SIAM J. SCI. COMPUT.
Vol. 33, No. 3, pp. 1077–1102
© 2011 Society for Industrial and Applied Mathematics

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

SENSITIVITY AND STABILITY OF RANKING VECTORS*

TIMOTHY P. CHARTIER†, ERICH KREUTZER†, AMY N. LANGVILLE‡, AND KATHRYN E. PEDINGS‡

Abstract. We conduct an analysis of the sensitivity of three linear algebra-based ranking methods: the Colley, Massey, and Markov methods. Our analysis employs reverse engineering, in that we start with a simple input ranking vector that we use to build a perfect season, and we then determine the output rating vectors produced by the three methods. This analysis shows that the PageRank rating vector is strongly nonuniformly spaced, while the Colley and Massey methods provide a uniformly spaced rating vector, which is more natural for a perfect season. We further extend our study of the sensitivity and rank stability of these three methods with a careful perturbation analysis of the same perfect season dataset. We find that the Markov method is highly sensitive to small changes in the data and show with an example from the NFL that the Markov method's ranking vector displays some odd unstable behavior.

Key words. Massey method, Colley method, Markov method, PageRank, ranking vector, rating vector, sensitivity, stability

AMS subject classifications. 60J22, 60J20, 65F05, 65F15, 65F99

DOI. 10.1137/090772745

1. Introduction. In the past decade, ranking methods have often been used for a variety of applications and hence are well studied. This is due in large part to the success that Google has had with its PageRank algorithm for ranking webpages. Today every serious search engine employs some sort of ranking method. While such methods have received great attention recently, ranking is an old problem with a long history. For instance, for centuries people have used mathematical methods to rank teams or individuals engaging in head-to-head contests such as cricket or chess tournaments. The study of these pairwise comparisons produced several early ranking methods [8, 16, 17, 24, 25, 37]. In another much earlier context, 13th and 18th century mathematicians struggled over the most precise method for ranking candidates in a political election [7, 29, 30].

In this paper, we conduct a careful investigation of the sensitivity and stability of three popular ranking methods: the Colley, Massey, and Markov methods. There are literally hundreds of ranking methods that we could have analyzed in this paper. Why were these three chosen for this paper? Our selection of methods hinged on three criteria: (1) the method must be based in linear algebra, which enables the perturbation analysis of section 5, (2) the method must have a simple elegant formulation with a closed-form solution, and (3) the method must have a history of success. As a result, these criteria excluded the Keener and Redmond methods [24, 36], two simple linear algebra methods that are historically outperformed by the Colley and Massey

*Submitted to the journal's Methods and Algorithms for Scientific Computing section October 1, 2009; accepted for publication (in revised form) February 16, 2011; published electronically May 5, 2011.
http://www.siam.org/journals/sisc/33-3/77274.html

†Department of Mathematics, Davidson College, Davidson, NC 28035-6908 (tichartier@davidson.edu, erkreutzer@davidson.edu). The research of these authors was supported in part by a research fellowship from the Alfred P. Sloan Foundation.

‡Department of Mathematics, College of Charleston, Charleston, SC 29401 (langvillea@cofc.edu, kepeding@edisto.cofc.edu). The research of these authors was supported in part by NSF grant CAREER-CCF-0546622.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

methods. One measure of the success of the Colley and Massey methods is that they make up two of the six computer methods used by the Bowl Championship Series to rank NCAA football teams.[1] On the other hand, the random walker method of Callaghan, Mucha, and Porter [10] and the LRMC method of Kvam and Sokol [26] seem to perform well and are based in linear algebra, yet are too complicated for a complete perturbation analysis. Thus, from the sports ranking methods, the Colley and Massey methods [15, 31] fit our criteria.

The third and final ranking method that we include comes from webpage ranking. The Markov method is our name for the method described in [21, 23] that uses Markov chains to rank items, which could range from political candidates to sports teams to webpages. In fact, our Markov method is a more general version of the PageRank algorithm [9] that Google uses for ranking webpages. Hundreds of papers have been written on PageRank. Recently, PageRank has been applied to many other areas, including the ranking of species [2], the ranking of genes [18], and the ranking of social networks. In section 4.5, we describe the Markov method, which applies PageRank to rank sports teams. However, in section 5.3, we show that a direct application of PageRank is not always appropriate when translating the method to a new application area. Instead, modeling changes should be made for the specific application.

In this paper, we directly apply the PageRank method for ranking webpages to sports ranking problems, with only slight modifications. We do not address modeling questions, and we do not consider other Markov chain models that may be derived specifically for sports ranking problems. We analyze sensitivity of rating and ranking vectors for a perfect tournament. While we do not consider Markov models different from PageRank and while we analyze sensitivity only in the context of a perfect tournament, we believe that our conclusions on sensitivity of PageRank ranking and rating vectors may extend to more general sports ranking problems and to other application domains for the PageRank method, and possibly also to other Markov-type models for ranking. We believe that our findings may also be significant, for example, for the issue of link spamming in web search.

We note that with some careful thought, it is possible to modify any of these methods to fit nearly any particular ranking application. For instance, in [14], we show how the originally conceived sports ranking methods of Colley and Massey can be tailored to rank webpages or even movies. Thus, the conclusions of our sensitivity and stability analysis hold for applications of these three methods in other ranking settings.

Our investigation of these three methods is concerned with questions of sensitivity and stability [1, 3, 4, 27, 28, 32, 33]. Because such questions typically depend on the dataset under consideration, we restrict ourselves to one particular dataset with very special structure. As a consequence, our special dataset allows us to produce closed-form expressions of the ranking vectors, which, in turn, enables precise conclusions about a method's sensitivity and stability under certain perturbations. In this paper, we make a distinction between a rating vector and a ranking vector. We call a method sensitive if small changes in the input data create big changes in the output *rating* vector. And we say a method is rank instable if small changes in the input data create big changes in the *ranking* vector.

Most ranking methods, including the three we study here, produce a rating vector of numerical scores, which, when sorted, produces a ranking vector of integer rank

[1] The information revealed about the remaining four methods is not thorough enough to reproduce their implementation.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

values. As such, a ranking vector for a collection of n items is simply a permutation of the integers 1 through n.

2. Related work. The literature for ranking methods can be divided into categories according to the application. For instance, there is a body of literature for webpage ranking and another, quite separate, one for sports ranking. In the sports ranking context, very little, if any, analysis has been conducted on the sensitivity or stability of the rankings. A natural question in this application is, How many times or by how many points does team i need to beat team j in order to jump k positions up the ranked list? Coaches may be very interested in this information, hoping to motivate their players to achieve invitations to end-of-year tournament or bowl games. Further, because the Colley and Massey methods are two of the six computer ranking methods factored into NCAA college football's BCS rankings, which affect bowl invitations, which, in turn, generate millions of dollars for colleges, it is essential that computer ranking methods be accurate and stable—justification for the careful sensitivity and stability studies we conduct in section 5.

On the other hand, a good deal of research has been done for the web context. There the focus has been on the PageRank vector, with the competing HITS vector receiving some, though much less, attention. For instance, Ng, Zheng, and Jordan [32, 33] conduct an experimental study of the stability of rating vectors to random perturbations to the web graph. In particular, they compute bounds on the L_1 and L_2 norm difference between the original rating vector and the perturbed rating vector.

Avrachenkov and Litvak [3] focus only on webpages and determine how the PageRank of one webpage is affected by rank-one updates to the Markov matrix. In this sense, our analysis of section 5 is similar, though more general, because we assess the effect of rank-one updates on all teams with all three rating methods under consideration.

It appears that Lee and Borodin [27], followed by Lempel and Moran [28], were the first to make the distinction between the stability of the *rating vector* and the *ranking vector*. It is a much harder (though more useful) problem to answer questions about the sensitivity and stability of the ranking vector as opposed to the rating vector. After all, the fact that the values in the rating vector may change does not imply that the ordering of the elements in the ranking vector will change. Using an analysis quite different from the typical derivative or norm difference analysis, Lempel and Moran studied the rank stability of the usual set of popular webpage ranking algorithms: PageRank, HITS, and SALSA. First, they focus on an original graph and a perturbed graph. The restriction is that the two graphs can differ only by k edges. Then, they use Kendall's tau to measure the difference between two ranked lists. If this difference approaches 0 as the size of the graphs approaches infinity, the algorithm creating the ranking is said to be rank stable. According to this definition, Lempel and Moran found that PageRank and HITS were not rank stable on the class of authority-connected graphs (i.e., irreducible cocitation matrices). Further, they describe a very interesting example graph for which the change of a single edge of a low ranking webpage causes a complete reversal in the entire ranking! Last, we note that Govan, Langville, and Meyer [22] created a sports ranking method based on the HITS algorithm, but it includes a nonlinear step and thus does not meet our linear-algebra criterion for inclusion in this study of ranking methods.

3. The perfect season. Like most [1, 3, 4, 27, 28], we make the problem of answering questions of sensitivity and stability tractable by restricting our study to a special case. In this paper, we use a perfect season to uncover the differences in

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

the three ranking methods of Colley, Massey, and Markov. A perfect season occurs when every team plays every other team exactly once and there are no upsets. For example, a win-loss matrix for a perfect season among four teams could take the form

$$\begin{array}{c c} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \end{array}$$

where a 1 indicates a win. Thus, team 1 beat all its opponents, team 2 beat teams 3 and 4, team 3 beat only team 4, and team 4 lost to all its opponents. In a perfect season, the ranking of teams is clear.

In this paper, for each method, we build the data for a perfect season from the ranking vector $(\,1\;2\;\cdots\;n\,)$, where n is the number of items to be ranked. Thus, team 1 is first place, team 2 is second place, and so on. Though this ranking may seem overly specific, it is actually very general because every ranking vector is simply a permutation of this very specific ranking vector. Hence, the matrices and vectors associated with the three ranking methods are simply permutations of the matrices and vectors studied in this paper. We have three reasons for restricting our study to the perfect season. First, it provides wonderful structure for the matrices and vectors required of the three ranking methods. Second, this structure makes the perturbation analysis of section 5 possible. And third, the perfect season is the most well-behaved of all possible systems. If a method cannot demonstrate appropriate behavior, which we later link to insensitivity, then what hope does it have for ill-behaved imperfect seasons?

The idea behind our analysis of the perfect season comes from the observation that a ranking vector is often used in a predictive sense. A higher ranked team is typically predicted to beat a lower ranked team in future matchups. Thus, in this paper, we conduct some reverse engineering as shown in Figure 3.1.

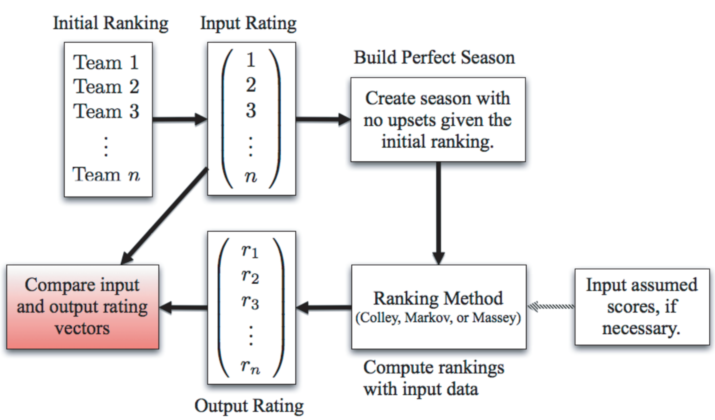

FIG. 3.1. *Flow chart of analysis of three ranking methods on the perfect season.*

In particular, we begin with an input ranking vector and an input rating vector that are identical, i.e., $\mathbf{r} = (\,1\;2\;\cdots\;n\,)$. Notice that the input rating is uniformly

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

spaced with increments of exactly 1. Given this specially structured rating vector, we work backward and reverse engineer the issue by seeing which method produces a rating, and hence ranking, that maintains the properties of the input vector. Of course, as expected, the three ranking methods output the correct ranking of teams. However, when we look beyond the ranking vector to the rating vector, we start to see some surprising differences between the input and the output ratings given by various methods. See sections 4 and 5 for more on this.

In section 4, we define each method and apply it to the special perfect season dataset. Our findings reveal some interesting and important differences between the three methods, namely, that the Markov method is the only one to produce a nonuniformly spaced rating vector. In section 5, we conduct an extensive perturbation analysis of each method, wherein we determine how much the ranking is affected by a small perturbation to the perfect season. Such sensitivities to small perturbations of the perfect season are exaggerated only when imperfect seasons are studied. Section 6 demonstrates this point with a practical scenario from a real-world (i.e., imperfect) season.

4. Applying the methods to the perfect season.

4.1. The Colley method in general. The Colley method for ranking items was designed by Wesley Colley in 2002 [15]. Colley originally had sports teams in mind as the items under consideration, so for consistency we too will use the sports context throughout this paper. Colley's method can be succinctly summarized with one linear system, $\mathbf{Cr} = \mathbf{b}$, where $\mathbf{r}_{n\times 1}$ is the unknown Colley rating vector, $\mathbf{b}_{n\times 1}$ is the right-hand-side vector defined as $b_i = 1 + \frac{1}{2}(w_i - l_i)$, and $\mathbf{C}_{n\times n}$ is the Colley coefficient matrix defined as

$$\mathbf{C}_{ij} = \begin{cases} 2 + t_i, & i = j, \\ -n_{ij}, & i \neq j. \end{cases}$$

The scalar n_{ij} is the number of times teams i and j played each other, t_i is the total number of games played by team i, w_i is the number of wins for team i, and l_i is the number of losses for team i. It can be proved that the Colley system $\mathbf{Cr} = \mathbf{b}$ always has a unique solution since $\mathbf{C}_{n\times n}$ is invertible.

4.2. The Colley method for the perfect season. To apply the Colley method to the perfect season, we must use the rating vector $\mathbf{r} = (\,1 \quad 2 \quad \cdots \quad n\,)$ to create initial game data of pairwise matchups. Thus, we assume that each team played every other team exactly once so that $n_{ij} = 1 \,\forall\, i \neq j$ and $t_i = n - 1$. In matrix notation, the Colley coefficient matrix $\mathbf{C}$ can be expressed as a rank-one update to the scaled identity matrix. That is,

$$\mathbf{C} = (n+2)\,\mathbf{I} \;-\; \mathbf{e}\,\mathbf{e}^T.$$

Of course, in a perfect season, a higher ranked team will beat a lower ranked team, which means that the simple structure of a perfect season implies that team i beats team j whenever $i < j$. This win-loss information is used to build the Colley right-hand-side vector $\mathbf{b}$. In fact, $b_i = 1 + 1/2\,(n - 2i + 1)$.

Because the perfect season has such nice structure, it creates a Colley system with equally nice structure. Using the Sherman–Morrison formula for inversion under a rank-one update, we find that this system can be solved with a closed-form expression

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

for the Colley ranking vector $\mathbf{r}$:

$$\mathbf{r} = \mathbf{C}^{-1}\mathbf{b} = \frac{1}{(n+2)}\,\mathbf{b} \;+\; \frac{n}{2(n+2)}\,\mathbf{e} = \frac{1}{2(n+2)}\begin{pmatrix} 2n+1 \\ 2n-1 \\ 2n-3 \\ \vdots \\ 2n-2i+3 \\ \vdots \\ 3 \end{pmatrix}.$$

As the elements in $\mathbf{r}$ are in descending order, the output Colley ranking is consistent with the input ranking. Further, because the increment between successive elements in the output Colley rating vector is constant at $1/(n+2)$, the output Colley rating vector also maintains the *uniform spacing* that is fundamental to the input rating. This uniformity is not trivial—not all ranking methods maintain this uniformity, and this distinction explains the sensitivity results that we obtain in section 5.

A few other properties of the Colley method are worth mentioning now, as they too relate to the sensitivity results we obtain in later sections. The Colley method is said to be *bias-free*, which means that it is impervious to the bias that creeps into the ratings when strong teams "run the score up" on weak opponents. This bias-free property is a direct consequence of the fact that the Colley method considers only wins and losses, not game scores (as most other ranking methods do, including the Massey and Markov methods discussed next). Last, the Colley method has a *conservation* property. The sum of all Colley ratings is always $n/2$. Thus, if one team's rating improves, another's must suffer.

4.3. The Massey method in general. The Massey method for ranking items was created by Kenneth Massey in 1997 [31]. Like Colley, Massey had sports teams in mind when he developed his model. Massey's model revolves around the rule that the difference in the ratings of two teams i and j, denoted $r_i - r_j$, represents the point differential in a matchup of these two teams. Also like Colley, the Massey method can be succinctly summarized with one linear system

$$\mathbf{Mr} = \mathbf{p},$$

where the Massey coefficient matrix $\mathbf{M}$ is related to the Colley matrix $\mathbf{C}$ by the formula $\mathbf{M} = \mathbf{C} - 2\,\mathbf{I}$ so that

$$\mathbf{M}_{ij} = \begin{cases} t_i, & i = j, \\ -n_{ij}, & i \neq j. \end{cases}$$

The Massey right-hand-side vector $\mathbf{p}$ is a vector of cumulative point differentials. That is, p_i is the total number of points team i scored on all opponents minus the total number of points opponents scored against team i. It can be proved that $\mathbf{M}$ is singular since $\mathrm{rank}(\mathbf{M}) = n - 1$. As a result, an adjustment is made to ensure nonsingularity. Any row of $\mathbf{M}$ is replaced with a row of all 1's, and the corresponding entry in $\mathbf{p}$ is set to 0. This new constraint forces the ratings to sum to 0. Following Massey's advice, we apply this nonsingularity adjustment to the last row, creating an adjusted linear system which we denote by $\bar{\mathbf{M}}\mathbf{r} = \bar{\mathbf{p}}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

4.4. The Massey method for the perfect season. We start as we did for the Colley method. Each team plays every other team exactly once, resulting in $n_{ij} = 1 \, \forall \, i \neq j$ and $t_i = n - 1$. In matrix notation, the Massey coefficient matrix $\mathbf{M}$ can be expressed as a rank-one update to the scaled identity matrix so that

$$\mathbf{M} = n\,\mathbf{I} \;-\; \mathbf{e}\,\mathbf{e}^T.$$

In our analysis, we again use an input rating vector $(\,1\;2\;\cdots\;n\,)$. As we did with the Colley system, we assume a higher ranked team beats a lower ranked team. When this rule is applied to the perfect season, we have that team i beats team j whenever $i < j$. Recall that the right-hand-side vector $\mathbf{p}$ of the Massey system uses more than just the simple win-loss information of the Colley system. Specifically, implicit in the Massey method are point differentials equaling $j - i$, where $i < j$. Such an assumption is part of the Massey method's derivation and as such are the point differentials in our perfect season. As a result, $p_i = \frac{n(n-2i+1)}{2}$.

The nonsingularity-adjusted Massey matrix $\bar{\mathbf{M}}$ is a rank-two update to $\mathbf{M}$:

$$\bar{\mathbf{M}} = n\,\mathbf{I} - \mathbf{e}\,\mathbf{e}^T + \mathbf{e}_n\,\mathbf{d}^T,$$

where $\mathbf{d}^T = (\,2 \quad 2 \quad \cdots \quad 2 \quad -n+2\,)$ and $\mathbf{e}_n$ is the nth row of the $n \times n$ identity matrix. With the Sherman–Morrison–Woodbury inversion formula for a rank-two update, we find that

$$\bar{\mathbf{M}}^{-1} = \frac{1}{n}\,(\mathbf{I} + \mathbf{e}\,\mathbf{e}_n^T - \mathbf{e}_n\,\mathbf{e}^T),$$

which is an arrow matrix (i.e., containing nonzeros only on the diagonal, the last row, and the last column). The expression for $\bar{\mathbf{M}}^{-1}$ enables us to compute an expression for the Massey ranking vector $\mathbf{r}$ associated with the perfect season:

$$\mathbf{r} = \bar{\mathbf{M}}^{-1}\bar{\mathbf{p}} = \frac{1}{n}\,\bar{\mathbf{p}} \;-\; \frac{n-1}{2}\,\mathbf{e}_n = \frac{1}{n}\,\mathbf{p} = \frac{1}{2}\begin{pmatrix} n-1 \\ n-3 \\ n-5 \\ \vdots \\ n-2i+1 \\ \vdots \\ -n+1 \end{pmatrix}.$$

Thus, the increment between successive elements in the Massey rating vector is exactly 1. In summary, when applied to the perfect season, the Massey method, like the Colley method, returns a rating that is *uniformly spaced* and whose ranking matches that of the input ranking.

4.5. The Markov method in general. Due to the success of Google and their PageRank algorithm, all major search engines today employ ranking techniques. PageRank's influence has spread beyond the Web to many other applications. For instance, PageRank has been used to rank genes [18] and species [2] and members in social networks. For sports, the application of PageRank appeared in two recent theses [21, 23]. For most of these applications, few or no modifications are made to PageRank. Instead it is applied in a very straightforward and direct manner. In section 5, we show that this direct application can be a mistake for some applications. These

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

results also explain why the Markov method has performed poorly, despite various voting methods, in [21] and [23] for ranking sports teams. There is an interesting relationship between the Markov method and a recent method by Park and Newman [35]. The Park and Newman method creates two measures for each team, a win measure and a loss measure, which are then combined to create an overall ranking of the teams. The Markov ranking is analogous to a normalized version of the Park and Newman win measure.

When the items to be ranked are not webpages, we refer to the application of PageRank to these items as the Markov method. The main idea behind PageRank and the more general Markov method is voting. To parallel the description of the Colley and Massey methods, we again take sports teams as the items of comparison. In this case, a weaker team casts votes for each stronger team it faces. In fact, for each game, the loser votes for the winner, in effect, declaring with their vote that "yes, they are the better team." There are several methods for vote casting. For instance, a losing team may cast one vote for a team it is beaten by, or it may cast a number of votes equal to the number of points it was beaten by. Regardless of the vote casting method, the result is a team-team matrix that is row normalized in order to enforce stochasticity, which creates the transition matrix for a Markov chain. The stationary vector of this chain is the Markov rating vector. It turns out that the elements of the stationary vector correspond to the proportion of time one would visit a particular team if one took a random walk on the graph defined by the voting matrix. The higher the rating (i.e., the higher the value in the stationary vector), the more frequent the visits to that team and hence the greater the relative importance of that team.

Unlike the Colley and Massey methods, which are both linear systems, the Markov method is an eigensystem. As a result, it must store all information, including matchup and win-loss or point score information, in the matrix. Compare this with the Colley and Massey methods, which, in addition to a matrix, can store information in the right-hand-side vector as well.

4.6. The Markov method for the perfect season. As mentioned in the previous section, we have some flexibility when applying the Markov method to the perfect season. There we discussed two scenarios for creating the Markov voting matrix $\mathbf{V}$. In the first method, a losing team casts a single vote for each team that beats it. Thus, $\mathbf{V}$ is a strictly lower triangular matrix of all ones. In fact, because this Markov method uses only wins and losses, it is a direct competitor to the Colley method. On the other hand, the second method of Markov voting, in which a losing team casts a number of votes equal to the number of points it was beaten by, is a direct competitor to the Massey method, which also uses point differential information. In this case, $\mathbf{V}$ is a strictly lower triangular matrix with 1's on the -1 (or sub-) diagonal, 2's on the -2 diagonal, 3's on the -3 diagonal, and so on. We computed the stationary vector for the stochasticized version of both matrices. Naturally, it was easier to do the computations on the simpler voting matrix of 1's. Thus, in this section, we present the results for the simpler case and make the results for the more complicated case available upon request. In fact, the simpler case gives a very clear presentation of our main point: when applied to the perfect season, the Markov method produces a rating vector that is nonuniform. The second voting matrix reaches the same conclusion. The only difference between the two voting matrices is that the second requires more complex calculations and results in a rating vector for which this nonuniformity is even more pronounced.

We now use an $n = 5$ example and the voting matrix $\mathbf{V}$ of 1's to demonstrate

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

this nonuniformity:

$$\mathbf{V} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

Normalizing the rows produces

$$\bar{\mathbf{V}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 \end{pmatrix}.$$

Because the first team is undefeated, the Markov chain has an absorbing state. Again, there are several strategies for handling undefeated teams. However, to keep with the direct application of PageRank, we employ its "dangling node" adjustment. That is, we add a row of $1/5$ to the first row. Thus, the matrix $\bar{\mathbf{V}} + 1/n\,\mathbf{e}_1\,\mathbf{e}^T$ is stochastic and hence the transition matrix for an aperiodic irreducible Markov chain with a unique, positive stationary vector. Though we do not encounter the problem of cycles in the perfect season, we still employ its solution of full teleportation in order to show the effect of the teleportation parameter α as it moves over its domain from 0 to 1. Thus, we examine the matrix $\alpha(\bar{\mathbf{V}} + 1/n\,\mathbf{e}_1\mathbf{e}^T) + (1-\alpha)/n\,\mathbf{e}\mathbf{e}^T$. We have set the so-called personalization vector, denoted $\mathbf{v}^T$, to the uniform stochastic vector $1/n\,\mathbf{e}^T$. While this is a traditional and simple choice for $\mathbf{v}^T$, there are certainly many others. For example, modelers attempt to choose a personalization vector that forces the rating vector $\mathbf{r}$ to have certain desirable properties (such as spam resistance, robustness, or uniform spacing, which we argue for in section 4.7). However, in general, this is a very difficult task. It can be argued that the simple teleportation vector of $\mathbf{v}^T = 1/n\,\mathbf{e}^T$ is a reasonable modeling approach for sports because as the saying goes "on any given Sunday," every team, even poor ones, has at least a small chance of winning if game-changing events such as injuries, weather, referees, or chance align favorably.

Our goal at this point is to find the stationary vector $\mathbf{r}^T$ of this Markov matrix. The vector $\mathbf{r}^T$ must satisfy the eigensystem equation $\mathbf{r}^T(\alpha\bar{\mathbf{V}} + \alpha/n\ \mathbf{e}_1\,\mathbf{e}^T + (1-\alpha)/n\ \mathbf{e}\mathbf{e}^T) = \mathbf{r}^T$, which can be rearranged with some algebra to $\mathbf{r}^T = (\alpha r_1 + 1 - \alpha)/n\,\mathbf{e}^T(\mathbf{I} - \alpha\bar{\mathbf{V}})^{-1}$. Every Markov system has an additional constraint that $\mathbf{r}^T\mathbf{e} = 1$, which is simply a normalization step since $\mathbf{r} > \mathbf{0}$. In our case, this normalization is unnecessary. In fact, because we are concerned only with the relative position and size of elements in $\mathbf{r}$, the scalar $(\alpha r_1 + 1 - \alpha)/n$ is irrelevant. Thus, our focus is on $\mathbf{e}^T(\mathbf{I} - \alpha\bar{\mathbf{V}})^{-1}$, the column sums of $(\mathbf{I} - \alpha\bar{\mathbf{V}})^{-1}$. Fortunately, the structure of $\bar{\mathbf{V}}$ produces a nicely structured $(\mathbf{I} - \alpha\bar{\mathbf{V}})^{-1}$, which can be seen clearly from the $n = 5$ example:

$$(\mathbf{I} - \alpha\bar{\mathbf{V}})^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{\alpha}{1} & 1 & 0 & 0 & 0 \\ \frac{\alpha(\alpha+1)}{1\cdot 2} & \frac{\alpha}{2} & 1 & 0 & 0 \\ \frac{\alpha(\alpha+1)(\alpha+2)}{1\cdot 2\cdot 3} & \frac{\alpha(\alpha+2)}{2\cdot 3} & \frac{\alpha}{3} & 1 & 0 \\ \frac{\alpha(\alpha+1)(\alpha+2)(\alpha+3)}{1\cdot 2\cdot 3\cdot 4} & \frac{\alpha(\alpha+2)(\alpha+3)}{2\cdot 3\cdot 4} & \frac{\alpha(\alpha+3)}{3\cdot 4} & \frac{\alpha}{4} & 1 \end{pmatrix}.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Focusing on the unscaled unnormalized rating vector $\mathbf{r}^T = \mathbf{e}^T(\mathbf{I} - \alpha\bar{\mathbf{V}})^{-1}$ reveals a nice recursion. For n items, $r_n = 1$ and r_{n-1} can be written in terms of r_n as follows:

$$r_{n-1} = 1 + \frac{\alpha}{n-1} = \left(\frac{\alpha + n - 1}{n-1}\right)(1) = \left(\frac{\alpha + n - 1}{n-1}\right) r_n.$$

Similarly, r_{n-2} can be written in terms of r_{n-1} as follows:

$$\begin{aligned} r_{n-2} &= 1 + \frac{\alpha}{n-2} + \left(\frac{\alpha}{n-2}\right)\left(\frac{\alpha + n - 2}{n-1}\right) \\ &= \left(\frac{\alpha + n - 2}{n-2}\right)\left(1 + \frac{\alpha}{n-1}\right) = \left(\frac{\alpha + n - 2}{n-2}\right) r_{n-1}. \end{aligned}$$

In general, we can prove that

$$r_{j-1} = \left(\frac{\alpha + j - 1}{j-1}\right) r_j.$$

Now we vary the teleportation parameter α and examine the rating vector $\mathbf{r}^T$. We start at the extreme end of no teleportation (i.e., $\alpha = 1$). In general, for n items,

$$\mathbf{r}^T = (\, n \quad n/2 \quad n/3 \quad \cdots \quad n/(n-1) \quad 1\,),$$

which, when scaled and normalized, is

$$\mathbf{r}^T = 1/H(n)\ (\,1 \quad 1/2 \quad 1/3 \quad \cdots \quad 1/(n-1) \quad 1/n\,),$$

where $H(n)$ is the nth partial sum of the harmonic series. While the Markov rating does get the correct rank order, it does not display the uniform spacing that the Colley and Massey methods do for the perfect season. In fact, at the tail of the distribution, the spacing is given by $r_n - r_{n-1} = 1/(n-1)$, while at the head, the spacing is given by $r_1 - r_2 = n/2$. The increment between r_j and r_{j+1} is $\frac{n}{j(j+1)}$. As the length of the vector grows (i.e., the number of teams, n, increases), the Markov increment approaches 0. There are few elements in the head of the Markov curve and many in the tail, making a long-tailed distribution of ratings. When applied to webpage ranking, Markov ratings are PageRanks, which have been shown repeatedly on a variety of real (imperfect) graphs to follow a scale-free or power law distribution [5, 6, 34]. As n, the number of items under comparison, grows, it gets harder and harder to distinguish between elements in the tail, and one must be very careful and precise in the numerical calculation of these elements. When normalization is applied to $\mathbf{r}^T$, the elements get scrunched together even tighter so that the Markov increment between r_j and r_{j+1} is $\frac{1}{H(n)j(j+1)}$. At the head of the Markov rating, the increment is $\frac{1}{2H(n)}$, while at the tail, the increment shrinks to a tiny $\frac{1}{H(n)n(n-1)}$, which is smaller than $O(\frac{1}{n^2})$. Thus, in the tail the Markov increment shrinks to 0 quickly at an inverse quadratic rate. As a result, the normalized Markov rating $\mathbf{r}^T$ reaches the same findings as the unnormalized rating vector. The findings are simply more exaggerated with the normalized vector. Therefore, without loss of generality and for ease of exposition, in what follows we use the unnormalized $\mathbf{r}^T$ for the Markov rating.

Next we consider the other extreme, full teleportation (i.e., $\alpha = 0$). In this case, the Markov system is governed completely by the uniform teleportation matrix $1/n\,\mathbf{e}\mathbf{e}^T$. The increment between any two successive elements is 0 at the head and tail

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

and all elements in between because $r_j = 1/n \, \forall j$. Of course, the full teleportation Markov system is also useless in terms of ranking. Fortunately, we can vary the Markov system between these two extremes of full and no teleportation. Figure 4.1 compares the perfect season, Colley, Massey, and Markov ratings for the perfect season with $n = 100$ teams. We normalized the perfect season, Colley, and Massey ratings in order to show them on the same graph on the left-hand side of Figure 4.1. The right graph of Figure 4.1 shows the Markov ratings for the perfect season for several fixed values of α; $\alpha = .1$, $\alpha = .5$, and $\alpha = .9$. We normalized the ratings as probability vectors in order to show them on the same graph. Notice that α can be used to control the degree of the bend of the Markov ratings on the perfect season. As $\alpha \to 1$, we see the pronounced power law curve, which becomes much less pronounced as $\alpha \to 0$. This ability to control the shape of the ratings curve with α will become important when we move on to the issue of sensitivity in section 5. Figure 4.1 makes the major finding from the previous analysis clear. Of the three methods, the Markov method is the only one to produce a nonuniform vector.

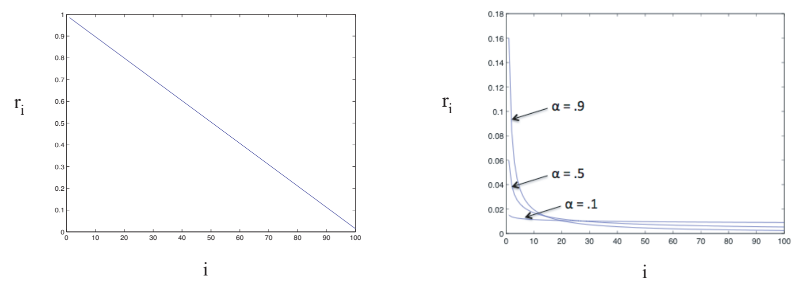

FIG. 4.1. *(Left) normalized perfect season, Colley, and Massey ratings and (right) Markov ratings for $\alpha = .1$, $\alpha = .5$, and $\alpha = .9$ (top curve from left) for $n = 100$ teams.*

4.7. Summary of section 4. When applied to the perfect season, the Colley and Massey methods produce a rating vector with the same uniform structure as the input rating. On the other hand, the Markov method produces a nonuniformly spaced vector with a long tail. The bend of the Markov curve and length of the long tail can be controlled by the Markov teleportation parameter α. We have made the point earlier that all three ranking methods can and have been applied in other settings. For instance, when the Markov method of this section is applied to the Web, it is called PageRank. For imperfect graphs, PageRank exhibits a long-tailed distribution similar to the perfect season. In section 5.3, we connect the sensitivity issues of the method to this long-tailed distribution of ratings. One might argue that the users of PageRank need not worry about sensitivity in the tail of the distribution because search engines rely on top-k lists and thus rarely delve into this long tail. However, this argument is flawed for two reasons. First, search engines pull webpages from this long tail to answer esoteric search queries. Second, this tail sensitivity is precisely what makes link spamming so effective. We will return to this issue again in section 5.3.

5. Perturbation analysis for the perfect season. The previous section presented some clear differences between the three ranking methods. In particular, we found that the *Markov method produces a rating that is not uniformly spaced* and in-

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

stead follows the so-called scale-free power law. In this section, we examine in greater depth the consequences of nonuniformity in the output rating vector. We use perturbation analysis to determine the precise effect that a small change in the input data has on the rating vector. In particular, we study the effect of an upset, i.e., a lone deviation from the perfect season, on the rating vector produced by the three methods. Giving away the ending now, by the end of this section we will find that the *Markov method is very sensitive in its tail*.

5.1. Perturbation analysis for the Colley method. In this section, we study the effect of a small change to the Colley system on the rating vector $\mathbf{r}$. Because the Colley system was built within the context of ranking, the type of perturbation we make must satisfy the constraints of that method. Thus, we define a *Colley-constrained perturbation* as follows.

DEFINITION 5.1 (Colley-constrained perturbation). *Allow team j an additional ϵ of a win against team i, where $j > i$. These two teams have played an additional ϵ of a game. The perturbed Colley matrix $\tilde{\mathbf{C}}$ and right-hand-side vector $\tilde{\mathbf{b}}$ are related to the original matrix $\mathbf{C}$ and vector $\mathbf{b}$ by*

$$\begin{aligned}\tilde{c}_{ij} &= c_{ij} - \epsilon,\\ \tilde{c}_{ji} &= c_{ji} - \epsilon,\\ \tilde{c}_{ii} &= c_{ii} + \epsilon,\\ \tilde{c}_{jj} &= c_{jj} + \epsilon\end{aligned} \qquad \textit{and} \qquad \begin{aligned}\tilde{b}_i &= b_i - \epsilon/2,\\ \tilde{b}_j &= b_j + \epsilon/2.\end{aligned}$$

Or, in compact notation,

$$\tilde{\mathbf{C}} = \mathbf{C} + \epsilon(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T \qquad \textit{and} \qquad \tilde{\mathbf{b}} = \mathbf{b} - \epsilon/2\,(\mathbf{e}_i - \mathbf{e}_j).$$

Given this particular perturbation, we work to compute $\tilde{\mathbf{r}} = \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{b}}$. Employing the Sherman–Morrison inversion formula for a rank-one update, we find that

$$\tilde{\mathbf{C}}^{-1} = \mathbf{C}^{-1} - \frac{\epsilon}{(n+2)(n+2+2\epsilon)}\,(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T$$

and

$$\tilde{\mathbf{r}} = \mathbf{r} + \frac{\epsilon(i-j-n/2-1)}{(n+2)(n+2+2\epsilon)}\,(\mathbf{e}_i - \mathbf{e}_j).$$

Notice that for the perfect season a Colley perturbation between teams i and j affects only the ratings of these two teams. As a result, we say that Colley perturbations to the perfect season have an *isolated effect*.

5.1.1. The case of fractional values of $\boldsymbol{\epsilon}$. In this section, our goal is to find the value of ϵ that causes a change in the original rank ordering. The event is that a lower ranked team j upsets a higher ranked team i in an additional ϵ of a game. Though we can study changes in the ranking of both team i and team j, here we focus only on team j's ability to move up the ranked list. What ϵ is necessary to move the jth team k positions up the ranked list? That is, what ϵ forces $\tilde{r}_j > \tilde{r}_{j-k}$? Isolating the j and $j-k$ elements of $\tilde{\mathbf{r}}$ and using algebra, we discover that any value of ϵ following the rule below achieves the goal:

$$\boxed{\text{Colley perturbation:} \quad \epsilon > \frac{k(n+2)}{(j-i)+n/2+1-2k}.}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Remark 1. For a one position move up the ranked list, i.e., $k = 1$, as $(j - i) \to n$ and $n \to \infty$, $\epsilon \to 2/3$. In other words, if the two teams involved in the perturbation are very unevenly matched (i.e., the difference in their ratings is large) and the collection of teams grows, the size of the perturbation ϵ required to move team j one position up the ranked list approaches the scalar $2/3$.

Remark 2. For a one position move up the ranked list, i.e., $k = 1$, as $(j - i) \to 0$ and $n \to \infty$, $\epsilon \to 2$. On the other hand, if the two teams involved in the perturbation are closely matched and the collection of teams grows, the size of the perturbation ϵ required to move team j one position up the ranked list approaches the scalar 2. Thus, it takes a bigger perturbation ϵ to change the rank ordering when teams are more evenly matched. Conversely, when a weak team suddenly beats a strong team, it takes less to move that team one position up the ranked list. In either case, the good news is that the value of ϵ does not get infinitesimally small and the Colley method is not sensitive to small changes.

Example. For $n = 10$, when team $j = 7$ upsets team $i = 3$, the Colley perturbation formula finds that any $\epsilon > \frac{12k}{10-2k}$ forces team j to jump k positions up the ranked list. We demonstrate this for $k = 1$ with the ranking vectors for the perfect season and an ϵ-perturbed season, denoted by $\mathbf{r}$ and $\mathbf{r}_\epsilon$, below:

$$\mathbf{r} = \frac{1}{24}\begin{pmatrix} 21\\ 19\\ 17\\ 15\\ 13\\ 11\\ 9\\ 7\\ 5\\ 3 \end{pmatrix} \quad \text{and} \quad \mathbf{r}_{\epsilon=1.5} = \frac{1}{24}\begin{pmatrix} 21\\ 19\\ 15\\ 15\\ 13\\ 11\\ 11\\ 7\\ 5\\ 3 \end{pmatrix}.$$

Since $k = 1$, any $\epsilon > 1.5$ forces team $j = 7$ to move one position up and team $i = 3$ to move one position down. With $\epsilon = 1.5$, there are ties in the ratings, as shown in the $\mathbf{r}_{\epsilon=1.5}$ vector above.

5.1.2. The case of $\epsilon = 1$. While the previous analysis is interesting, it does not respect the discrete nature that the ranking problem takes on in most settings. For example, in the sports ranking context, it does not make sense to talk about an additional noninteger or fractional ϵ of a win of one team over another. Though there are some ranking contexts for which fractional values of ϵ do make sense (e.g., new links in the webpage setting), it is much more common to consider discrete integer values for this parameter. Thus, we consider the most natural integer case, $\epsilon = 1$, where the weaker team j beats the stronger team i once for $j > i$.

When team j upsets team i, its rating increases by $\frac{(j-i)+n/2+1}{(n+2)(n+4)}$ and team i's rating decreases by this same amount. If this increment is larger than the usual Colley increment of $1/(n+2)$, team j moves one position up the ranked list and team i moves one position down. We consider two types of possible upsets, one at each extreme. First, we consider the extreme upset that occurs when the worst team upsets the best team, i.e., $j = n$ and $i = 1$. In this upset, for any $n > 8$, the worst place team will move up one position and the first place team will move down one position. The second type of upset is at the other extreme, the very mild "upset" that occurs when j beats i for $j = i + 1$. In this scenario, the increase in team j's rating is not enough

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

to move it even one position up the ranked list. We also note that it is impossible in either extreme upset scenario for team j to move more than one position up the list. *In the perfect season scenario with the Colley method, the best team j can hope to improve is a single rank position, even if it defeats the first place team.*

5.2. Perturbation analysis for the Massey method. To study the effect of a small change to the Massey input on the Massey output rating vector, we begin with the definition of an allowable perturbation.

DEFINITION 5.2 (Massey-constrained perturbation). *Allow team j an additional ϵ of a win against team i, where $i < j < n$. In this new matchup, team j outscores team i by β points. Thus, the perturbed Massey matrix $\tilde{\mathbf{M}}$ and right-hand-side vector $\tilde{\mathbf{p}}$ are related to the original nonsingularity-adjusted matrix $\bar{\mathbf{M}}$ and vector $\bar{\mathbf{p}}$ by*

$$\begin{aligned} \tilde{m}_{ij} &= \bar{m}_{ij} - \epsilon, \\ \tilde{m}_{ji} &= \bar{m}_{ji} - \epsilon, \\ \tilde{m}_{ii} &= \bar{m}_{ii} + \epsilon, \\ \tilde{m}_{jj} &= \bar{m}_{jj} + \epsilon \end{aligned} \qquad \textit{and} \qquad \begin{aligned} \tilde{p}_i &= \bar{p}_i - \beta, \\ \tilde{p}_j &= \bar{p}_j + \beta. \end{aligned}$$

Or, more compactly,

$$\tilde{\mathbf{M}} = \bar{\mathbf{M}} + \epsilon(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T \qquad \textit{and} \qquad \tilde{\mathbf{p}} = \bar{\mathbf{p}} - \beta\,(\mathbf{e}_i - \mathbf{e}_j).$$

Invoking the Sherman–Morrison update formula once again, we find that

$$\tilde{\mathbf{M}}^{-1} = \bar{\mathbf{M}}^{-1} + \frac{\epsilon}{n(n+2\epsilon)}(\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T$$

and

$$\tilde{\mathbf{r}} = \mathbf{r} + \frac{-\beta - \epsilon\,(j-i)}{n+2\epsilon}\,(\mathbf{e}_i - \mathbf{e}_j).$$

Notice that, like a Colley perturbation, a Massey perturbation to the perfect season affects only the rating of the two teams i and j involved in the perturbation. As a result, the Massey perturbations to the perfect season also have an *isolated effect*.

5.2.1. The case of fractional values of ϵ. With this closed-form expression for $\tilde{\mathbf{r}}$ we can now find the value of ϵ that causes a change in the rank ordering of these n items. The event is that a lower ranked team j upsets a higher ranked team i in an additional ϵ of a game. Though we can study changes in the ranking of both team i and team j, here we focus only on team j's ability to move up the ranked list. What ϵ is necessary to move the jth team k positions up the ranked list? That is, what ϵ forces $\tilde{r}_j > \tilde{r}_{j-k}$? Isolating the j and $j-k$ elements of $\tilde{\mathbf{r}}$ and using algebra, we discover that any value of ϵ or β following the rule below achieves our goal:

$$\boxed{\text{Massey perturbation:} \qquad \epsilon > \frac{kn-\beta}{j-i-2k} \qquad \text{or} \qquad \beta > kn - \epsilon\,(j-i-2k).}$$

Notice that a Massey-constrained perturbation between teams i and j involves two related parameters ϵ and β. In order to simplify the analysis and make clear conclusions, we fix one parameter at a time and reach the following conclusions.

Remark 3. *If $\epsilon = 0$, $\beta = kn$.* If no additional games are played, yet the point differential between teams i and j is increased by $\beta = kn$ or more points in favor of

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

team j, then team j moves k positions up in the ranking and the ranking is altered. Notice that, in this scenario, the value of β is independent of i and j.

Remark 4. *If* $\epsilon = 1$*,* $\beta = k(n+2) - (j-i)$. If one additional game is played between teams i and j and team j wins by $\beta = k(n+2) - (j-i)$ or more points, then team j moves k positions up the ranking. The size of the point differential parameter β depends on the position of the two teams. If the two teams are very evenly matched so that $(j-i) = 1$, then $\beta = k(n+2) + 1$. On the other hand, if the two teams are very unevenly matched so that $j - i = n - 1$ (in other words, j pulls off a magnificent upset), then $\beta = (k-1)(n-1) + 3k$. Thus, it takes only a point differential of 3 points of weak team j over strong team i to propel team j one position up the ranked list. Similarly, it takes a point differential of $n + 5$ points of weak team j over strong team i to propel team j two positions up the ranked list.

In all these cases, notice that the size of β grows linearly with n. Thus, if the collection of teams grows (i.e., $n \to \infty$), then $\beta \to \infty$, showing that it takes a larger and larger value of β to affect the rank ordering. Thus, *for the perfect season, the Massey model is more rank stable than the Colley method.*

5.2.2. The case of $\epsilon = 1$. Again, it is more natural to consider perturbations ϵ that are integer, so we begin with the $\epsilon = 1$ case, in which team j upsets the stronger team i, where $j > i$. When team j upsets team i, its rating increases by $\frac{\beta+(j-i)}{n+2}$ and team i's rating decreases by this same amount. If this increment is larger than the usual Massey increment of 1, team j moves one position up the ranked list and team i moves one position down. We consider two types of possible upsets, one at each extreme. First, consider the huge upset that occurs when the worst team upsets the best team, i.e., $j = n$ and $i = 1$. In this upset, for any $\beta > 3$, the worst place team will move up one position and the first place team will move down one position. Notice that this finding is independent of n. Further, notice that the point differential in the upset of j over i must be greater than 3 points for team j to move just one position up the ranked list.

This point differential parameter β in the Massey model gives additional interpretability that is absent from the bias-free Colley method, which considers only wins and losses. In particular, for the perfect season, if we allow β to increase to greater values, we find that it is possible for the weaker team j to upset the stronger team i by a big margin of victory and propel itself several positions up the ranked list. Specifically, if $\beta > (k-1)(n-1) + 3k$, team $j = n$ can move k positions up the ranked list with its upset of team $i = 1$.

The second type of upset is at the other extreme, the very mild "upset" that occurs when j beats i for $j = i + 1$. In this scenario, if $\beta > n + 1$, team j moves one position up the ranked list. Compare this lower bound on β with the $\beta > 3$ lower bound from the huge upset case. Thus, when the upset is barely an upset ($j = i + 1$), team j must beat team i by a rather large amount to move just one position up the list, which reinforces the finding about the stability of the Massey method. Furthermore, for team j to jump k positions up the list, $\beta > k(n+2) - 1$, which means that team j must dramatically run up the score against team i in order to jump multiple positions up the list.

5.3. Perturbation analysis for the Markov method. To study the effect of a small change to the Markov input on the Markov rating vector, we begin with the definition of an allowable perturbation.

Definition 5.3 (Markov-constrained perturbation). *Allow team j an additional ϵ of a win against team i, where $1 < i < j \leq n$. Thus, team i casts ϵ votes for team j.*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

This creates a perturbed Markov voting matrix $\tilde{\mathbf{V}}$ *that is related to the original voting matrix* $\bar{\mathbf{V}}$ *by*

$$\tilde{\mathbf{V}} = \bar{\mathbf{V}} + \left(\frac{\epsilon}{i-1+\epsilon}\right) \mathbf{e}_i \left(\mathbf{e}_j^T - \frac{1}{i-1}\mathbf{f}^T\right),$$

where $\mathbf{f}^T = \begin{array}{c} \begin{array}{cccccccc} 1 & 2 & \dots & i-1 & i & \dots & n \end{array} \\ \begin{pmatrix} 1 & 1 & \dots & 1 & 0 & \dots & 0 \end{pmatrix} \end{array}.$

After one last application of the Sherman–Morrison update formula, use of recursive formulas, and pages of algebra, we arrive at the messy but useful equations below:

$$(\mathbf{I} - \alpha\tilde{\mathbf{V}})^{-1} = (\mathbf{I} - \alpha\bar{\mathbf{V}})^{-1} + \frac{\gamma\alpha\,(\mathbf{I} - \alpha\bar{\mathbf{V}})^{-1}\,\mathbf{e}_i\mathbf{d}^T(\mathbf{I} - \alpha\bar{\mathbf{V}})^{-1}}{1 - \gamma\alpha\,\mathbf{d}^T(\mathbf{I} - \alpha\bar{\mathbf{V}})^{-1}\mathbf{e}_i}$$

and

$$\begin{aligned} \tilde{\mathbf{r}}^T &= \mathbf{e}^T\,(\mathbf{I} - \alpha\tilde{\mathbf{V}})^{-1} \\ &= \mathbf{r}^T + \beta\, r_i \mathbf{h}^T, \end{aligned}$$

where the vector $\mathbf{h}^T$ and the scalars γ, β, and τ are defined as

$$\begin{aligned} \mathbf{h}^T = (\; & \overset{1}{\tfrac{1}{i-1}\tfrac{\alpha+i-2}{i-2}\tfrac{\alpha+i-3}{i-3}\cdots\tfrac{\alpha+2}{2}\tfrac{\alpha+1}{1}\,\tau} \;\Big|\; \overset{\dots}{\dots} \;\Big|\; \overset{i-k}{\tfrac{1}{i-1}\tfrac{\alpha+i-2}{i-2}\tfrac{\alpha+i-3}{i-3}\cdots\tfrac{\alpha+i-k}{i-k}\,\tau} \;\Big|\; \overset{\dots}{\dots} \;\Big| \\ & \overset{i-3}{\tfrac{(\alpha+i-3)(\alpha+i-2)}{(i-3)(i-2)(i-1)}\,\tau} \;\Big|\; \overset{i-2}{\tfrac{(\alpha+i-2)}{(i-2)(i-1)}\,\tau} \;\Big|\; \overset{i-1}{\tfrac{1}{j-1}\,\tau} \;\Big|\; \overset{i}{\tfrac{\alpha}{i}\tfrac{\alpha+i}{i+1}\cdots\tfrac{\alpha+j-2}{j-1}\,\tau} \;\Big| \\ & \overset{i+1}{\tfrac{\alpha}{i+1}\tfrac{\alpha+i+1}{i+2}\cdots\tfrac{\alpha+j-2}{j-1}} \;\Big|\; \overset{\dots}{\dots} \;\Big|\; \overset{j-2}{\tfrac{\alpha}{j-2}\tfrac{\alpha+j-2}{j-1}} \;\Big|\; \overset{j-1}{\tfrac{\alpha}{j-1}} \;\Big|\; \overset{j}{1} \;\Big|\; \overset{j+1}{0} \;\Big|\; \overset{\dots}{\dots} \;\Big|\; \overset{n}{0}\;), \end{aligned}$$

with $\gamma = \frac{\epsilon}{i-1+\epsilon}$, $\beta = \frac{\gamma\alpha}{1-\gamma\alpha(\frac{\alpha}{i}\cdot\frac{\alpha+i}{i+1}\cdot\frac{\alpha+i+1}{i+2}\cdots\frac{\alpha+j-2}{j-1})}$, and $\tau = \frac{\alpha(\alpha+i-1)(\alpha+i)\cdots(\alpha+j-2)}{i(i+1)\cdots(j-1)} - 1.$

The first observation regards which teams are affected by the event involving teams i and j. Recall that the Colley and Massey perturbations to the perfect season affect only the ratings of teams i and j, the teams involved in the event. Contrast this with the changes induced by the Markov method. Examining $\mathbf{h}^T$ shows that for $0 \le \alpha < 1$, a *Markov perturbation to the perfect season affects the ratings of all teams from* 1 *through* j. When $\alpha = 1$, the messy formula for $\mathbf{h}^T$ simplifies a bit and we conclude that a *Markov perturbation to the perfect season affects the ratings of all teams from* i *to* j. The formula for $\tilde{\mathbf{r}}^T$ above shows that the ratings for all teams from i to j are affected (in fact, boosted in most cases). Further, after normalization is applied so that $\tilde{\mathbf{r}}^T\mathbf{e} = 1$, the ratings of other teams (i.e., $k \ne [i, i+1, \dots, j-1, j]$) are diminished. This finding seems ripe for complaint from coaches. In a theretofore perfect season, if the last place team $j = n$ upsets team i, the ratings of all teams ranked above i are reduced by the unrelated event. In summary, one lone event can impact the rating of every other team because of the interdependent nature of the Markov system.

5.3.1. The case of fractional values of $\boldsymbol{\epsilon}$ when $\boldsymbol{\alpha = 1}$. With this closed-form expression for $\tilde{\mathbf{r}}$ we can now find the value of ϵ that causes a change in the rank ordering of these n items. Because of the complexity in the formula for the Markov perturbation, in this section we study the algebraically simplest case, $\alpha = 1$,

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

and conduct our analysis. In the later experimental section, we fix α at other values in its domain and conduct an experimental sensitivity analysis. For our theoretical sensitivity analysis, we study the $\alpha = 1$ case and begin with the event that a lower ranked team j upsets a higher ranked team i in an additional ϵ of a game. Though we can study changes in the ranking of teams i, j, and everything between, here we focus only on team j's ability to move up the ranked list. What ϵ is necessary to move the jth team k positions up the ranked list? That is, what ϵ forces $\tilde{r}_j > \tilde{r}_{j-k}$? After some algebra, we discover that any value of ϵ following the rule below achieves our goal:

$$\boxed{\text{Markov perturbation:} \quad \epsilon = \begin{cases} \dfrac{k(i^2-i)}{j(j-k)-k(i-1)}, & j-k < i, \\ \dfrac{k(i^2-i)}{j(j-k-1)-k(i-1)}, & i \le j-k < j. \end{cases}}$$

Remark 5. *For $k = 1$, if $i = j - 1$, $\epsilon = 1$.* When the two teams involved in the perturbation are closely matched and the collection of teams grows, the size of the perturbation ϵ approaches the scalar 1. Thus, it takes a bigger perturbation ϵ to change the rank ordering when teams are more evenly matched. Conversely, as the next remark shows, when a weak team suddenly beats a strong team, it takes a small perturbation to move that team up the ranked list.

Remark 6. *For $k = 1$, if $i = 2$ and $j = n$, $\epsilon = \frac{2}{n^2-2n-1}$. As $n \to \infty$, $\epsilon \to 0$.* When the two teams are very unevenly matched (j pulls off a magnificent upset), then it takes only a small fractional ϵ of a win of lowly team j over mighty team i to propel team j up the ranked list.

Notice that the size of ϵ is quadratically inversely proportional to n. Thus, as the collection of teams grows (i.e., $n \to \infty$), it takes a smaller and smaller value of ϵ to affect the rank ordering, especially in the tail of the Markov rating vector. Thus, *the Markov model is very sensitive to perturbations from the perfect season.*

5.3.2. The case of $\epsilon = 1$ when $\alpha = 1$. Again, it is more natural to consider perturbations ϵ that are integer, so we analyze the $\epsilon = 1$ case, in which team j upsets the stronger team i, where $j > i$. The question is, How far up the ranked list can team j propel itself with an upset over the higher ranked team i? Specifically, can team j jump k positions up the ranked list? The analysis for the Markov method is a bit more complicated because, unlike the Colley and Massey methods, a perturbation between teams i and j affects, at the very least, all teams in between these two. Using the rating perturbation formula derived above, we find that if

$$\frac{j(j-1)}{i^2+j+1} > k,$$

team j will jump k positions up the ranked list with its upset of team i.

As before, we consider the two extreme cases of upsets. In the huge upset case, the last place team $j = n$ upsets the first place team $i = 1$ and can jump

$$\left\lfloor \frac{n(n-1)}{n+2} \right\rfloor$$

positions up the list. Depending on the value of n, this jump can cover nearly the entire list! See Table 5.1. Because this implementation of the Markov method does

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Table 5.1
Jumps of size p in the Markov rating when team $j = n$ upsets team $i = 1$.

n	k
5	2
10	7
50	47
100	97
500	497

not consider point differentials, the last place team could upset the first place team by only 1 point (in a fluke event) and jump to nearly the top of the list, arguably unjustifiably so.

In the other extreme upset case, the mild upset that occurs when team $j = i + 1$ upsets its neighbor in the ranked list, it is impossible for team j to move even one position up the ranked list because

$$\frac{i^2 + i}{i^2 + i + 2} < 1.$$

These remarks have special significance when the items to be ranked are webpages. In this case, an upset occurs whenever a low ranked webpage gets an inlink from a higher ranked webpage. Once normalized, this lone inlink corresponds to a perturbation with a fractional value of ϵ. This event is precisely the mechanism behind several common link spamming techniques that aim to unjustly boost the rank of a particular page for mercantile reasons. The remarks of this section explain why rank-boosting and spamming techniques such as link farming are so effective in the Markov method of ranking. In the Markov method, if a low ranked page can convince a high ranked page to hyperlink to it, then this endorsement can catapult the lowly page several rungs up the ranked list. Our results indicate that Massey and Colley may be more resistant to link spamming than PageRank. Investigating and comparing the sensitivity of Colley, Massey, and PageRank against webpage link spamming would be an interesting topic of study that can be pursued in future work, motivated by the sensitivity issues that we have identified in the current paper.

5.4. Summary of the comparisons.

5.4.1. Isolated response to a perturbation. At this point we use numerical examples to summarize the findings of this section on perturbations to the perfect season. Our first major finding is that, for the perfect season, the Colley and Massey methods have an *isolated* response to a single perturbation between teams i and j. On the other hand, for the Markov method, the rating of every team will change as a result of the same perturbation. Consequently, the *ranking* of more than just two teams may change in the Markov system.

An example visually makes this very clear. Consider a league of $n = 10$ teams with the perturbation event that team $j = 10$ beats team $i = 1$ by $\beta = 3$ points. Figure 5.1 shows the ratings, and hence rankings, for all three methods of the 10 teams both before and after the perturbation. The black bars show the team rating for the perfect season, i.e, before the perturbation. The grey bars show the ratings after the dramatic upset. We highlight a few observations to be made from Figure 5.1:

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

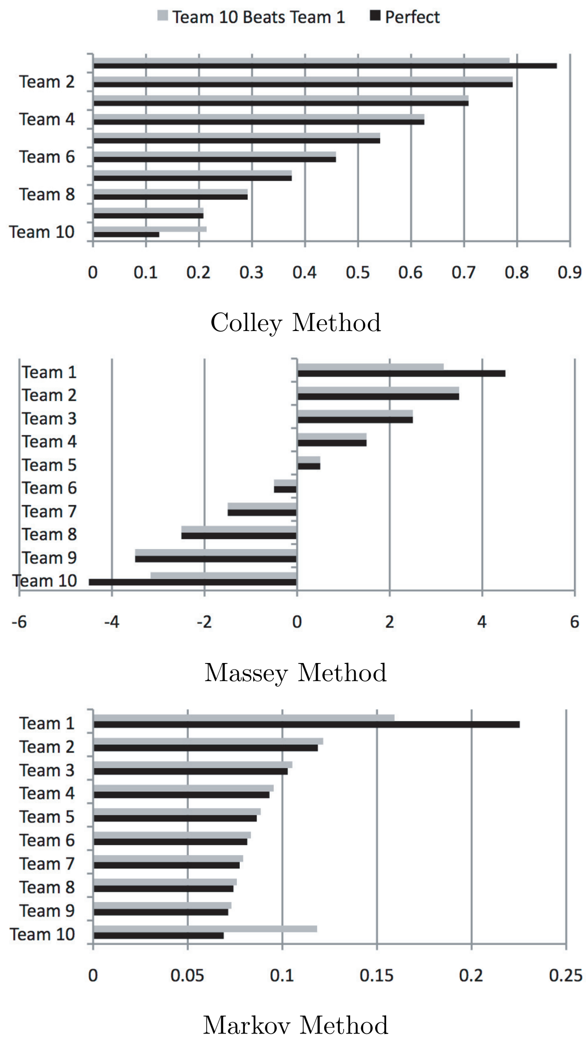

FIG. 5.1. *Comparison of methods on perturbation from the perfect season: team* 10 *upsets team* 1 *by* 3 *points. Black bars show the preperturbation ratings, and grey bars show the postperturbation ratings.*

1. Notice the nice step size pattern in the black bars for the Colley and Massey ratings of the perfect season. Compare this with the power law distribution of the Markov ratings on the perfect season. These were our findings from section 4.
2. Comparing the ranks of the teams before and after the perturbation, we see that for the Colley and Massey methods, team 1 drops below team 2 and team 10 rises above team 9. However, for the Markov method, team 10 is able to rise all the way above team 3 with their lone upset of the first place team. These were the section 5 findings regarding the sensitivity and rank stability of the methods.
3. Finally, notice that the Markov rating for every team does change slightly as a result of the lone event, whereas, in the Massey and Colley systems, only the ratings of the teams in the perturbation are affected.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

(a) Values of ε for Colley *as k = 1...10 and j = k+1...10*

k \ j	3	4	5	6	7	8	9	10
1	2.4	2	1.7	1.5	1.3	1.2	1.1	1
2	8	6	4.8	4	3.4	3	2.7	2.4
3		18	12	9	7.2	6	5.1	4.5
4			48	24	16	12	9.6	8
5				∞	60	30	20	15
6					∞	∞	72	36
7						∞	∞	∞
8							∞	∞
9								∞
10								

(b) Values of ε for Massey when β=5 *as k = 1...10 and j = k+1...10*

k \ j	3	4	5	6	7	8	9	10
1	∞	∞	5	2.5	1.7	1.3	1	1.8
2	∞	∞	∞	∞	15	7.5	5	3.8
3		∞	∞	∞	∞	∞	25	12.5
4			∞	∞	∞	∞	∞	∞
5				∞	∞	∞	∞	∞
6					∞	∞	∞	∞
7						∞	∞	∞
8							∞	∞
9								∞
10								

(d) Values of ε for Markov when α=1 *as k = 1...10 and j = k+1...10*

k \ j	3	4	5	6	7	8	9	10
1	1	.29	.14	.09	.06	.04	.03	.03
2	4	2	.5	.25	.15	.11	.08	.06
3		6	3	.67	33	.21	.14	.11
4			8	4	.8	.40	.25	.17
5				10	5	.91	.45	.29
6					12	6	1	.50
7						14	7	1.1
8							16	8
9								18
10								

(c) Values of β for Massey when ε=3 *as k = 1...10 and j = k+1...10*

k \ j	3	4	5	6	7	8	9	10
1	13	10	7	4	1	Cer	Cer	Cer
2	29	26	23	20	17	14	11	8
3		42	39	36	33	30	27	24
4			55	52	49	46	43	40
5				68	65	62	59	56
6					81	78	75	72
7						94	91	88
8							107	104
9								120
10								

FIG. 5.2. *Comparison of size of perturbation parameter ϵ on Colley, Massey, and Markov methods for fixed $i = 2$ and $n = 10$. For example, the $(2, 6)$-entry of Table* (a) *means that a value of $\epsilon > 4$ causes team $j = 6$ to jump $k = 2$ positions up the ranked list with an upset of team $i = 2$.*

5.4.2. Impact of the perturbation parameter. The tables of Figure 5.2 provide side-by-side comparisons of the impact of the size of the perturbation on the three methods. Consider again a league of $n = 10$ teams with the perturbation event that team j upsets team $i = 2$. For each method in turn, we consider the size of the perturbation parameter ϵ needed to move team j up the list k rank positions with their upset over team $i = 2$.

First, we consider the Colley method. From the Colley perturbation formula, we see that the size of the Colley perturbation ϵ depends on four parameters: i, j, k, and n. First, there is the event—lower ranked team j upsets team i ϵ times. Next, there is k, the number of positions up the list that team j hopes to jump as a result of these ϵ upsets of i. Last, there is n, the number of teams in the league. In this example, we fix two of the parameters, $i = 2$ and $n = 10$. Now, we vary both j and k and determine ϵ. The rounded results are displayed in table (a) of Figure 5.2. The (1,4)-element of this table means that in order for team $j = 4$ to move $k = 1$ position up the ranked list, ϵ must be greater than 2. In other words, team 4 must upset team $i = 2$ twice. Similarly, the (1,10)-element says that team 10 will move one position up the ranked list if it upsets team 2 once. Notice that elements in the table get bigger as we move down and to the right. For example, the (6,9)-element says that

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

team 9 must upset team 2 a total of 72 times in order to jump six positions up the list. Further, the infinite entries describe impossible events. Because there are very few values in the table that are less than or equal to 1, this example shows that the Colley method is very stable in the face of lone, single game upsets.

Next we conduct a similar analysis for the Massey method. However, the Massey method has a few more parameters to consider. The Massey perturbation formula contains 6 parameters (ϵ, β, n, i, j, and k) that we can vary in our analysis. To simplify the analysis, we fix several parameters, $i = 2$, $n = 10$, and the point differential $\beta = 5$. Table (b) of Figure 5.2 displays ϵ, the number of times team j would need to upset team $i = 2$ in order to achieve the desired outcome of jumping k positions up the ranked list. The $(2, 9)$-entry means that if team $j = 9$ beat team $i = 2$ a total of $\epsilon = 5$ times by a total of $\beta = 5$ points, it would jump $k = 2$ positions up the ranked list. The position and abundance of entries with value infinity show that it is impossible for low ranked teams to make big rank jumps in the Massey method.

Next we vary a different Massey parameter. We fix $n = 10$, $i = 2$, $\epsilon = 3$ and this time focus on β, the cumulative point differential, while varying j and k. For example, the interpretation of the $(2, 4)$-entry of table (c) of Figure 5.2 is that if team $j = 4$ beat team $i = 2$ a total of $\epsilon = 3$ times by a cumulative point differential of $\beta = 26$ points, it would jump $k = 2$ positions up the ranked list. The `Cer` entries describe certain events. In other words, if team $j = 8$ beats team $i = 2$ a total of $\epsilon = 3$ times, it is guaranteed to move $k = 1$ position up the ranked list, regardless of the cumulative point differential β. Notice that the entries become large in size at the lower right of the table, meaning that in order for a low ranked team to achieve a large rank jump, not only must it upset the higher ranked team $\epsilon = 3$ times, but it must do so handily, by a very large point differential.

Last, we consider the impact of the perturbation parameter ϵ on the Markov method. The Markov perturbation formula contains 4 parameters (ϵ, i, j, and k) that we can vary in our analysis. As usual, we fix parameters $n = 10$ and $i = 2$ and focus on ϵ, the number of times team j needs to beat team $i = 2$ in order to jump k positions up the ranked list. Table (d) of Figure 5.2 displays ϵ as j and k vary. The (6,10)-element of the matrix says that if team $j = 10$ wins an additional .50 of a game against team $i = 2$, then it will jump $k = 6$ positions in the ranking.

It is striking to compare the magnitude of the elements in this Markov table to those of the Massey and Colley tables. Team 10 does not even have to win an entire game to jump 6 positions in the ranking, which shows the extreme sensitivity of the Markov method to small perturbations in the data and the reason link spamming is so effective on the Web. Contrast this with the Colley and Massey methods, where it is rarely ever even possible to make a 6-position rank jump.

6. Situational/scenario analysis. The analysis from sections 1–5 dealt only with the perfect season. Of course, perfect seasons are a true rarity in sports. In practice, nearly every season turns out to be imperfect with inconsistencies and upsets, which is part of the lure for sports fans. Consequently, in this section, we move from the most well-behaved system created by a perfect season to consider realistic seasons that are rife with deviations from perfection. Our major finding, that Markov rankings are sensitive to upsets from the most well-behaved system, is only more pronounced the further we move from the perfect season. To demonstrate this point, we conduct scenario analysis, a process that is described below for a general season and then applied to one particular football season.

Scenario analysis typically begins as the regular season comes to a close and

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

football teams jockey for position for playoff slots (e.g., in the NFL) or invitations to prestigious bowl games (e.g., in NCAA football). Some teams are lock-ins, while others are "on the bubble." A bubble team has the potential to receive an invitation, provided they finish the season well and key opponents finish poorly. This is typically the part of the season in which coaches and fans engage in scenario analysis. For instance, with one game left in the regular season, a coach many wonder how his team's ranking will change as a result of the outcome of the final game. In particular, the coach of a bubble team might ask, How many points must we beat this final opponent by in order to move to rank position k (where k presumably guarantees an invitation)? Conversely, a coach might wonder if it is possible to rest his starting players to keep them injury-free, and potentially even lose the game, and still receive an invitation.[2]

Monday Night Football games in the NFL fit nicely into both scenario analysis and our perturbation analysis. For Monday Night games, all league games have been played that week but one—the Monday Night event. Thus, we compute rankings for all teams both before and after the Monday Night event and observe the changes induced by the lone game, which can be mathematically modeled as a rank-one update to the coefficient matrix. Unlike the BCS, such rankings would not bear weight on the NFL's ranking of the teams. Nonetheless, rankings can be used to predict future outcomes. We found a particularly dramatic upset that occurred during the Monday Night game on 10/13/2008, where very low ranked Cleveland beat the high ranked NY Giants by a score of 35 to 14. Figure 6.1 uses bipartite graphs to show the difference in the rankings of each method both before and after the Monday Night event.

The Colley rankings show that Cleveland moves up 1 position with its upset, while the NY Giants drop 8 positions. Because we are now in the imperfect season case, the effect of this Monday Night event is no longer isolated and a few other teams (e.g., Seattle, Green Bay, San Francisco, and New England) are also affected in minor ways.

The Massey rankings reveal similar findings. Cleveland's rank is boosted, while the Giants' drops. However, because point scores are considered and the upset was a dramatic 35-14 score, the rank changes are more dramatic. Cleveland jumps up 3 positions, while the Giants drop 10 positions. And as happened with the Colley rankings, the ranks of some other teams are affected as well. Recall that the Massey method uses point score information, while the Colley method uses only win-loss information, which explains the differences in the two rankings shown above.

The Markov rankings are the most interesting. They demonstrate our findings regarding the method's sensitivities and further reveal some very odd behavior. The most disturbing observation from Figure 6.1 is the fact that for the $\alpha = 1$ Markov method the Giants' rank actually increases one position as a result of their loss to the lowest ranking team of Cleveland! This bizarre behavior is a result of the Markov method's extreme interdependence of its rankings.

We also provide the Markov ratings for three other values of the teleportation parameter α. Recall our finding from section 4.6 that α can control the bend of the Markov power law curve. As $\alpha \to 0$, the Markov curve is less pronounced, while as $\alpha \to 1$, the tail of the curve grows, and this tail was the main contributor to the sensitivity issues of section 5.3. For the perfect season, while any value of $\alpha > 0$

[2]This scenario analysis is reminiscent of the analysis many students conduct at the end of a semester. Many students ask, What grade must I make on the final exam to get an A in the class? Or, conversely, what grade must I make to keep a C?

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

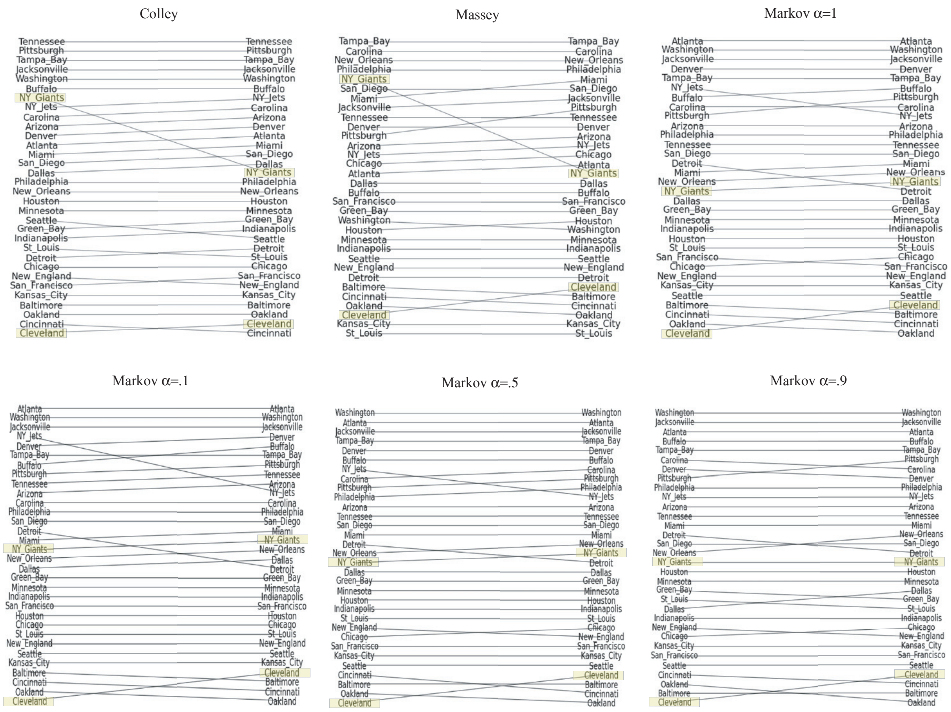

FIG. 6.1. *Colley, Massey, and Markov rankings before (left) and after (right) the Monday Night game.*

maintains the correct ranking, a very low value of α makes the method less sensitive. However, for imperfect seasons, choosing the most appropriate value of α is a delicate balancing act. Ideally, one should set α low enough to reduce sensitivity, yet high enough to maintain the ranking given by the dominance relationships. If α is too low, then the ranking is skewed too heavily toward the uniform ranking from the $1/n\, \mathbf{e}\mathbf{e}^T$ half of the Markov matrix. The Markov bipartite graph for $\alpha = .1$ in Figure 6.1 shows this. Its rankings, both pre- and postperturbation, vary wildly from the Colley and Massey rankings. If α is too high, then the ranking leans toward the power law ranking from the $\bar{\mathbf{V}} + 1/n\, \mathbf{e}_1\mathbf{e}^T$ half of the Markov matrix and is more susceptible to sensitivity. Hand tuning with trial and error to find a suitable teleportation parameter α is impractical. Search engines face the same dilemma—some argue for high values of α and others for low values [19].

7. Implications for ranking applications.

7.1. Top-k lists. Some ranking applications are very concerned with items ranked at the top of the lists and are less concerned or completely unconcerned with the remaining items. For instance, consider the ranked list associated with an election. In most political elections, only the top ranked, or first place, item matters as that candidate becomes the elected official. In other smaller elections, for instance for school organizations, sometimes the top three vote getters fill the three open positions of president, vice president, and secretary, in which case the focus is on the top-k list,

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

where $k = 3$.

Web search engines are another application of top-k rankings. For a query on "2008 U.S. presidential election," first all webpages using this phrase are ranked according to some predetermined measure. Then the top-k pages are presented to the user. For example, the first page of Google results present the top-10 pages, the second page the next 10, and so on. Even though users typically scan only the first few pages of search results, Google needs accuracy in their PageRank measure well beyond just a top-30 list. This is because the scores for only relevant pages (i.e., those using the search terms) are pulled from the full PageRank vector and sorted. For esoteric queries on obscure terms, it often happens that the relevant pages all have PageRanks in the tail of the vector. Thus, sensitivities in the tail of the PageRank vector can impact the sorting of these PageRank scores.

7.2. Full lists. There are some instances in sports tournaments that require accuracy in the full ranking of the teams. Consider the ESPN Challenge associated with the annual NCAA basketball tournament. Here participants must complete a bracket predicting the outcome of 68 matchups among the 68 teams (from over 340 NCAA teams) that are invited to attend the tournament. Because both automatic and invited bids are used, the teams range greatly in ability. Consequently the 68 tournament teams are not the top 68 teams in the nation that year. In fact, sometimes teams in the tail of a rating distribution appear in the tournament as automatic bids. Thus, a participant submitting a bracket in such a sports ranking application, unlike the ranking for most webpage queries, is concerned with the full ranked list, including values in the tail of the distribution. All must be accurate.

Here is a scenario that helps explain the issue. Suppose a team from a weak conference plays an early season game against a team from a strong conference and just barely pulls off an upset that is considered by nearly all fans and analysts to be a fluke. The coach for the team from the strong conference could have used this nonconference game to try out various defenses or player rotations or perhaps to discipline a few players. Regardless, the outcome was a surprising win for the team from the weak conference. Recall that such a small perturbation to the Markov system causes the rating of all teams rated between the mighty team i and the weaker team j to change. Thus, this fluke could cause the teams ranked slightly above j to move significantly (though arguably unjustly) up the ranking.

Sensitivity to such fluke events makes the Markov method perform poorly when ranking full lists as required by the ESPN Tournament Challenge. In fact, repeated experiments substantiate this claim. The Markov method does a poor job at predicting winners of games when compared to other ranking methods, particularly the Massey and Colley methods [13, 11, 12, 14, 20, 21, 23].

8. Conclusion and future work. Ranking methods rate items based on a set of numerical data. In some cases, such datasets are very large. Search engines rate, and hence rank, billions of webpages. Sports enthusiasts rank tens or hundreds of teams depending on the context. With such a large amount of data, one might understandably simply employ a method to produce a ranking without sifting through the data to verify that one agrees with the method's conclusion. Indeed, removing the need for such hands-on data mining is part of the purpose of such numerical methods. Still, a ranking method is valuable only if found useful. For instance, a search engine could develop a new ranking method, but it would be deemed helpful only if one found the rankings to be useful. In this paper, we have seen that the Colley and Massey methods are insensitive, while the Markov method is sensitive, particularly in

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

its tail. Because the effects of this sensitivity can result in questionable rankings, the Markov method should be carefully studied and, when possible, tailored to a particular application before being used. For instance, in section 6, imagine the public reaction if a loss by the Giants to the much lower ranked Cleveland on that Monday night in 2008 resulted in a higher ranking for the Giants in the newspaper on Tuesday morning. Clearly, a ranking was produced, but its usefulness is highly debatable.

This paper underscores a major difference in the Colley and Massey methods in comparison to the Markov method. As noted at the beginning of this paper, Colley and Massey have widespread use in sports ranking, whereas the Markov method has widespread use in ranking webpages. Investigating and comparing the sensitivity of Colley, Massey, and PageRank against webpage link spamming would be an interesting topic of study that can be pursued in future work, motivated by the sensitivity issues that we have identified in the current paper. For example, how would the Colley and Massey methods perform in ranking webpages? Could they deliver rankings that are less sensitive to the types of anomalies we and others have perceived in PageRank? Might such rankings be perceived as more useful?

REFERENCES

[1] D. Achlioptas, A. Fiat, A. Karlin, and F. McSherry, *Web search via hub synthesis*, in Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science, 2001, pp. 611–618.

[2] S. Allesina and M. Pascual, *Googling food webs: Can an eigenvector measure species' importance for coextinctions?*, PLoS Computational Biology, 5 (2009), e1000494.

[3] K. Avrachenkov and N. Litvak, *The effect of new links on Google PageRank*, Stoch. Models, 22 (2006), pp. 319–331.

[4] Y. Azar, A. Fiat, A. R. Karlin, F. McSherry, and J. Saia, *Spectral analysis of data*, in Proceedings of the 33rd Annual ACM Symposium on Theory of Computing, 2001, pp. 619–626.

[5] A.-L. Barabási, *Linked: The New Science of Networks*, Plume, New York, 2003.

[6] A.-L. Barabási, R. Albert, and H. Jeong, *Scale-free characteristics of random networks: The topology of the World-Wide Web*, Phys. A, 281 (2000), pp. 69–77.

[7] J. C. Borda, *Memoire sur les elections au scrutin*, in Histoire de l'Académie Royale des Sciences, Paris, 1781.

[8] R. A. Bradley and M. E. Terry, *Rank analysis of incomplete block designs:* I. *The method of paired comparisons*, Biometrika, 39 (1952), pp. 324–345.

[9] S. Brin, L. Page, R. Motwami, and T. Winograd, *The PageRank Citation Ranking: Bringing Order to the Web*, Technical report 1999-0120, Computer Science Department, Stanford University, Stanford, CA, 1999.

[10] T. Callaghan, P. J. Mucha, and M. A. Porter, *Random walker ranking for NCAA division I-A football*, Amer. Math. Monthly, 114 (2007), pp. 761–777.

[11] T. P. Chartier, E. Kreutzer, A. N. Langville, and K. E. Pedings, *Bracketology: How can math help?*, in Mathematics and Sport, Dolciani Math. Exp. 43, J. Gallian, ed., Mathematical Association of America, Washington, DC, 2010, pp. 55–70.

[12] T. P. Chartier, E. Kreutzer, A. N. Langville, and K. E. Pedings, *Sports ranking with nonuniform weighting*, J. Quant. Anal. Sports, to appear.

[13] T. P. Chartier, E. Kreutzer, A. N. Langville, K. Pedings, and Y. Yamamoto, *Minimum violations sports ranking using evolutionary optimization and binary integer linear program approaches*, in Proceedings of the Tenth Australian Conference on Mathematics and Computers in Sport, A. Bedford and M. Ovens, eds., MathSport (ANZIAM), New South Wales, Australia, 2010, pp. 13–20.

[14] T. P. Chartier, A. N. Langville, and P. Simov, *March madness to movies*, Math Horizons, April (2010), pp. 16–19.

[15] W. N. Colley, *Colley's Bias Free College Football Ranking Method: The Colley Matrix Explained*, http://www.colleyrankings.com/matrate.pdf (2002).

[16] H. A. David, *Ranking from unbalanced paired-comparison data*, Biometrika, 74 (1987), pp. 432–436.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[17] H. A. David, *The Method of Paired Comparison*, Oxford University Press, New York, 1988.
[18] C. Gao, X. Dang, Y. Chen, and D. Wilkins, *Graph ranking for exploratory gene data analysis*, BMC Bioinformatics, 10 (2009), article 519.
[19] D. Gleich, *Models and Algorithms for PageRank Sensitivity*, Ph.D. thesis, Stanford University, Stanford, CA, 2009.
[20] N. Goodson and C. Stephenson, *March Mathness: Sports Ranking Using Linear Algebra*, 2008.
[21] A. Y. Govan, *Ranking Theory with Application to Popular Sports*, Ph.D. thesis, North Carolina State University, Raleigh, NC, 2008.
[22] A. Y. Govan, A. N. Langville, and C. D. Meyer, *Offense-defense approach to ranking team sports*, J. Quant. Anal. Sports, 5 (2009), pp. 1–17.
[23] L. Ingram, *Ranking NCAA Sports Teams with Linear Algebra*, Master's thesis, College of Charleston, Charleston, SC, 2007.
[24] J. P. Keener, *The Perron–Frobenius theorem and the ranking of football teams*, SIAM Rev., 35 (1993), pp. 80–93.
[25] M. G. Kendall and B. B. Smith, *On the method of paired comparisons*, Biometrika, 31 (1940), pp. 324–345.
[26] P. Kvam and J. S. Sokol, *A logistic regression/Markov chain model for NCAA basketball*, Naval Res. Logist., 53 (2006), pp. 788–803.
[27] H. C. Lee and A. Borodin, *Perturbation of the Hyperlinked Environment*, Lecture Notes in Comput. Sci. 2697, Springer-Verlag, Heidelberg, 2003.
[28] R. Lempel and S. Moran, *Rank-stability and rank-similarity of link-based web ranking algorithms in authority-connected graphs*, in Proceedings of the Second Workshop on Algorithms and Models for the Web-Graph (WAW 2003), Budapest, Hungary, 2003.
[29] R. Llull, *Artifitium electionis personarum*, 1283.
[30] R. Llull, *De arte eleccionis*, 1283.
[31] K. Massey, *Statistical Models Applied to the Rating of Sports Teams*, Master's thesis, Bluefield College, Bluefield, VA, 1997.
[32] A. Y. Ng, A. X. Zheng, and M. I. Jordan, *Link analysis, eigenvectors and stability*, in Proceedings of the Seventh International Joint Conference on Artificial Intelligence, 2001.
[33] A. Y. Ng, A. X. Zheng, and M. I. Jordan, *Stable algorithms for link analysis*, in Proceedings of the 24th Annual International ACM SIGIR Conference, 2001.
[34] G. Pandurangan, P. Raghavan, and E. Upfal, *Using PageRank to characterize Web structure*, in Proceedings of the Eighth Annual International Computing and Combinatorics Conference (COCOON), 2002.
[35] J. Park and M. E. J. Newman, *A network-based ranking system for US college football*, J. Stat. Mech. Theory Exp., October (2005), P10014.
[36] C. Redmond, *A natural generalization of the win-loss rating system*, Math. Mag., 76 (2003), pp. 119–126.
[37] T. H. Wei, *The Algebraic Foundations of Ranking Theory*, Ph.D. thesis, Cambridge University, Cambridge, UK, 1952.

Downloaded 11/06/12 to 128.233.210.97. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.