Crowd-ranking

DOI:
10.1007/s12351-017-0324-7

Document status and date:
Published: 01/03/2020

Document Version:
Author's version before peer-review

Title: Crowd-Ranking: A Markov-based method for ranking alternatives

Abstract

Many ranking algorithms rank a set of alternatives based on their performance in a set of pairwise comparisons. In this study, a special scenario is observed in which the objective is to rate and rank a set of groups in a traditional recruiting situation, in which the groups extend offers to the set of individuals, and the individuals will select one of their available offers. The new ranking method, Crowd-Ranking, uses collective wisdom and decision-making in conjunction with Markov chains to create competitive matches between alternatives and ultimately provide a ranking of the alternatives. First, the method is evaluated by its performance in a perfect season scenario. Next, it is applied to the case of NCAA football recruiting in the power conferences (ACC, Big Ten, Big 12, Pac 12 and SEC) in the Football Bowl Subdivision. For the Big Ten conference, the method performs significantly better than popular existing services at predicting future team performance based on past recruiting rankings. For a comprehensive national ranking of the power conferences, there is no statistically significant difference between Crowd-Ranking and the other methods.

Keywords

Decision making, Markov chains, ranking, crowds

1. Introduction

In general, for a set of n alternatives, the rank of an alternative is its relative importance to the other alternatives in the set. Often, a ranking method will first produce ratings of the alternatives. Next, sorting the alternatives in order of decreasing ratings will provide a ranking.

Many ranking algorithms use pairwise comparisons between the alternatives to determine ratings and rankings. Kendall and Smith (1940) introduced the Method of Paired Comparisons, which developed rankings of alternatives based on a set of pairwise preferences between alternatives. Arpad Elo's Rating System (1961) is another example of a rating method that uses the results of head-to-head competitive matches to provide ratings, and has been used primarily in the rating of chess players. The Analytic Hierarchy Process (Saaty 1980, 1986) is a popular ranking method that uses pairwise comparisons to populate a reciprocal dominance matrix, in which the dominant eigenvector of the matrix is the rating of the alternatives.

In addition to using pairwise comparisons, some ranking methods use Markov chains (David 1970) to develop ratings. Google uses the PageRank method (Brin, Page 1998) to rank its webpages when returning search results, which contains a series of pairwise comparisons embedded in its algorithm. There also exist methods that use pairwise competitive matches to estimate conditional probabilities (see Kvam and Sokol (2006), Callaghan et al. (2007)) that ultimately provide ratings. There is also the Markov method (Govan 2008) that directly uses Markov chains to rate its alternatives by connecting them through a voting process based on pairwise results.

The Crowd-Ranking method is an extension of the Markov method. It uses the basic theory of the method as the foundation of the ranking algorithm. The basic idea of the Markov method can be described by *voting*, in that the weaker alternative will place a vote for the stronger alternative.

These votes populate a dominance matrix in which the steady-state probability vector is the rating of the alternatives. There are many ways to determine the voting scheme for the method. In Crowd-Ranking, we use a simple approach in that the losing alternative places one vote for the winning alternative.

The remainder of this article is organized as follows. Section 2 will provide the problem definition regarding the ranking method. Section 3 will define the ranking method and algorithm. Next, in Section 4, a perfect-case scenario is observed to validate the theoretical basis of the ranking method. Section 5 will provide a case study application of Crowd-Ranking to NCAA football recruiting from 2003 – 2013, and compare its predictive power to other leading recruiting ranking services. Last, Section 6 provides a conclusion and discussion points.

2. Problem definition

This study is to focus on a specific scenario, in which a set of groups are recruiting a set of individuals to join the groups. The main objective is to rank the groups in order of the quality of individuals that they ultimately obtain. This problem can be thought of as a typical recruiting process, where groups recruit individuals to join them.

The scenario is defined below:

1) There are n groups, $\{G_1, G_2, \ldots, G_n\}$, that extend offers to a set of m individuals, $\{I_1, I_2, \ldots, I_m\}$. An individual group can extend offers up to the entire set of individuals.
2) Each individual, however, may only select one group to join.

Many recruiting applications fit this defined scenario. For example, a university can be defined as a group that extends admittance to a set of individual students, and the students can only select one university to attend. Another example is to consider a company (group) that extends job offers to a set of individual prospective employees (individuals), and the prospective employees may only select one job offer. Last, consider the case of collegiate athletic recruiting, specifically, in NCAA football. The college football teams (groups) will extend scholarship offers to a set of individual recruits (individuals), and the recruits may only select one team from the set of offers.

It is important to note that we are ranking one specific crop of individuals for each group, so we won't necessarily have a comprehensive ranking of the groups as a whole. This is because a group is comprised of several crops of individuals over time. However, we could use these individual rankings to feed into an encompassing ranking. For example, in college football, the set of five recruiting classes that comprise a team could be combined to obtain one final team ranking (as will be seen in Section 5 when measuring predictability).

3. Crowd-Ranking methodology

The Crowd-Ranking method relies on a dual-level decision process between two parties: the groups and the individuals. The groups are the entities being ranked, and must decide on the set of individuals that will receive an offer. The individuals are a large set of decision-makers that select a group based on their available offers.

The fundamental input of the ranking algorithm relies on two forms of data:

- An individual's set of offers.
- An individual's selection of a group from its set of offers.

Once the data is obtained, the first step in the ranking algorithm is to develop the voting matrix. When an individual selects a group, which is said to be the "winning" group, the remaining groups that offered the individual and were not selected are said to be the "defeated" groups. The set of defeated groups will place a "vote" for the winning group, and this will be done for the set of all individuals.

The voting matrix, V, will take the following form:

$$V = \begin{bmatrix} 0 & v_{12} & v_{13} & \cdots & v_{1n} \\ v_{21} & 0 & v_{23} & \cdots & v_{2n} \\ v_{31} & v_{32} & 0 & \cdots & v_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & v_{n3} & \cdots & 0 \end{bmatrix} \tag{1}$$

For Equation (1), the following expressions are appropriate:

$$v_{ij} = total\ number\ of\ votes\ from\ G_i\ to\ G_j, \forall\ i \neq j$$

$$v_{ij} = 0, \forall\ i = j$$

Each matrix entry indicates the total number of preferences from individuals regarding the groups. For example, the entry $v_{32} = 7$ indicates that there were seven individuals that selected group G_2 that had offers from group G_3. In turn, the entries of the voting matrix can be thought of as a series of pairwise comparisons that display the relative dominance among the groups.

One can also think of the voting matrix as a collection of wins and losses in terms of individuals between all of the groups in their respective contests, with a contest being a head-to-head matchup between two groups that offered an individual, and the winner of the matchup being the selected group. There is always one winner per individual, the selected group, and there can be any number of losers, depending on how many offers were received by that individual. Notice that a strong individual with many offers will have a larger impact on the rankings than an individual with only a few offers. Also, the quality of the offers are important, meaning offers from groups with higher rankings will carry more weight than from groups with lower rankings.

Notice that this voting system is unlike other systems used in the theory of social choice, because the voters (in our case, individuals) do not necessarily have the option to vote for all of the alternatives (in our case, groups). If it were the case where each individual could vote for any of the alternatives, then there exists many different ranking schemes to address this problem (for example, see Balinski (2014)).

Based on the voting matrix in Equation (1), we can use the principles from the Markov method to rank the relative dominance of the groups (Govan 2008, Langville 2012). The Markov method assumes that the voting matrix can be represented as a Markov chain. If a *random walk* is taken on the Markov graph, the long-run proportion of time spent at each group will be the rating of that group's strength.

The next step is to normalize the rows of the voting matrix in order to obtain a transition probability matrix. The matrix, P, will take the following form:

$$P = \begin{bmatrix} 0 & \frac{v_{12}}{\sum_{j=1}^{n} v_{1j}} & \frac{v_{13}}{\sum_{j=1}^{n} v_{1j}} & \cdots & \frac{v_{1n}}{\sum_{j=1}^{n} v_{1j}} \\ \frac{v_{21}}{\sum_{j=1}^{n} v_{2j}} & 0 & \frac{v_{23}}{\sum_{j=1}^{n} v_{2j}} & \cdots & \frac{v_{2n}}{\sum_{j=1}^{n} v_{2j}} \\ \frac{v_{31}}{\sum_{j=1}^{n} v_{3j}} & \frac{v_{32}}{\sum_{j=1}^{n} v_{3j}} & 0 & \cdots & \frac{v_{3n}}{\sum_{j=1}^{n} v_{3j}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{v_{n1}}{\sum_{j=1}^{n} v_{nj}} & \frac{v_{n2}}{\sum_{j=1}^{n} v_{nj}} & \frac{v_{n3}}{\sum_{j=1}^{n} v_{nj}} & \cdots & 0 \end{bmatrix} \tag{2}$$

For Equation (2), the following expressions are appropriate:

$$p_{ij} = \frac{v_{ij}}{\sum_{j=1}^{n} v_{ij}}, \forall\, i \neq j$$

$$p_{ij} = 0, \forall\, i = j$$

$$\sum_{j=1}^{n} p_{ij} = 1, \forall\, i$$

Each matrix entry indicates the probability that the group will vote for the respective group. For example, the entry $p_{32} = .25$ indicates that group G_3 will vote for G_2 25% of the time.

The row-normalized matrix in Equation (2) is now a transition probability matrix from a Markov chain. We can then find the dominant eigenvector of the matrix, which is equivalent to the steady-state probability vector of the matrix. A necessary condition to obtain this steady-state probability vector is to have an ergodic Markov chain. For an ergodic Markov chain, one can also find the solution to Equation (3), which indirectly yields the steady-state probability vector.

$$\lim_{n\to\infty} (P)^n = \begin{bmatrix} \pi_1 & \pi_2 & \pi_3 & \cdots & \pi_n \\ \pi_1 & \pi_2 & \pi_3 & \cdots & \pi_n \\ \pi_1 & \pi_2 & \pi_3 & \cdots & \pi_n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \pi_1 & \pi_2 & \pi_3 & \cdots & \pi_n \end{bmatrix} \tag{3}$$

Thus, the steady-state probability value for each group becomes a rating of the relative dominance of that group. There exist other ranking methods, which use the dominant eigenvector of a non-negative matrix to develop a ranking (Saaty 1987, Keener 1993). A matrix that gives pairwise dominances of its alternatives will yield an eigenvector solution that displays the relative dominance among its alternatives.

The vector, $\boldsymbol{\pi}$, will be not only be the steady-state probability vector of the matrix P, but it will also be the ratings of the respective groups. The vector can be expressed as follows:

$$\boldsymbol{\pi} = \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_n \end{bmatrix} \tag{4}$$

Notice that the entries in the vector in Equation (4) will always sum to a value of one.

A critical element of this model is that the groups possess a collective wisdom on the quality of individuals that they are offering. In turn, the quality of an individual is weighted by the quality of its offers. For example, if a group obtains an individual with offers from other groups with high ratings, it will gain votes from each of those groups. Because those top groups will now give higher transition probabilities to the winning group, the steady-state probability value and thus rating for that group will increase. On the flip side, if a group obtains an individual with no offers from other groups, it will not have an impact on their rating. This is because the model is built on the concept that the reward of an individual is based on how much other groups desire that individual. There have been recent studies, for example, Herm, Callsen-Bracker, and Kreis (2014), which have shown the value of using crowds to predict in sports.

4. Crowd-Ranking method for perfect season

To validate the ranking method itself, a "perfect season" scenario will be introduced and reviewed. The purpose of this is to show that the ranking method is performing and providing results as expected. This approach is very similar to the approach from Chartier et al. (2011) when reviewing the sensitivity and stability of various ranking methods.

In defining a perfect season, it is assumed that there is a specific ordering of both the groups and individuals, and the best groups offer and obtain the best individuals. An intuitive explanation of a perfect season is to have a series of matches with no "upsets," where an upset is defined as a weaker group defeating a stronger group in a match.

To start, assume the following preference relationship holds true for the set of all groups:

$$G_i > G_{i+1} \; \forall \; i \; \in \{1, \ldots, N-1\}$$

Also, assume that the following preference relationship holds true for the set of all individuals:

$$I_i > I_{i+1} \; \forall \; i \; \in \{1, \dots, M - 1\}$$

In this example, for simplicity, say that both the number of groups and individuals are equal to 5, and that the following set of offers and selections were made based on the above preference relationships. An offer is indicated by a lower-case "x" and the selection of an offer is indicated by a capital and bold "**Y**."

	G_1	G_2	G_3	G_4	G_5
I_1	**Y**	x	x	x	x
I_2		**Y**	x	x	x
I_3			**Y**	x	x
I_4				**Y**	x
I_5					**Y**

Table 1: List of offers and selections for perfect season

As discussed before, there are no upsets in Table 1. Group 1 defeated all of the other groups for the best individual, Group 2 defeated the lower three groups for the 2^{nd} best individual, etc.

An issue with the resulting voting matrix is that an undefeated group creates an absorbing state, because there were no losses and hence no votes. If you were to normalize the rows of the voting matrix, the first row would still contain all zeroes. A popular strategy for handling an undefeated group is obtained from the PageRank method and its "dangling node" adjustment (Langville and Meyer 2006). This is also the strategy used by Chartier et al. (2011) when observing the Markov method for a perfect season. The adjustment is to add a value of $1/n$ to the row in the transitional probability matrix of the absorbing state. Thus, the resulting matrix will be stochastic and can be solved to obtain a steady-state probability vector.

For Crowd-Ranking, the dangling node adjustment is needed when a group obtains all of the individuals to whom they make offers. Essentially, this group would never have "lost" because no individual that they offered selected another group, and so they will never vote for another group.

The final ratings produce a coherent ranking of the five groups to what was expected from the initially assumed preferences. Generally, for *n* groups, the rating for a perfect season with the Markov method is given as follows (Chartier et al., 2011), where $H(n)$ is the *n*th partial sum of the harmonic series:

$$\boldsymbol{\pi} = \frac{1}{H_n}\begin{bmatrix} 1 \\ 1/2 \\ 1/3 \\ \vdots \\ 1/n \end{bmatrix}$$

In turn, for a perfect season, it can be seen that the Crowd-Ranking method produces a "correct" ranking in terms of the initial preference relationships for all of the groups. However, it should be noted that the Crowd-Ranking method is an extension of the Markov method, and that the Markov method can be extremely sensitive to upsets, especially in its tail (Chartier et al., 2011).

5. Application of Crowd-Ranking to NCAA football recruiting

In this section, we apply Crowd-Ranking to NCAA football recruiting in the Football Bowl Subdivision (FBS) during the years 2003 – 2013. In this study, the groups were the football teams' recruiting classes, and the individuals were the prospective football recruits.

Due to limited available data, we consider only teams in the power conferences (ACC, Big Ten, Big 12, Pac 12 and SEC) and Notre Dame. We will do rankings by each individual conference and a comprehensive national ranking. In the conference rankings, we only consider teams that did not switch conferences during the study timeframe.

The input data was available through Rivals (www.rivals.com), a subscription-based online recruiting website. Rivals is a leading service for online recruiting news and information, reaching an estimated 28.39 million unique users in March 2011, according to its website. The data for smaller conferences (such as the Mid-American Conference) was not available and thus we could not include those conferences in the study.

The Rivals website provides the data in its individual recruit pages of all of the teams that have offered a football scholarship to that particular recruit. The recruit will select one team by the end of the recruiting process. The selected team will receive votes from all other teams that extended this recruit a scholarship offer. As explained in Section 3, each individual recruit can consist of up to *n-1* competitive matches among the teams. Table 2 is an example of the voting matrix for Big Ten teams during the 2012 recruiting season.

	Ill.	**Ind.**	**Iowa**	**Mich.**	**MSU**	**Minn.**	**NU**	**OSU**	**PSU**	**Purd.**	**Wisc.**
Ill.	0	5	6	16	8	4	4	6	4	5	4
Ind.	3	0	7	10	5	1	5	9	0	4	4
Iowa	3	1	0	10	4	3	4	6	5	1	4
Mich.	0	0	2	0	2	1	3	11	1	1	4
MSU	0	1	2	14	0	2	1	10	1	1	4
Minn.	2	3	5	3	3	0	1	7	0	2	3
NU	1	1	3	3	2	0	0	3	2	2	2
OSU	0	0	1	6	2	2	1	0	0	0	1
PSU	0	0	1	6	1	0	0	7	0	1	0
Purd.	0	3	4	4	5	0	1	8	0	0	1
Wisc.	3	0	4	5	4	1	0	5	1	1	0

Table 2: Voting Matrix for Big Ten teams in 2012

Each matrix entry indicates a vote from one team to another. For example, the entry (Ill., MSU) = 8 means that there were 8 recruits that selected MSU (Michigan State) that had scholarship offers from Ill. (Illinois). Notice that the voting matrix is not a symmetrical matrix.

The next step is to normalize the rows of the voting matrix so that the sum of each row in the voting matrix is equal to a value of one. Table 3 is an example of the row-normalized matrix for Big Ten teams during the 2012 recruiting season.

	Ill.	**Ind.**	**Iowa**	**Mich.**	**MSU**	**Minn.**	**NU**	**OSU**	**PSU**	**Purd.**	**Wisc.**
Ill.	0	0.081	0.097	0.258	0.129	0.065	0.065	0.097	0.065	0.081	0.065
Ind.	0.063	0	0.146	0.208	0.104	0.021	0.104	0.188	0	0.083	0.083
Iowa	0.073	0.024	0	0.244	0.098	0.073	0.098	0.146	0.122	0.024	0.098
Mich.	0	0	0.080	0	0.080	0.040	0.120	0.440	0.040	0.040	0.160
MSU	0	0.028	0.056	0.389	0	0.056	0.028	0.278	0.028	0.028	0.111
Minn.	0.069	0.103	0.172	0.103	0.103	0	0.034	0.241	0	0.069	0.103
NU	0.053	0.053	0.158	0.158	0.105	0	0	0.158	0.105	0.105	0.105
OSU	0	0	0.077	0.462	0.154	0.154	0.077	0	0	0	0.077
PSU	0	0	0.063	0.375	0.063	0	0	0.438	0	0.063	0
Purd.	0	0.115	0.154	0.154	0.192	0	0.038	0.308	0	0	0.038
Wisc.	0.125	0	0.167	0.208	0.167	0.042	0	0.208	0.042	0.042	0

Table 3: Row-Normalized Matrix for Big Ten teams, 2012

The values in Table 3 are the probabilities that each team will vote for the other teams. For example, the entry (Ill., MSU) = 0.129 means that Ill. will vote for MSU 12.9% of the time.

The above row-normalized matrix is now analogous to the one-step transition probability matrix for a Markov chain, and can be used to obtain the rating vector of the teams.

Table 4 is the steady-state probability vector and resulting ranking for the 2012 Big Ten recruiting classes. Notice that, as expected, the sum of the team ratings is equal to one.

Rank	Team	Rating
1	Mich.	0.2344
2	OSU	0.2246
3	MSU	0.1062
4	Wisc.	0.0939
5	Iowa	0.0922
6	NU	0.0649
7	Minn.	0.0627
8	PSU	0.0361
9	Purd.	0.0359
10	Ill.	0.0276
11	Ind.	0.0215

Table 4: Rankings and Ratings of Big Ten team recruiting classes, 2012

The recruiting class rankings were calculated each year from 2003 – 2013 for the ACC, Big Ten, Big 12, Pac 12, and SEC conferences. In addition to the conference rankings, we did a national ranking of all teams in the power conferences and Notre Dame for the years 2003 – 2013.

5.1 Comparison of Crowd-Ranking method to other recruiting rankings

We evaluate the Crowd-Ranking method with Rivals (www.rivals.com), Scout (www.scout.com), and 247Sports (www.247sports.com), which are leading providers of recruiting news, information, and rankings for NCAA football. ESPN is another major recruiting service, but has data dating back to only 2006, and thus was not included in this study.

The three services generally use an additive linear function to obtain team rankings, with the primary input being individual players' ratings. As an example, Rivals first calculates the total points each team obtains summing up the individual points for the top 20 recruits. Rivals has a team of analysts that assign a point value to each recruit. The ratings for the Rivals method have been normalized out of 100 total points for comparison. Table 5 is an example of the comparison between Rivals and Crowd-Ranking for the 2012 Big Ten recruiting season.

Rivals Ranking Method			Crowd-Ranking Method		
Rank	Team	Rating	Rank	Team	Rating
1	OSU	23.14	1	Mich.	23.44
2	Mich.	20.71	2	OSU	22.46
3	Purd.	9.76	3	MSU	10.62
4	MSU	8.51	4	Wisc.	9.39
5	Iowa	8.28	5	Iowa	9.22
6	PSU	6.88	6	NU	6.49
7	Wisc.	5.70	7	Minn.	6.27
8	NU	5.10	8	PSU	3.61
9	Ill.	4.46	9	Purd.	3.59
10	Ind.	4.23	10	Ill.	2.76
11	Minn.	3.23	11	Ind.	2.15

Table 5: Comparison of Rivals Ranking to Crowd-Ranking, 2012

From Table 5 it can be seen that both ranking methods perform similar to one another not only in rankings, but in the distribution of ratings. Table 6 shows the difference in ranking for each team between these two methods for the Big Ten recruiting season in 2012.

Team	Rivals	Crowd-Ranking
Ill.	9	10
Ind.	10	11
Iowa	5	5
Mich.	2	1
MSU	4	3
Minn.	11	7
NU	8	6
OSU	1	2
PSU	6	8
Purd.	3	9
Wisc.	7	4

Table 6: Crowd-Ranking vs. Rivals, 2012

Note that there were a few teams with significant differences in rankings. For example, in 2012, the Rivals method ranked Purdue (Purd.) 3^{rd} and Wisconsin (Wisc.) 7^{th}, while the Crowd-Ranking method ranked Wisconsin 4^{th} and Purdue 9^{th}. The reason behind this difference is that in 2012, Purdue had 26 recruits and Wisconsin had only 12 recruits. This illustrates that although Purdue had a large number of recruits, they were not necessarily recruits offered by other Big Ten teams. Wisconsin, on the other hand, had a smaller number of recruits in 2012, but many other quality Big Ten teams offered the recruits.

The same ranking analysis was performed for the ACC, Big 12, Pac 12 and SEC conferences from 2003 – 2013 for Rivals, Scout, and 247Sports, and is discussed in Section 5.2. In Section 5.3, we perform the analysis for a comprehensive national ranking that included the power conferences and Notre Dame.

5.2 Predictive power of Crowd-Ranking method, by conference ranking

In this section, we examine the predictive power of the four ranking methods: Crowd-Ranking, Rivals, Scout, and 247Sports, by conference. Since conferences does not issue their own set of rankings, we will examine the results of all conference games (excluding teams that entered or exited the conference during the study period) from 2007 – 2013, and compare Crowd-Ranking to Rivals, Scout, and 247Sports in their ability to accurately pick the winner.

First, we establish a composite ranking for each team in each season. Since college football players can remain on their team for up to five years, we used the sum of the previous five recruiting class rankings to obtain a composite team ranking. For example, in 2009, the composite ranking for Illinois will be the sum of their recruiting class rankings from 2005 – 2009, since those are the classes that directly impact the team.

First, we examine the total number of games accurately predicted by each ranking method. Next, we will see how many games that Crowd-Ranking predicted correctly and the opposing recruiting service ranking method did not predict correctly, and how many games that the opposing recruiting service ranking method predicted correctly and Crowd-Ranking did not predict correctly. This is important because many of the games will involve both ranking methods predicting either correctly or incorrectly, but we are concerned with when one method out-performs the other method. This approach is similar to Kvam and Sokol (2006) when comparing the predictability of several ranking methods. Also, when a ranking method has the same ranking for both teams (the event of a tie), we will not consider this to be a correct prediction.

Tables 7 – 11 show results for the Crowd-Ranking (CR) versus the Rivals, Scout, and 247Sports methods for all conference games played from 2007 – 2013.

ACC	Rivals vs CR	Scout vs CR	247Sports vs CR
2007	3 vs 6	2 vs 7	4 vs 5
2008	2 vs 5	2 vs 5	2 vs 6
2009	4 vs 5	6 vs 5	5 vs 5
2010	6 vs 4	8 vs 5	8 vs 5
2011	3 vs 3	3 vs 3	2 vs 4
2012	5 vs 3	6 vs 3	5 vs 2
2013	5 vs 0	4 vs 0	3 vs 0
Total Score	28 vs 26	31 vs 28	29 vs 27
McNemar's Chi sq.	0.0185	0.0678	0.0179
p-value	**0.8918**	**0.7946**	**0.8936**

Table 7: ACC game results, 2007 – 2013

Big 10	Rivals vs CR	Scout vs CR	247Sports vs CR
2007	2 vs 3	1 vs 2	3 vs 7
2008	2 vs 2	1 vs 3	2 vs 6
2009	1 vs 4	0 vs 2	1 vs 4
2010	2 vs 3	1 vs 2	2 vs 3
2011	2 vs 3	2 vs 2	3 vs 2
2012	0 vs 2	0 vs 1	0 vs 2
2013	0 vs 4	0 vs 3	1 vs 1
Total Score	13 vs 26	5 vs 15	12 vs 25
McNemar's Chi sq.	3.6923	4.0500	3.8919
p-value	**0.0547**	**0.0442**	**0.0485**

Table 8: Big 10 game results, 2007 – 2013

Big 12	Rivals vs CR	Scout vs CR	247Sports vs CR
2007	1 vs 1	0 vs 1	2 vs 1
2008	0 vs 4	0 vs 1	0 vs 3
2009	2 vs 1	0 vs 0	3 vs 1
2010	2 vs 2	2 vs 0	2 vs 2
2011	2 vs 2	2 vs 0	2 vs 1
2012	2 vs 2	2 vs 2	2 vs 1
2013	3 vs 1	2 vs 1	3 vs 1
Total Score	12 vs 13	8 vs 5	14 vs 10
McNemar's Chi sq.	0	0.3077	0.3750
p-value	**1**	**0.5791**	**0.5403**

Table 9: Big 12 game results, 2007 – 2013

Pac 12	Rivals vs CR	Scout vs CR	247Sports vs CR
2007	5 vs 0	0 vs 1	2 vs 2
2008	2 vs 2	1 vs 3	2 vs 3
2009	1 vs 1	2 vs 1	2 vs 2
2010	1 vs 0	1 vs 0	2 vs 0
2011	0 vs 1	1 vs 0	0 vs 1
2012	1 vs 2	2 vs 1	1 vs 2
2013	1 vs 0	2 vs 0	2 vs 0
Total Score	11 vs 6	9 vs 6	11 vs 10
McNemar's Chi sq.	0.9412	0.2667	0
p-value	**0.3320**	**0.6056**	**1**

Table 10: Pac 12 game results, 2007 – 2013

SEC	Rivals vs CR	Scout vs CR	247Sports vs CR
2007	7 vs 3	7 vs 3	7 vs 4
2008	4 vs 7	4 vs 6	5 vs 8
2009	4 vs 3	3 vs 4	4 vs 3
2010	3 vs 2	3 vs 2	3 vs 2
2011	5 vs 2	7 vs 2	5 vs 3
2012	2 vs 2	2 vs 4	2 vs 2
2013	1 vs 4	3 vs 5	1 vs 5
Total Score	26 vs 23	29 vs 26	27 vs 27
McNemar's Chi sq.	0.0816	0.0727	0.0185
p-value	**0.7751**	**0.7874**	**0.8918**

Table 11: SEC game results, 2007 – 2013

The Crowd-Ranking method was more successful at predicting the winner in the Big Ten conference at a statistically significant level versus Rivals, Scout, and 247Sports. In fact, there was only one season where one of the three methods predicted more outcomes than Crowd-Ranking (247Sports in 2011). However, in the ACC, Big 12, Pac 12, and SEC conferences, there is not a statistically significant difference in any of the ranking methods versus Crowd-Ranking. We define statistically significant by the p-value obtained from McNemar's test. This is the same approach used by Kvam and Sokol (2006) when comparing ranking methods and their ability to predict the winner of a match.

In addition, Crowd-Ranking was strong at ranking the eventual Big 10 champion. Table 12 shows the Big Ten champions from 2007 – 2013, and the rankings given to those teams by both methods.

Year	Big Ten Champs	Scout	247Sports	Rivals	CR
2007	OSU	2	8	2	1
2008	OSU, PSU	2, 3	6, 2	2, 3	1, 3
2009	OSU	1	4	1	1
2010	OSU, MSU, Wisc.	1, 5, 4	3, 5, 6	1, 5, 8	1, 5, 5
2011	Wisc.	6	5	8	4
2012	Wisc.	7	8	8	5
2013	MSU	4	4	3	3

Table 12: Crowd-Ranking vs. Rivals to predict Big Ten champion(s), 2007 – 2013

Notice that the Crowd-Ranking method ranked the eventual Big Ten champion higher than all other three methods in 2007, 2011, and 2012. In 2009 and 2013, Crowd-Ranking ranked the eventual champion equally as high as at least one competing ranking, and no other methods ranked the champion higher. In 2008 and 2010, there were conference co-champions, but Crowd-Ranking generally performed as least as well as the other methods in ranking the eventual champion.

In summary, Crowd-Ranking was significantly better than Rivals, Scout, and 247Sports at predicting future team performance based on recruiting class rankings in the Big Ten conference.

In the ACC, Big 12, Pac 12 and SEC conferences, there is no significant difference in the predictive performance of Crowd-Ranking and the other ranking methods. Ultimately, we considered various factors but were unable to conclude a reason that the Crowd-Ranking method performed so well in the Big Ten conference in comparison to the other conferences.

5.3 Predictive power of Crowd-Ranking method, by national ranking

In this section, we examine the predictive power of the four recruiting ranking methods but by a *national* ranking. We will consider all teams in the power conferences with the addition of Notre Dame (65 teams). The process is identical as in Section 5.2 in that we compile recruiting rankings for Crowd-Ranking each year, obtain the composite team recruiting ranking, and then compare it to the other recruiting ranking methods.

National Rank	Rivals vs CR	Scout vs CR	247Sports vs CR
2007	26 vs 18	27 vs 15	23 vs 21
2008	19 vs 17	19 vs 19	22 vs 20
2009	19 vs 22	23 vs 22	22 vs 22
2010	23 vs 25	27 vs 26	20 vs 17
2011	26 vs 32	26 vs 26	17 vs 21
2012	21 vs 23	21 vs 24	22 vs 18
2013	13 vs 20	23 vs 18	16 vs 15
Total Score	147 vs 157	166 vs 150	142 vs 134
McNemar's Chi sq.	0.2664	0.7120	0.1775
p-value	**0.6057**	**0.3988**	**0.6735**

Table 12: Game results, 2007 – 2013

As seen in Table 12, there is no significant difference between Crowd-Ranking and the other three recruiting ranking services in their ability to predict game results. Although Crowd-Ranking performs slightly better than Rivals, and slightly worse than Scout and 247Sports, the p-values of their significances are relatively large and thus offer no reliable conclusion to the “better” method.

Upon analyzing the results from Crowd-Ranking, we noticed that the voting matrices for national data were relatively sparse, meaning that many teams in power conferences do not recruit as often against teams from other conferences. For example, a regional program from the Pac-12 may not recruit against a regional program from say, the ACC. The method must rely on the stronger teams to connect the clusters of conference teams. For example, USC from the Pac-12 and Florida St. from the ACC are both national powers in recruiting and may recruit against each other for the top national recruits.

For a Markov-based ranking method to be effective there should be strong connectivity amongst the nodes, and in this application with national data that does not appear to be the case. However, with this shortcoming, Crowd-Ranking still performed no worse than the other three methods in predicting future games. In future applications, we recommend that users be cautious when using Crowd-Ranking with a weakly connected Markov graph.

5.4 Potential Biases in the Data

In this study, the accuracy of the reported scholarship offers of a specific recruit is a primary input of the model. In turn, the results of the comparison above rely on the quality of data posted on the Rivals website regarding scholarship offers received by the recruits.

Also, occasionally a high-quality player accepts an offer early in the recruitment process. In this case, the player may not receive additional offers. This is not to say the player was not coveted by other teams, but just that the recruitment process ended before they could obtain more offers.

Last, for the rankings done by conference, teams that recruit nationally and out of their respective conference's region will be subject to bias because not as many conference teams will recruit those players. For example, a team on the outer footprint of the conference region may recruit heavily in that region and often against schools from another conference. Some quality players from that region might not receive offers due to their proximity from the conference footprint, and in turn, said team will not receive as many votes. This issue of course is not applicable to the national ranking, in which all teams factor in the analysis.

6. Conclusion

In conclusion, a new approach, the Crowd-Ranking method, was proposed for a special recruiting problem. The method reflects the decisions of two stakeholders, the groups and the individuals, and considers both quantity and quality when producing a ranking. This method can be applied to any traditional recruiting situation, in which a set of groups extend offers to multiple individuals, and the individuals can select one group to join. The method uses Markov chains to rank the groups based on the weighted preferences of the individuals.

Based on our application in Big Ten football recruiting, the Crowd-Ranking method performed better than Rivals, Scout, and 247Sports in both predicting future performances in Big Ten football games, and predicting the eventual Big Ten champion(s). In the ACC, Big 12, Pac 12 and SEC conferences, there was no statistical significance between Crowd-Ranking and other recruiting services' rankings in their predictive power. For our national ranking, there was no statistically significant difference between Crowd-Ranking and the other recruiting services in predicting the winner of games.

Future work could involve a deeper analysis as to why Crowd-Ranking performed significantly in the Big Ten conference. For example, is there a difference in recruiting strategy in that conference compared to others? For the national ranking, a future study could address the concern of weak connectivity in the Markov graph. We cannot force teams to recruit more players out of their conference region, but we could modify the method to address the shortcoming. Last, future work could involve comparisons against other popular ranking methods, such as Sagarin, national polls (ESPN, AP, Harris), Massey (1997), and Colley (2002). These comparisons can show us how much information is attributed to the collective wisdom of coaches. However, the recruiting class ranking should only be a portion of the study, because there are many other factors included in national polls – such as, strength of coaches, returning players' experience, and injuries.

Acknowledgements

The authors would like to thank Thomas Hagebols for his contribution in gathering the data for the national ranking. The authors would also like to thank anonymous referees for helping to improve the quality of the manuscript.

References

Balinski M, Laraki R (2014) Judge: Don’t Vote!. Operations Research 62(3): 483-511.

Brin S, Page L (1998) The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems 33: 107-117

Callaghan T, Mucha PJ, Porter M A (2007) Random walker ranking for NCAA division I-A football. The American Mathematical Monthly 114(9): 761-777

Chartier TP, Kreutzer E, Langville AN, Pedings KE (2011) Sensitivity and stability of ranking vectors. SIAM Journal of Scientific Computing 33(3): 1077-1102

Colley WN (2002) Colley’s bias free college football ranking method: The Colley matrix explained. www.colleyrankings.com/matrate.pdf

David HA (1970) Ranking the players in round robin tournament. Institute of Statistics Mimeo Series No. 718.

Elo AE (1961) The New U.S.C.F. Rating System. Chess Life 16: 160-161

Govan AY (2008) Ranking Theory with Application to Popular Sports. Dissertation, North Carolina State University.

Herm S, Callsen-Bracker H, Kreis H (2014) When the crowd evaluates soccer players’ market values: Accuracy and evaluation attributes of an online community. Sports Management Review 17(4): 484-492

Keener JP (1993) The Perron-Frobenius theorem and the ranking of football teams. SIAM Review 35(1): 80-93

Kendall MG, Smith BB (1940) On the method of paired comparisons. Biometrika 31: 324-345

Kvam P, Sokol JS (2006) A logistic regression/Markov chain model for NCAA basketball. Naval Research Logistics 53: 788-803

Langville AN, Meyer CD (2006) Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press, Princeton

Langville AN, Meyer CD (2012) Who's #1? The Science of Rating and Ranking. Princeton University Press, Princeton

Massey K (1997) Statistical models applied to the rating of sports teams. Bachelor's Thesis, Bluefield College.

Rivals. Retrieved January 2017 from website, http://www.rivals.com

Saaty TL (1980) The Analytic Hierarchy Process. McGraw-Hill, New York

Saaty TL (1986) Axiomatic Foundation of the Analytic Hierarchy Process. INFORMS 32(7): 841-855

Saaty TL (1987). Rank according to Perron: A new insight. Mathematics Magazine 60(4): 211-213

Scout. Retrieved January 2017 from website, http://www.scout.com

247Sports. Retrieved January 2017 from website, http://www.247sports.com