Bracketology: How can math help?

Tim Chartier Erich Kreutzer Amy Langville Kathryn Pedings

Abstract

Every year, people across the United States predict how the field of 65 teams will play in the Division I NCAA Men's Basketball Tournament by filling out a tournament bracket for the postseason play. This article discusses two popular rating methods that are also used by the Bowl Championship Series, the organization that determines which college football teams are invited to which bowl games. The two methods are the Colley Method and the Massey Method, each of which computes a ranking by solving a system of linear equations. The article also discusses how both methods can be adapted to take late season momentum into account. All the methods were used to produce brackets in 2009 and their results are given, including a mathematically-produced bracket that was better than 97% of the nearly 4.5 million brackets submitted to ESPN's Tournament Challenge.

1 Introduction

Every year around the beginning of March, students and faculty at 65 schools along with a large portion of the United States get excited about the prospects of March Madness, the Division I NCAA Men's Basketball Tournament. At the start of the tournament each of the 65 teams has a chance of ending the season crowned the champion. With the excitement of watching basketball comes the adventure of attempting to predict the outcome of the tournament by filling out a bracket. The NCAA estimates that 10% of the nation will fill out a bracket [4]. Be it participating in a small contest with friends or submitting your bracket to the nationwide ESPN Tournament Challenge, everyone wants to do well. How can we use mathematics to help? That is the goal of this paper—to describe several mathematical methods that can improve your chances of winning the office sports pool and possibly even the $1 million prize for the ESPN Tournament Challenge.

Sports teams are often ranked according to winning percentage, which is easily calculated for the ith team as $p_i = w_i/t_i$, where w_i is the total number of wins for team i and t_i is the total number of games it played. To make our computations concrete, we will examine data from the 2009 season of the Southern Conference of Division I NCAA Men's Basketball. The standings at the end of regular season are in Table 1. Ranking the teams by winning percentage produces the results seen in the table.

While Davidson College is ranked first according to winning percentage, the teams two regular season conference losses were to the College of Charleston and the Citadel. Should Davidson still be ranked first or should the losses change the ranking? Davidsons winning percentage would be the same whether one of its losses came from second place College of Charleston or last place Furman University. Should the College of Charleston and the Citadel be tied for second place just because their winning percentages are the same? Could examining against whom teams won and lost serve to predict teams' future performance better?

These issues are reasons why many rating systems incorporate more than winning percentages. This article discusses two popular rating methods used by the Bowl Championship Series, the organization that determines which college football teams are invited to which

Team	Record	Rating	Rank
Davidson College	18–2	0.90	1
College of Charleston	15–5	0.75	2
The Citadel	15–5	0.75	2
Wofford College	12–8	0.60	4
Univ. of Tennessee at Chattanooga	11–9	0.55	5
Western Carolina Univ.	11–9	0.55	5
Samford Univ.	9–11	0.45	7
Appalachian State Univ.	9–11	0.45	7
Elon Univ.	7–13	0.35	9
Georgia Southern Univ.	5–15	0.25	10
Univ. of North Carolina at Greensboro	4–16	0.20	11
Furman Univ.	4–16	0.20	11

Table 1: The regular 2009 season standings and winning percentage ratings for the teams in the Southern Conference of Division I NCAA Men's Basketball.

bowl games [5]. The two methods are the Colley method and the Massey method. The Colley method was created by astrophysicist Wesley Colley who saw problems with ranking by winning percentage and developed a new method [1]. The Massey method started as Ken Massey's undergraduate honors math project and eventually made its way into the BCS [3]. The main idea behind this method is to use a least-squares approximation to find a rating vector.

Each method calculates ratings for all teams that can then be used to complete a bracket for the March Madness tournament by choosing the higher rated team as the winner for each matchup in the bracket. The goal of this paper is to examine if a mathematical strategy to completing a bracket can outperform the typical sports fan's bracket.

2 Colley Method

Ranking teams by winning percentage is a common ranking method even in professional sports. Wesley Colley proposed applying Laplace's rule of succession, which transforms the standard winning percentage into

$$r_i = \frac{1 + w_i}{2 + t_i}. \tag{1}$$

This minor change may appear to be of little help, but Colley used it as a stepping stone to a more powerful result. Before making such a step, let's apply this formula to the season's data in Table 1. Using (1) gives the ratings in Table 2. The rankings of the teams do not change but now the average of the ratings is 0.5, which we will now use in the derivation of the Colley method.

Team	Record	Rating	Rank
Davidson	18–2	0.864	1
Charleston	15–5	0.727	2
Citadel	15–5	0.727	2
Wofford	12–8	0.591	4
Chattanooga	11–9	0.545	5
W. Carolina	11–9	0.545	5
Samford	9–11	0.455	7
App. State	9–11	0.455	7
Elon U	7–13	0.364	9
Georgia Southern	5–15	0.273	10
UNC Greensboro	4–16	0.227	11
Furman	4–16	0.227	11

Table 2: The regular 2009 season standings and ratings as calculated with (1) of the teams in the Southern Conference of Division I NCAA Men's Basketball.

An alternative way to write the number of wins for a team is

$$w_i = \frac{w_i - l_i}{2} + \frac{w_i + l_i}{2} = \frac{w_i - l_i}{2} + \frac{t_i}{2},$$

and

$$t_i/2 = \frac{1}{2}\overbrace{(1 + 1 + \cdots + 1)}^{\text{total number of games}} = \left(\frac{1}{2} + \frac{1}{2} + \cdots + \frac{1}{2}\right).$$

At the beginning of the season, all ratings are 1/2, and they hover around 1/2 as the season proceeds. So,

$$\frac{1}{2}\,(\text{total games}) \approx (\text{sum of opponents' ranks for all games played}),$$

or

$$\frac{1}{2}t_i \approx \sum_{j \in O_i} r_j,$$

where O_i is the set of opponents for team i. Substituting this back into our equation for w_i we get

$$w_i \approx \frac{w_i - l_i}{2} + \sum_{j \in O_i} r_j. \tag{2}$$

This substitution is approximate as the average over all opponents' ratings may not be 1/2 since, for one thing, every team may not play every other team.

Assuming equality in (2) and inserting this into (1) produces

$$r_i = \frac{1 + (w_i - l_i)/2 + \sum_{j \in O_i} r_j}{2 + t_i}. \tag{3}$$

The advantage of this representation is that we now have the interdependence of our teams' ratings. That is, team i's rating depends on the ratings r_j of all its opponents. This procedure for computing ratings is called the Colley method.

While each r_i can be computed individually using (3), an equivalent formulation uses a linear system

$$C\mathbf{r} = \mathbf{b}, \tag{4}$$

where C is the so-called Colley matrix. How is this linear system derived from (3)? It is easiest to see the contributions of (3) if it is written as

$$(2 + t_i)r_i - \sum_{j \in O_i} r_j = 1 + \frac{1}{2}(w_i - l_i). \tag{5}$$

The vector $\mathbf{b}$ has components $b_i = 1 + \frac{1}{2}(w_i - l_i)$. The diagonal elements of the Colley coefficient matrix C are $2 + t_i$ and the off-diagonal elements c_{ij}, for $i \neq j$, are $-n_{ij}$, where n_{ij} is the number of games between teams i and j.

Let's simplify by taking a subset of the Southern Conference and rank only Davidson, Charleston, the Citadel, Wofford, and Chattanooga according to their 2009 regular season records against each other. Their performance results in the linear system

$$\begin{pmatrix} 10 & -2 & -2 & -2 & -2 \\ -2 & 9 & -2 & -2 & -1 \\ -2 & -2 & 9 & -2 & -1 \\ -2 & -2 & -2 & 10 & -2 \\ -2 & -1 & -1 & -2 & 8 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{pmatrix} = \begin{pmatrix} 3 \\ 1.5 \\ 0.5 \\ 0 \\ 0 \end{pmatrix},$$

where r_1, r_2, r_3, r_4 and r_5 are the ratings for Davidson, Charleston, the Citadel, Wofford and Chattanooga, respectively. Solving this linear system yields Colley ratings of $r_1 = 0.667, r_2 = 0.554, r_3 = 0.464, r_4 = 0.417$ and $r_5 = 0.398$, which maintains the property that the average of the ratings is 1/2. The Colley matrix provides matchup information. For example, the (2,5)-entry of C means that team 2, Charleston, played team 5, Chattanooga, only once while Wofford played each of its opponents twice that season.

To emphasize interdependence, let's add a fictional game to this linear system. Suppose Chattanooga and Charleston play one more time with Chattanooga winning. Then the linear system becomes

$$\begin{pmatrix} 10 & -2 & -2 & -2 & -2 \\ -2 & 10 & -2 & -2 & -2 \\ -2 & -2 & 9 & -2 & -1 \\ -2 & -2 & -2 & 10 & -2 \\ -2 & -2 & -1 & -2 & 9 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 0.5 \\ 0 \\ 0.5 \end{pmatrix}.$$

The new ratings are $r_1 = 0.667, r_2 = 0.500, r_3 = 0.458, r_4 = 0.417, r_5 = 0.458$. The additional game affects the ratings of more than just the two teams playing, yet the property that the average ratings are 1/2 is maintained.

Let's return to the regular season results for the entire Southern Conference. Solving the 12×12 linear system produces the ratings in Table 3. The ranking from winning percentages are listed for comparison.

The Colley method is unaffected by differences in final scores. Nowhere is the game score considered. In the Colley method, a win is a win regardless of the score. The Massey method is a ranking method that includes game scores in the ratings.

Team	Record	Colley Rating	Colley Rank	Winning % Rank
Davidson	18–2	0.82	1	1
Charleston	15–5	0.74	2	2
Citadel	15–5	0.68	3	2
Wofford	12–8	0.57	4	4
Chattanooga	11–9	0.56	5	5
W. Carolina	11–9	0.53	6	5
Samford	9–11	0.47	7	7
App. State	9–11	0.46	8	7
Elon U	7–13	0.40	9	9
Georgia Southern	5–15	0.28	10	10
UNC Greensboro	4–16	0.26	11	11
Furman	4–16	0.25	12	11

Table 3: The regular 2009 season standings and the Colley ratings of teams in the Southern Conference of Division I NCAA Men's Basketball.

3 Massey Method

If Davidson beat Elon by 10 points and Elon beat Furman by 5 points, could we then predict that Davidson would beat Furman by 15 points? If this were true, sports wouldn't be as much fun to watch but it would make sports ranking easy. Transitivity will rarely, if ever, hold perfectly, but assuming that it holds approximately is the foundation of the Massey method. Let r_1, r_2 and r_3 be the ratings for Davidson, Elon, and Furman and let's compute these ratings so that they can predict outcomes of future games. For our small example, $r_1 - r_2 = 10$. (Davidson beat Elon by 10 points.) The difference in the ratings r_1 and r_2 is the margin of victory achieved by the higher rated team. We also have $r_2 - r_3 = 5$. (Elon beat Furman by 5 points.) If we add these two linear equations, we create the prediction $r_1 - r_3 = 15$. (Davidson beats Furman by 15 points.) How do we compute the ratings r_1, r_2 and r_3?

For this small example, we have the linear equations $r_1 - r_2 = 10$ and $r_2 - r_3 = 5$; there are infinitely many solutions: for any nonnegative integer c, $r_1 = 15 + c, r_2 = 5 + c$ and $r_3 = c$ is a solution. In practice, the situation is reversed. There are typically more games than teams, so we have a system with more equations than variables that in general will have no solution. For instance, suppose we add an additional game in which Furman beats Davidson by 1 point. Thus, $r_3 - r_1 = 1$. Along with our earlier linear equations this forms the linear system, $M_1\mathbf{r} = \mathbf{p}_1$ or

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} = \begin{pmatrix} 10 \\ 5 \\ 1 \end{pmatrix}.$$

You might think that a system of three equations and three unknowns would have a unique solution. However, regardless of the number of games played m between n teams, the matrix M_1 of dimensions $m \times n$ never has full rank. Because of the symmetry of matchups, any set of $n-1$ columns can be used to build the remaining column, which means that the rank of M_1 is at most $n-1$. Since we cannot find an exact solution to the system, we will find an approximate solution. We will use the method of least squares to find the vector $\mathbf{r}$ such that the length of the vector $\mathbf{p}_1 - M_1\mathbf{r}$, which is the residual error, is minimized.

To do this, we first compute $M_1^T M_1 \mathbf{r} = M_1^T \mathbf{p}_1$, which we denote as $M_2\mathbf{r} = \mathbf{p}_2$,

$$\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} = \begin{pmatrix} 9 \\ -5 \\ -4 \end{pmatrix},$$

where $M_2 = M_1^T M_1$ and $\mathbf{p}_2 = M_1^T \mathbf{p}_1$ is the cumulative point differential vector. Once again, $M_2\mathbf{r} = \mathbf{p}_2$ is a singular system since it can be proven that the rank of M_2 is $n-1$. To create a nonsingular system $M\mathbf{r} = \mathbf{p}$, we need only take one additional step and replace a row in the matrix M_2 by a row of ones to form M, the Massey matrix, and the corresponding entry in the vector $\mathbf{p}_2$ with a zero to form $\mathbf{p}$. This step implies that the sum of the ratings should be 0. Solving our new linear system,

$$\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} = \begin{pmatrix} 9 \\ -5 \\ 0 \end{pmatrix},$$

produces the desired ratings. This method of rating is called the Massey method. While any row of M_2 can be replaced by a row of ones, replacing the last row is the convention established by Ken Massey in his undergraduate honors thesis for Bluefield College in Virginia. Solving this linear system we find that $r_1 = 3, r_2 = -1.667$, and $r_3 = -1.333$. Therefore, our method predicts Davidson (team 1) will beat Elon (team 2) by $3 - (-1.667) = 4.667$ points.

As with (4) for the Colley method, we can create the Massey linear system directly from a season's data. In all but the last row, the diagonal elements of the Massey matrix are the number of games played by team i and the off-diagonal elements m_{ij} are the negative of the number of games played between teams i and j. Notice that $C = M_2 + 2I$, which is convenient when building the two systems' ratings. The vector $\mathbf{p}_2$ is the sum of the point differentials in team i's games where a win gives a positive differential and a loss, a negative one.

Having seen the Massey method on a small example, let's return to the regular season results for the entire Southern Conference. We again have a 12×12 linear system to solve and this system, created according to the Massey method, results in the ratings found in Table 4. While both methods produce ratings that are interdependent, there are differences in the rankings. Chattanooga drops from fifth to eighth when ratings are produced by the Massey method. Given the interdependence of the ratings, such a drop can be difficult to fully explain. Part of the explanation is Chattanooga's big losses to highly ranked teams, including a 22 point loss to Davidson.

4 Weighting Methods

Neither the Colley nor Massey method takes into account when in the season a game occurs. This can be important. A star player may be injured half way through the season or a team may mature and improve over time. It seems especially valuable for March Madness predictions to boost the rating of a team that has won its last ten games even if

Team	Record	Massey Rating	Massey Rank	Colley Rank
Davidson	18–2	13.99	1	1
Charleston	15–5	5.42	2	2
Citadel	15–5	5.32	3	3
Wofford	12–8	0.73	5	4
Chattanooga	11–9	-1.30	8	5
W. Carolina	11–9	2.67	4	6
Samford	9–11	0.00	6	7
App. State	9–11	-0.17	7	8
Elon U	7–13	-4.22	9	9
Georgia Southern	5–15	-7.73	11	10
UNC Greensboro	4–16	-6.57	10	11
Furman	4–16	-8.14	12	12

Table 4: The regular 2009 season standings and the Massey ratings of teams in the Southern Conference of Division I NCAA Men's Basketball.

it performed poorly at the beginning of the season. There are ways to adapt both methods to take momentum into account.

In the standard version of the Massey and Colley methods, a game at the beginning of the season contributes to a team's rating with the same weight as a game at the end of the season. What if we want to reward teams for playing well in the weeks leading up to the tournament? We can do this by weighting games according to the date they were played, so the outcome of recent games affects the ratings more than earlier ones.

4.1 Linear weighting and the Colley method

To weight the games, we want a function $g(t)$ that takes as input t, the number of days into the season at which the game occurs, and returns a real positive weight as output. A simple weighting function is a linear weighting

$$g(t) = (t - t_0)/(t_f - t_0) = t/t_f, \tag{6}$$

where $t_0 = 0$ represents opening day for the season and t_f is the total number of days in the season. The weights in (6) for games played by Davidson College against opponents in the Southern Conference are plotted in Figure 1 (b). See Figure 1 (a) for uniform weighting under the standard version of the Colley method. In the linearly weighted Colley method, the weights lie between 0 and 1 with the highest weight occurring on the last day of the season. Games that occur on the same day are given the same weight, and games on opening day are given no weight ($g(t) = 0$). Uncomfortable with that choice? How about t_0 is 1 on opening day, which alters the weights on all games and could result in a different ranking. Have another idea? This could lead to a personalized bracket!

An important step in deriving (5) in the Colley method was the hovering of the ratings about 1/2. An interested reader may wish to step through the derivation of the Colley method to include a weighting function. We supply only an intuitive sense of why one would expect such derivation to hold. In the standard version of the Colley method, the outcome of a game increments one team's number of wins by 1 and the number of losses for the other team by 1. In the weighted version of the method, the outcome of a game increments the number of wins and losses of the associated teams by $g(t)$, so the important hovering about 1/2 still holds and the derivation that produced (5) is mirrored in the steps that produce an algorithm with a weighting function.

For a weighted method, the total games for team i, t_i, is altered. Instead of adding 1 to t_i for each game played we add $g(t)$. The alterations to the linear system $C\mathbf{r} = \mathbf{b}$ of the Colley method are minor. As before, the diagonal element of the Colley matrix c_{ii} is $2 + t_i$ but now t_i is this accumulation of games that includes weighting. The off-diagonal elements c_{ij} for $i \neq j$ equals $-n_{ij}$ where n_{ij} is the number of weighted games played between teams i and j. As with total games, n_{ij} is incremented by $g(t)$ for a game played between i and j at time t. Finally, b_i, which is an element of the vector $\mathbf{b}$ corresponding to team i, is $1 + \frac{1}{2}(w_i - l_i)$, where a game played on day t contributes $g(t)$ of a win to the winner and $g(t)$ of a loss to the loser.

4.2 Linear weighting in the Massey method

What alterations can incorporate weights into the Massey method? We begin with the initial Massey system $M_1\mathbf{r} = \mathbf{p}_1$ rather than the simplified system $M\mathbf{r} = \mathbf{p}$. Instead of applying least squares to $M_1\mathbf{r} = \mathbf{p}_1$, we use a weighted least squares method. Weighted least squares allows us to incorporate a measure of how much each game should count. If we have a low weighting for a game, say $r_i - r_j = 10$, then the algorithm will pay less attention to making that equation true when compared to a game with a higher weight. This works exactly as we would like in that if we weight more recent games heavily, the ratings that result from the least squares will give a rating that more closely reflects the outcome of the more recent games.

A weighted least squares method can be run by adding a weighting matrix W to the computation. Before we had M_1 which had each game on a different row. In the unweighted method, we calculated $M_1^T M_1 \mathbf{r} = M_1^T \mathbf{p}$ which led us to the final $M\mathbf{r} = \mathbf{p}$ system. We now add the weighting matrix, a square matrix whose diagonal entry w_{jj} corresponds to the weight of the jth game and has a value corresponding to the weighting function of the game. We then solve $M_1^T W M_1 \mathbf{r} = M_1^T W \mathbf{p}_1$.

4.3 Alternative weightings – when life isn't linear

If we are to weight games, how best to do it? While the linear model in (6) weights games at the end of the season higher than earlier games, we found this method was not as good as other choices. An alternative is logarithmic weighting

$$g(t) = \ln\left(1 + \frac{t}{t_f}\right), \tag{7}$$

which, without additional modifications, does not range from 0 to 1. As defined in (7), $g(t)$ produces weights as seen in Figure 1 (c) for games played by Davidson College. The function could also be defined so that it plateaus toward the end of the season and weights the more recent games nearly evenly.

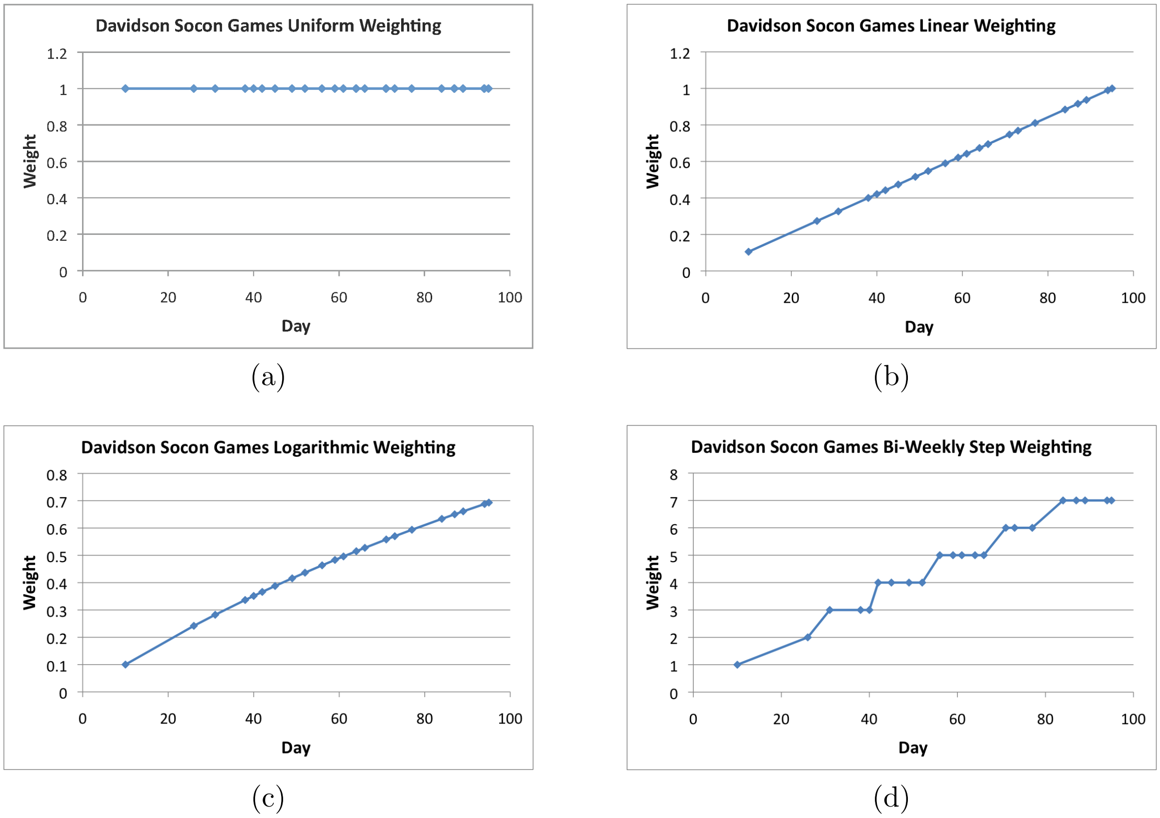

Figure 1: Weightings ((a) uniform, (b) linear, (c) logarithmic, and (d) bi-weekly step functions) of games that Davidson played against Southern Conference opponents during the 2009 regular season. The scale of the y-axis differs on each graph.

Another option is to give the same weight to temporally similar games with a step function as is done with the following bi-weekly step function

$$g(t) = \lfloor (t - t_0)/14 + 1 \rfloor = \left\lfloor \frac{t}{14} + 1 \right\rfloor,$$

where $\lfloor x \rfloor$ represents the greatest integer less than or equal to x. This function divides the season into groups of 14 days and weights each group more than the previous two week period as seen in Figure 1 (d). This step function weights all games with integral values greater than or equal to 1. If a team wins a game on day 72 of the season, $g(72) = \lfloor (\frac{72}{14}) + 1 \rfloor = \lfloor 5.14{+}1 \rfloor = 6$. This win is equivalent to 6 wins in the unweighted (or uniformly weighted) formulation of the Colley method. We will see that the bi-weekly step function produced impressive results in predicting the results during March Madness in 2009.

Although our weighting methods depend on time, other non-temporal weightings are possible. Home court advantage can be accounted for if every away win is weighted higher than a neutral win, which is weighted higher than a home win. Another possibility is placing a small weight on games in which key players were injured. Weighting is an area of active research.

5 2009 Results

In this paper, we described two ranking methods, the Colley and Massey methods, and four weighting schemes (no or uniform weighting, linear, logarithmic, and step weighting),

Mathematically produced brackets			Non-mathematically produced brackets		
Algorithm	**Score**	**Percent**	**Name**	**Score**	**Percent**
Colley No Weighting	940	62^{nd}	President Barak Obama	1230	80^{th}
Colley Linear	940	65^{th}	Mike Greenberg (sports analyst)	1060	70^{th}
Colley Logarithmic	980	65^{th}	Mike Golic (sports analyst)	800	43^{rd}
Colley Bi-Weekly Step	1420	97^{th}	Dwyane Wade (NBA star)	800	43^{rd}
Massey No Weighting	1300	88^{th}			
Massey Linear	1220	79^{th}			
Massey Logarithmic	1240	81^{st}			
Massey Bi-Weekly Step	1220	79^{th}			

Table 5: 2009 ESPN Tournament Challenge results (score and percentile) for the eight ranking methods and several non-mathematical submissions.

creating a total of $2 \cdot 4 = 8$ possible ratings that can be produced. Many more combinations exist, involving many other rating methods, weighting functions, and data inputs. The possibilities are limitless, which is part of the reason why in the past few years we have submitted several brackets to ESPN's Tournament Challenge. We enjoy seeing how brackets built from mathematical models compare both against each other and against non-mathematical submissions from other fans. The process for using a mathematical method to fill out a bracket is simple. First, we use all the data on games prior to the March Madness tournament to rate the teams. Then for each March Madness matchup in Round 1, we predict that the higher rated team will win. We do the same with the Round 2 matchups that we created, and so on for each round until we have completed a valid bracket. We follow the same process for each mathematical rating method that we submit to the ESPN Challenge.

For the 2009 March Madness tournament, we submitted several brackets. How did they fare? Out of the nearly 4.5 million brackets submitted, the Colley method with bi-weekly step weighting reached the 97th percentile, i.e., that bracket was better than 97% of the other brackets submitted. The previous year the bracket for the Massey method with logarithmic weighting was in the 99th percentile, putting it in the top 1000 of all 3.6 million brackets submitted. Table 5 provides a side-by-side comparison of our mathematical submissions alongside the non-mathematical submissions of some other famous fans.

While most of our mathematical submissions outperformed many of the non-mathematical submissions, one year the Colley method gave the best ESPN score yet the previous year the Massey method was the best. This is an issue with sports ranking. Although methods generally perform well from year to year, because of the inherent randomness of sports, it is difficult for a method to consistently take the title of the best mathematical method.

6 Concluding Remarks

Part of the popularity of March Madness relates to the difficulty of predicting the outcome of sports games. The linear models of this paper have performed quite well in producing brackets for March's basketball tournament. While the algorithms account for the strength of a team's opponents, there are countless possible weighting methods. What weighting would you use? Would you want to weight home versus away games, games in which key players were injured, or other statistics available about team performance? With a little thought and some luck you too could design a variation on these methods to fill out your own bracket and maybe even win next year's bracket challenge!

References

[1] W. N. Colley, Colley's Bias Free College Football Ranking Method: The Colley matrix explained, (2002) available at: http://www.colleyrankings.com/method.html.

[2] A. N. Langville and C. D. Meyer, *Recommendation Systems: Ranking, Clustering, Searching, and Visualizing Data*, Princeton University Press, Princeton, forthcoming.

[3] K. Massey, Statistical models applied to the rating of sports teams, undergraduate honors thesis, Bluefield College, 1997.

[4] T. Otterman, NCAA March Madness can cause lifetime gambling problems, *U.S. News & World Report*, March 17, 2009.

[5] Bowl Championship Series, available at: http://www.bcsfootball.org.

Authors

TIMOTHY P. CHARTIER is an Associate Professor of Mathematics at Davidson College. He is a recipient of the Henry L. Alder Award for Distinguished Teaching by a Beginning College or University Mathematics Faculty Member from the Mathematical Association of America. As a researcher, Tim has worked with both the Lawrence Livermore and Los Alamos National Laboratories on the development and analysis of computational methods to increase the efficiency and robustness of numerical simulation on the lab's supercomputers, which are among the fastest in the world. Tim's research with and beyond the labs was recognized with an Alfred P. Sloan Research Fellowship. In his time apart from academia, Tim enjoys the performing arts, mountain biking, nature walks and hikes, and spending time with his wife and two children.

Department of Mathematics, Davidson College, Davidson, NC, 28036, tichartier@davidson.edu

ERICH KREUTZER is a student at Davidson College majoring in mathematics with a concentration in computer science. He is currently the president of the Bernard Society of Mathematics at Davidson. In the summer of 2007, Erich participated in the Google Summer of Code working on the Adium instant messenger application. His areas of interest include Mac OS X application development, sports ranking, data mining, and mathematical modeling. Outside of school Erich enjoys cycling, frisbee golf, and spending time with friends.

Department of Mathematics, Davidson College, Davidson, NC, 28036, erkreutzer@davidson.edu

AMY N. LANGVILLE is an Associate Professor of Mathematics at the College of Charleston. Her award-winning book, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, coauthored with Carl Meyer, explains and analyzes several popular methods for ranking webpages. Dr. Langville's research deals with ranking and clustering items using matrix decompositions. Her primary areas are numerical linear algebra, computational algorithms, and mathematical modeling and programming. Amy has consulted on six industrial projects with companies such as Fortune Interactive, Piffany, and the SAS Institute. She also regularly consults with law firms and is called to serve as a technical expert on patent infringement cases concerning webpage ranking systems. Off campus, Amy enjoys biking, basketball, and the outdoors.

Department of Mathematics, College of Charleston, Charleston, SC 29401, langvillea@cofc.edu

KATHRYN PEDINGS graduated from the College of Charleston in 2008 with a B.S. in Mathematics and a minor in Secondary Education. She is currently working towards a Master's in Mathematics at the College of Charleston and conducting research on sports ranking with Dr. Amy Langville while teaching full time at a local public high school.

Department of Mathematics, College of Charleston, Charleston, SC 29401, kepeding@edisto.cofc.edu