SIAM REVIEW
Vol. 35, No. 1, pp. 80–93, March 1993

004

THE PERRON–FROBENIUS THEOREM AND THE RANKING OF FOOTBALL TEAMS*

JAMES P. KEENER†

Abstract. The author describes four different methods to rank teams in uneven paired competition and shows how each of these methods depends in some fundamental way on the Perron–Frobenius theorem.

Key words. Perron–Frobenius theorem, paired comparisions, ranking, orderings

AMS(MOS) subject classifications. 15-01, 15A48

1. Introduction. Throughout the fall of every year, arguments rage over which is the best college football team. The AP and UPI polls add to the confusion because they are based on votes which are certainly not objective. Many newspapers publish one or more additional indices that rank the top football teams, but these are not understood or accepted by the general public as easily as the polls, because they are usually based on "mathematical formulas." Given the general level of appreciation of mathematics among sports fans, these rankings are usually shrouded in mystery.

I first became interested in the problem of ranking football teams a few years ago when the football team at a rival campus won the national championship because it was the only undefeated team in the country. I wanted to know if a mathematically based ranking scheme would agree with the conclusions of the UPI and AP voters. What I learned (beyond what I hoped I would find!) is that a number of ranking schemes rely in some fundamental way on the Perron–Frobenius theorem, and that with the problem of ranking of teams in uneven paired competition I had discovered a marvelous way to motivate students to learn about a beautiful theorem that has in recent times fallen into relative obscurity.

An uneven paired competition is one in which the outcome of competition between pairs of teams (also called paired comparisons) is known, but the pairings are not evenly matched. That is, the competition is not a round robin in which each team is paired with every other team an equal number of times.

A good ranking scheme has a large number of potential uses. For example, it could be used to rank football teams, to create a tennis ladder, or to determine the research strength of mathematics departments. However, ranking schemes remove some, but not all, subjectivity, and different ranking schemes can give vastly different answers about who is number one, depending on the factors that are emphasized by the scheme.

This paper is about the ranking methods that I use. I use them not because they solve with certainty the problem of which team is number one, but because the mathematics is fun and well motivated. These methods are excellent vehicles by which to introduce students to interesting and important mathematical ideas, including the Perron–Frobenius theorem, the power and inverse power methods for finding eigenvalues of a matrix, and fixed point theorems for nonlinear maps. I find that the few minutes I spend each week during the fall collecting and entering data for my computer program are justified by the increased student interest in the mathematics of the methods generated by my weekly posted rankings. It is not difficult for students to write their own computer program to test some of these ideas on their favorite competition.

*Received by the editors January 2, 1992; accepted for publication (in revised form) August 25, 1992.
†Department of Mathematics, University of Utah, Salt Lake City, Utah 84112.

In this paper four different ranking schemes are described. The first, in §2, formulates the ranking problem as a linear eigenvalue problem and makes direct use of the Perron–Frobenius theorem. In §3, a nonlinear generalization of the first method is described. This method makes use of successive approximations to find a fixed point of a nonlinear map. The third and fourth methods attempt to assign a probability to the outcome of a contest, and make indirect use of the Perron–Frobenius theorem. Finally, in §6 we show the results of these four schemes when applied to the 1989 NCAA football schedule.

2. The direct method. The first method we describe is perhaps the most direct ranking method. To each participant in a contest we wish to assign a score that is based on the interactions with other participants. The assigned score should depend on both the outcome of the interaction and the strength of its opponents. If we suppose there is a vector of ranking values $\mathbf{r}$, with positive components r_j indicating the strength of the jth participant, then we define a score for participant i as

$$s_i = \frac{1}{n_i} \sum_{j=1}^{N} a_{ij} r_j, \tag{2.1}$$

where a_{ij} is some nonnegative number depending on the outcome of the game between participant i and participant j, N is the total number of participants in the competition, and n_i is the number of games played by participant i. The matrix A with entries a_{ij} is often called a preference matrix. For example, for football we could pick a_{ij} to be 1 if team i won the game, $\frac{1}{2}$ if the game ended in a tie, and zero otherwise. The division by n_i is to prevent teams from accumulating a large score by simply playing extra games.

Now we propose that the strength (or rank) of a participant should be proportional to its score, that is,

$$A\mathbf{r} = \lambda \mathbf{r}, \tag{2.2}$$

where A is the matrix with entries a_{ij}/n_i. In other words, the ranking vector $\mathbf{r}$ is a positive eigenvector of the positive matrix A.

The Perron–Frobenius theorem tells us when this problem has a solution, as follows.

THEOREM. *If the (nontrivial) matrix A has nonnegative entries, then there exists an eigenvector $\mathbf{r}$ with nonnegative entries, corresponding to a positive eigenvalue λ. Furthermore, if the matrix A is irreducible, the eigenvector $\mathbf{r}$ has strictly positive entries, is unique and simple, and the corresponding eigenvalue is the largest eigenvalue of A in absolute value (i.e., is equal to the spectral radius of A).*

To clarify the nomenclature, we refer to a vector with nonnegative entries as a nonnegative vector, and a vector with positive entries as a positive vector. We also introduce a partial order on the set of nonnegative vectors by saying that $\mathbf{p} > \mathbf{q}$ whenever $\mathbf{p} - \mathbf{q}$ is a positive vector and $\mathbf{p} \geq \mathbf{q}$ whenever $\mathbf{p} - \mathbf{q}$ is nonnegative.

The following are equivalent ways to describe an irreducible matrix.

(i) A is irreducible if for any two numbers i and j there is an integer $p \geq 0$ and a sequence of integers $k_1, k_2, \ldots, k_p$, so that the product $a_{ik_1} a_{k_1 k_2} \cdots a_{k_p j} \neq 0$.

(ii) A is irreducible if there is no permutation that transforms the matrix A into a block matrix of the form

$$\begin{matrix} A_{11} & A_{12} \\ 0 & A_{22} \end{matrix},$$

with A_{11} and A_{22} square matrices.

(iii) The nonnegative matrix A is irreducible if for any $\mathbf{r} \geq 0, A\mathbf{r} > 0$.

For paired competitions, if we take $a_{ij} = 0$ or 1 for a loss or a win, respectively, then the matrix A is irreducible if there is no partition of the teams into two sets S and T such that no team in S plays any team in T or every game between one team from S and one team from T resulted in a victory for the team in S. In particular, for this preference matrix to be irreducible, there can be no winless teams.

The proof of this theorem, while found in a number of older books, has not been included in most recent linear algebra books, so we include it in the appendix for completeness.

To calculate the eigenvector $\mathbf{r}$ we can use another powerful idea, namely, the power method [12], [10]. Since the ranking vector $\mathbf{r}$ is a simple eigenvector and corresponds to the largest eigenvalue of A, it follows that

$$\lim_{n\to\infty} \frac{A^n\mathbf{r}_0}{|A^n\mathbf{r}_0|} = \mathbf{r} \tag{2.3}$$

for any nonnegative vector $\mathbf{r}_0$.

Now comes the important question of how to pick the entries of the matrix A, and here there is room for subjectivity. We suggested earlier the choice $a_{ij} = 1$ if team i beat team j, $a_{ij} = \frac{1}{2}$ if team i and j tied, and $a_{ij} = 0$ otherwise. With this choice, if we guess an initial ranking vector $\mathbf{r}_0$ with all entries equal to one, then the ith component of $A\mathbf{r}_0$ is the winning percentage for team i. The ith component of vector $A^2\mathbf{r}_0$ is the average winning percentage of the teams that team i defeated. In some sense, $A^2\mathbf{r}_0$ contains information about the strength of schedule. I have heard it suggested by a nationally prominent football coach that $A^2\mathbf{r}_0$ should be used to determine a national champion. While this is a better ranking than the winning percentage $A\mathbf{r}_0$, it places a very high premium on strength of schedule. Of course, he did not express his scheme in mathematical notation, and, therefore, did not see the obvious generalization of using $A^n\mathbf{r}_0$ with large n. We now know that in the limit of n going to infinity, $A^n\mathbf{r}_0/|A^n\mathbf{r}_0|$ converges to the unique positive eigenvector of A, and this eigenvector gives a positive ranking for teams.

The idea of using the matrix A to find a ranking vector has been around for some time. Kendall and Babington Smith [6] considered the ranking $\mathbf{r} = A\mathbf{r}_0$, and the idea of powering the matrix A to find a ranking vector was initiated by Wei [13] and Kendall [5], and revisited often [1], [4], [9].

This simple choice for the entries a_{ij} leaves much to be desired. It is adequate for sports such as baseball where teams play each other often during a season. If teams play each other more than once, then a_{ij} is the total number of victories of team i over team j. With an increasing number of games, a_{ij} becomes a better indicator of the comparative strength of the two teams. But in football where teams play each other only once per season, there is information in the game score that is discarded when credit is given only for the win. For example, under this simple scheme, whether a score is nearly even or quite lopsided, all of the credit for the win goes to the winner. Also, a winless team has rank zero and, therefore, contributes nothing to the score of its opponents, and a matrix with a winless team is not irreducible. In fact, beating a winless team is more harmful than not playing that team at all because the winning team earns no points and its average point earning decreases.

A better method is to distribute the one point per game between two competing teams in a continuous, rather than discrete way. One way to assign a value to a_{ij} is to distribute the one point on the basis of the game score. If team i scored S_{ij} points and team j scored S_{ji} points in their encounter, we might award $a_{ij} = S_{ij}/(S_{ij} + S_{ji})$ points to team i. This is slightly unfair because in a close defensive game with final score 3–0, the winner takes all, even though the two teams were evenly matched. To prevent this, we might consider an award of $a_{ij} = (S_{ij} + 1)/(S_{ij} + S_{ji} + 2)$ to team i, for example.

With such a scheme there is another weakness, namely, for a good team to show its dominance and get an appropriate score for the win, it can show no mercy. To avoid having teams run up a score to improve their ranking, the one point could be distributed in a nonlinear way. For example, the choice

$$
\begin{aligned}
a_{ij} &= h\left(\frac{S_{ij}+1}{S_{ij}+S_{ji}+2}\right), \\
h(x) &= \tfrac{1}{2} + \tfrac{1}{2}\mathrm{sgn}\left(x - \tfrac{1}{2}\right)\sqrt{|2x-1|}
\end{aligned}
\tag{2.4}
$$

has the features that it is continuous, $h(\frac{1}{2}) = \frac{1}{2}$, and away from $x = \frac{1}{2}$, h goes rapidly to zero or 1. A sketch of $h(x)$ is shown in Fig. 1. With an award distribution as in (2.4), to obtain a good score it is important to win, but not as useful to run up the score.

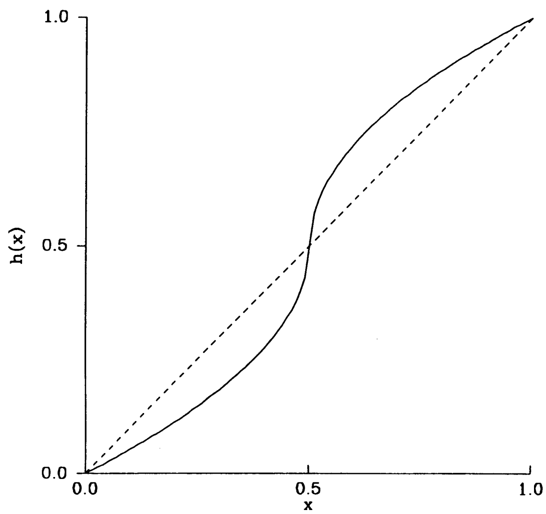

FIG. 1. *Plot of $h(x)$ as a function of x(solid curve) and the line $y(x) = x$(dashed) shown for comparison.*

3. A nonlinear scheme. Although the Perron–Frobenius scheme seems well motivated, after examining the results for a number of years its weaknesses became apparent. (By weakness, I mean that coaches and fans object to certain features of the ranking, not that there is a mathematical deficiency.) With this method, strength of schedule is quite important. If a strong team plays mostly weak opponents, with few strong opponents, it cannot earn a high ranking. This is because a team can never earn enough points playing weaker opponents to increase its earned score. Of course, this is not all bad, since simply because a team is undefeated does not mean it should have the highest rank, particularly if it did not play a difficult schedule. We have found that there is often an enormous difference in difficulty of schedule between some of the top ranked football teams.

There are a number of ways to address this problem. We could use this scheme to determine a national champion anyway and hope that coaches will eventually come to understand that to earn a high ranking they cannot pad their team's schedule with weak opponents. This might also force some conferences with only one or two strong teams to consider realignment.

But another dilemma exists, and that is if a team does reasonably well against strong opponents, even though it may lose many or even most of its games, it can still earn a high ranking. For example, with the above linear method it is not unusual to find teams with losing records ranked among the top twenty-five teams. The reason for this is the decision implicit in the scheme to base ranking on a point system whereby one must earn points to improve one's rank. The teams that can optimize earning points are by definition the better teams. This may not be all bad, because some teams that are indeed very good nonetheless have losing records.

Since it is not likely that anytime in the near future coaches will be motivated by this ranking scheme to adjust their schedules, we decided to generalize this method to avoid the "problem" that a strong team with a weak schedule may be underrated. The idea is to calculate the rank for each team as

$$r_i = \frac{1}{n_i} \sum_{j=1}^{N} f(e_{ij} r_j), \tag{3.1}$$

where e_{ij} is a number that is determined from the outcome of the game between team i and team j, r_j is again the positive rank of team j, and f is some continuous monotone increasing function with $f(0) = 0$, and $f(\infty) = 1$. The advantage of this method is that now a team can earn up to a maximum of one point for each game it plays either by doing well against a highly ranked team, or by clobbering a poor team, but at least there is a way to have a weak schedule and still earn a good score.

We can again use interesting mathematics to conclude that a positive ranking vector $\mathbf{r}$ exists. If we define the nonlinear function of $\mathbf{r}$,

$$F_i(\mathbf{r}) = \frac{1}{n_i} \sum_{j=1}^{N} f(e_{ij} r_j), \tag{3.2}$$

then F is a bounded, nonlinear map of the positive orthant into itself. If we further suppose that $f(0) > 0$, and that $f(x)$ is a strictly concave function satisfying $f(tx) > tf(x)$ for all $t, 0 < t < 1$, then there is a unique fixed point of the map $F(\mathbf{r})$ in the positive orthant that can be found by successive approximation starting with any positive vector $\mathbf{r}_0$, whereby

$$\lim_{n \to \infty} F^n(\mathbf{r}_0) = \mathbf{r}. \tag{3.3}$$

The assumption $f(0) > 0$ implies that a team earns something just for showing up. Concavity is not strictly necessary to have a reasonable ranking, but it does guarantee a unique ranking vector. The proofs of these facts are relatively simple, and are relegated to the appendix. The equation $F(\mathbf{r}) = \mathbf{r}$ is a nonlinear eigenvector problem for which we seek a positive eigenvector, so this result can be viewed as a nonlinear generalization of the Perron–Frobenius theorem.

For this problem we found, after considerable experimentation, that

$$f(x) = \frac{.05x + x^2}{2 + .05x + x^2}, \tag{3.4}$$

and

$$e_{ij} = \frac{5 + S_{ij} + S_{ij}^{2/3}}{5 + S_{ji} + S_{ij}^{2/3}}, \tag{3.5}$$

work reasonably well. By reasonably well, we mean only that it gave results that aroused the ire of fewer people than did the linear method. A plot of $f(x)$ is shown in Fig. 2. The function $f(x)$ in (3.4) is not strictly concave, and neither is $f(0) > 0$, but it is close enough that the iterations (3.3) converge to a useful ranking vector. The function e_{ij} is shown plotted in Fig. 3 as a function of score S_{ij} for different values of S_{ji} fixed at $0, 10, 20, 30$, and 50. Note that $e_{ij} = 1$ when $S_{ij} = S_{ji}$, that e_{ij} is an increasing function of S_{ij} and a decreasing function of S_{ji}.

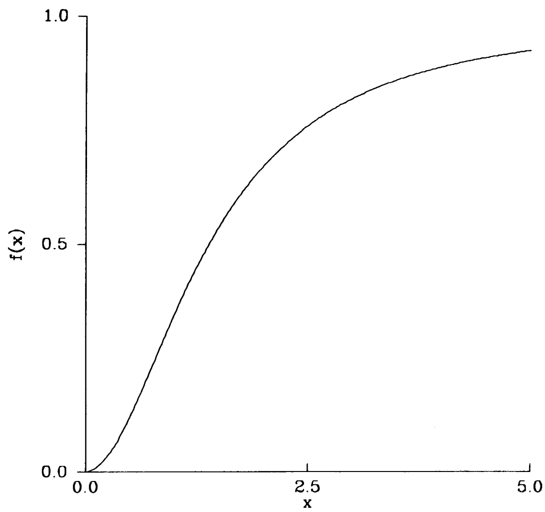

FIG. 2. *Plot of the function $f(x)$ as a function of x.*

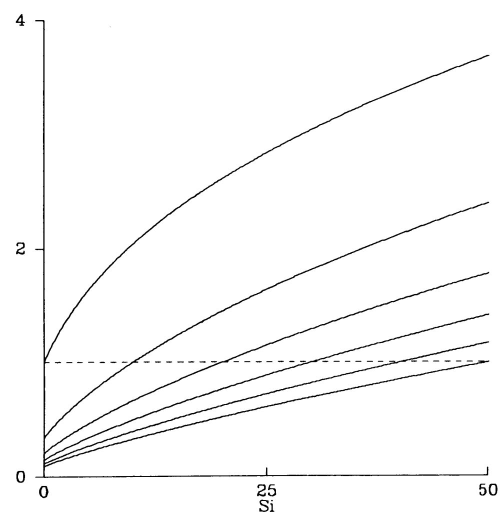

FIG. 3. *Plots of $e_{ij}(S_{ij}, S_{ji})$ as a function of S_{ij} with $S_{ji} = 0, 10, 20, 30, 40,$ and 50. Since $e_{ij}(x, x) = 1$, the value of S_{ji} can be identified by its intersection with the level 1 (dashed line).*

This choice for the map F in (3.2) points out the subjective nature of the methods described so far. The methods that follow are less subjective because they have an improved theoretical basis.

4. Assessing the probability of winning. Many people like to use ranking systems to predict the outcome of games between rivals, and so determining the probability that team i will beat team j is of primary interest. To this end, it would be nice if the ranking vector $\mathbf{r}$ could be given some probabilistic interpretation.

Suppose the ranking vector $\mathbf{r}$ is defined so that the probability π_{ij} that team i beats team j is

$$\pi_{ij} = \frac{r_i}{r_i + r_j}. \tag{4.1}$$

Since $\pi_{ij} + \pi_{ji} = 1$, it follows that

$$\pi_{ji} r_i - \pi_{ij} r_j = 0. \tag{4.2}$$

Unfortunately, we do not know π_{ij}, but if we did, we could find $\mathbf{r}$.

The relationship (4.2) between probability and the ranking vector is one of many possibilities in the class of so-called linear models having the form $\pi_{ij} = \prod(v_i - v_j)$, where $\mathbf{v}$ is the ranking vector [11]. The identification $r = e^v$ shows that (4.2) is a linear model. Other possibilities for the function $\prod$ are the Heaviside function, or

$$\prod(v) = \int_0^v e^{-x^2} dx. \tag{4.3}$$

The model (4.3) (due to Mosteller [8]) is motivated by the idea that the ith team has an actual performance that is a random variable with mean v_i and variance σ^2, σ being the same for each team. Then, if $\sigma = 1$, the probability π_{ij} that team i beats team j is

$$\pi_{ij} = \frac{1}{\sqrt{2\pi}} \prod \left(\frac{v_i - v_j}{\sqrt{2}} \right).$$

An interesting mathematical problem is to use statistical tests to determine the best linear model $\prod$. Bradley [2] gives a test of the hypothesis that the model (4.1) (known as the Zermelo model [14]) is correct.

If we use game scores to estimate π_{ij}, a reasonable estimate for π_{ij} is

$$\pi_{ij} = \frac{S_{ij}}{S_{ij} + S_{ji}}, \tag{4.4}$$

and (4.2) becomes

$$S_{ji} r_i - S_{ij} r_j = 0. \tag{4.5}$$

If teams i and j do not play each other, we take $S_{ij} = S_{ji} = 0$. Since, in any season there are many more games than there are teams, (4.5) gives many more equations than there are unknowns. Perhaps we can find the "best" solution of the overdetermined system (4.5) using a least squares method.

The least squares solution of all of the equations of the form (4.5) is trivial, $\mathbf{r} = \mathbf{0}$, and this is not the desired solution. Instead, we seek to minimize the squared error subject to the constraint that $\mathbf{r}$ has norm 1. Thus, using the Lagrangian multipliers, we seek to minimize

$$\sum_{ij}(S_{ji}r_i - S_{ij}r_j)^2 - \mu\left(\sum_{i=1}^{N} r_i^2 - 1\right). \tag{4.6}$$

After differentiating (4.6) with respect to $\mathbf{r}$, we find that a minimum occurs only if $\mathbf{r}$ satisfies the matrix equation

$$B\mathbf{r} = \mu\mathbf{r}, \tag{4.7}$$

where the matrix B has entries b_{ij} given by

$$b_{ii} = \sum_k S_{ik}^2, \quad b_{ij} = -S_{ij}S_{ji}, \quad i \neq j. \tag{4.8}$$

To understand the solution properties of (4.7), we notice some important properties of the matrix B. The matrix B is invertible whenever the columns of the matrix associated with (4.5) are linearly independent, and it is reasonable to assume that this occurs naturally with enough games. The matrix B has positive diagonal and nonpositive off diagonal entries.

For some number $\lambda_0 > 0$, the shifted matrix $B' = B + \lambda_0 I$ is diagonally dominant. Then, for the vector $\mathbf{r}_0$ with all entries equal to $1, B'\mathbf{r}_0$ has all positive entries. Now, notice what happens to the faces of the positive orthant under transformation by B'. If $\mathbf{r}_j$ has all entries positive except its jth entry which is zero, then the jth component of $B'\mathbf{r}_j$ is negative or zero. We will assume that there are enough entries in the matrix B so that the jth component of $B'\mathbf{r}_j$ is strictly negative, and then none of the faces of the positive orthant are invariant. In other words, the boundary of the positive orthant is mapped by the matrix B' to the exterior of the positive orthant. Since there is at least one vector, namely $\mathbf{r}_0$, that maps from the positive orthant into the positive orthant, it follows that B' maps the positive orthant to a cover of the positive orthant. Necessarily, B'^{-1} maps the positive orthant *into* the positive orthant and is therefore a positive map, meaning that its nonzero entries are positive. We conclude from the Perron–Frobenius theorem that B'^{-1} has a positive eigenvector $\mathbf{r}$, and that its corresponding eigenvalue is the largest in absolute value of the eigenvalues of B'^{-1}. As a result, $\mathbf{r}$ is the unique positive eigenvector of B', and the corresponding eigenvalue is the smallest eigenvalue in absolute value of B'.

The vector $\mathbf{r}$ is easily calculated by the inverse power method, since

$$\lim_{n\to\infty} \frac{(B + \lambda_0 I)^{-n}\mathbf{r}_0}{|(B + \lambda_0 I)^{-n}\mathbf{r}_0|} = \mathbf{r}. \tag{4.9}$$

Of course, we should never calculate the inverse of $B + \lambda_0 I$ explicitly, but rather calculate its LU decomposition, and then perform the inverse iteration using forward and backward substitution.

5. A maximum likelihood estimate. Suppose that the probability that team i beats team j is π_{ij}, and that the outcome of the contest between team i and team j is given by a_{ij}. For now we will take $a_{ij} = 1$ if team i beat team j, and zero otherwise. If the result of the contest between two teams is a Bernoulli trial with the outcome determined by the values π_{ij}, then the probability of the event a_{ij} is

$$P = \prod_{i<j} \binom{a_{ij} + a_{ji}}{a_{ij}} \pi_{ij}^{a_{ij}} \pi_{ji}^{a_{ji}}. \tag{5.1}$$

We now suppose that the ranking vector $\mathbf{r}$ has the property that

$$\pi_{ij} = \frac{r_i}{r_i + r_j}, \tag{5.2}$$

so the probability that the outcome is represented by a_{ij} is

$$P(\mathbf{r}) = \prod_{i<j} \binom{a_{ij} + a_{ji}}{a_{ij}} \left(\frac{r_i}{r_i + r_j}\right)^{a_{ij}} \left(\frac{r_j}{r_i + r_j}\right)^{a_{ji}}. \tag{5.3}$$

Since the outcome a_{ij} is known to have occurred, we pick $\mathbf{r}$ so that $P(\mathbf{r})$ is as large as possible. The resulting vector $\mathbf{r}$ is called the maximum likelihood solution.

The problem of choosing $\mathbf{r}$ to maximize $P(\mathbf{r})$ is quite old. This model and an iterative method for its solution was first proposed by Zermelo in 1926 [14] and then rediscovered by Ford in 1955 [7] and is often called a Bradley–Terry model [3], [11]. We give a new proof of existence and uniqueness of the solution here.

With the choice (5.2) and since the matrix A is fixed, it is equivalent to maximize the function

$$F_A(\mathbf{r}) = \prod_{i<j} \left(\frac{r_i}{r_i + r_j}\right)^{a_{ij}} \left(\frac{r_j}{r_i + r_j}\right)^{a_{ji}}, \tag{5.4}$$

or

$$\ln F_A(\mathbf{r}) = \sum_{i<j} (a_{ij}(\ln r_i - \ln(r_i + r_j)) + a_{ji}(\ln r_j - \ln(r_i + r_j))). \tag{5.5}$$

To show that a maximum exists, we assume that the matrix A is irreducible. Clearly, the function $F_A(\mathbf{r})$ is continuous and bounded on the interior of the positive orthant. While it is not defined on the faces of the positive orthant, if A is irreducible we can define $F_A = 0$ on the faces of the positive orthant as the continuous extension of $F_A(\mathbf{r})$. That is, if $\mathbf{r}_0$ is on a face of the positive orthant, then one of its elements, say r_i, is zero, and another of its elements, say r_j, is nonzero. Because the matrix A is irreducible, there is a sequence of indices $i_0, i_1, \ldots, i_k$, with $i_0 = i$ and $i_k = j$ with the property that $a_{i_p i_{p+1}} > 0$ for $p = 0, 1, \ldots, k-1$. Necessarily, there are consecutive integers $m < n$ for which $r_{i_m} = 0$, and $r_{i_n} > 0$.

We write

$$F_A(\mathbf{r}) = \left(\frac{r_{i_m}}{r_{i_m} + r_{i_n}}\right)^{a_{i_m i_n}} \phi(\mathbf{r}), \tag{5.6}$$

and observe that $\phi(\mathbf{r})$ is positive and bounded in the interior of the positive orthant. It follows that $\lim_{r \to r_0} F_A(\mathbf{r}) = 0$. As thus extended, the function $F_A(\mathbf{r})$ is continuous and bounded on the closed and bounded set $\sum = \{\mathbf{r} | r_i \geq 0, \sum_i r_i = 1\}$, $F_A(\mathbf{r})$ is strictly positive on the interior of the set $\sum$, and is zero on the boundary of the set $\sum$. It follows that $F_A(\mathbf{r})$ attains a maximum on the interior of the positive orthant. (This part of the proof is from Ford's work [7].)

To find an extremum we differentiate the function $\ln F_A(\mathbf{r})$ to find

$$\frac{\partial}{\partial r_k} \ln F_A(\mathbf{r}) = \frac{\alpha_k}{r_k} - \sum_j \frac{A_{jk}}{r_j + r_k}, \tag{5.7}$$

where $\alpha_k = \sum_j a_{jk}$, and $A_{jk} = a_{jk} + a_{kj}$. Consequently, the maximizing vectors $\mathbf{r}$ must satisfy the nonlinear system of equations

$$\frac{\alpha_k}{r_k} - \sum_j \frac{A_{jk}}{r_j + r_k} = 0. \tag{5.8}$$

Zermelo [14] and Ford [7] used an iterative method to solve (5.8). In my opinion, it is just as easy to solve (5.8) by integrating the system of differential equations

$$\frac{dr_k}{dt} = \frac{\alpha_k}{r_k} - \sum_j \frac{A_{jk}}{r_j + r_k}, \tag{5.9}$$

using one's favorite numerical integrator, starting from any initial point in the interior of the positive orthant. We are assured that the solution of the differential equation system (5.9) will approach a steady state because it is a gradient system, and along trajectories

$$\begin{aligned} \frac{d \ln F_A(\mathbf{r})}{dt} &= \sum_k \frac{\partial}{\partial r_k}(\ln F_A(\mathbf{r}))\frac{dr_k}{dt} \\ &= \sum_k \left(\frac{\partial}{\partial r_k}(\ln F_A(\mathbf{r}))\right)^2, \end{aligned} \tag{5.10}$$

which is positive except at an extremum of $\ln F_A(\mathbf{r})$. Hence $F_A(\mathbf{r})$ increases along trajectories of (5.9).

Finally, we can show that the maximum of $\ln F_A(\mathbf{r})$ is unique. We calculate the Hessian H of $\ln F_A(\mathbf{r})$ at any extremum to be $H = (h_{ik})$, where

$$h_{ik} = \frac{\partial \ln F_A(\mathbf{r})}{\partial r_i \partial r_k} = \left(-\frac{\alpha_i}{r_i^2} + \sum_j \frac{A_{ij}}{(r_i + r_j)^2}\right)\delta_{ik} + \frac{A_{ik}}{(r_i + r_j)^2}. \tag{5.11}$$

By virtue of (5.8),

$$\alpha_i = \sum_j A_{ij}\left(\frac{r_i}{r_i + r_j}\right) > \sum_j A_{ij}\left(\frac{r_i}{r_i + r_j}\right)^2, \tag{5.12}$$

and the off-diagonal elements of H are positive. Observe also that H has a null space, since $H\mathbf{r} = 0$ for any vector $\mathbf{r}$ satisfying (5.8). This null space results from the invariance of (5.4) under changes of the scale of $\mathbf{r}$.

Now we want to find the eigenvalues of H. Notice that for any sufficiently large positive λ_0, the matrix $-H + \lambda_0 I$ is diagonally dominant. However, because all of its off-diagonal elements are negative, the matrix $-H + \lambda_0 I$ maps the boundary of the positive orthant to the exterior of the positive orthant. It follows from our friend the Perron–Frobenius theorem that $(-H+\lambda_0 I)^{-1}$ is a map of the positive orthant into itself, having a unique positive eigenvector with corresponding eigenvalue μ_1, with $\mu_1 > \mu_2 \geq \cdots \geq \mu_n$. (This is the same application of the Perron-Frobenius theorem as used in §4.) Therefore, the eigenvalues of H are

$$\lambda_0 - \frac{1}{\mu_1} > \lambda_0 - \frac{1}{\mu_2} \geq \cdots \geq \lambda_0 - \frac{1}{\mu_n}. \tag{5.13}$$

The maximizing vector $\mathbf{r}$ satisfying (5.8) also satisfies $H\mathbf{r} = 0$ so that $\mathbf{r}$ is an eigenvector of $(-H + \lambda_0 I)^{-1}$ as well, and being positive, must correspond to its largest eigenvalue. It follows that $\lambda_0 - (1/\mu_1) = 0$, and the remaining eigenvalues of H must be strictly negative. Thus, on the surface $\sum$, all extrema for $\ln F_A(\mathbf{r})$ are local maxima. We conclude that there is, therefore, exactly one extremum.

The motivation for this model was based on the assumption that the numbers a_{ij} were integers. But clearly, there is nothing in the proof of existence and uniqueness that forces this requirement. For the purpose of ranking football teams it is preferable to use a different determination for a_{ij}. A choice that works well is

$$a_{ij} = \frac{S_{ij}}{S_{ij} + S_{ji}}. \tag{5.14}$$

6. Putting it all together. There are 106 Division I-A college football teams in the United States, which during each season play about 570 games, including bowl games. Schedules for the coming season and results from the previous season are available annually in the NCAA Football book (available from NCAA Publications, Mission, KS 66201).

In Table 1, we present the results of the above ranking schemes for the 1989 season. In Table 1 there are eight columns. The top 40 teams are ordered in the table according to percentage of wins. W-L-T refers to the win-lose-tie record for the 1989–90 season (including bowl games). Columns labelled 1–4 show the integer rank of the team for methods 1–4, respectively, and the columns labeled UPI and AP are the final poll results for those teams that were ranked.

For this table, methods 1–4 are defined as follows:

(1) The direct linear method based on the eigenvalue problem (2.2) with entries a_{ij} chosen using (2.4);

(2) The nonlinear method (3.1) with $f(x)$ satisfying (3.6) and scoring factors e_{ij} satisfying (3.7);

(3) The least squares estimate of probabilities (4.6);

(4) The maximum likelihood method (Bradley–Terry model) (5.3) with entries a_{ij} satisfying (5.15).

What can we conclude from all of this? First, there is no unique way to devise a ranking scheme. The different ranking schemes give different rankings because they weigh important factors differently. Each of the schemes proposed here have strengths and weaknesses, but invariably when a method is tweaked to get rid of some "undesirable" feature, another "counterintuitive" result shows up. After studying these methods for awhile, it is also apparent that intuition is not a good guide to determining a ranking. With 106 teams there are just too many factors to consider. On the other hand, the numbers are not biased; they simply report the results of the algorithm.

It is interesting to compare the results of the ranking algorithms with the UPI and AP polls. First, it is obvious that there is much more variation between the ranking schemes than between the polls, suggesting that the two polls are not independent. Second, there are noted differences between the polls and the ranking schemes. For example, counting the number of teams whose poll rankings do not lie within the range of rankings from the four mathematical schemes, we find nine teams for whom the polls are "too high" and four teams for whom the polls are "too low."

TABLE 1
Ranking of NCAA *Division* I-A *football teams for* 1989 *season*.

Team	W-L-T	#1	#2	#3	#4	UPI	AP
Notre Dame	12-1-0	3	6	5	3	3	2
Miami (Fla)	11-1-0	8	2	1	1	1	1
Tennessee	11-1-0	4	10	8	11	5	5
Colorado	11-1-0	11	5	10	7	4	4
Fresno State	10-1-0	48	20	47	43		
Florida St.	10-2-0	2	3	2	5	2	3
Auburn	10-2-0	6	8	4	8	6	6
Alabama	10-2-0	9	11	6	12	7	9
Arkansas	10-2-0	10	16	11	17	13	13
Michigan	10-2-0	16	12	12	9	8	7
Illinois	10-2-0	20	18	20	14	10	10
Nebraska	10-2-0	29	9	13	13	12	11
Clemson	9-2-0	5	7	9	6	11	12
Houston	9-2-0	15	1	3	2		14
USC	9-2-1	1	4	7	4	9	8
Northern Illinois	7-2-0	65	52	59	68		
BYU	10-3-0	22	21	27	25	18	
Virginia	9-3-0	13	22	19	19	15	18
Texas Tech	9-3-0	21	33	25	27	16	19
Hawaii	9-3-1	33	28	46	41		
Eastern Michigan	6-2-1	79	49	83	90		
Penn State	8-3-1	19	15	18	18	15	15
Pittsburgh	8-3-1	41	36	37	34	19	17
Syracuse	7-3-0	45	39	38	45		
West Virginia	7-3-1	37	27	35	38		
Washington	8-4-0	7	13	14	10	19	20
Arizona	8-4-0	12	24	28	20		
Oregon	8-4-0	17	19	23	21		
Texas A & M	8-4-0	18	17	16	16		20
Duke	8-4-0	26	38	34	37		
Michigan State	8-4-0	32	14	15	15	16	16
Ohio State	8-4-0	51	42	48	30		
Air Force	8-4-1	30	30	32	33		
Mississippi	7-4-0	24	34	21	31		
Oklahoma	7-4-0	57	29	44	29		
Ball State	6-3-2	81	59	90	89		
Georgia Tech	6-4-0	25	40	30	42		
Arizona State	6-4-1	39	57	52	44		
Virginia Tech	6-4-1	36	35	42	26		
Florida	7-5-0	27	25	22	23		
Number of upsets		102	110	108	100		

There are other ways that one might try to rank teams. For example, a method that is unrelated to those presented here is to try to minimize the number of upsets. An upset occurs when a team is ranked higher than a team to which it lost. At the bottom of the columns in Table 1 are listed the number of upsets for each of the algorithms used here. The number of upsets cannot be zero since there is no well ordering, but by assigning an objective function that measures the degree of an upset, we can devise algorithms to find the best ranking with respect to that particular measure.

Appendix A. Proof of the Perron–Frobenius theorem. Let $\sum$ be the set of all nonnegative vectors with Euclidean norm one. For each vector $\mathbf{s}$ in the set $\sum$ let σ^* be the positive number for which $A\mathbf{s} \leq \sigma\mathbf{s}$ whenever $\sigma \geq \sigma^*$. If $\mathbf{s}$ has zero entries then σ^* may be infinite. Since $\sum$ is a closed and bounded set, the smallest value of σ^* is attained for some vector $\mathbf{s}^*$ in $\sum$. We claim that $\mathbf{s}^*$ is a positive eigenvector of A.

Suppose that $A\mathbf{s}^* \leq \sigma^*\mathbf{s}^*$ but $\mathbf{s}*$ is not an eigenvector of A. Then some, but not all, of the relations in the statement $A\mathbf{s}^* \leq \sigma^*\mathbf{s}^*$ are equalities. (If there were no equalities, the number σ^* would be incorrectly chosen.) After permutation, we can write the relations

$As^* \leq \sigma^* s^*$ in the form

$$
\begin{aligned}
A_{11}\mathbf{s}_1 + A_{12}\mathbf{s}_2 &< \sigma^* \mathbf{s}_1, \\
A_{21}\mathbf{s}_1 + A_{22}\mathbf{s}_2 &= \sigma^* \mathbf{s}_2.
\end{aligned}
\tag{A.1}
$$

Since A is irreducible, A_{21} is not identically zero, so we can reduce at least one component of the vector $\mathbf{s}_1$, thereby changing at least one of the equalities to a strict inequality, without changing any of the original strict inequalities. After this change in $\mathbf{s}^*$ we rescale the vector to have norm one. Proceeding inductively, we can continue to modify the vector $\mathbf{s}^*$ until all of the relations in $A\mathbf{s}^* \leq \sigma^*\mathbf{s}^*$ are strict inequalities, but of course, this contradicts the definition of σ^*; so we are done.

To prove uniqueness, we note that a nonnegative eigenvector $\mathbf{r}$ must have all positive entries. Suppose there are two linearly independent eigenvectors of $A, \mathbf{r}_1$ and $\mathbf{r}_2$, satisfying $A\mathbf{r}_1 = \lambda_1\mathbf{r}_1$, and $A\mathbf{r}_2 = \lambda_2\mathbf{r}_2$, and suppose that $\mathbf{r}_1$ has strictly positive entries. If the entries of $\mathbf{r}_2$ are all of one sign, then without loss of generality they can be taken as positive. The vector $\mathbf{r}(t) = r_1 - t\mathbf{r}_2$ has nonnegative entries for all t in some range $0 \leq t \leq t_0$ with $t_0 > 0$, and $\mathbf{r}(t_0)$ has some zero entries but is not identically zero, while for $t > t_0$, $\mathbf{r}(t)$ has some negative entries. Then $A\mathbf{r}(t_0) = \lambda_1(\mathbf{r}_1 - t_0\lambda_2/\lambda_1\mathbf{r}_2)$ has only positive entries. By the maximality of t_0, it must be that $|\lambda_2| < |\lambda_1|$. But if both $\mathbf{r}_1$ and $\mathbf{r}_2$ have only positive entries, we can interchange them in the above argument to conclude that $|\lambda_1| < |\lambda_2|$. This is, of course, a contradiction. We conclude that the positive eigenvector is unique and all other eigenvectors have eigenvalues that are smaller in absolute value. A minor modification of this argument shows that the largest eigenvalue is simple. For, if $\mathbf{r}_2$ is a generalized eigenvalue of A satisfying $A^k\mathbf{r}_2 = \lambda_1^k\mathbf{r}_2$ for some $k > 1$, then $A^k\mathbf{r}(t_0) = \lambda_1^k\mathbf{r}(t_0)$ is strictly positive, contradicting the definition of t_0.

Appendix B. Proof of the nonlinear fixed point theorem (nonlinear generalization of the Perron–Frobenius theorem). Suppose F is a positive, monotone, and strictly concave mapping of a finite-dimensional space to itself. That is, $F(\mathbf{r}) > 0$ for all $\mathbf{r} > 0, F(\mathbf{p}) > (\geq)F(\mathbf{q})$ whenever $\mathbf{p} > (\geq)\mathbf{q}$, and $F(t\mathbf{r}) > tF(\mathbf{r})$ for $0 < t < 1$.

To see that there is at least one positive fixed point, let $\mathbf{r}_0$ have all entries equal to 1 and notice that $F(\mathbf{r}_0) < 1$. Define the sequence of vectors $\mathbf{r}_k$ by successive approximation

$$
\mathbf{r}_k = F(\mathbf{r}_{k-1}),
\tag{B.1}
$$

and notice that $\mathbf{r}_k < \mathbf{r}_{k-1}$. The monotone decreasing sequence of vectors $\{\mathbf{r}_k\}$ is bounded below by $F(\mathbf{0}) > \mathbf{0}$, and therefore converges to some positive vector $\mathbf{r}$. Since F is continuous, $\mathbf{r}$ is a fixed point of F.

The positive fixed point $\mathbf{r}$ is unique. If not, there is a positive vector $\mathbf{q}$ satisfying $F(\mathbf{q}) = \mathbf{q}$. Since $\mathbf{r} \neq \mathbf{q}$, one of the inequalities $\mathbf{r} \leq \mathbf{q}$ and $\mathbf{q} \leq \mathbf{r}$ must fail to hold. Without loss of generality suppose that $\mathbf{q} \leq \mathbf{r}$ does not hold. Now, there is a maximal t_0 with $0 < t_0 < 1$ so that $t\mathbf{q} \leq \mathbf{r}$ for all t in $0 \leq t \leq t_0$. Therefore,

$$
\mathbf{r} = F(\mathbf{r}) \geq F(t_0\mathbf{q}) > t_0F(\mathbf{q}) = t_0\mathbf{q},
\tag{B.2}
$$

contradicting the maximality of t_0.

Acknowledgment. Thanks to Joe Keller for introducing me to this fascinating topic over ten years ago, and to Fred Phelps for his ideas on how to define a nonlinear ranking scheme.

REFERENCES

[1] C. Berge, *The Theory of Graphs and Its Applications*, Wiley and Sons, New York, 1962.

[2] R. A. Bradley, *Incomplete block rank analysis: on the appropriateness of a model for the method of paired comparisons*, Biometrics, 10 (1954), pp. 375–390.

[3] R. A. Bradley and M. E. Terry, *The rank analysis of incomplete block diagrams*. I. *The method of paired comparisons*, Biometrica, 39 (1952), pp. 324–345.

[4] H. A. David, *The Method of Paired Comparisons*, Griffin, London, 1969.

[5] M. G. Kendall, *Further contributions to the theory of paired comparisons*, Biometrics, 11 (1955), p. 43.

[6] M. G. Kendall and B. Babington Smith, *On the method of paired comparisons*, Biometrica, 31 (1939), p. 324.

[7] L. R. Ford, Jr., *Solution of a ranking problem from binary comparisons*, Amer. Math. Monthly, 64 (1957), pp. 28–33.

[8] F. Mosteller, *Remarks on the method of paired comparisons*: I. *The least squares solution assuming equal standard deviations and equal correlations*, 16 (1951), pp. 3–9.

[9] T. L. Saaty, *Rank according to Perron*: *a new insight*, Math. Magazine, 60 (1987), p. 211.

[10] G. W. Stewart, *Introduction to Matrix Computations*, in Computer Science and Applied Mathematics, W. Rheinboldt, ed., Academic Press, New York, 1973.

[11] M. Stob, *A supplement to "A mathematicians guide to popular sports*," Amer. Math. Monthly, 91 (1984), pp. 277–281.

[12] R. S. Vargas, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1962.

[13] T. H. Wei, *The algebraic foundations of ranking theory*, Cambridge University Press, London, 1952.

[14] E. Zermelo, *Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Warscheinlichkeitsrechnung*, Math. Z., 29 (1926), pp. 436–460.