The 23rd ECMI Modelling Week
European Student Workshop
on Mathematical Modelling
in Industry and Commerce

Wrocław University of Technology, Wrocław, Poland, August 23-30, 2009

Report of project group 9 on

# Ranking American Football Teams

The 23$^{rd}$ ECMI Modelling Week

Wrocław, Poland

# Ranking American Football Teams

Raphael Ruben Kohler, Kaiserslautern University of Technology
Pablo Barbero, Autonomous University of Barcelona
Qixiao Yu, Eindhoven University of Technology
Milana Gatarić, University of Novi Sad
Carsten Vogel, Dresden University of Technology
Kamil Bogus, Wrocław University of Technology
Dr. Mason A. Porter, University of Oxford *Supervisor*

October 2009

**Abstract**

The current official ranking system BCS (Bowl Championship Series) for American college football games, based on expert opinion and computer algorithms, arouses a lot of controversy. People doubt whether BCS is fair to universities in "non-BCS" conferences. In this report, we will design a simple, transparent ranking system that attempts to be fair to all teams depending on game data of the recent years. First, a naive method, tree concept, is introduced, in which a main criterion dominates the result. Then a more robust and flexible method, balance concept, is used. It includes more data with different weights. Results are tested and compared using both the methods for seasons 2005–2008.

# Contents

# 1 Background and problem

American college football[1] is extremely popular in U. S. At stake is the prestige of the universities and sponsor interests, and for many players it is the beginning of a great sporting career. Adding to the enormous popularity of games (often in the stadium there are about 10 thousand fans) and from the pageantry of game played we can conclude that college football is a big business in the U. S.

The college football games are scheduled on weekends. Every year, around 120 university teams are divided into different conferences, playing games to achieve the title of the best U.S. college football team. Due to time and cost, American college football teams will not play with every other team in one season. Each team plays 11-14 games, with the vast majority of them playing 12-13 games. After each season, top teams will be selected to play the bowl games[2]. Since 1998, the bowl games have been determined by Bowl Championship Series (BCS)[3].

The game itself is based on relatively clear rules[1]. Unfortunately one cannot say the same for the complicated official ranking system BCS, which relies on a combination of polls and computer selection methods to determine team rankings. Before 2004, The BCS combined a number of factors, a final point total was created and the teams that received the 25 lowest scores were ranked in descending order. Factors includes poll average, computer average of the rankings of a team in three different computer polls, strength of schedule and losses. However, controversy was created by the voters in the AP poll naming USC as the No. 1 ranked team at the end of year 2003. Supporters of USC and the media in general criticized the fact that human polls were not weighted more heavily than computer rankings. A new formula came into being and has been used since 2004. New factors are: AP poll, Coaches' poll and Computer average of six algorithms. However, controversies are still there.

A large part of the computer selection methods is not made public, this is one cause for controversies and inconsistencies with the position of team in ranking. There exist many myriad ranking designed by fans, whose goal is to provide a more reliable method of ranking teams. One example is the ”random-walker rankings” studied by applied mathematicians Thomas Callaghan, Peter Mucha, and Mason Porter that employs the science of complex networks.[4].

Some of the methods are good, while some eliminate even cases like this:

| Rank | Team | # games - # wins |
|---|---|---|
| 1 | Florida | 13-1 |
| 2 | Texas | 12-1 |
| 3 | USC | 12-1 |
| 4 | Oklahoma | 12-2 |
| 5 | Utah | 13-0 |
| 6 | Alabama | 12-2 |
| 7 | Penn State | 11-2 |
| ⋮ | ⋮ | ⋮ |

Florida has lost one game and has the same number of wins as Utah but nevertheless it is ranked higher. Is such a situation fair? Does the ranking of a team represent its strength to a certain extent?

This project aims to answer such questions, which requires the creation of an alternative rating system which should will be fair from our perspective. For this purpose, we used data from the 2005–2008 seasons, including around 120 teams and about 700 games. The highlights of this work was the selection of appropriate criteria. Among them the ones we used are:

- number of wins
- number of losses

- the ratio of losses to the number of wins
- margin per every single game
- results in their own stadion (home) and at their opponent's place (away)

In this report, the two methods of ranking are obtained :

◇ The tree-concept is based on simple classification, i.e. we start at the bottom of a tree and then we slowly branch the teams with respect to different criteria. At the root, all the teams are equally ranked (this might be valid until the first branch comes). For example, we might start with number of wins as first criterion. All teams is split into clusters of teams with the same number of wins (strong branches). This is continued until each cluster (i.e. each subbranch) consists of one and only one team (i.e. each position in the ranking is clear, no team has the same position).
◇ The balance-concept based on a weight function. This model is more complex. Imagine, each team has its own board on a multi-board balance. On this board, you can see some stones with different sizes. Each stone reflects one criterion and its size corresponds with the value of the team in this criterion and the weight of this criterion.

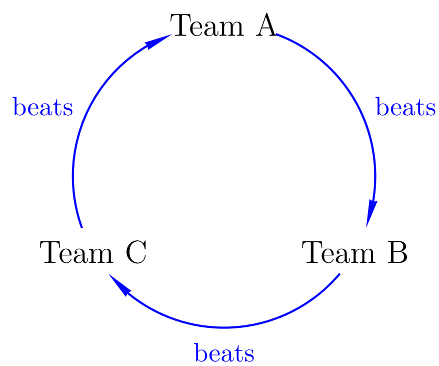
A detailed description of these two methods is given in later parts of this report in Section 2.1 and 2.2. The reference ranking used was the official rankings from the years 2005–2008.

# 2 Approaches

## 2.1 Tree-concept

After getting familiar with our problem and considering separately different criteria in order to decide between two teams which is the better, we came across the following question: which algorithm could be used to rank teams using our criteria? Such an algorithm must be able to minimize pathologies that naturally could appear.

The following example will clarify what we understand by a pathology.




Suppose that team A beats team B, team B beats team C and team C beats team A. It is difficult to say what's the better team between these three teams only based on this data. Hereby more information, more data is required. Luckily we could get all kinds of information we need about the games from internet. Here comes the question: how can we use these data? Which data should be used and what criteria decide a good team over a bad team?

### 2.1.1 Idea and algorithms

Before ranking the teams we decided to rank the criteria under our personal point of view (the first one should be the most important one and so on). In this way we made a list of criteria $c_1, c_2, c_3, \ldots, c_n$. This kind of sorted criteria allowed us to use them in a particular way, first use only $c_1$ to split the set of all teams $T$, in several subsets, $T_{1i}$, afterwards use $c_2$ and divide this $T_{1i}, i = 1, 2, \ldots$ into more subsets, $T_{2j}, j = 1, 2, \ldots$, and continue doing so until we eliminate all pathologies created by ambiguous results of the criteria applied. Our tree-concept is derived from this idea. A figure below shows a schematic view of our first approach.

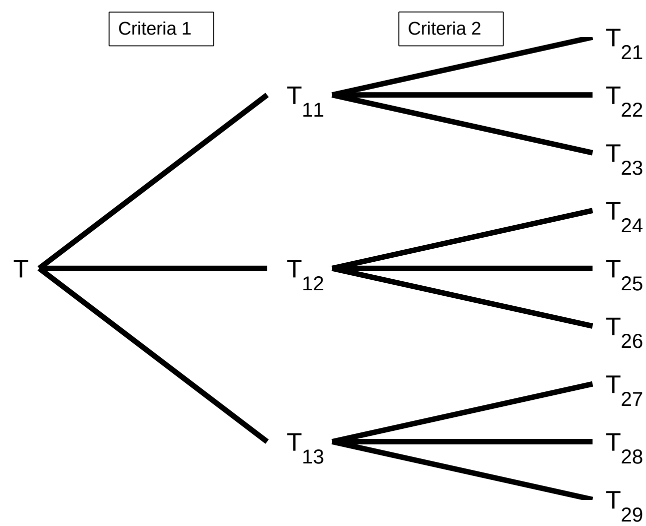



Figure 1: Using criterion 1 and 2 to split set of teams and sort them accordingly

Here we use these notations:

$T$ is the set of all teams.

$T_{1i}, i = \{1, \ldots, k_1\}$ are the sets (which are subsets of $T$) after apply c1 in T.

$\vdots$

$T_{ni}, i = \{1, \ldots, k_n\}$ are the sets (which one is subset of $T_{(n-1)j}$ for some j).

So, obviously we have, $T \supseteq T_{1i} \supseteq \ldots \supseteq T_{nj}$.

This idea is widely used in other sport game ranking systems. One example is the UEFA European Football Championship[5]. In the qualifying phase teams are divided into groups. Teams in the same group play with each other home and away. The points are awarded as three for a win, one for a draw, and none for a loss. If one or more teams having equal points after all matches have been played, other criteria, such as number of points obtained, are used to distinguish the sides. Here, the first criterion is the points, the second criterion is number of points obtained.

Notice that the above algorithm doesn't allow to consider different criteria in one row (i.e. one has to choose one and only one criterion in each level). This is an important point which has a lot of implications but roughly speaking we can conclude that the choice and hierarchy of the criteria has a big impact in the final ranking. Hence the following questions arise naturally: how to choose and to sort the criteria, and also how to improve our criteria through the use of mathematical tools?

This, of course, is a subjective topic, however one can argue about it. We will work with the following approach:

**(i)** Modify the criteria in order to avoid the highest number of pathologies. It can be accomplished maximizing the number of subsets after apply one criteria.

**(ii)** Try to be sensible, and take into account more relative data than absolute ones.

Particular case as a example:

| Absolute | Relative |
|---|---|
| "Number of wins" | "Number of wins" per "games played" |

Analysis:

i) Taking into account that each team has played from 10 to 13 matches. Therefore we shall assume that the average of matches played per team is 12. The following things are worth noting:

- *Number of wins* takes integer values, from 0 to 12.
- *Number of wins* per *games played* takes rational values, from 0 to 1.
- *Number of wins* criterion split $T$ in a maximum number of 13 subsets (number of victories goes from 0 to 12).
- *Number of wins* per *games played* criterion split $T$ in a maximum of 37 subsets.

A small explanation of where this 37 arise from could be necessary, let us define $W$ and $M$ as the sets of possible number of wins and possible number of matches played. Thus, $W = \{0, \ldots, 12\}$ and $M = \{10, \ldots, 13\}$. Now defining

$$W/M = \{a/b \mid a \in W, b \in M \text{ and } a/b \in \mathbb{Q}\},$$

it is obvious that 37 should be the $W/M$ cardinality. Notice that $\frac{0}{n} = 0$, $\frac{11}{11} = \frac{12}{12} = \frac{13}{13}$ and $\frac{5}{10} = \frac{6}{12}$.

ii) Assume that team A has won 6 and has played 10, and team B has won 7 and played 13. Which is stronger?

Number of wins says B.

Number of wins per games played says A.

Moreover, consider this other two data in order to have the criteria: in the family of absolute ones is *sum of scored points*, in the family of relative ones is *sum of scored points* per *total scores* (allowed and scored).

Assume now that these teams, A and B, have the following statistics:

A has a sum of scored points in 10 matches of 119 and has allowed 99 scores whereas team B has a sum of scored points in 13 matches of 118 and has allowed 14.

| | Score | Points allowed |
|---|---|---|
| Team A | 119 | 99 |
| Team B | 118 | 14 |

Which team is stronger now?
The criterion of sum of scored points suggests B.
The criterion of sum of scored points per total scores suggests A.

Obviously, team A is stronger than team B (it leads both categories).

### 2.1.2 Results

The implementation of this algorithm with different criteria is a quick job, and we got easily the different rankings. At this point we should check their validity somehow. The idea is the next one. In order to be sure that our theoretical speculations give us reasonable models and furthermore to find out new attributes of our rankings or to inspire our mind with the conviction of reaching new speculations, we took a new approach: use mathematical tools to compare our rankings arisen from our algorithms with each other, and, moreover, compare them individually with the official ranking, which we take as a reference ranking. Here we should notice that we do not think the official rank as a *correct* one, it is not curve fitting. The official ranking is just a benchmark for us for now, since it collected all kinds of opinions and was a reasonable but not precise ranking.

We will show some plots and comments. The $x$ axis of these plots is labeled with teams, from 1 to 120, sorted according the official ranking[8]. Each one of these team ranking values is assigned to another one, which corresponds to the position of this team in our ranking. Obviously, the closer a ranking is to the official one, the more similar to the straight line $y = x$ is the plot. So in order to measure the differences between two rankings we can use statistical tools like linear regression, correlation coefficient, least square error (LSE), etc. Here we use correlation [6] and least square error (LSE)[7] to compare our results with official ranking.

The first two showed figures are comparisons between our ranking taking data from 2005 season and the official ranking of the same season. In these first rankings we only used two criteria to simplify a bit. In the first ranking we choose as criterion 1 the number of wins and as criterion 2 the total score achieved, which might be classified as absolute ones. In the second ranking we used as criterion 1 the number of wins over number of games played and as criterion 2 the total margin (total score achieved minus total points allowed), which might be classified as relatives criteria. So, this could be useful, for instance, to compare the relative and the absolute criteria throughout different years taking as reference the official ranking.

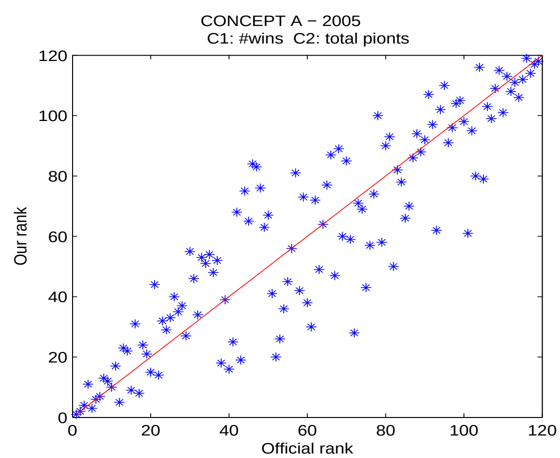



Figure 2: Comparison of our ranking and official ranking 2005

Criterion 1: number of wins

Criterion 2: total score

LSE = 173.4589

RankCorrelation = 0.8929

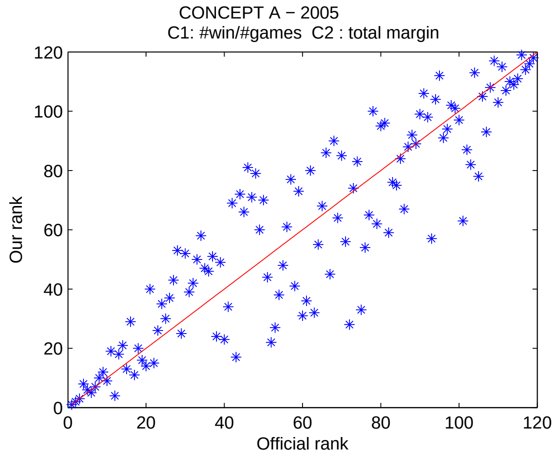



Figure 3: Comparison of our ranking and official ranking 2005

Criterion 1: number of wins / number of games

Criterion 2 : total margin

LSE = 173.7930

RankCorrelation = 0.8925

To be confidence with the results got we can examine other years under the same conditions. For example if we take the 2006 season we get the following plots:

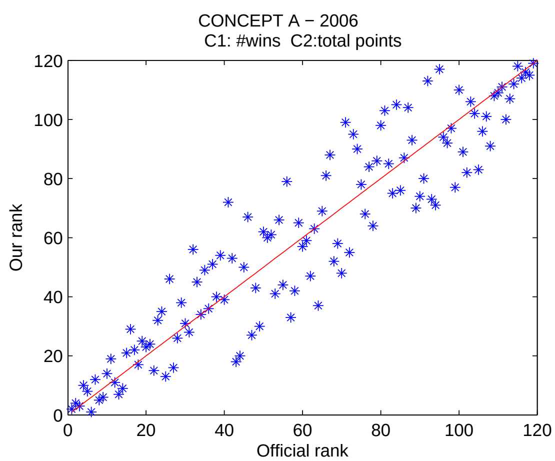



Figure 4: Comparison of our ranking and official ranking 2006

Criterion 1: number of wins

Criterion 2: total score

$$\text{LSE} = 140.8332$$

$$\text{RankCorrelation} = 0.9294$$

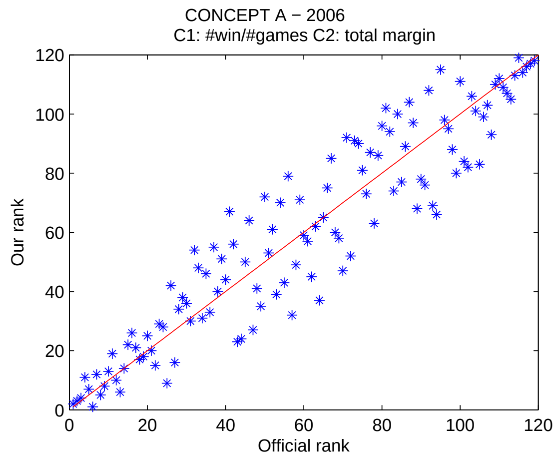



Figure 5: Comparison of our ranking and official ranking 2006

Criterion 1: number of wins / number of games

Criterion 2 : total margin

$$\text{LSE} = 136.0294$$

$$\text{RankCorrelation} = 0.9341$$

Examining the above figures, we could see that the ranking we got using tree-concept is close to the official ranking, especially for the top ranking teams and bottom ranking teams. The criteria we have selected decide whether a team is good (w.r.t. to the official rankings) or not to some extent. The result from Figure 2 and Figure 3 are similar, and the same with Figure 4 and Figure 5. It shows that taking #wins or #win/#games will give very similar results. The reason is that they are very related criteria: number of games played varies very less, from 11-14. So both of them are suitable for criterion 1.

Moreover, the methodology we follow should work with a larger number of seasons analyzed but not only with two seasons. Anyway these figures were useful to clarify that our first method is reasonable, though no result proves it to be a fair good one.

### 2.1.3 Advantages and drawbacks

Tree-concept is an intuitive idea to ranking teams. That empowered tree-concept with the following advantages: it can be quickly implemented into algorithms on computer; there are not many parameters in the model. The only thing we need to decide is which criterion should be the most important one, which should be the second, and so on. That is much easier to decide than the parameter in the following concept. We will see that soon.

However, as a naive concept, tree-concept also has its drawbacks. First of all, the teams sorted according to criterion 1 cannot flip between different sets $T_{1i}$, i. e. criterion 1 dominates the ranking. As the above example shows, we say the more games a team won, the better the team was. Is it true? Let's first see an example of Ohio State University football team in year 2005.

Ohio State University team won 10 games out of 12 games in that season. But the official ranking in 2005 listed it as No. 4, higher than other teams which have won 11 games, or which have got more margin of a game. Other ranking systems also list this team very high. According to our criteria mentioned above, this team should not list so high, but why both official ranking and other ranking system made similar decisions? Let us have a closer look at all the game results of Ohio State University in season 2005:

| Date | Win/Loss | Score | Opponent |
|---|---|---|---|
| 03-Sep-05 | Win | 34:14 | Miami OH |
| **10-Sep-05** | **Loss** | **22:25** | **Texas** |
| 17-Sep-05 | Win | 27:6 | San Diego St |
| 24-Sep-05 | Win | 31:6 | Iowa |
| **08-Oct-05** | **Loss** | **10:17** | **Penn State** |
| 15-Oct-05 | Win | 35:24 | Michigan St |
| 22-Oct-05 | Win | 41:10 | Indiana |
| 29-Oct-05 | Win | 45:31 | Minnesota |
| 05-Nov-05 | Win | 40:2 | Illinois |
| 12-Nov-05 | Win | 48:7 | Northwestern |
| 19-Nov-05 | Win | 25:21 | Michigan |
| 02-Jan-06 | Win | 34:20 | Notre Dame |

Table 1: Ohio State University, 2005

In 2005, Ohio State University only lost to Texas and Pennsylvania State University, whose official rankings are No. 1 and No. 3 in that season, and we can also find that the margin of the two games are very small, which implies that these teams might be well-matched in strength. From this example, we see that not only the number of wins that decides whether a team is good or not, but also the opponent's strength.

Secondly, tree-concept also has the drawback that there are fewer choices to modify and optimize with respect to the official ranking. Explicitly, the opimization is with respect to the correlation to the official ranking, i.e. maximization of the correlation which is smaller or equal to 1. The flexibility of this concept is too small for deeper studying.

We should look for a better concept, a more complicated one.

## 2.2 Balance-concept

### 2.2.1 Idea and algorithms

Balance-concept is our second approach to the problem of ranking football teams . It weights the considered criteria and ranks the teams with respect to a ranking value which is the sum of the weighted criteria in our model. For example, if we consider criteria as are number of wins, number of loss and scored points, and if we weight every of these criteria with factors $\omega_1$, $\omega_2$, and $\omega_3$, we will get the ranking value for each team in function $f$:

$$f = \omega_1 \times \#\text{wins} + \omega_2 \times \#\text{loss} + \omega_1 \times \text{scored points}.$$

While tree-concept does not allow criterion 2 to flip teams which are already ranked distinctively by criterion 1, balance-concept gives more freedom. For example, it allows flipping and treating different criteria equally important. It is also possible to consider much more factors (in tree-concept several level of subrankings might not lead to any further changes in the final ranking).

To take a closer look, let us examine the following example.

| | Number of wins | Number of loss |
|---|---|---|
| Team A | 12 | 20 |
| Team B | 11 | 5 |

Table 2: tree-concept will rank Team A higher than Team B

With a "traditional" way of ranking (i. e tree-concept, like it is often used in soccer leagues etc.) taking the number of wins in the first step and the number of loss in the second step, Team A will be clearly ahead of Team B. However, if we weight the number of wins with 0.2 and the number of losses with 0.8 (i. e. normalized weights), then Team B will be ahead.

| | Win | Loss | Ranking value |
|---|---|---|---|
| Team A | 12×0.2 | 20×0.8 | -13.6 |
| Team B | 11×0.2 | 5×0.8 | -1.8 |

Table 3: Give weight also to number of losses will change the rank

This concept is actually a generalization of tree-concept. To see this, let us consider a ranking with respect to tree-concept that considers the number of wins as the first criterion and the number of losses as the second one. Now we have to find a suitable weight vector $w = (w_{\#\text{win}}, w_{\#\text{loss}})$ to be able to get ranking values which leads to the same ranking as tree-concept suggests. We will get this by choosing the weights in a way that the number of wins get much higher importance than the number of loss, i. e. $w_{\#\text{win}} \gg w_{\#\text{loss}}$.

The implementation of this ranking procedure is very easy. After we have extracted the number of games, wins etc. in a matrix $M$, we multiplied this matrix (each row corresponds to the statistics of one team at the point of the ranking) with the weight vector and sorted the outcome in a descending order. In the following example, the first team (first row) has $x$ games at the point of our ranking with $y$ won games etc. The 7 columns contain the following values respectively:

1. Total number of games (at the moment of the ranking procedure)
2. Number of won games
3. Number of lost games
4. Ratio of won games over total number of games
5. Scored points
6. Allowed points (i. e. points made by the opponent)
7. Margin of win (i. e. difference between scored points and allowed points)

Obviously, this is a highly correlated data. For example, a team with a high number of won games has probably a high number of scored points (this is necessary to win the football games). The matrix looks like Table 4 (here UNC and NCS stands for North

Carolina University and North Carolina State University respectively).

| Team | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Virginia | 12 | 5 | 7 | 0.4167 | 181 | 214 | -33 |
| UNC | 12 | 3 | 9 | 0.2500 | 216 | 366 | -150 |
| Maryland | 13 | 9 | 4 | 0.6923 | 284 | 284 | 0 |
| Florida St | 13 | 7 | 6 | 0.5385 | 345 | 258 | 87 |
| Duke | 12 | 0 | 12 | 0.0000 | 179 | 406 | -227 |
| NCS | 12 | 3 | 9 | 0.2500 | 210 | 262 | -52 |
| Wake Forest | 14 | 11 | 3 | 0.7857 | 302 | 215 | 87 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

Table 4: Data matrix (unscaled).

An important point should be stressed: With this matrix we work already with highly correlated data. The ratio of won games over total number of games and margin of win are simple functions of factors of won and lost games and respectively the allowed and made scored points. This should be born in mind in any interpretation, since a ranking which considers apparently a high number of “different but strong correlated” factors won’t differ much from a ranking with respect to corresponding reduced core factors.

To make the comparison easier, we scale every factor values to the unit interval $[0, 1]$: Table 5. But we were aware that this means a refining of weighting values.

| Team | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Virginia | 0 | 0.3846 | 0.5833 | 0.4167 | 0.1320 | 0.1779 | 0.4978 |
| UNC | 0 | 0.2308 | 0.7500 | 0.2500 | 0.2020 | 0.5589 | 0.3223 |
| Maryland | 0.5000 | 0.6923 | 0.3333 | 0.6923 | 0.3380 | 0.3534 | 0.5472 |
| Florida St | 0.5000 | 0.5385 | 0.5000 | 0.5385 | 0.4600 | 0.2882 | 0.6777 |
| Duke | 0 | 0 | 1.0000 | 0 | 0.1280 | 0.6591 | 0.2069 |
| NCS | 0 | 0.2308 | 0.7500 | 0.2500 | 0.1900 | 0.2982 | 0.4693 |
| Wake Forest | 1.0000 | 0.8462 | 0.2500 | 0.7857 | 0.3740 | 0.1805 | 0.6777 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

Table 5: Data matrix (scaled).

Again, we compared the outcome (i. e. our algorithmic ranking by personal selection of the criteria) with the official data. In the Figure 6 one can find the position of the teams in the official ranking on the $x$ axis and their rank in our ranking on the $y$ axis. For example, given the weight vector $w = [0, 1, 0, 0, 1, 0, 0]$ the team on the first place in the official data got the 42$^{\text{th}}$ rank. A probable reason might be the scoring rate of this team, since we considered this in our ranking model equally with the number of wins.

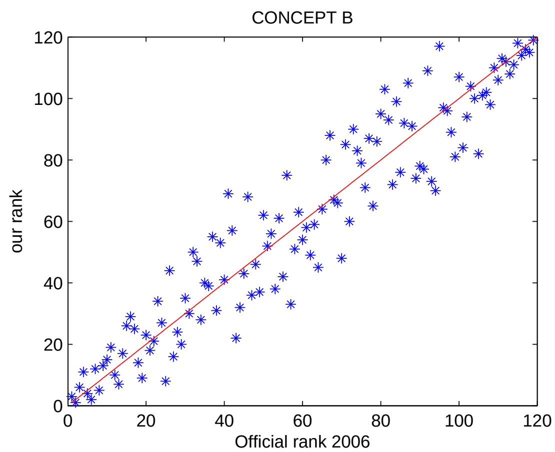



Figure 6: Comparison of the rank we got and the official rank, 2006

### 2.2.2 Home and away

After we have seen the differences in the ranking positions for teams which developed over the season, we want to examine the differences with respect to the place where the game is played, home or away, and whether there is such a relation between wins and home/away or not.

After examining 119 teams in season 2006, we get the result in figure 7. In this figure, $x$ axis listed all 119 teams of that season, $y$ axis is the value of the ratio

$$\frac{\#\text{ wins at home}}{\#\text{ total wins}}. \tag{1}$$

If the ratio is > 0.5 (corresponds to the red spots above the blue line in the figure 7 below), it means that team won more games at home than away. The larger the ratio is, the more wins were gained while at home. Otherwise, the team won more while away.

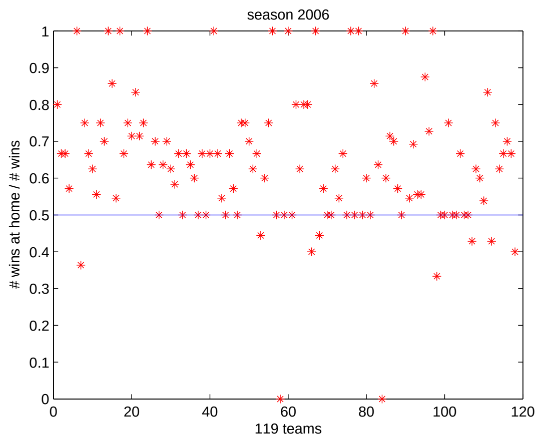



Figure 7: number of wins at home over total number of wins, 2006

The above figure shows that most of the teams have the ratio > 0.5, i. e. most of the teams won more games at home. Actually, 85 out of 119 teams' ratios (71.43% of the teams) are > 0.5. Similar results appeared not only in season 2006, but also the other seasons, see Table 6.

| Season | # (ratio> 0.5) | # teams | # percentage |
|---|---|---|---|
| 2005 | 83 | 119 | 69.75% |
| 2006 | 85 | 119 | 71.43% |
| 2007 | 91 | 120 | 75.83% |

Table 6: The number of teams that had won more games at home than away over different seasons

Clearly, a team has an advantage if it plays in his own stadium. Then it knows the environment and the players are used to play on the court. Even more important, the home team gets support by their own fans in usually higher number than the guests. It is quite expensive and sometimes difficult for fans to support their team when it plays in another stadium, sometimes the teams have to travel thousands of kilometers. That's why the number of fans for the home team is much higher than the number of supporters for the guest team. This can have a big impact on the game, because the guest might feel uncomfortable in the unknown stadium. The home fans will shout to confuse the guests in critical situations and in general they will try to create an atmosphere which makes it difficult for the guest players to stay focused and concentrated.

With these things in mind we want to examine the results of the teams in their own stadium and at their opponent's place. More accurately spoken, we want to weight these results because obviously a win in the own stadium is easier to achieve than at the unknown stadium with many opposite fans.

Let $W_H^i, W_A^i$ denote number of wins at home, number of wins away for team $i$ respectively. Let $(\alpha, \beta)$ be the weight vector. So if we give different values to $\alpha$ and $\beta$, ranking according to the value

$$y_i = W_H^i \times \alpha + W_A^i \times \beta, \quad i \in \mathbb{N} \tag{2}$$

will give different ranking results.
Since

$$W_H^i + W_A^i = \#wins, \tag{3}$$

we should normalize the weight vector, i. e.

$$\alpha + \beta = 1. \tag{4}$$

So the next question is how to decide the value for $\alpha$ and $\beta$ in order to get our ranking results. Our approach will be illustrated in the following section.

### 2.2.3 Optimization with respect to the correlation

After we have seen, how a different weighting of the results at home and on the road can influence the ranking, we want to try to approximate the official ranking, i. e. the ranking published by the National College Football Association each week with our approach.

The correlation gives us a measure how close our ranking (our ranking plot) reaches the official plot. Unfortunately, some parts of the algorithm for the computation of the official ranking are not published. That's why it is not possible to understand the calculation of the official ranking fully. With our approach and a very high correlation we could just guess that the results at home and on the road the ranking of the team might influence.

We got the following correlation data:

| Year | max.Correlation | $\alpha$ value |
|---|---|---|
| 2005 | 0.900912 | 0.500 |
| 2006 | 0.935657 | 0.650 |
| 2007 | 0.914605 | 0.585 |
| 2008 | 0.945975 | 0.575 |

Table 7: Optimization with official ranking as benchmark

From the result we can see that official rankings do not treat home and away very differently. One possible reason is that people think better teams should always beat worse teams.

### 2.2.4 Point margin per game

Now we will consider more factors than what are in Matrix $M$.

The motivation is very simple: We imagine the following situation where Team A had 8 games won and 5 games lost. Assume that the lost games are equally distributed over the whole season (here with a total of 13 games). Team B had big problems at the beginning, they lost the first games. But afterwards they won almost every game (only 1 lost). Which team is better now? Of course, like for the whole problem, both possible answers, i. e. Team A is better or respectively Team B is better, can be well-supported by arguments. However, let us consider the second answer. It might be reasonable that Team B is better, because teams need usually some time, the new players have to be integrated and the new strategies should become familiar. In the former discussed rankings, this fact is not considered. The first games counts already fully and that is why a team might have a better ranking only due to the fact that they won easier games at the beginning etc. But Team B might be actually the better team after this time of getting used to the new season. See Table 8.

If we want to consider this point now, we can do this in different ways. For example, we could give the factor "game won" a higher value at the end of the season. Another approach is giving the point margin a higher value at the end of the season. As a point margin, we understand the difference between the scored points and the allowed points in one game. This means that teams which scored very high at the end of the season benefit in the ranking.

One problem arises with this approach: Often the teams have not the same amount of played games. We could set the missing games with a point margin of 0, i. e. a tied game. But then a good team which usually wins very high has a disadvantage. We tried to decrease the impact of this problem by giving the earlier games lower weights, i. e. the scored points in games played or not have not a very strong impact on the ranking value.

| games | team A | team B | team C |
|---|---|---|---|
| $1^{th}$ | 0 | 3 | 0 |
| $2^{th}$ | 0 | 5 | -22 |
| $3^{th}$ | 24 | -1 | -17 |
| $4^{th}$ | 28 | -11 | -34 |
| $5^{th}$ | 40 | 8 | -20 |
| $6^{th}$ | 22 | 25 | -28 |
| $7^{th}$ | 17 | 8 | 2 |
| $8^{th}$ | 5 | -2 | -8 |
| $9^{th}$ | 9 | 13 | 11 |
| $10^{th}$ | **-10** | **7** | **14** |
| $11^{th}$ | **1** | **34** | **27** |
| $12^{th}$ | **6** | **5** | **13** |
| last | **-5** | **9** | **19** |

Table 8: matrix of point margin per game, strength of teams vary from the beginning of the season to the end

In general, the consideration of point margins per game gives us a strongly increased amount of data points which are considered for the ranking value. The more data we can consider, the more the application of statistical methods is justified. Unfortunately, we could not implement deeper methods due to time restrictions.

To reduce the necessary weighting parameters, we separated the played games in three parts, i. e. the first five games of the season, the most recent four games and the four games in between. In our modeling ansatz, the weight for the first part was the smallest.

### 2.2.5 Strength of scheduling

Before we end our attempts to give different factors an impact in our ranking model, we want to see how we give the current position of the team in the ranking (or its last year's position in the final ranking) an importance.

To see that, let us recall the example in section 2.1.3, where Ohio State University team won 10 games in season 2005, but was ranked higher than the teams who won 11 games.

From this example, we can see that not only the results of a game is important, but also the opponent's strength will have a large impact on the game result. Therefore, it is necessary to include the strength of the schedule into our ranking system.

Let us assume one team wins against a team with a higher ranking in previous year. Then, this win might be much more important than a win against a team on the last place in this league. Why? Winning against a stronger team is typically much harder than winning over a weaker team.

So even the game margin is very less, we would like to add weight to this win, i. e. give "bonus" to the weaker winning team while ranking. On the other hand, if a strong team, which in our context means ranking high in previous year's game, lose to a weaker team, we would also like to "punish" this strong team. The more difference of the two teams' ranking are, the more bonus or punishment we give to the ranking value. For example, if a team ranks 50, it would get more bonus if it wins a top 10 team than it wins a team ranking 40. Also, this team will get less punishment if it loses to the team with ranking 60 than losing to teams in the last ranking. This is exactly our idea which we try to implement in our ranking model.

We implement the idea as follows: the bonus value (if it is punishment, we would treat it as negative bonus) should be proportional to the difference in previous year's ranking. More precisely, if team $i$ wins team $j$ in one game, the bonus for team $i$ should be

$$\text{bonus}_i = k_1 \times (R_i - R_j), \quad if \quad R_i > R_j \tag{5}$$

where $R_i, R_j$ is the ranking of team $i$ and team $j$ respectively. Notice, here a stronger team has a higher ranking, i. e. $R_{\text{strong}} < R_{\text{weak}}$.

Analogously, if team $i$ loses to team $j$, the bonus for team $i$ should be

$$\text{bonus}_i = k_2 \times (R_i - R_j), \quad if \quad R_i < R_j. \tag{6}$$

Now, we would also use the bonus as a criterion to get our ranking in balance-concept.

One thing should be mentioned is that this criterion "strength of schedule" could not be used alone to get a ranking, since we only add value to the wins over stronger teams and the losses to the weaker team. So it is only an important minor factor. Figure 8 shows that if we ranking only according to the strength of scheduling, it is essentially uncorrelated to the official ranking. Similar results are obtained for other seasons.

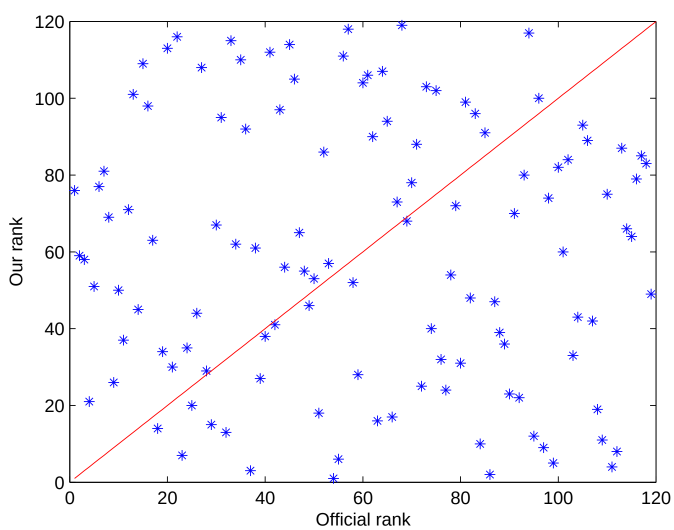



Figure 8: Correlation of this ranking and official ranking is only -0.0550

# 3 Conclusion

We hope that the last pages showed the complexity and variety of the problem about ranking football teams. One of the most important conclusions should be stressed: There is *not that* ranking what means that whatever ranking somebody chooses, it depends on his personal preferences about the factors (which one should be considered, how strong should be every single weight etc.). Therefore, his preferences lead to a model with respect to this parameters.

Besides no ranking can bear in mind every possible factor, so one will always have to choose suitable criteria for a ranking. In the past years the range of knowledge gathered by the mankind increases very fast. This will continue and will even fasten up. The main problem is to select suitable information from this data, to give them a certain priority (i. e. a ranking). The performance of computers will always be a limit and that is why mathematicians will have to look for efficient algorithms.

# References


[1] http://en.wikipedia.org/wiki/American_football

[2] http://en.wikipedia.org/wiki/Bowl_game#Post-season_bowls

[3] http://en.wikipedia.org/wiki/Bowl_Championship_Series

[4] Callaghan, Thomas; Mucha, Peter J.; and Porter, Mason A. [2007]. Random Walker Ranking for NCAA Division I-A Football. American Mathematical Monthly, Vol. 114, No. 9: 761-777.

[5] http://en.wikipedia.org/wiki/UEFA_European_Football_Championship

[6] http://en.wikipedia.org/wiki/Correlation

[7] http://en.wikipedia.org/wiki/Least_squares

[8] BCS Official Rankings http://www.mratings.com/cf/compare.htm