conferences, and the reputations of the various conferences can have a large influence on the team rankings, particularly in the human polls. Jeff Sagarin's NCAA football ratings published in USA Today [20] include conference rankings as well as rankings of individual teams. There is a great deal of interest in ranking methods for college football. Kenneth Massey maintains an extensive web site [14] that provides information on an astonishing number of ranking methods in addition to his own. For instance, in the 2008 season, Massey compared the rankings produced by 113 different methods [13].

A natural approach to developing a mathematical ranking method is to create a matrix whose entries are determined in some way by the results of games played between teams. Keener [8] discussed a method in which the ranking of teams is derived from the computation of the dominant eigenvector of a non-negative matrix. Massey [10] took an approach that involves the solution of a least squares problem. Colley [4] developed a method that also involves the solution of a linear system of equations. The Massey and Colley methods are both used by the BCS, as is Sagarin's method. Matrix methods based on Markov Chains have been proposed by Redmond [17] and Govan et al. [5]. The latter method is a generalization of the Google PageRank algorithm developed by Page et al. [16].

In this paper we propose a Markov method that is also similar to the PageRank idea. Our method takes advantage of the block structure that is particularly evident in college football due to the low number of games played (relative to other sports such as basketball) and the fact that most teams are affiliated with conferences. For the most part, teams within conferences all play each other (some larger conferences are further divided into two divisions), and there are comparatively fewer interconference games. Because of this, a matrix whose entries are determined by the outcomes of games can be partitioned very nicely into a block structure in which the dense diagonal blocks represent conference games, and the sparse off-diagonal blocks correspond to inter-conference matchups. Figure 1 shows the structure of the adjacency matrix that corresponds to the 2006 Division I college football season, where each nonzero entry represents a game between two teams.

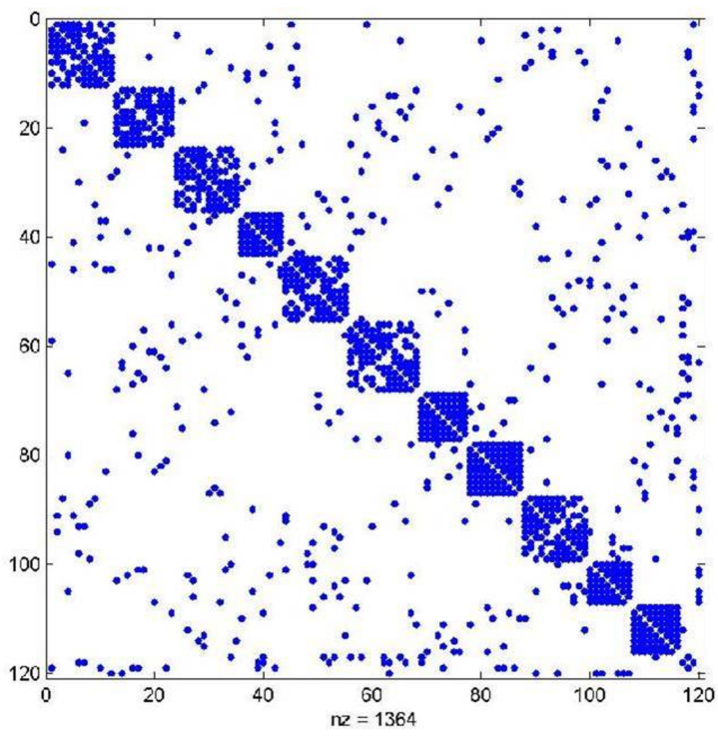

Figure 1: Adjacency Matrix for 2006 Season

To exploit this natural block structure, we propose the use of the method of stochastic complementation developed by Meyer [15]. This method provides a natural way to subdivide the original Markov chain into several smaller chains, each of which would represent a single conference. The stationary distribution vector (or steady state vector) of each smaller chain can be computed independently. These vectors can then be combined to produce the stationary distribution vector of the original Markov chain. This process involves the computation of a coupling matrix that represents yet another Markov chain in which all of the teams in a particular conference are aggregated into a single state. A benefit of this approach is that, in addition to ranking the individual teams, the stationary distribution vector of the coupling matrix provides a natural way to rank the conferences.

2 Overview of the method

The basic idea of our Markov chain approach to team ranking is straightforward. Each state in the Markov chain corresponds to one football team, and transitions are only allowed between teams that have played each other. Our ranking method can be thought of as an iterative procedure that occurs at the end of the season. Suppose that we have a large number of college football fans who (as improbable as this sounds) have no preference for which team they support. Initially we distribute the fans equally among all available teams. At each iteration of our method, a fan can choose to either remain with the same team or change to another team. The probability that a fan remains with a given team should be higher for a winning team than it is for a losing team. If a fan does not remain with the same team, it might seem logical to assign the fan (with equal probability) to any of the teams to whom the current team lost. Markov chain theory guarantees that if the transition matrix is irreducible and aperiodic, then the iterative process will always converge to a unique steady state vector [3, p. 222]. The percentage of fans assigned to each team would therefore represent that team's rating. We sort the rating vector in decreasing order, and then each team's position in the sorted vector represents its ranking.

A potential problem with this approach is that if there are undefeated teams, they would represent absorbing states in our Markov chain. In the case where we had only one undefeated team, all of the fans would end up assigned to that single team. Our ranking vector would not be particularly useful, because the undefeated team would be ranked first, with everyone else tied for second. We therefore make the following adjustment: we allow transitions between teams who played each other in both directions.

The guiding principles that we used in developing our ranking method are easy to describe qualitatively: the probability of a fan remaining with a team should be higher for a winning team than for a losing team, and if a fan changes teams, transitions to teams that defeated the current team should occur with higher probability than transitions to teams who lost to the current team. Determining the specific probabilities that should be used became the central focus of our research.

Algorithm 1 describes our method for constructing the transition matrix.

Input: List of results of all games between a set of teams
Output: A stochastic matrix with transition probabilities based on game results
Let N = maximum number of games played by any team;
Let p = a value in the interval $(0.5, 1.0)$;
foreach *Game* k **do**
  Let i = index of team that won game k;
  Let j = index of team that lost game k;
  $a_{ji} = p$;
  $a_{ij} = 1 - p$;
end
foreach *Team* i **do**
  $a_{ii} = N - \sum_{j \neq i} a_{ij}$;
end
A = A/N;

Algorithm 1: Construction of transition matrix

Algorithm 1 assumes that the same two teams never play more than once. In college football, exceptions to this can occur, particularly in large conferences that have a conference championship game between the winners of the two divisions after the regular season games, but before the bowl games. Rematches could also occur in bowl games, but our interest was in determining rankings before the bowl games began so that we could compare our results to the official BCS standings that were used to select the top two teams. In both 2006 and 2008, one conference championship game was a rematch, and in 2007 there were 3 rematches. We dealt with rematches in the following way: if two teams played each other twice with the same winner each time, we made no adjustment to our transition matrix. If the teams split the two games, then we reasoned that transitions between the two teams should be equally likely in either direction, so we set $a_{ij} = a_{ji} = 0.5$ regardless of the value of p that we were using.

3 Illustrative Examples

Example 1. Our first example demonstrates the construction of the transition matrix. It involves only six teams, so we do not deal with partitioning the Markov chain into smaller chains. We assume that each team plays only 3 games, and the hypothetical results of these games are summarized in Table 1.

Team	Wins	Losses
1	4, 5, 6	none
2	3	4, 5
3	4,6	2
4	2	1, 3
5	2, 6	1
6	none	1, 3, 5

Table 1: Game Results for Example 1

To construct the transition matrix, we follow the steps below. For purposes of this example we chose $p = 0.8$.

- **Step 1:** If i beats j, set $a_{ji} = 0.8$ and set $a_{ij} = 0.2$, producing the following matrix:

$$\begin{pmatrix} 0 & 0 & 0 & 0.2 & 0.2 & 0.2 \\ 0 & 0 & 0.2 & 0.8 & 0.8 & 0 \\ 0 & 0.8 & 0 & 0.2 & 0 & 0.2 \\ 0.8 & 0.2 & 0.8 & 0 & 0 & 0 \\ 0.8 & 0.2 & 0 & 0 & 0 & 0.2 \\ 0.8 & 0 & 0.8 & 0 & 0.8 & 0 \end{pmatrix}$$

- **Step 2.** In this step we adjust the diagonal entries so that all row sums in the matrix are constant. While teams never play themselves, the diagonal entry represents the probability that a fan will stay with the current team rather than changing to a new one. We accomplish this as follows. For each row of the matrix, set $a_{ii} = N - \sum_{i \neq j} a_{ij}$, where N is the total number of games played by each team. In this example, $N = 3$. In the event that each team does not play the same number of games, this value just needs to be large enough to ensure that the diagonal entry is non-negative. In our experiments on actual data, we used $N = 13$ since teams normally play 12 regular season games, and a few teams also play a conference championship game. After this adjustment, all rows sums are the same. We then divide the entire matrix by N (3 in this example) to create a stochastic matrix. Our matrix now looks like this:

$$\begin{pmatrix} 0.8000 & 0 & 0 & 0.0667 & 0.0667 & 0.0667 \\ 0 & 0.4000 & 0.0667 & 0.2667 & 0.2667 & 0 \\ 0 & 0.2667 & 0.6000 & 0.0667 & 0 & 0.0667 \\ 0.2667 & 0.0667 & 0.2667 & 0.4000 & 0 & 0 \\ 0.2667 & 0.0667 & 0 & 0 & 0.6000 & 0.0667 \\ 0.2667 & 0 & 0.2667 & 0 & 0.2667 & 0.2000 \end{pmatrix}$$

Note: In the methods described by Keener, if i defeats j, then a_{ij} is close to 1, and a_{ji} is close to zero. Here, we have those two entries transposed. Since our method uses Markov chains, it bears some resemblance to the methods used for Google's PageRank algorithm as described in Langville and Meyer [9, pp. 31–34]. In that work, a_{ij} is set to some nonzero value whenever there is a link from page i to page j. In other words, page i is recommending page j as an "authority" by virtue of its link. The analogy here is that team j "recommends" team i if j was defeated by i.

The steady state vector for this Markov chain is given below:

$$\begin{pmatrix} 0.4515 & 0.0858 & 0.1240 & 0.1021 & 0.1741 & 0.0625 \end{pmatrix}$$

According to these values, the teams should be ranked in the following order: $1, 5, 3, 4, 2, 6$. This is consistent with the teams' win-loss records. Team 1 was undefeated and Team 6 was winless. Teams 5 and 3 both won two games, but Team 5 lost only to Team 1. Teams 4 and 2 each had only one win, but Team 4 defeated Team 2. Note that Team 2 actually defeated a higher-ranked team (Team 3). We adopt the terminology used by Ali et al. [1] and refer to this situation as a "violation." It is virtually impossible to develop a ranking system that avoids violations. Although we can define a relation on a set of teams according to who defeated whom, such a relation is clearly not transitive. In this example, 3 beat 4, 4 beat 2, but 2 beat 3. Next, we present a series of examples that demonstrate how the method of stochastic complementation can be used to rank conferences, as well as teams.

Example 2. We assume that we have 10 teams divided into two 5-member conferences. Each team plays every other team in their own conference, but only one team from the other conference, for a total of 5 games. Our hypothetical results are given in Table 2.

West	Conference Wins	Non-conference Wins	East	Conference Wins	Non-conference Wins
1	2, 3, 4, 5	6	6	7, 8, 9, 10	none
2	3, 4, 5	7	7	8, 9, 10	none
3	4, 5	none	8	9, 10	3
4	5	9	9	10	none
5	none	10	10	none	none

Table 2: Game Results for Example 2

We deliberately constructed this example so that within each conference, there is a clear ordering $(1 > 2 > 3 > 4 > 5)$ and $(6 > 7 > 8 > 9 > 10)$ so that we can focus our attention on comparing the two conferences. Using Algorithm 1 with $p = 0.80$ we obtain the following transition matrix:

$$A = \begin{pmatrix} 0.80 & 0.04 & 0.04 & 0.04 & 0.04 & 0.04 & 0 & 0 & 0 & 0 \\ 0.16 & 0.68 & 0.04 & 0.04 & 0.04 & 0 & 0.04 & 0 & 0 & 0 \\ 0.16 & 0.16 & 0.44 & 0.04 & 0.04 & 0 & 0 & 0.16 & 0 & 0 \\ 0.16 & 0.16 & 0.16 & 0.44 & 0.04 & 0 & 0 & 0 & 0.04 & 0 \\ 0.16 & 0.16 & 0.16 & 0.16 & 0.32 & 0 & 0 & 0 & 0 & 0.04 \\ 0.16 & 0 & 0 & 0 & 0 & 0.68 & 0.04 & 0.04 & 0.04 & 0.04 \\ 0 & 0.16 & 0 & 0 & 0 & 0.16 & 0.56 & 0.04 & 0.04 & 0.04 \\ 0 & 0 & 0.04 & 0 & 0 & 0.16 & 0.16 & 0.56 & 0.04 & 0.04 \\ 0 & 0 & 0 & 0.16 & 0 & 0.16 & 0.16 & 0.16 & 0.32 & 0.04 \\ 0 & 0 & 0 & 0 & 0.16 & 0.16 & 0.16 & 0.16 & 0.16 & 0.20 \end{pmatrix}$$

To exploit the natural block structure induced by the conference scheduling, we partition A as follows:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where each submatrix in this example is 5×5. In general, only the diagonal blocks need to be square, with the size of each conforming to the size of that particular conference.

The stochastic complements are computed according to the formulas below (see Meyer for further details):

$$S_{11} = A_{11} + A_{12}(I - A_{22})^{-1} A_{21}$$

$$S_{22} = A_{22} + A_{21}(I - A_{11})^{-1} A_{12}$$

For this example, the matrices turn out to be:

$$S_{11} = \begin{pmatrix} 0.8279 & 0.0461 & 0.0412 & 0.0428 & 0.0421 \\ 0.1745 & 0.6994 & 0.0412 & 0.0428 & 0.0421 \\ 0.2327 & 0.2036 & 0.4594 & 0.0539 & 0.0504 \\ 0.1752 & 0.1691 & 0.1618 & 0.4518 & 0.0422 \\ 0.1752 & 0.1691 & 0.1618 & 0.1642 & 0.3298 \end{pmatrix}$$

and

$$S_{22} = \begin{pmatrix} 0.7533 & 0.0633 & 0.0867 & 0.0497 & 0.0469 \\ 0.2156 & 0.6011 & 0.0867 & 0.0497 & 0.0469 \\ 0.1711 & 0.1656 & 0.5800 & 0.0419 & 0.0414 \\ 0.2133 & 0.1867 & 0.2133 & 0.3400 & 0.0467 \\ 0.2133 & 0.1867 & 0.2133 & 0.1711 & 0.2156 \end{pmatrix}$$

For each stochastic complement, we can compute the unique stationary distribution vectors which satisfy $v_1 S_{11} = v_1, \quad v_1 e = 1$ and $v_2 S_{22} = v_2, \quad v_2 e = 1$, where e is a column vector of all 1's. In this case, our results are as follows:

$$v_1 = \begin{pmatrix} 0.5208 & 0.2310 & 0.1013 & 0.0866 & 0.0603 \end{pmatrix}$$

and

$$v_2 = \begin{pmatrix} 0.4461 & 0.2175 & 0.2041 & 0.0772 & 0.0551 \end{pmatrix}$$

Note that in each vector, the entries appear in decreasing order, consistent with the ordering that was determined by the results of the conference games.

Our primary interest here is in finding the coupling matrix C, whose entries are given by $c_{ij} = v_i A_{ij} e$. In our example, the coupling matrix is:

$$C = \begin{pmatrix} 0.9478 & 0.0522 \\ 0.1355 & 0.8645 \end{pmatrix}$$

and its stationary distribution vector is

$$w = \begin{pmatrix} 0.7221 & 0.2779 \end{pmatrix}$$

In this simple example, the Western conference has a higher rating (0.7221) than the Eastern conference (0.2779), which seems reasonable, given that in head to head competition, Western conference teams won 4 of 5 games. In the next couple of examples, we will leave the conference results unchanged, but adjust the inter-conference results to see the effect upon the stationary vector of the coupling matrix.

Example 3. In this example, the results of the conference games are unchanged, but this time, teams from the Western Conference win 3 of the 5 inter-conference games as shown in Table 3.

West	Conference Wins	Non-conference Wins	East	Conference Wins	Non-conference Wins
1	2, 3, 4, 5	6	6	7, 8, 9, 10	none
2	3, 4, 5	none	7	8, 9, 10	2
3	4, 5	8	8	9, 10	none
4	5	none	9	10	4
5	none	10	10	none	none

Table 3: Game Results for Example 3

In the interest of space we do not display the 10×10 stochastic matrix or the stochastic complements. Adjusting the off-diagonal blocks as noted above leads to the following new coupling matrix:

$$C = \begin{pmatrix} 0.9314 & 0.0686 \\ 0.1087 & 0.8913 \end{pmatrix}$$

whose stationary distribution vector is given by

$$w = \begin{pmatrix} 0.6132 & 0.3868 \end{pmatrix}$$

Note that the Western conference still receives a higher rating than the Eastern conference (0.6132 vs. 0.3868), but the difference is less pronounced than we saw in Example 2.

Example 4. In this final example, we again only change the inter-conference results as shown in Table 4.

West	Conference Wins	Non-conference Wins	East	Conference Wins	Non-conference Wins
1	2, 3, 4, 5	none	6	7, 8, 9, 10	1
2	3, 4, 5	none	7	8, 9, 10	2
3	4, 5	8	8	9, 10	none
4	5	9	9	10	none
5	none	10	10	none	none

Table 4: Game Results for Example 4

Adjusting the off-diagonal blocks as noted above leads to the following coupling matrix:

$$C = \begin{pmatrix} 0.8850 & 0.1150 \\ 0.0657 & 0.9343 \end{pmatrix}$$

The new stationary distribution vector is given by

$$w = \begin{pmatrix} 0.3636 & 0.6364 \end{pmatrix}$$

Note that in this case, the Eastern conference receives the higher rating and is therefore ranked first. Although they only won 2 of the 5 games, their 2 wins came at the expense of the strongest teams in the Western conference. So this method gives more weight to 2 wins against strong teams, as opposed to 3 wins against weaker teams. We hesitate to over-analyze these examples, whose primary purpose is to illustrate how the method works. In the next section, we discuss our experiences in applying this method to actual data.

4 Experimental Results

We tested our ranking method using the complete game results for all 120 college teams in the Division I Football Bowl Subdivision (FBS) from the 2006, 2007 and 2008 seasons. (Note: in 2006, there were only 119 FBS teams, as Western Kentucky did not join this subdivision until 2007.)

Teams in the FBS occasionally play teams from lower divisions. We did not have ready access to complete game data for the teams from the lower divisions, and made the decision to discard game results involving non-FBS teams. In most cases, the FBS teams won these games as expected, with some notable exceptions. For instance, traditional power Michigan lost to Appalachian State in 2007. We recognize that in a few cases, teams who lost to non-FBS teams would benefit with a higher ranking than they would have received had we considered these results. However, we felt that discarding these games would have a minimal effect on our results, since for the most part they affected only lower-ranked teams. Table 5 shows for each year the number of games played by FBS teams against non-FBS opponents, the number of losses to such teams, and the highest BCS rank of a team from the FBS that had such a loss.

Year	Games	Losses	Rank
2006	78	7	63
2007	80	9	29
2008	86	2	108

Table 5: Games involving non-FBS teams

Data for this study was obtained from an Internet site maintained by James Howell [6]. We wrote C++ code to extract the information that we needed from the raw data and to create the transition matrix. Matlab was used for all of the matrix computations. The main focus of our effort was to experiment with the values of the transition probabilities that we assigned for wins and losses. In the examples above, we used $p = 0.80$. In our experiments, we tested values of p between 0.55 and 0.95 in increments of 0.05. Note: a value of $p = 0.50$ would not be useful, because it would mean that transitions between teams would occur in either direction with equal likelihood. The resulting steady state vector would therefore be the vector of all ones (normalized by the number of teams.) In other words, all 120 teams would be ranked equally.

Google's PageRank algorithm inspired one other variation that we tested. In the PageRank algorithm, transitions between states (web pages) occur primarily due to explicit links between pages. It is recognized, however, that surfers sometimes do not follow links, but have the ability to jump to any other web page simply by typing in a new URL. We wanted to incorporate this idea into our ranking method in accordance with the old adage that "on any given day" any team is capable of defeating any other team. To implement this, transition matrix A is constructed according to Algorithm 1. A second matrix G is created, with each entry equal to $1/N$ where N is the number of teams. In matrix G, a transition from any team to any other team is equally likely. The final transition matrix is the weighted sum

$$\alpha A + (1 - \alpha)G.$$

Langville and Meyer [9, p. 41] report that in the case of ranking web pages, Google gets the best results using $\alpha = 0.85$. In our tests, the values of α that we used ranged from 0.80 to 1.00 in increments of 0.05. For values of $\alpha < 1$, we could use $p = 1$ in the construction of matrix A since the incorporation of the G matrix would prevent absorbing states.

Our rankings were based on the results of games played through the end of the regular season, including conference championships for those conferences that have them, but not the bowl games. We then counted the number of violations, that is, the number of regular season games in which a game was won by a lower-ranked team. Our assumption is that, given a choice between two ranking methods, the one that produces the lower number of violations is preferred.

The graphs shown in Figures 2, 3 and 4 summarize the performance of the ranking methods that we tested on actual data from the 2006, 2007 and 2008 seasons. The horizontal axis in each graph shows the values of p that we tested (before the rows of the matrix were normalized.) The vertical axis shows the number of violations. Each graph includes 5 sets of data, corresponding to the values of α that we tested.

We make two observations about our results from these three seasons. First of all, the method generally produced fewer violations when lower values of p were used. In each of the three years, the fewest violations were produced using either $p = 0.55$ or $p = 0.60$. Secondly, it appears that using the G matrix to allow transitions between all teams generally increased the number of violations, although in 2006, the combination of $p = 0.55$ and $\alpha = 0.95$ produced the best results.

Table 6 compares the performance of our best method for each year against the final BCS standings as reported on Massey's web site at the end of the regular season but before the bowl games [11, 12, 13]. We report the number of violations in the rankings produced by each method. We were also interested in how accurately each method could "predict" the winners

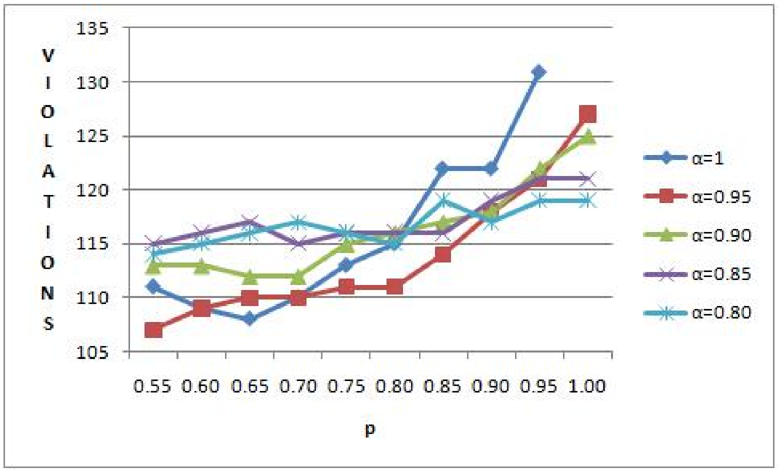

Figure 2: Test results from 2006 season

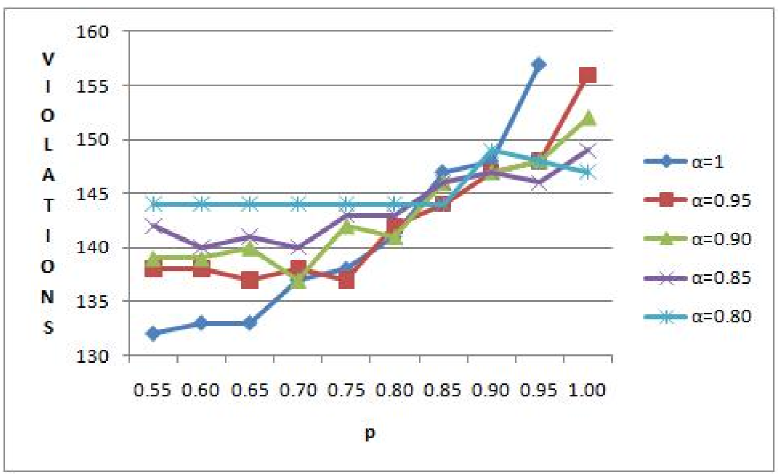

Figure 3: Test results from 2007 Season

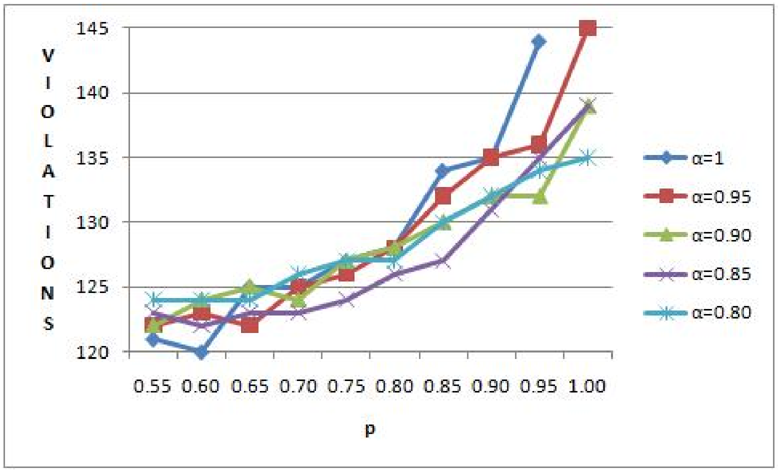

Figure 4: Test results from 2008 Season

Year	Method	Number of Games	Violations	Number of Bowls	Bowl Upsets
2006	$p = 0.55, \alpha = 0.95$	682	107	32	11
	BCS		107		11
2007	$p = 0.55, \alpha = 1.0$	686	132	32	10
	BCS		128		14
2008	$p = 0.60, \alpha = 1.0$	683	120	34	17
	BCS		125		19

Table 6: Performance Comparison for Ranking Methods

of the bowl games that are played after the regular season games. If a team wins a bowl game against a higher-ranked opponent, we refer to this as an "upset." It might appear that an upset is just another name for a violation, but there is one difference – the results of the bowl games were not considered when the rankings were calculated.

It appears that generally, our method produces comparable results to the BCS rankings, but there is no clear pattern. Not surprisingly, both methods did a better job on the regular season games than on the bowl games. The number of violations was consistently in the range of 16–19%. By contrast, the number of bowl upsets ranged from 31% to 56%. In 2006, the two methods had the same number of violations and failed to predict 11 bowl games correctly (out of 32.) In 2007, our best method produced more violations than the BCS rankings, but we had fewer bowl upsets. In 2008, our best method produced fewer violations than the BCS rankings. In that year, neither method did a particularly good job of predicting the bowl game results, although we did slightly better than the BCS, predicting 50% correctly (17 out of 34.)

We were also curious to compare the teams that would have been selected to play in the BCS championship game based on our results. Tables 7, 8 and 9 list the top 5 teams as ranked by our best method, and also shows how those teams fared in the official BCS rankings published at the end of the regular season.

Team	Our rank	BCS rank
Florida	1	2
Ohio State	2	1
Southern California	3	4
Michigan	4	3
Louisville	5	6

Table 7: 2006 Team Rankings

Team	Our rank	BCS rank
Missouri	1	6
Virginia Tech	2	3
Georgia	3	5
Louisiana State	4	2
Ohio State	5	1

Table 8: 2007 Team Rankings

Team	Our rank	BCS rank
Oklahoma	1	1
Florida	2	2
Texas	3	3
Texas Tech	4	7
Utah	5	6

Table 9: 2008 Team Rankings

In 2006, our method would have selected the same two teams (Florida and Ohio State) for the BCS championship game, although we had the top two teams reversed, and as it turns out, Florida won the game. In 2008, the correlation was even stronger – our method agreed with the BCS rankings on the top three teams. In 2007, the results were quite different. Our top two teams (Missouri and Virginia Tech) were ranked 6th and 3rd, respectively, by the BCS. It is worth noting, however, that Missouri lost to Oklahoma in their conference championship game. Prior to that loss, Missouri had been ranked first in the BCS. Missouri's only other loss was also to Oklahoma, earlier in the season. In that light, the results of our ranking method are more understandable – Missouri's loss in the conference championship game provided no new information.

We were curious to explore other ways of handling rematches in cases where the same team won both games. This situation did not occur in either 2006 or 2008. In 2007, however, there were two instances. In addition to Oklahoma defeating Missouri twice, Central Florida also defeated Tulsa twice. We decided to re-run our method (with $p = 0.55$ and $\alpha = 1.00$) using the following adjustment: rather than ignoring the conference championship game, we simply doubled the probabilities in the appropriate locations in the transition matrix. This resulted in a corresponding reduction in the diagonal entry for each of the teams involved in these rematches, but the value of the diagonal is no different than it would have been if the conference championship game had not been a rematch. The adjusted results are shown in Table 10.

Team	Our rank	BCS rank
Virginia Tech	1	3
Georgia	2	5
Louisiana State	3	2
Missouri	4	6
Ohio State	5	1

Table 10: 2007 Team Rankings Adjusted for Multiple Losses to Same Team

The revised rankings actually showed a very modest reduction in the number of violations (132 to 131) with no change in the number of bowl games predicted correctly. Missouri dropped from first to fourth, while Oklahoma improved from 14th to 8th. One might argue that Oklahoma should be ranked ahead of Missouri by virtue of their two wins, but it is worth considering that Oklahoma lost two games in 2007 to Colorado and Texas Tech, and that Missouri defeated both of these teams.

The final variation that we considered was to construct the transition matrix using variable probabilities based on the final score of each game. Our method for computing transition probabilities is similar to an approach described by Keener. Assuming that Team i loses to Team j by a score of s_j to s_i, the general formula is given by

$$p_{ij} = \frac{s_j + A}{s_i + s_j + 2A}$$

For the transition in the opposite direction, we used $p_{ji} = 1 - p_{ij}$ as in our earlier experiments. The idea is to have the transition probabilities close to 0.5 for closely contested games. At the opposite extreme, in the case of a blowout victory, we would want p_{ij} close to 1 and p_{ji} close to zero. The purpose of the parameter A is to prevent transition probabilities of 0 and 1 in the event of a shutout. We experimented with several values of A, but in these experiments we held $\alpha = 1$. In other words, we did not include the possibility of direct transitions between teams that had not played each other. Our results are shown in Figure 5.

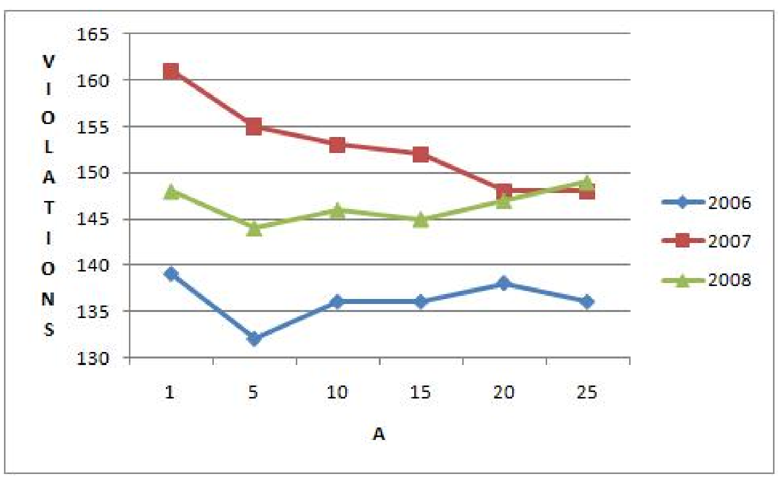

Figure 5: Test results using variable probabilities

No clear pattern emerges from analyzing the results from these three seasons. In 2006 and 2008, $A = 5$ produced the best results, but in 2007, the fewest violations using variable probabilities were produced with $A = 20$ and $A = 25$. Table 11 compares the best results obtained for each year using fixed or variable probabilities.

Year	Method	Violations	Bowl Upsets
2006	fixed, $p = 0.55, \alpha = 0.95$	107	11
	variable, $A = 5$	132	11
2007	fixed, $p = 0.55, \alpha = 1.0$	132	10
	variable, $A = 20$	148	7
2008	fixed, $p = 0.60, \alpha = 1.0$	120	17
	variable, $A = 5$	144	13

Table 11: Performance of Variable Probability Methods

The use of variable probabilities always resulted in a higher number of violations. Somewhat surprisingly, the variable probability approach did result in fewer bowl upsets in both 2007 and 2008.

5 Conference rankings

In this section, we discuss the conference rankings produced by our ranking method. We focus only on the method that gave the best results in the previous section as determined by the number of violations. Of the 120 teams currently in the Division 1 Football Bowl Subdivision, 116 of them belong to one of 11 conferences. The remaining 4 teams (Army, Navy, Notre Dame and Western Kentucky) are not affiliated with a conference. For purposes of this analysis, we considered these four independents as their own conference, bringing the total number of FBS conferences to 12.

To develop a ranking of the conferences in each of the three years that we studied, we started with the transition matrix created with the combination of p and α that had produced the fewest violations in the team rankings. Using the method of stochastic complementation, we aggregated the original 120 teams (119 teams in 2006) into their respective conferences, producing a 12×12 coupling matrix. The coupling matrix is also stochastic, and our idea was to use the steady state vector of the coupling matrix as the conference rating vector.

It is important to understand the relationship between the steady state vector of the original transition matrix (the team rating vector) and the steady state vector of the coupling matrix. Suppose that π is the team rating vector, and that it is partitioned as follows:

$$\pi = (\pi_1 \quad \pi_2 \quad \dots \quad \pi_{11} \quad \pi_{12})$$

In this vector, π_1 is the subvector that contains the individual ratings of the 12 teams in the Atlantic Coast Conference (ACC), π_2 contains the ratings of the 11 teams in the Big Ten Conference, and so on. The subvector π_{11} corresponds to the 9 teams in the Western Athletic Conference (WAC), while π_{12} corresponds to the independent teams (3 in 2006, 4 in 2007 and 2008). Let ξ be the steady state vector of the coupling matrix. There is a very simple relationship between ξ and π, as described in Meyer. The entry ξ_j, which is a single number corresponding to the j-th aggregated state, is simply the sum of the elements in π_j! Theoretically, it would be possible to perform the necessary computations without explicitly forming the stochastic complements as we demonstrated in Example 2. The steady state vector for the 120-state Markov chain could be computed and then partitioned to find the steady state vector for the aggregated 12-state chain.

In practice, we found that the steady state vector for the coupling matrix, on its own, did not provide satisfactory conference ratings. Since the rating of each conference is the sum of the individual ratings of its teams, this method would favor conferences with a large number of teams, such as the Mid-American Conference (MAC) which has 13 teams, while penalizing smaller conferences. The Big East and Sun Belt Conferences have only eight members each. The solution to this problem was simple. We simply divided each entry in the steady state vector by the number of teams in the corresponding conference, so that the result could be thought of as the rating of an average team from each conference. This produced much better results. For instance, in 2008, we found that before normalization, the MAC was ranked fifth, despite the fact that MAC teams had only won 14 of 44 games against teams from other FBS conferences. After normalization, the MAC was ranked tenth, although it was still ahead of Conference USA and the Sun Belt Conference, both of which had even lower non-conference winning percentages.

In Tables 12, 13 and 14, we report our conference rankings for 2006, 2007 and 2008. To provide some context, we included the non-conference win-loss record for each of the conferences, although we only included games in which both teams were from the FBS. The non-conference winning percentage only provides a very rough guide to the strength of a conference. If the strongest team from one conference defeats the weakest team from another, that provides little insight into the relative strength of the two conferences. We also included the bowl record for each conference, again recognizing that this would only reflect the performance of the better teams in the conference. The number of teams in a conference earning bowl bids would provide some indication of the strength of a conference. Finally, we wanted to compare our conference rankings to those determined by Jeff Sagarin as reported in USA Today [18, 19, 20]. Sagarin uses two different methods for ranking conferences. His Simple Average method (SA) computes the average rating for each team in the conference. The Central Mean (CM) approach computes a weighted average, with more weight given to teams in the middle of the conference. For the most part, both methods give very similar results so we report only the SA rankings here. Note: Sagarin's rankings include teams from the Football Championship Subdivision (formerly known as Division I-AA) which we excluded here. Finally, the Sagarin rankings reported here were determined using the results of the bowl games. We were unable to find archived versions of his weekly rankings from earlier in the seasons that we studied. We were curious to compare our method more directly, so we recomputed our conference rankings by including both regular season and bowl games. The last column in the table shows our updated rankings.

Conference	Our Rank before bowls	Non-conf record	Bowl record	Sagarin Rank	Our Rank after bowls
PAC 10	1	17-9	3-3	3	2
SEC	2	33-7	6-3	1	1
Big East	3	26-8	5-0	2	3
Big 10	4	27-10	2-5	4	4
Independents	5	17-11	0-2	8	6
Big 12	6	23-14	3-5	5	7
ACC	7	21-18	4-4	6	5
MWC	8	12-18	3-1	7	8
WAC	9	11-20	3-1	9	9
CUSA	10	13-28	1-4	10	10
MAC	11	6-38	1-3	11	11
Sun Belt	12	5-30	1-1	12	12

Table 12: 2006 Conference Rankings

Conference	Our Rank before bowls	Non-conf record	Bowl record	Sagarin Rank	Our Rank after bowls
SEC	1	31-8	7-2	1	1
Big East	2	23-11	3-2	4	4
ACC	3	26-15	2-6	5	6
PAC 10	4	19-10	4-2	2	2
Big 10	5	29-7	3-5	6	5
Big 12	6	29-11	5-3	3	3
MWC	7	15-16	4-1	7	7
Independents	8	12-24	0-1	8	9
WAC	9	7-20	1-3	9	10
Sun Belt	10	8-30	1-0	11	8
CUSA	11	10-30	2-4	10	11
MAC	12	12-39	0-3	12	12

Table 13: 2007 Conference Rankings

Not surprisingly, our post-bowl rankings correlate more closely with the Sagarin rankings than our pre-bowl rankings. In reviewing our conference rankings, it is clear that, to achieve a high ranking, the combined strength of all of the teams in a conference is more important than having a few very highly ranked teams. For instance, in 2008, the top two teams came from the Big 12 (Oklahoma) and the SEC (Florida), but our method ranked these conferences second and third. The highest ranked conference (according to us) turned out to be the ACC which had 10 of its 12 teams earn bowl bids. According to our method, the ACC had nine teams in the top 30, and the lowest-ranked ACC team (Duke) was 61st. By contrast, the SEC had only three teams in our top 30, and the lowest-ranked SEC team (Mississippi State) was ranked 86th.

6 Conclusions

We believe that the method of stochastic complementation produces a viable ranking method for college football teams. The number of violations produced by our method compares favorably to the number of violations produced by the actual BCS standings. Our method also provides a natural way to rank conferences and gives a theoretical justification for using average team ratings to determine conference ratings. There are certainly

Conference	Our Rank before bowls	Non-conf record	Bowl record	Sagarin Rank	Our Rank after bowls
ACC	1	23-11	4-6	3	1
Big 12	2	28-10	4-3	2	3
SEC	3	28-11	6-2	1	2
Big East	4	22-12	4-2	5	5
Big 10	5	23-12	1-6	6	7
MWC	6	19-10	3-2	7	6
PAC 10	7	12-17	5-0	4	4
WAC	8	11-19	1-4	9	8
Independents	9	14-26	1-1	10	10
MAC	10	14-30	0-5	11	11
CUSA	11	11-30	4-2	7	9
Sun Belt	12	10-27	1-1	12	12

Table 14: 2008 Conference Rankings

additional questions that could be investigated. Our experiments involved three different parameters: the value of the fixed transition probability p, the value of the weighting factor α for the random transition matrix G, and the parameter A used in our computation of variable probabilities. Other combinations of these parameters could easily produce variations on the basic method presented here that might lead to even better performance. In particular, it would be of interest to develop other methods of computing variable transition probabilities that account for factors such as margin of victory and home field advantage.

7 Acknowledgements

The authors would like to thank Anjela Govan and Carl Meyer for their comments and encouragement. Kenneth Massey provided helpful information about his own ranking method and about the BCS rankings. We are also appreciative of the data sets provided by James Howell. Finally, we would like to thank the SUNY Cortland Undergraduate Research Council for providing a 2008 Summer Research Fellowship that supported the work of the second author.

References

[1] Iqbal Ali, Wade D. Cook, and Moshe Kress. On the minimum violations ranking of a tournament. *Management Science*, 32(6):660–672, 1996.

[2] BCS standings. http://www.bcsfootball.org/bcsfb/standings.

[3] Abraham Berman and Robert J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*, volume 9 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1994.

[4] W. N. Colley. Colley's bias free college football ranking method: The Colley matrix explained. http://www.colleyrankings.com/matrate.pdf, 2002.

[5] A. Govan, C. D. Meyer, and R. Albright. Generalizing Google's PageRank to rank National Football League teams. In *Proceedings of the SAS Global Forum 2008*, 2008.

[6] James Howell. James Howell's College Football Scores. http://homepages.cae.wisc.edu/ dwilson/rsfc/history/howell/.

[7] L. R. Ford Jr. Solution of a ranking problem from binary comparisons. *American Mathematical Monthly*, 64(8):28–33, 1957.

[8] James P. Keener. The Perron-Frobenius theorem and the ranking of football teams. *SIAM Review*, 35(1):80–93, 1993.

[9] Amy N. Langville and Carl D. Meyer. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, Princeton, NJ, 2006.

[10] Kenneth Massey. Statistical models applied to the rating of sports teams. Master's thesis, Bluefield College, 1997. http://www.masseyratings.com/theory/massey97.pdf.

[11] Kenneth Massey. College Football Ranking Comparison. http://www.masseyratings.com/cf/arch/compare2006-14.htm, 2006.

[12] Kenneth Massey. College Football Ranking Comparison. http://www.masseyratings.com/cf/arch/compare2007-14.htm, 2007.

[13] Kenneth Massey. College Football Ranking Comparison. http://www.masseyratings.com/cf/arch/compare2008-15.htm, 2008.

[14] Kenneth Massey. College Football Ranking Summary. http://www.masseyratings.com/cf/compsum.htm, 2009.

[15] Carl D. Meyer. Stochastic complementation, uncoupling Markov chains and the theory of nearly reducible systems. *SIAM Review*, 31(2):240–272, 1989.

[16] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: bringing order to the web. Technical report, Department of Computer Science, Stanford University, 1998.

[17] C. Redmond. A natural generalization of the win-loss rating system. *Mathematics Magazine*, 76(2):119–126, 2003.

[18] Jeff Sagarin. Jeff Sagarin NCAA Football Ratings. http://www.usatoday.com/sports/sagarin/fbc06.htm, 2007.

[19] Jeff Sagarin. Jeff Sagarin NCAA Football Ratings. http://www.usatoday.com/sports/sagarin/fbc07.htm, 2008.

[20] Jeff Sagarin. Jeff Sagarin NCAA Football Ratings. http://www.usatoday.com/sports/sagarin/fbc08.htm, 2009.