Journal of Quantitative Analysis in Sports

Volume 5, *Issue* 1 | 2009 | *Article* 4

Offense-Defense Approach to Ranking Team Sports

Anjela Y. Govan* Amy N. Langville† Carl D. Meyer‡

*North Carolina State University, anjela.govan@gmail.com
†College of Charleston, langvillea@cofc.edu
‡North Carolina State University, meyer@math.ncsu.edu

Offense-Defense Approach to Ranking Team Sports*

Anjela Y. Govan, Amy N. Langville, and Carl D. Meyer

Abstract

The rank of an object is its relative importance to the other objects in the set. Often a rank is an integer assigned from the set 1,...,n. A ranking model is a method of determining a way in which the ranks are assigned. Usually a ranking model uses information available on the objects to determine their respective ratings. The most recognized application of ranking is the competitive sports. Numerous ranking models have been created over the years to compute the team ratings for various sports. In this paper we propose a flexible, easily coded, fast, iterative approach we call the Offense-Defense Model (ODM), to generating team ratings. The convergence of the ODM is grounded in the theory of matrix balancing.

*Our special thanks go to Luke Ingram who has worked on developing ODM in his Masters thesis with Amy Langville.

1 Introduction

The *rank* of an object is its relative importance to the other objects in the set. Often a rank is an integer assigned from the set $\{1, 2, ..., n\}$. Ideally an assignment of available ranks ($\{1, 2, ..., n\}$) to n objects is one-to-one. However in certain circumstances it is possible that more than one object is assigned the same rank. A *ranking model* is a method of determining a way in which the ranks are assigned. Typically a ranking model uses information available to determine a *rating* for each object. The ratings carry more information than the ranks because they provide us with the degree of relative importance of each object. Once we have the ratings the assignment of the ranks can be as simple as sorting the objects in the descending order of the corresponding ratings. The ranking models can be used for a number of applications such as sports, web search, literature search, etc. This paper concentrates on ranking teams (or players) in sports that have paired comparisons (games) that produce scores for each team. The first section of this article introduces the Offense-Defense Model (ODM).

A natural application of sports ranking models is predictions of game outcomes. The game prediction in the second section of this article were done using four models, the Offense-Defense Model, the Colley Matrix Method by Colley (2002), the Keener Perron vector approach by Keener (1993), and the Massey Least Squares model by Massey (1997). Both the method by Colley and the Massey's model are part of the BCS rankings. Following are brief introductions to the last three methods.

The Colley Matrix Method uses only the number of wins and losses and the number of games played for each team, assuming no ties. The essence of the Colley Matrix Method is to formulate and solve a system of linear equations $\mathbf{Cr} = \mathbf{b}$ to obtain the rating of each team. Each equation in the system corresponds to a team and the simultaneous solution to the system provides the rating score for each team. The algorithm is based on a result from probability called Laplace's Rule of Succession Ross (2006, p.108). The rule is used to approximate probabilities of boolean events (in our case, the probability of winning or losing a game). The existence and uniqueness of the solution is based on a theory of a special type of nonnegative matrices, M-matrices.

James P. Keener proposed his ranking method based on the theory of nonnegative matrices in 1993. Specifically, Keener makes use of properties of the Perron vector guaranteed by the Perron-Frobenius Theorem, see Meyer (2001, p.673). The nonnegative score matrix $\mathbf{K}$ is formed using Laplace's Rule of Succession and a smoothing function. The substance of this ranking algorithm is in determining an eigenvector of the matrix $\mathbf{K}$ corresponding to the dominant eigenvalue of $\mathbf{K}$. Other examples of ranking models that use the Perron-Frobenius Theorem are found in publications by Kleinberg (1999) and Saaty (1987).

Kenneth Massey created a least squares approach to computing ratings. Computing team ratings amounts to solving a least squares problem $\mathbf{L}^T\mathbf{L}\mathbf{x} = \mathbf{L}^T\mathbf{b}$ where $\mathbf{L}$ is a game-by-team matrix and each (i, j) indicates whether team j won (set to be 1), lost (set to be -1) or did not play (set to be 0). Vector $\mathbf{b}$ is formed by using score differences of the corresponding games. Another ranking model using a least squares approach can be found in Stefani and Clarke (1992).

There are numerous other ranking models that fit into many different categories. One type is the Markov Chains based ranking models for example Govan et al. (2008), Kvam and Sokol (2006), Redmond (2003), and Brin and Page (1998). Another is the Sinkhorn-Knopp based ranking models such as one by Smith (2005) and ODM. Yet another type uses logistic regression, for examples see Clarke and Dyte(2000) or Holder and Nevill (1997).

2 Offense-Defense Model

For a set of teams engaged in a competitive sport, the first objective of this article is to present a model for rating the overall strength of each team relative to the others. While there are numerous factors that might be taken into account, our approach is to characterize "strength" by combining each team's relative offensive and defensive prowess in a non-linear fashion.

To compute offensive and defensive ratings we start with the assumption that larger offensive ratings correspond to greater offensive strength, i.e. the capability of producing larger scores. On the other hand, smaller defensive ratings will correspond to greater defensive strength, i.e., a low defensive rating indicates that it is hard for the opposition to run up a large score.

If teams i and j compete, then let $\mathbf{A} = [a_{ij}]$ be such that a_{ij} is the score that team j generated against team i (set $a_{ij} = 0$ if the two teams did not play each other). Alternately, a_{ij} can be thought of as the number of points that team i held team j to. In other words, depending on how it is viewed, a_{ij} simultaneously reflects relative offensive and defensive capability. To utilize this feature we define the *offensive rating* of team j to be the combination

$$o_j = a_{1j}(1/d_1) + ... + a_{nj}(1/d_n),$$

where d_i is the *defensive rating* of team i that is defined to be

$$d_i = a_{i1}(1/o_1) + ... + a_{in}(1/o_n).$$

Since o_j's and d_i's are interdependent, these values will have to be determined by a successive refinement technique that is described below. For example, consider a league consisting of three teams, called team 1, team 2, and team 3. Suppose that initially team 1 has the strongest offense and defense, team 2 has the next strongest offense and defense, and team 3 has the worst offense and defense. Then $o_1 \geq o_2 \geq o_3$, and $d_1 \leq d_2 \leq d_3$. Suppose that some initial estimates of these quantities are made (or guessed). Given that each team played every other team exactly once let us examine the refined offensive and defensive ratings for team 3. The new offensive rating for team 3 is

$$o_3 = a_{13}(1/d_1) + a_{23}(1/d_2).$$

Since team 1 has strong defense then d_1 is relatively small and therefore team 3's offense will be rewarded for scoring high against a strong opponent, i.e., points a_{13} produced against team 1 will be divided by a smaller number thus boosting the offensive rating of team 3.

The new defensive rating of team 3 is

$$d_3 = a_{31}(1/o_1) + a_{32}(1/o_2).$$

Team 1 has the strongest offense, so the defense of team 3 is less penalized for allowing team 1 to score more points. On the other hand, team 2's offense is less impressive, and therefore defense 3 must hold offense 2 to fewer points to avoid an increase in d_3's rating.

This intuition leads to the following general rule for successive refinement. Given $\mathbf{A} = [a_{ij}]$, initialize $d_i = 1$ for all i so that $\mathbf{d}^{(0)} = (\ 1 \ \ 1 \ \ \dots \ \ 1\ )^T$. Define $\mathbf{o}^{(1)} = \mathbf{A}^T(1/\mathbf{d}^{(0)})$, where $1/\mathbf{d}^{(0)}$ denotes the column vector $1/\mathbf{d}^{(0)} = (\ 1/d_1 \ \ \dots \ \ 1/d_n\ )^T$. Now successively refine the defensive and offensive values by the following iterative procedure

$$\begin{aligned} \mathbf{o}^{(k)} &= \mathbf{A}^T \frac{1}{\mathbf{d}^{(k-1)}}, \\ \mathbf{d}^{(k)} &= \mathbf{A} \frac{1}{\mathbf{o}^{(k)}}. \end{aligned} \tag{1}$$

2.1 Convergence of Offense-Defense model

The equations (1) for computing the offensive and defensive ratings are in fact equivalent to a row-column scaling of the matrix $\mathbf{A}$. That is, provided that the iterative procedure converges, the entries of $1/\mathbf{o}$ normalize columns of $\mathbf{A}$ and entries of $1/\mathbf{d}$ normalize rows of $\mathbf{A}$ so that $\mathbf{A}(1/\mathbf{o}) = \mathbf{e}$ and $\mathbf{A}^T(1/\mathbf{d}) = \mathbf{e}$, where $\mathbf{e}$ is a vector of all 1's. In other words, $\mathfrak{D}(1/\mathbf{d})\mathbf{A}\mathfrak{D}(1/\mathbf{o})$ is a doubly stochastic matrix where $\mathfrak{D}(\mathbf{x})$ denotes the diagonal matrix

$$\mathfrak{D}(\mathbf{x}) = \begin{pmatrix} x_1 & & & \\ & x_2 & & \\ & & \ddots & \\ & & & x_n \end{pmatrix}.$$

Scaling of a square non negative matrix $\mathbf{A}$ so that row and column sums of $\mathbf{A}$ are 1 is a special case of matrix balancing. In a general case the rows and columns of $\mathbf{A}$ do not have to add to the same scalar, see Schneider and Zenios (1994) for details.

The convergence of the Offense-Defense model is guaranteed by the Sinkhorn-Knopp Theorem Sinkhorn and Knopp (1967). In order for the set of equations (1) to converge, the matrix $\mathbf{A}$ has to have *total support*. A nonnegative square matrix $\mathbf{A}$ is said to have total support if $\mathbf{A} \neq 0$ and if every positive element of $\mathbf{A}$ lies on a positive diagonal. A diagonal of the matrix $\mathbf{A}$ is a set of elements $\{a_{1,\sigma(1)}, ..., a_{n,\sigma(n)}\}$, where σ is a permutation of $\{1, ..., n\}$.

In 1967 Sinkhorn proved the convergence of the matrix scaling method given that matrix $\mathbf{A}$ is positive. An alternative proof was given in 1998 by Borobia (1998). The following theorem states the conditions for convergence of the scaling method when $\mathbf{A}$ is nonnegative.

Theorem 2.1 *(Sinkhorn-Knopp, 1967) For each nonnegative matrix* $\mathbf{A}$ *with total support there exists a unique doubly stochastic matrix* $\mathbf{S}$ *of the form* $\mathbf{DAE}$ *where* $\mathbf{D}$ *and* $\mathbf{E}$ *are unique (up to a scalar multiple) diagonal matrices with positive main diagonal. Matrices* $\mathbf{D}$ *and* $\mathbf{E}$ *are obtained by alternatively normalizing columns and rows of* $\mathbf{A}$ *using the 1-norm.*

If a matrix $\mathbf{A}$ *has support (has at least one positive diagonal), then the iterative procedure of alternatively normalizing columns and rows of* $\mathbf{A}$ *converges. However the sequences of the diagonal matrices produced for this normalization do not necessarily converge.*

The matrix scaling (see Schneider and Zenios, 1994, Kalantari et al., 1993, and Rothblum et al., 1994) procedure in the Sinkhorn-Knopp theorem is one of the methods for matrix balancing and in some papers is referred to as a $\mathbf{RAS}$ method. Matrix scaling and its convergence has received increased attention in the past several decades. Rates of convergence Knight (2008), Kalantari et al. (1997), and Soules (1991), algorithms Knight and Ruiz (2007), Ruiz (2001), and multidimensional scaling Franklin and Lorenz (1989) are just a few related developing areas. We make use of the Sinkhorn-Knopp convergence result for the nonnegative matrices in the following theorem.

Theorem 2.2 *Given that the score matrix* $\mathbf{A}$ *has total support, the entries in the vectors* $\mathbf{o}$ *and* $\mathbf{d}$ *are the reciprocals of the main diagonal elements of the matrices* $\mathbf{D}$ *and* $\mathbf{E}$*.*

Proof. The iterative method used to obtain matrices $\mathbf{D}$ and $\mathbf{E}$ using $\mathbf{D}_0 = \mathbf{E}_0 = \mathbf{I}$ is

$$\begin{aligned} \mathbf{c}_k^T &= \mathbf{e}^T \mathbf{D}_{k-1} \mathbf{A} \mathbf{E}_{k-1}, \\ \mathbf{r}_k &= \mathbf{D}_{k-1} \mathbf{A} \mathbf{E}_k \mathbf{e}, \end{aligned}$$

where $\mathbf{D}_k = [\mathfrak{D}(\mathbf{r}_k)]^{-1}$, and $\mathbf{E}_k = [\mathfrak{D}(\mathbf{c}_k)]^{-1}$. Given that matrix $\mathbf{A}$ has total support, this iteration converges, and the final matrices $\mathbf{D} = [\mathfrak{D}(\mathbf{r})]^{-1}$ and $\mathbf{E} = [\mathfrak{D}(\mathbf{c})]^{-1}$ are unique. Consequently $\mathbf{S} = \mathbf{DAE} = [\mathfrak{D}(\mathbf{r})]^{-1}\mathbf{A}[\mathfrak{D}(\mathbf{c})]^{-1}$, where $\mathbf{S}$ is doubly stochastic. Using $\mathfrak{D}(\mathbf{x})\mathbf{e} = \mathbf{x}$ and $\mathbf{e}^T\mathfrak{D}(\mathbf{x}) = \mathbf{x}^T$ with $\mathbf{Se} = \mathbf{e}$ and $\mathbf{e}^T\mathbf{S} = \mathbf{e}^T$ produces

$$\begin{aligned} \mathbf{c} &= \mathbf{A}^T [\mathfrak{D}(\mathbf{r})]^{-1} \mathbf{e}, \\ \mathbf{r} &= \mathbf{A} [\mathfrak{D}(\mathbf{c})]^{-1} \mathbf{e}. \end{aligned}$$

By setting $\mathbf{c} = \mathbf{o}$, $\mathbf{r} = \mathbf{d}$, and noting that the inverse of a diagonal matrix with positive diagonal is obtained by taking the element-wise reciprocals of the diagonal entries, the above equation set is rewritten as

$$\mathbf{o} = \mathbf{A}^T \frac{1}{\mathbf{d}}, \qquad \mathbf{d} = \mathbf{A}\frac{1}{\mathbf{o}},$$

which is the limit of the original definition of the offense-defense (1) model with $\mathbf{A}$ having total support. Thus the theorem is proven.

In practice there is no guarantee that the score matrix $\mathbf{A}$ will have total support. To compensate, a rank-one perturbation can be incorporated to define a new matrix

$$\mathbf{P} = \mathbf{A} + \epsilon \mathbf{e}\mathbf{e}^T.$$

This is used in place of $\mathbf{A}$ so that convergence of (1) is ensured.

This perturbation adds ϵ to all elements of $\mathbf{A}$, and if ϵ is sufficiently small, then its effect on the model should not be significant. However, adding ϵ may affect convergence rates. The method is finalized below

Definition 2.3 *The Offense-Defense Model (ODM)*

The offensive ratings o_i and the defensive ratings d_i are entries of the vectors $\mathbf{o}$ and $\mathbf{d}$ that are limits of $\mathbf{o}^{(k)}$ and $\mathbf{d}^{(k)}$, as $k \to \infty$, in which

$$\begin{aligned} \mathbf{o}^{(k)} &= \mathbf{P}^T \frac{1}{\mathbf{d}^{(k-1)}}, \qquad \text{where } \mathbf{d}^{(0)} = \mathbf{e}, \\ \mathbf{d}^{(k)} &= \mathbf{P}\frac{1}{\mathbf{o}^{(k)}}, \end{aligned} \tag{2}$$

where

$$p_{ij} = \begin{cases} a_{ij} + \epsilon & \text{if team } i \text{ played team } j, \\ \epsilon & \text{team } i \text{ and } j \text{ did not play}, \end{cases}$$

and $\mathbf{e}$ is a vector of all 1's.

The second part of the Sinkhorn-Knopp Theorem states that if a matrix $\mathbf{A}$ has support, then the sequences of the diagonal matrices produced for the matrix normalization do not necessarily converge. Almost always matrices created using sports team data for a season will have support (this happens when each team plays at least once or until each team acquires a nonzero score), but it is not given that enough games played translates to total support of the matrix $\mathbf{A}$.

It is straightforward to check whether a matrix $\mathbf{A}$ has support — all row and column sums have to be positive. Checking for existence of total support in a large matrix is a nontrivial issue. One may check that each diagonal of $\mathbf{A}$ is either zero or positive.

Alternatively one can check whether there is a positive diagonal to which each nonzero element of $\mathbf{A}$ belongs, as each element of $\mathbf{A}$ belongs to multiple diagonals. This is the reason why it is computationally simpler to perturb each element of $\mathbf{A}$ to force total support rather than perturbing select elements.

Given that $\mathbf{A}$ has support, the matrix $[\mathfrak{D}(\mathbf{d}^{(k)})]^{-1}\mathbf{A}[\mathfrak{D}(\mathbf{o}^{(k)})]^{-1}$ will converge to a stochastic matrix as k tends to infinity, but the entries in $\mathbf{o}$ and $\mathbf{d}$ will either converge to zero or grow without bound. Some numerical testing suggests that the relative order of the entries in both $\mathbf{o}$ and $\mathbf{d}$ stabilize after only a few iterations. Since the ranking is assigned based on the relative order of the vector values, we might not have to perturb the matrix $\mathbf{A}$ after all.

2.2 Rank Aggregation

The Offense-Defense model assigns two rating scores, offensive and defensive, to each team. In most ranking applications the preference is to have a single rating score for each team. This can be accomplished by aggregating the corresponding offensive and defensive rating scores.

Rank aggregation is a function which uses several ratings (or ranks) obtained using various models as an input to produce a single rating (or rank) of each team as an output. The simplest aggregation function that can be applied to the Offense-Defense model is $r_i = o_i/d_i$, i.e., the overall rating score of team i is its offensive rating divided by its defensive rating. This allows us to retain the "large value is better" interpretation of the ratings. This rating score is a reciprocal of the Sinkhorn rating score proposed by Smith (2005).

Another aggregation method is called the "Rank Aggregation model." As an input it takes several ranked lists of n teams. We then form a directed graph where all the teams are represented by the graph nodes. The edges point from lower ranked teams to higher ranked teams and each weight can be defined as

$$w_{ij} = \text{number of ranked lists having } i \text{ below } j$$

or

$$w_{ij} = \text{sum of rank differences from lists having } i \text{ below } j$$

Then we can apply our favorite ranking method to compute the ratings and form the aggregated ranked list. In essence each ranked list can represent a round robin tournament where the rank of the team is a score and the team with the lowest score wins. In general the goal of rank aggregation is to minimize the effect of the outliers, i.e., the teams that are assigned a rank by one method that differs significantly from the ranks assigned by the majority of the other methods.

2.3 Score Matrix

It is not a certainty that a better team will always accumulate the highest score during a given game. Hence we can take into account other statistics as alternative strength indicators of a given team. For example, in football each team accumulates total yardage in each game. We can form a score matrix $\mathbf{A}$ using the total yards statistic and compute the offensive and defensive strength of a team according to this statistic. One can compute multiple offensive and defensive ratings of a given team based on several statistics (e.g. game scores, total rushing yards, total first downs, etc.) and aggregate them into a single offensive and a single defensive scores.

2.4 Hyperlink-Induced Topic Search model (HITS)

The Offense-Defense model is in part inspired by the ranking algorithm HITS created by Jon Kleinberg in 1998 for web ranking Kleinberg (1999). The origins of web-page ranking lie in ranking methods for scientific journals via citations Pinski and Narin (1976), Geller (1978). According to HITS a web page P_i has two rating scores associated with it, an authority, a_i, and a hub h_i. The world wide web is represented as a directed graph with web pages being nodes and hyperlinks as directed edges. The rating scores are computed as

$$\begin{aligned} a_i &= l_{1i}h_1 + ... + l_{ni}h_n, \\ h_i &= l_{i1}a_1 + ... + l_{in}a_n. \end{aligned}$$

where $l_{ij} = 1$ if P_i has a hyperlink to P_j and $l_{ij} = 0$ otherwise. The process of computing authority and hubs ratings is again iterative and can be written using the adjacency matrix $\mathbf{L}$ of the directed web graph

$$\begin{aligned} \mathbf{a}^{(k)} &= \mathbf{L}^T\mathbf{h}^{(k-1)}, \\ \mathbf{h}^{(k)} &= \mathbf{L}\mathbf{a}^{(k)}, \end{aligned}$$

which simplifies to

$$\begin{aligned} \mathbf{a}^{(k)} &= \mathbf{L}^T\mathbf{L}\mathbf{a}^{(k-1)}, \\ \mathbf{h}^{(k)} &= \mathbf{L}\mathbf{L}^T\mathbf{h}^{(k-1)}. \end{aligned} \tag{3}$$

Analogous to the Offence-Defense model, HITS assumes each object has two ratings. However, unlike the offense and defense ratings both authority and hub values are "good" when they are high. Hence there are no reciprocals in the computation, and, providing that the iterative procedure (3) converges, vectors $\mathbf{a}$ and $\mathbf{h}$ are eigenvectors of $\mathbf{L}^T\mathbf{L}$ and $\mathbf{L}\mathbf{L}^T$ respectively. In other words HITS is a linear model while our ODM formulation is non-linear.

3 Game Prediction

3.1 Data Gathering

Now we proceed to the second objective of this article which is a validation of the proposed Offense-Defense Model. One way to validate a ranking model is by means of game predictions. This could amount to periodically computing the overall team ratings and use them to predict the outcomes of the upcoming games. The process could be as simple as comparing current team ratings and assigning the team with the higher rating as the winner of an upcoming match. The predictions are then compared to the actual results. We will refer to this approach as "foresight predictions." A somewhat different way of game prediction is to use the entire season of data to compute the ratings score, and then use these rating scores to predict the outcomes of the games of that season. In other words, if we have all the information available about a given season, what is the maximum accuracy that we can achieve? This approach will be referred to as "hindsight predictions."

The complexity of the rating model may require a substantial effort in data gathering and processing. If the model uses only win-loss counts and possibly scores of the individual games played, then the data is relatively easy to acquire. Many sports websites contain lists of NFL or National Collegiate Athletic Association (NCAA) games along with scores for a number of seasons, see for example http://www.mratings.com/.

If there is a need for individual game statistics such as rushing or passing yards, then the data gathering process may involve parsing data files of the game box score such as ones found on the John M. Troan and ESPN websites (see references). Two main attributes are required from the potential data website sources. One is the reliability of the information, and the second is the stability of the file format used to showcase the information. In the case of our game prediction experiments we chose to use game scores to generate the offensive and defensive team ratings. All the data files were downloaded and parsed, using Perl software, from two sources; the John M. Troan website for the NFL and the ESPN website for NCAA football and basketball.

3.2 Game Prediction Results

In this section we present the results of the game predictions for three sports; football (NFL), college football (NCAA football), and college basketball (NCAA basketball). Different team sports present different challenges to game predictions. For example, the NFL tends to have a very regular schedule with a sufficient interplay within divisions as well as without. The NFL currently consists of only 32 teams that are relatively close to each other in "quality." This makes for a narrow difference in the ratings values. A small number of teams corresponds to a small size matrix and hence fast Matlab computations for the game predictions. In contrast, NCAA football currently has 120 Division I-A teams, and the team quality ranges wildly. The game data for the NCAA is very sparse since there are relatively few games played in comparison to the number of teams. There is a small interplay between divisions, and Division I-A teams tend to play I-AA teams during the beginning of the season. Finally, NCAA basketball currently has 341 teams in its Division I. On average each Division I team tends to play about 30 games per season, and although during the 2007-2008 season teams played more than 5000 games, the resulting matrices are sparse. There is also a great fluctuation in relative strength of the teams.

Both foresight and hindsight game prediction results are presented for the NFL. The first type recomputes the rating scores weekly and uses them to predict the winner of the upcoming games. These predictions for the NFL are done for the regular season games starting with week 1 through the Super Bowl. In order to ensure that the models have enough game data the pre-season game results are taken into account during the first three weeks of the regular season. After three weeks enough games have been played, so the pre-season data is henceforth discarded. In case two teams play each other multiple times, the scores produced by each team are accumulated, although another approach would be to average them. The number of games predicted correctly by each of the four models (Colley, Keener, Massey, and ODM) is converted into percentage since over the years the total number of games played by NFL teams has changed. The second type of game predictions is more theoretical in nature than practical because it employs hindsight. The ratings used to do game predictions are computed using the game outcomes for the entire season. Since different values of the ODM parameter ϵ did not produce significant changes in the predictions, all the results assume a tolerance $tol = 0.01$ and $\epsilon = 0.001$. This gives us an idea of just how well the ranking methods can do in an ideal case. There is a substantial increase in the accuracy of predictions as evidenced in Table 3.2 and Figure 3.2.

	Colley	Keener	Massey	ODM	Colley	Keener	Massey	ODM
2001	57.92	58.69	60.23	60.62	72.97	72.97	69.88	69.88
2002	59.18	58.43	60.3	63.3	68.16	68.91	67.04	68.54
2003	63.3	58.05	64.04	61.05	75.66	73.78	71.91	72.28
2004	61.8	59.93	62.17	58.43	74.16	70.79	67.42	68.54
2005	61.8	62.55	65.17	64.04	73.03	75.66	75.28	76.4
2006	58.8	57.68	60.3	58.05	72.66	69.29	71.16	70.04
2007	66.67	62.55	68.16	68.91	75.66	76.03	73.41	72.28

Table 1: Foresight/Hindsight game prediction percentages for the NFL.

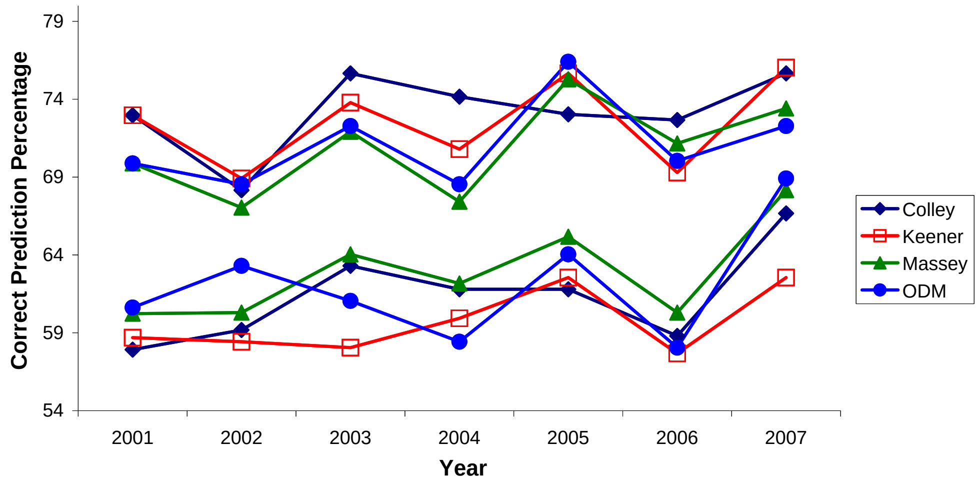

Figure 1: Foresight/Hindsight game prediction percentages for the NFL.

Both foresight and hindsight predictions are done for NCAA football. Foresight game predictions are done for the regular season of NCAA football starting with week 6 all the way through the Bowl games. Since NCAA football does not have a pre-season, the game predictions can not start until week 6 of the regular season. Only Division I-A teams are considered. The games played between I-A and I-AA are discarded. Both foresight and hindsight prediction results for NCAA football are in Table 3.2 and the following Figure 3.2.

	Colley	Keener	Massey	ODM	Colley	Keener	Massey	ODM
2003	66.3	70.29	69.62	69.18	82.04	76.72	77.38	76.05
2004	66.14	63.68	67.71	67.49	81.17	77.8	76.91	79.15
2005	67.34	64.21	67.79	64.43	81.66	77.63	76.06	74.27
2006	68.74	65.74	73.23	71.95	82.23	78.37	77.09	77.3
2007	67.1	68.82	69.89	68.6	79.35	76.77	77.42	75.48

Table 2: Foresight/Hindsight game prediction percentages for NCAA football.

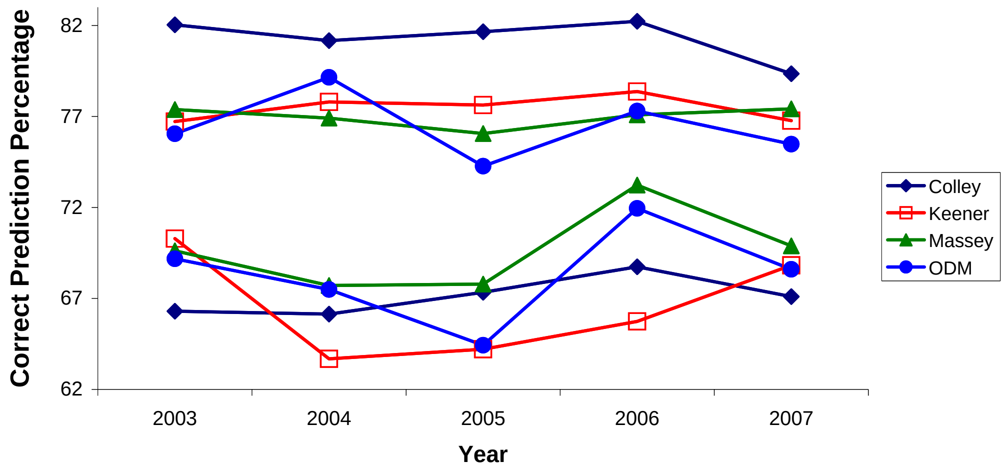

Figure 2: Foresight/Hindsight game prediction percentages for NCAA football.

The largest example of the game prediction in this paper is done with NCAA basketball. Since teams may play more than once in any one week period, the game predictions have to be done daily. As with previous sports, the total number of games played varies from year to year, so the predictions are converted to percentages. Only the games between the Division I teams were used. The games of the Division I teams with teams outside of the Division I were disregarded. Since there is no pre-season, foresight predictions wait until game day 26 to start the predictions and are predicted through the tournament. The viable starting point may be earlier for some seasons depending on the number of games discarded in the initial game days. The hindsight predictions used the entire season together with the tournament to compute the ratings, and then these ratings were used to predict games daily for the same season. Finally, given the tolerance $tol = 0.01$ for ODM convergence, the NCAA basketball foresight and hindsight game prediction results are shown in Table 3.2 and Figure 3.2.

	Colley	Keener	Massey	ODM	Colley	Keener	Massey	ODM
2001	68.60	64.60	69.65	70.03	76.00	70.03	74.40	74.11
2002	69.02	64.79	70.13	70.03	75.97	70.42	74.50	74.37
2003	68.92	64.92	70.19	70.22	76.90	70.13	75.83	75.66
2004	68.65	65.27	70.50	70.12	76.27	70.65	74.99	74.97
2005	66.98	64.44	68.95	69.56	75.99	69.32	74.80	74.57
2006	68.37	64.84	70.02	69.69	76.37	69.88	74.83	74.77
2007	68.28	64.91	70.13	70.07	76.05	69.82	74.92	74.83

Table 3: Foresight/Hindsight game prediction percentages for NCAA basketball.

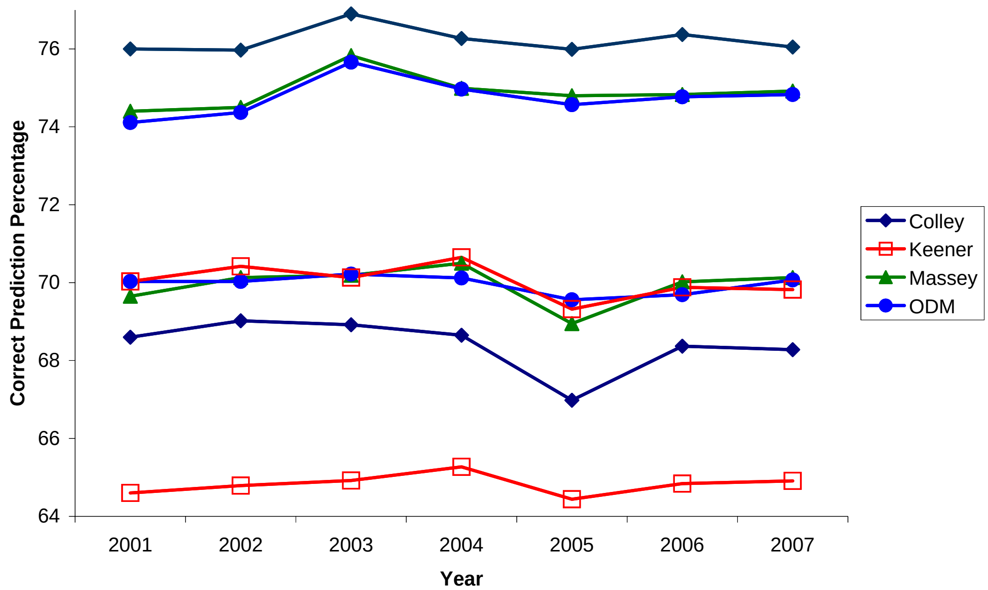

Figure 3: Foresight/Hindsight game prediction percentages for NCAA basketball.

There is a clear and expected increase in the prediction accuracy if we use the entire season's data to compute the ratings. Prediction results for all three sports yield an interesting observation that ODM does better as a foresight predictor relative to the other ranking models than it does as a hindsight predictor. Furthermore both Colley and Keener excel in hindsight but not in foresight predictions. Another point of interest is the computation time. Given that foresight predictions of NCAA basketball games is the largest of the examples used, it is subsequently used to compute the total cpu time expended and the number of iterations used by ODM. Provided that the tolerance for the ODM is set to be $tol = 0.01$, Table 3.2 illustrates the total cpu time expended for each of the methods. Table 3.2 considers the total number of iterations performed by ODM for different values of ϵ.

	Colley	Keener	Massey	ODM $\epsilon = 10^{-3}$	ODM $\epsilon = 10^{-4}$	ODM $\epsilon = 10^{-5}$
2001	1.78125	162.90625	2.0625	0.90625	0.953125	0.9375
2002	1.5625	161.234375	2.015625	0.84375	0.921875	1.1875
2003	1.671875	167.609375	2.125	1.03125	1.0625	1.125
2004	1.796875	170.046875	2.09375	1.03125	0.90625	0.875
2005	1.84375	174.96875	2.03125	1.109375	1.03125	1.140625
2006	1.96875	181.5	2.1875	0.90625	0.953125	1.03125
2007	2.03125	207.09375	2.453125	1.140625	1.1875	1.203125

Table 4: Total cpu time (sec) for each of the methods on NCAA basketball.

	$\epsilon = 0.001$	$\epsilon = 0.0001$	$\epsilon = 0.00001$
2001	1576	1577	1577
2002	1690	1866	2174
2003	1622	2000	2619
2004	1515	1605	1745
2005	1748	1997	2402
2006	1421	1465	1532
2007	1555	1638	1776

Table 5: Total number of iterations used by ODM for each of the seasons on NCAA basketball.

Finally, consider the Rank Aggregation model described in section 2.2. We use rank differences for the weights and the aggregated ratings are produced using the ODM. If we were to be concerned with game predictions only, then there is another simple aggregation approach that is worth mentioning. Given five ranked lists (Aggregated Rank, Colley, Keener, Massey, and ODM) we predict the game outcomes using each of the lists, and then use these prediction to form an overall prediction. If a majority of the ranked lists predicts team A to beat team B in a given game, then the overall prediction is team A wins. If there is a tie in predictions (two of the ranked lists predict A to win and one predicts a tie), then the user decides which method breaks the tie. In the following experiments the ODM is chosen to be the tie breaker. The foresight game prediction results for the aggregated methods as well as other four ranking models of NCAA basketball appears in Table 3.2 and the graph 3.2.

	Aggregated Predictions	Aggregated Rank	Colley	Keener	Massey	ODM
2001	69.49	68.95	68.60	64.60	69.65	70.03
2002	70.16	69.85	69.02	64.79	70.13	70.03
2003	70.05	69.57	68.92	64.92	70.19	70.22
2004	69.89	69.43	68.65	65.27	70.50	70.12
2005	68.73	67.93	66.98	64.44	68.95	69.56
2006	69.28	68.42	68.37	64.84	70.02	69.69
2007	69.76	69.51	68.28	64.91	70.13	70.07

Table 6: Foresight game prediction percentages for NCAA basketball.

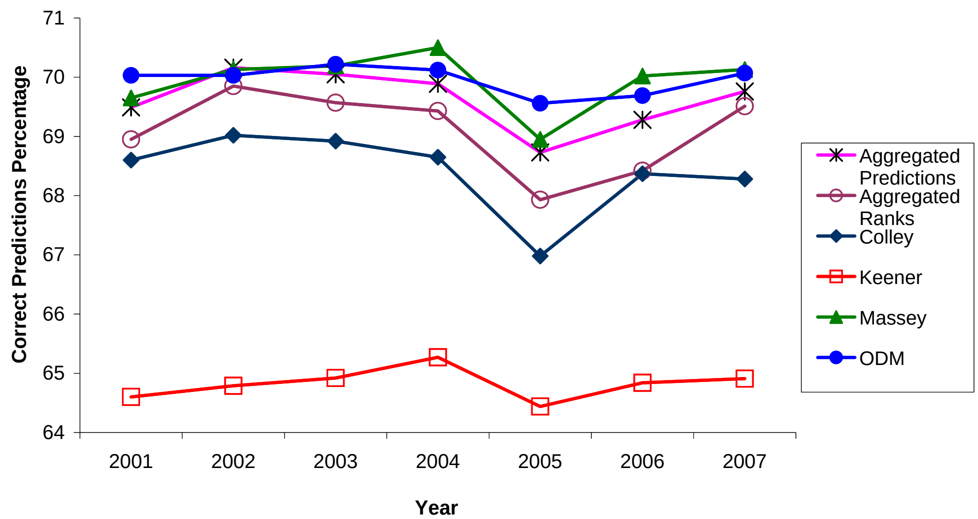

Figure 4: Foresight game prediction percentages for NCAA basketball.

As before the tolerance for ODM is set to be $tol = 0.01$, and predictions start with the game day 26 until the end of the NCAA basketball tournaments. As mentioned before, the goal of aggregation it to deemphasize the effect of the outliers, and thus it provides a greater level of confidence.

4 Conclusion

The Offense-Defense Method iteratively produces two ratings for team i; an offensive o_i and a defensive d_i. A simple way to combine them to produce an overall rating is $r_i = o_i/d_i$. The convergence of the method is guaranteed and is fast provided that the score matrix $\mathbf{A}$ has total support. The ODM can be used to make game predictions, and it tends to produce as good or better results as the leading sports ranking models in less cpu time. Uses of ranking methods, such as the ODM, are not limited to predicting outcomes of the games in competitive team sports. Other ranking applications (such as web page ranking) that can be described as weighted directed graphs can benefit from these models. Finally, the ODM lends itself to generalizations that can incorporate a variety of performance statistics, and it can be generalized to multiple dimensions beyond the two dimensions that have been discussed here.

References

[1] BOROBIA, A. Matrix scaling: A geometric proof of sinkhorn's theorem. *Linear Algebra and its Applications 268* (1998), 1–8.

[2] BRIN, S., AND PAGE, L. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems 33* (1998), 107–117.

[3] CLARKE, S. R., AND DYTE, D. Using official ratings to simulate major tennis tournaments. *International Transactions in Operational Research 7*, 6 (2000), 585–594.

[4] COLLEY, W. N. Colleys bias free college football ranking method: The colley matrix explained, 2002.

[5] COUR, T., SRINIVASAN, P., AND SHI, J. Balanced graph matching. In *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hofmann, Eds. MIT Press, Cambridge, MA, 2007.

[6] FRANKLIN, J., AND LORENZ, J. On the scaling of multidimensional matrices. *Linear Algebra and its Applications 114/115* (1989), 717–735.

[7] GELLER, N. L. On the citation influence methodology of pinski and narin. *Information Processing & Management 14* (1978), 93–95.

[8] GOVAN, A., MEYER, C. D., AND ALBRIGHT, R. Generalizing google's pagerank to rank national football league teams. In *Proceedings of the SAS Global Forum 2008* (2008).

[9] HOLDER, R. L., AND NEVILL, A. M. Modelling performance at international tennis and golf tournaments: Is there a home advantage? *The Statistician 46*, 4 (1997), 551–559.

[10] INGRAM, L. C. Ranking NCAA sports teams with Linear algebra. Master's thesis, College of Charleston, Charleston, SC 29424, Apr 2007.

[11] KALANTARI, B., AND KHACHIYAN, L. On the rate of convergence of deterministic and randomized ras matrix scaling algorithms. *Operations Research Letters 14* (1993), 237–244.

[12] KALANTARI, B., KHACHIYAN, L., AND SHOKOUFANDEH, A. On the complexity of matrix balancing. *Matrix Analysis and Applications 18*, 2 (1997), 450–463.

[13] KEENER, J. P. The perron-frobenius theorem and the ranking of football teams. *SIAM Review 35*, 1 (1993), 80–93.

[14] KLEINBERG, J. Authoritative sources in a hyperlink environment. *Journal of the ACM 46*, 5 (1999).

[15] KNIGHT, P. A. The sinkhorn–knopp algorithm: Convergence and applications. *SIAM Journal on Matrix Analysis and Applications 30*, 1 (2008), 261–275.

[16] KNIGHT, P. A., AND RUIZ, D. A fast algorithm for matrix balancing. In *Web Information Retrieval and Linear Algebra Algorithms* (2007).

[17] KVAM, P., AND SOKOL, J. S. A logistic regression/markov chain model for ncaa basketball. *Naval Research Logistics 53*, 8 (2006), 788–803.

[18] MASSEY, K. http://www.mratings.com/.

[19] MASSEY, K. Statistical models applied to the rating of sports teams, 1997.

[20] MEYER, C. D. *Matrix Analysis and Applied Linear Algebra*. SIAM, 2001.

[21] MEYER, C. D., AND LANGVILLE, A. N. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, 2006.

[22] NCAA. http://scores.espn.go.com/ncf/scoreboard.

[23] PAGE, L., BRIN, S., MOTWANI, R., AND WINOGRAD, T. The PageRank Citation Ranking: Bringing Order to the Web. Tech. rep., Department of Computer Science, Stanford University, Stanford, CA 94305, Jan 1998.

[24] PERFECT, H., AND MIRSKY, L. The distribution of positive elements in doubly-stochastic matrices. *Journal of the London Mathematical Society 40* (1965), 689–698.

[25] PINSKI, G., AND NARIN, F. Citation influence for journal aggregates of scientific publications: Theory, with applications to the literature of physics. *Information Processing & Management 12* (1976), 297–312.

[26] REDMOND, C. A natural generilazation of the win-loss rating system. *Mathematics Magazine 76*, 2 (2003), 119–126.

[27] ROSS, S. *A First Course in Probability*. Pearson Prentice Hall, 2006.

[28] ROTHBLUM, U. G., SCHNEIDER, H., AND SCNEIDER, M. H. Scaling matrices to prescribed row and column maxima. *SIAM Journal on Matrix Analysis and Applications 15*, 1 (1994), 1–14.

[29] RUIZ, D. A Scaling Algorithm to Equilibrate Both Rows and Columns Norms in Matrices. Tech. rep., Computational Science and Engineering Department, Science and Technology Facilities Council, Sep 2001.

[30] SAATY, T. L. Rank according to perron: A new insight. *Mathematics Magazine 60*, 4 (1987), 211–213.

[31] SCHNEIDER, M. H., AND ZENIOS, S. A. A comparative study of algorithms for matrix balancing. *Operations Research 38*, 3 (1994), 439–455.

[32] SINKHORN, R. Diagonal equivalence to matrices with prescribed row and column sums. *The American Mathematical Monthly 74*, 4 (1967), 402–405.

[33] SINKHORN, R. Continuous Dependence of A in the D1AD2 Theorems. In *Proceedings of the American Mathematical Society* (1972), vol. 32, pp. 395–398.

[34] SINKHORN, R., AND KNOPP, P. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal Of Mathematics. 21*, 2 (1967), 343–348.

[35] SMITH, W. D. Sinkhorn ratings, and new strongly polynomial time algorithms for sinkhorn balancing, perron eigenvector, and markov chains, 2005.

[36] SOULES, G. W. The rate of convergence of sinkhorn balancing. *Linear Algebra and its Applications 150* (1991), 3–40.

[37] STEFANI, R., AND CLARKE, S. Predictions and home advantage for australian rules football. *Journal of Applied Statistics 19* (1992), 251–261.

[38] TROAN, J. M. http://www.jt-sw.com/football/boxes/index.nsf.