THE RANKABILITY OF WEIGHTED DATA FROM PAIRWISE COMPARISONS

PAUL E. ANDERSON
Department of Computer Science and Software Engineering
California Polytechnic State University
San Luis Obispo, CA, USA

TIMOTHY P. CHARTIER
Department of Mathematics and Computer Science
Davidson College
Davidson, NC, USA

AMY N. LANGVILLE* AND KATHRYN E. PEDINGS-BEHLING
Mathematics Department
College of Charleston
Charleston, SC, USA

(Communicated by Michael W. Berry)

Abstract. In prior work [4], Anderson et al. introduced a new problem, the rankability problem, which refers to a dataset’s inherent ability to produce a meaningful ranking of its items. Ranking is a fundamental data science task with numerous applications that include web search, data mining, cybersecurity, machine learning, and statistical learning theory. Yet little attention has been paid to the question of whether a dataset is suitable for ranking. As a result, when a ranking method is applied to a dataset with low rankability, the resulting ranking may not be reliable.

Rankability paper [4] and its methods studied unweighted data for which the dominance relations are binary, i.e., an item either dominates or is dominated by another item. In this paper, we extend rankability methods to weighted data for which an item may dominate another by any finite amount. We present combinatorial approaches to a weighted rankability measure and apply our new measure to several weighted datasets.

1. Introduction. In [4], Anderson et al. posed the rankability problem as a fundamental yet little studied area of ranking. The objective in ranking is to sort objects in a dataset according to some criteria whereas the objective in rankability is to assess that dataset’s ability to produce a meaningful ranking of its items.

There are many ways to represent relationships between items that yield meaningful orderings. In this paper, we study the dataset’s ability to produce a meaningful ordering over all or a subset of the items using pairwise dominance information. Pairwise information about items may be represented by a directed graph or matrix [28]. Anderson et al. presented a rankability measure for unweighted (or uniformly

---

**Mathematics Subject Classification.** Primary: 90C08, 90C10, 52B12; Secondary: 90C35.

**Key words and phrases.** Ranking, rankability, linear program, integer program, combinatorial optimization, relaxation.

* Corresponding author: Amy N. Langville.
weighted) directed data. Ranking and rankability problems for unweighted data use binary dominance relations in a matrix $D$ where $d_{ij}$ is 1 if $i > j$ ($i$ dominates $j$) and 0, otherwise. A 1 in the $(i,j)$ position of the dominance matrix $D$ means that $i$ dominated $j$ by winning either a single event or the majority of its multiple events. Applications that create wins, losses, or draws without differential data create unweighted data. Binary survey data (product A is preferred over product B) is an example of unweighted data.

The purpose of this paper is to extend rankability to weighted dominance matrices. Many applications provide a margin of belief of how much better item $A$ is than item $B$. One obvious example is the final score in sports, which provides a margin of victory or a point differential from a matchup between the two teams. In this case, the teams are the items and the scores provide the dominance relationships between pairs of items. For another example, consider surveys that use star ratings (e.g., hotel A has 5 stars while hotel B received only 2 stars). In this case, the items are hotels and the score was 5 to 2. There are many ways to create a dominance matrix from such weighted data. Several are described in the next section.

2. Modeling data as a dominance matrix: Defining $D$. There is an art to modeling pairwise data to create a data matrix $D$ and this modeling can affect the success or failure of the ranking and, in turn, the rankability. Some domains have well-established methods for defining weighted pairwise connections between pairs (e.g., input-output matrices from economics [22]). For other applications, the transformations are not as well-defined and may vary by domain expert.

In this paper, we demonstrate how to prepare source data for rankability analysis where the pairwise relationship between items $i$ and $j$ is modeled as the degree of support (or contradiction) that item $i$ is ranked above item $j$. This pairwise dominance matrix $D$ is a square matrix and may be interpreted as an adjacency matrix for a weighted directed graph.

Tables 1, 2, and 3 show example data from economics, movies, and sports alongside the same data visualized as a weighted directed graph.

### Table 1. Sample Input/Output economic data based on Japan 2005 [22] (A) and its graphical representation in (B).

| From                  | To                        | Weight |
|-----------------------|----------------------------|--------|
| Plastic products      | Building construction      | 0.05   |
| Electricity           | Other business services    | 0.00   |
| Building construction | Electricity               | 0.04   |
| Railway transport     | Building construction      | 0.03   |
| Other business services | Building construction    | 0.06   |
| Other business services | Service industry        | 0.01   |
| Plastic products      |                            | 0.01   |

Sports data contains direct matchups between pairs of teams. Often, the data will include multiple matchups between the same pair of teams. When there are multiple matchups with $i > j$, $d_{ij}$ contains the aggregate, (e.g., sum, average, weighted average) of these events. Contrast this with movie rating data, for which direct matchups between movies must be built from the movie ratings. This is
Table 2. Sample of movie rating data based on MovieLens [19] (A) and graphical representation of user rating data transformed into pairwise comparisons (B).

(A)  

| User ID | Title        | Rating | Genre   |
|---------|--------------|--------|---------|
| 1       | Spellbound (1945) | 3   | Mystery |
| 1       | Cop Land (1997)    | 4   | Mystery |
| 1       | L.A. Confidential (1997) | 5   | Mystery |
| 2       | Spellbound (1945) | 4   | Mystery |
| 2       | Cop Land (1997)    | 2   | Mystery |

(B)  

Table 3. Sample of NCAA Men’s Basketball games is shown in (A) and the graphical representation of the aggregate dominance information, i.e., D is shown in (B).

(A)  

| Team 1 | Team 2 | Team 1 Score | Team 2 Score |
|--------|--------|--------------|--------------|
| 0      | UCLA   | Arizona      | 80           | 79           |
| 1      | UCLA   | Arizona      | 81           | 70           |
| 2      | Arizona| Delaware St  | 82           | 74           |
| 3      | Arizona| Delaware St  | 83           | 81           |
| 4      | Howard | Delaware St  | 93           | 84           |
| 5      | Howard | Delaware St  | 80           | 85           |
| 6      | Howard | Arizona      | 79           | 78           |

(B)  

typically done by finding, for each pair of movies, all users who rated both movies. If a user rated $i$ above $j$, then this is dominance information in the $i > j$ direction (i.e., $i$ is ranked above $j$). This can be seen in Table 2 where an edge weight of 2 is shown from Spellbound (1945) and Cop Land (1997) due the difference in rating from the second user.

In practice, data is often sparse when only direct information is used. Fortunately, indirect data can be added. Continuing with, but not limited to, our sports application, indirect dominance data is obtained through common opponents. In this case, suppose that team $i$ and team $j$ have never played and, hence, there is no direct dominance information between $i$ and $j$, but team $k$ has played both $i$ and $j$. As a result, this common opponent $k$ has valuable information about the relative strength of teams $i$ and $j$. There are three cases in which team $k$ believes that $i > j$:

1. team $k$ lost to $i$ but beat $j$,
2. team $k$ lost to both $i$ and $j$ but $k$ lost to $i$ by more than it lost to $j$, and
3. team $k$ beat both $i$ and $j$ and $k$ beat $j$ by more than it beat $i$.

Table 3 contains an example of indirect data. UCLA and Delaware State did not play each other directly yet a dominance assessment can be made through the common opponent of Arizona. Arizona believes that UCLA dominates Delaware State because UCLA beat Arizona who beat Delaware State (Case 1).
Each common opponent \( k \) in the set of all common opponents gets to cast votes, possibly weighted, for which team, \( i \) or \( j \), it believes is the stronger team. These indirect votes are then combined with the direct votes. Higher level indirect relations may be considered\(^1\) and if desired weighted with monotonically decreasing weights \( \alpha, \alpha^2, \alpha^3, \ldots \) for \( 0 < \alpha < 1 \), as, for example, is done by PageRank \([9, 10]\).

A user may also wish to provide a threshold for each of the three cases. In Case 1, a user may set a distance for the combined point differential that is deemed meaningful. An example is helpful. First, some notation: suppose \( s_{ij} \) is the number of points that team \( i \) scored against team \( j \). Then common opponent \( k \) only votes \( i > j \) when \( (s_{ik} - s_{ki}) - (s_{kj} - s_{jk}) \) is greater than some user-defined value, say 4, for this example. In this case, \( k \) does not cast a vote either way \((i > j \text{ or } j > i)\) if \( i \) beat \( k \) by two points and \( k \) beat \( j \) by 1. The spread must be 5 or more otherwise a spread this close is considered a tie in terms of the strength of the two teams. A user may set a different tie threshold for each case. Perhaps for Case 2, the user believes that the tie threshold for losses should be higher than, say, 10. Then in order for \( k \) to cast a vote that \( i > j \), \( i \) must have beat \( k \) by 10 more points than \( j \) beat \( k \). In general, we may define a spread threshold \((st)\) for the indirect case, a direct threshold \((dt)\) for the direct case, and a weighting parameter for combining indirect and direct dominance information \((w)\). In Section 9 we apply these ideas to building the dominance matrix \( D \) for the March Madness basketball tournament.

3. Overview of theory for weighted data. This paper’s rankability work extends two key ideas from unweighted data \([4]\) to weighted data.

- **Distance from perfection.** A scalar \( \delta \) is the distance that the input data \( D \) of pairwise dominance relations is from perfectly rankable data. This, of course, requires that we first define perfectly rankable data. Section 4 provides two such definitions.
- **Measure of uniqueness.** The scalar \( \rho \) is the number of rankings that are the optimal distance \( \delta \) from the given data \( D \). And the set of these rankings is denoted \( P \). We will later see in Section 6 that \( P \) is the set of all optimal rankings of an optimization model.

The distance from uniqueness \( \rho \) is an important yet seldom mentioned piece of the ranking. Consider the extreme cases. Perfectly rankable data is such in part because there is one and only one ranking of that data. On the other hand, for a constant uniform matrix \( D \), because each dominance relation has the same value, all possible \( n! \) rankings of the items are equally good (or bad) and thus the data is perfectly unrankable.

4. The standard of perfection for weighted data. In order to measure a distance from perfection for weighted data, we first need a definition of perfection for weighted data. In \([4]\), Anderson et al. define perfection for unweighted data as a dominance matrix in strictly upper triangular form (or a matrix that can be symmetrically reordered to such form). Is there an analogous standard of perfection for weighted data? Prior work by the linear ordering (LOP) community provides one answer (see Section 4.1) while prior work by Pedings et al. \([24]\) provides another (see Section 4.2).

\(^1\)We and others find that additional indirect relations are rarely worth the extra effort \([25]\).
4.1. **Triangularization and the degree of linearity.** For decades, the linear ordering community has been “triangularizing” weighted matrices like the dominance matrix $D$ of this paper. The goal in linear ordering is to find the ordering of items that maximizes the mass in the upper triangular part of the corresponding symmetrically reordered matrix. For the optimal ordering (or ranking), the fraction of the upper triangularized mass to the total mass is computed and is called the *degree of linearity*. A matrix with high degree of linearity has a high proportion of its mass in the upper triangle. This triangularization goal won Leontif a Nobel Prize when he applied it to the study of economic input-output matrices [26].

A matrix with maximal degree of linearity of 1 is a strictly upper triangular matrix and thus this becomes a definition of perfection for rankability. That is, a matrix that is (or can be reordered to be) in strictly upper triangular form is perfectly rankable. Returning to the sports application, this means that there exists an optimal ranking of teams in which the first place team was undefeated, the second best team was defeated only by the first team and defeated all others, and so on, down to the completely defeated last place team.

As Anderson et al. found with unweighted data, it is rare for real applications of weighted data to be (or be able to be reordered to be) perfectly triangularized. Thus, the degree of linearity can be used as a distance from perfection.

One formulation of the linear ordering problem (LOP) is Model (1) below.

$$
\begin{align*}
\max & \sum_{i \neq j} d_{ij} x_{ij} \\
\text{s.t.} & \quad x_{ij} + x_{ji} = 1 \quad \forall i < j \quad \text{(anti-symmetry)}
\end{align*}
$$

$$
\begin{align*}
x_{ij} + x_{jk} + x_{ki} & \leq 2 \quad \forall i < j, i < k, j \neq k \quad \text{(transitivity)} \\
x_{ij} & \in \{0, 1\} \quad \forall i \neq j \quad \text{(binary)}
\end{align*}
$$

The objective coefficients $d_{ij}$ are weights associated with the dominance relations between items $i$ and $j$. When the weight matrix has been transformed to normal form [28] (all entries are integral and nonnegative, diagonal elements are 0, and symmetric elements are complementary so that one is zero and the other, positive), the linear ordering problem is a weighted tournament with the goal of finding a minimum weight feedback arc set for this tournament [18, 28]. The distance $\delta$ from perfection is the sum of the components in $D$ minus the objective value of Model (1). Equivalently, the distance $\delta$ from perfection is the sum of the lower triangular elements of the optimally reordered dominance matrix. The number of optimal solutions is $\rho$, the distance from uniqueness. The lower the $\delta$ and $\rho$ are, the more rankable the data are.

4.2. **Hillside form and hillside count.** A stricter definition of perfection for weighted data comes from Pedings et al. [24]. They define a so-called **hillside form**.

**Definition 4.1.** A matrix $D$ is in hillside form if

$$
\begin{align*}
d_{ij} & \leq d_{ik} \quad \forall i \text{ and } \forall j \leq k \quad \text{(ascending order across the rows)} \\
d_{ij} & \geq d_{kj} \quad \forall j \text{ and } \forall i \leq k. \quad \text{(descending order down the columns)}
\end{align*}
$$

The name is suggestive as a 3D cityplot of a matrix in hillside form looks like a sloping hillside. The image on the right of Figure 1 shows a matrix in hillside form.
form. The matrix $D_1$ of weighted data below is in hillside form, while $D_2$ is not.

$$D_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 0 & 3 & 5 & 8 & 15 \\ 2 & 0 & 2 & 4 & 9 & \\ 3 & 0 & 0 & 3 & 6 & \\ 4 & 0 & 0 & 0 & 5 & \\ 5 & 0 & 0 & 0 & 0 & \end{pmatrix}$$

and

$$D_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 0 & 3 & 5 & 8 & 15 \\ 2 & 0 & 2 & 4 & 9 & \\ 3 & 7 & 0 & 3 & 4 & \\ 4 & 0 & 0 & 0 & 5 & \\ 5 & 0 & 0 & 0 & 0 & \end{pmatrix}$$

A matrix in hillside form (or one that can be symmetrically reordered to such form) has one unquestionable ranking of its items. For example, matrix $D_5$ says that not only is team 1 ranked above teams 2, 3, 4, and 5, but we expect team 1 to beat team 2 by some margin of victory, then team 3 by an even greater margin, and so on. For $n \times n$ matrices in hillside form, the ranking of the items is clear: $[1 \ 2 \ \cdots \ n]$. The next question becomes how to define distance from perfection, i.e., distance from hillside form. The Hillside Count method counts the number of violations of the hillside conditions of ascending rows and descending columns and denotes this as $\delta$, the distance from perfection. A matrix with more violations is farther from hillside form and thus less rankable than one with fewer violations. For example, the matrix $D_1$ above has 0 violations while $D_2$ has 7 violations. The 5 in the $(4,5)$ entry of $D_2$ accounts for one violation since it violates the hillside condition about the columns having descending values. The 7 in the $(3,1)$ entry accounts for the remaining 6 violations. There are two column-wise violations for the zeroes above and four row-wise violations because 7 is greater than the entries to the right. We will explain later that violations in the lower triangular are always weighted significantly larger than those in the upper triangular.

Often a matrix that appears to be non-hillside can be symmetrically reordered so that it is in hillside or near-hillside form. In fact, the non-hillside matrix $D_3$ shown below is the perfect hillside matrix $D_1$ when $D_3$ is reordered according to the vector $[4 \ 2 \ 5 \ 3 \ 1]$.

$$D_3 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 0 & 0 & 0 & 0 \\ 2 & 9 & 0 & 4 & 0 \\ 3 & 5 & 0 & 0 & 0 \\ 4 & 15 & 3 & 8 & 0 \\ 5 & 6 & 0 & 3 & 0 \end{pmatrix}$$

and reordered $D_3 = D_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 0 & 3 & 5 & 8 & 15 \\ 2 & 0 & 0 & 2 & 4 & 9 \\ 5 & 0 & 0 & 0 & 3 & 6 \\ 3 & 0 & 0 & 0 & 5 & \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$.

Typically after a data matrix has been reordered to be as close to hillside form as possible, violations remain. These violations are of two types: type 1 violations and type 2 violations. Type 1 violations violate transitivity in the ranking and manifest as nonzero entries in the lower triangular part of the reordered matrix. In the context of sports, type 1 violations correspond to upsets, i.e., when a lower ranked team beat a higher ranked team. On the other hand, type 2 violations violate the differentials required by hillside form. These violations occur in the upper triangular part of the matrix. In the context of sports, type 2 violations are weak wins, which occur when a high ranked team beats a low ranked team but by a smaller margin.

\[^2\] Another method, Hillside Amount, sums the total amount of violation from hillside form. That is, it sums the minimum amount that must be added or subtracted to elements of the reordered $D$ to make it hillside. Hillside Amount is an extension of the unweighted rankability model of [4] to weighted data and is described in [5].
of victory than expected. In the hillside method, an upset (i.e., type 1 violation) naturally accounts for more violations than a weak win (i.e., type 2 violation). In particular, violations further from the diagonal count more heavily than violations near the diagonal. It is possible to weight these two types of violations in other non-uniform ways if the modeler has a greater aversion to one type of violation over the other.

Finding the hidden hillside structure of a weighted dominance matrix was exactly the aim of [24]. The method of Pedings et al. finds a reordering of the items that when applied to the item-item matrix of weighted dominance data forms a matrix that is as close to hillside form as possible [24]. Figure 1 summarizes the method pictorially. The left is a cityplot of an $8 \times 8$ matrix in its original ordering of items. The right is a cityplot of the same data displayed with the optimal hillside ordering.

Pedings et al. use hillside form to find a minimum violations ranking of the items, the ranking with the minimum value $\delta$. In contrast, our goal is to focus on rankability information, rather than a ranking. Like Pedings et al. we use $\delta$, but we also use the other scalar $\rho$, the distance from uniqueness. In particular, $\rho$ is the number of rankings that, starting from $D$, are a distance of $\delta$ violations from hillside form.

Pedings et al. use the integer program of Model (2) to get $\delta$. The scalar $\rho$ is the number of optimal extreme point solutions of this integer program. Our contribution is a method for finding $\rho$ or information about $\rho$ without getting all optimal solutions of this model.

$$\begin{align*}
\min & \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} \\
\text{subject to} & \quad x_{ij} + x_{ji} = 1 \quad \forall i < j \quad \text{(antisymmetry)} \\
& \quad x_{ij} + x_{jk} + x_{ki} \leq 2 \quad \forall j \neq i, k \neq j, k \neq i \quad \text{(transitivity)} \\
& \quad x_{ij} \in \{0,1\} \quad \text{(binary)}
\end{align*}$$

The objective coefficient $c_{ij}$ is the number of hillside violations associated with item $i$ being ranked above item $j$. These objective coefficients $c_{ij}$ are built from the weighted input matrix $D$ of dominance relations and are defined as $c_{ij} := \# \{ k \mid d_{ik} < d_{jk} \} + \# \{ k \mid d_{ki} > d_{kj} \}$, where $\#$ denotes the cardinality of the corresponding set. Thus, for example, $\# \{ k \mid d_{ik} < d_{jk} \}$ is the number of teams receiving a lower
point differential against team $i$ than team $j$. Similarly, $\#\{k \mid d_{ki} > d_{kj}\}$ is the number of teams receiving a greater point differential against team $i$ than team $j$.3

For example, the dominance matrix $D_2$ above creates the hillside violation matrix $C_2$ below.

$$
C_2 = \begin{pmatrix}
1 & 2 & 3 & 4 & 5 \\
1 & 0 & 1 & 2 & 1 \\
2 & 5 & 0 & 1 & 0 \\
3 & 6 & 5 & 0 & 1 \\
4 & 6 & 6 & 5 & 0 \\
5 & 7 & 7 & 7 & 5
\end{pmatrix}.
$$

Entry $c_{1,2} = 1$ means that any ranking with team 1 above team 2 violates the dominance data once. This violation comes from indirect data, i.e., opponents who played both teams 1 and 2. In particular, team 3 played both teams 1 and 2 and from its outcomes (team 3 beat team 1 but lost to team 2), it believes $2 > 1$. Entry $c_{1,3} = 2$ means that any ranking with team 1 above team 2 violates the dominance data twice.

Figure 2 below is a pictorial representation of the difference between a less rankable (top half) weighted matrix and a more rankable (bottom half) matrix. The top half of Figure 2 corresponds to the 2008 Patriot league men’s college basketball season, which has rankability values of $\delta = 155$ and $\rho = 6$. The bottom half corresponds to the 2005 season, a much more rankable year with lower rankability values of $\delta = 92$ and $\rho = 4$.4 In each year, the left side shows the weighted dominance matrix $D$ with the original ordering and the right side shows an optimal hillside ordering output by the weighted rankability integer program of Model (2) above. In the top half, the less rankable year does not improve much from its original ordering to its optimal ordering. For that less rankable 2008 year, the right side, though optimal, is not great. Try as the integer program does, the data are just not very close to hillside form. Compare this with the more rankable 2005 data in the bottom half of Figure 2, a matrix that is much closer to hillside form. In other words, some data are just more rankable than others.

We conclude this section by noting that the two weighted rankability models, LOP Model (1) and Hillside Model (2) have the same constraint set and thus, the same polytope over which the objective function is optimized. Only the objective functions for the two models differ. This means that the theoretical results of the next section hold for both models. In what follows, we say rankability model and do not specify which, as either the LOP or Hillside may be used.

---

3The matrix $C = [c_{ij}]$ above counts hillside violations in a binary fashion, however, something more sophisticated can be done. For instance, weighted violations can be considered by summing the difference each time a hillside violation occurs. In this case, the entries of $C$ are defined as $c_{ij} := \sum_{k: d_{ik} < d_{jk}} (d_{jk} - d_{ik}) + \sum_{k: d_{ki} > d_{kj}} (d_{ki} - d_{kj})$. Going in the other direction, a user may decide to relax the hillside definition a bit so that weak wins are not penalized as much. In this case, the user can set a parameter $t$ for a so-called tie threshold, defining $c_{ij} := \#\{k \mid d_{ik} + t < d_{jk}\} + \#\{k \mid d_{ki} - t > d_{kj}\}$.

4Because the number of items and the number of matchups is roughly or exactly the same each season in this conference of college football, we can compare the two $(\delta, \rho)$ pairs without normalizing. For general comparisons, normalization is necessary. Anderson et al. [4] and Cameron et al. [11] propose a few normalizations. Such normalizations are usually domain-specific and require construction by a domain expert.
5. Related work. Some ranking methods produce both a ranking and a measure of the quality or confidence in that ranking that can possibly be related to a distance from perfection [20, 24, 13, 30, 7, 15, 3, 14, 12, 2, 28, 31]. For example, a ranking produced by a minimum violations or optimization method often produces an objective value that can be interpreted as $\delta$. For such minimum violations and optimization methods a distance from uniqueness $\rho$ or the set $P$ is typically not found, but usually can be found. In fact, the techniques of the next section for getting information about $\rho$ and $P$ can be adapted for many minimum violations and optimization methods, particularly those whose definition of decision variables are similar to those of the LOP or Hillside models above. Thus, rankability as defined in this paper is a framework that can be applied to many ranking methods. In so doing, the ranking community broadens its focus from simply the ranking produced and its quality to the (possible) non-uniqueness of the optimal rankings associated with the data.

6. Finding $\rho$ and $P$. In general, finding all optimal solutions for an integer program or linear program is known to be difficult. With default settings, solvers applied to the rankability integer program conclude with the optimal objective value $\delta$ and one solution matrix $X$ from which an optimal ranking can be built. However, most commercial solvers (e.g., Gurobi) have an option to output any other optimal solutions found along the way. When this option (e.g., in Gurobi, use the
PoolSearch option) is set, upon termination, the rankability integer program outputs $\delta$ and several $X$ matrices, each of which corresponds to an optimal ranking, and hence, a member of $P$. We call this set of rankings partial $P$ since we cannot be sure if it is the full set $P$, the set of all optimal rankings. If the full set of optimal rankings, $P$, is desired, programs such as SCIP are available [1]. These programs are tractable when the number of items is small or the number of optimal solutions is small. In practice, we are often only interested in a summary of all optimal solutions or a subset of that information. In this section, we show how to synthesize all optimal solutions into a solution matrix $X^f$ and then we further show this may be approximated by a solution matrix $X^* = \{1\}$ that is efficiently computed without obtaining a full or even a partial set $P$.

In practice we will not actually gather a full, or even a partial, $P$. For now to set the stage for the later approximation, suppose we did get the full set $P$. One method for summarizing the set $P$ is to calculate a fractional matrix $X^f$ whose element $X^f_{ij}$ measures the fraction of rankings in $P$ that rank item $i$ above item $j$. This matrix may be symmetrically reordered according to the vector $r$ obtained by sorting $X^f$ by its row sums. We demonstrate the calculation of the reordered matrix $X^f(r, r)$ on the 2005 season of the Big 12 conference in college football (Example 1.). Employing SCIP on Example 1, we find the set $P$ of all 8 optimal solutions, each having an optimal objective value of $\delta^* = 255$. These may be summarized to produce $X^f(r, r)$.
For Example 1, we built the exact fractional matrix $X^f$ from the full set $P$ of optimal rankings. This is intractable for many problems because, in practice, the set $P$ can be quite large. For example, SCIP was used to search for no more than 500,000 members of $P$ on an 8 core 64 GB virtual machine for the $n = 68$ 2002 NCAA March Madness season discussed in Section 9. SCIP took over 40 minutes to find these 500,000 optimal solutions. For the 2013 March Madness season, SCIP took over 8 hours to find a partial $P$ set of 500,000 members. In these cases, even if one had the computational resources to get all optimal solutions, this is rarely useful. What a user prefers instead is a summary of information about all solutions. We propose that a matrix called $X^*$ contains such summary information, approximates $X^f$, and can be found for the NCAA example above in less than 5 minutes. We now describe $X^*$. The theory presented in this section leads to a method that approximates the fractional elements of the exact $X^f$ matrix with the $X^*$ matrix of the linear programming (LP) relaxation of the weighted rankability model. This is done by employing the interior point, not the simplex, method.

First, we explain the interior point solver. An interior point solver is used because, when the polytope has an optimal face rather than an optimal point, its solution lies in the interior of the optimal face (and on or near the centroid if Mehrotra and Ye’s [29] interior point method is used). For our work, we prefer the optimal solution that is in the interior of the optimal face rather than an extreme point because, under certain conditions, it is a convex combination of all optimal extreme point solutions. Theorem 6.1 below shows that these optimal extreme points on the optimal face are the optimal rankings of the weighted rankability problem. In other words, the interior point solution can be considered a summary of all optimal rankings.

Interior point methods are designed for linear programs, not integer programs, so we solve the LP relaxation of the rankability problem. The LP weighted rankability polytope for the weighted rankability problem is defined as the anti-symmetry constraints ($x_{ij} + x_{ji} = 1$), the transitivity constraints ($x_{ij} + x_{jk} + x_{ki} \leq 2$), and the bound constraints ($0 \leq x_{ij} \leq 1$). Notice that the bound constraints are simply a relaxation of the binary constraints of the original integer program, and hence the name, LP relaxation. Even though the LP rankability polytope and the IP rankability polytope, defined as the convex hull of all feasible solutions of the rankability integer program, do not always define the same region, useful results regarding the IP rankability polytope can be gathered, as the following theoretical results show, from the LP rankability polytope. (Proofs are in the supplemental section.)

**Lemma 6.1.** Every ranking of a weighted rankability problem corresponds to a binary extreme point of the LP weighted rankability polytope.

**Corollary 1.** When $\delta_{I_P}^* = \delta_{I_P}^*$, every optimal ranking of a weighted rankability problem of Model (2) corresponds to a binary extreme point on the optimal face of the LP weighted rankability polytope.

When the LP relaxation of the interior point solver terminates, there are two options for the optimal objective value $\delta^*$ (integer and non-integer) and two options for the optimal solution matrix $X^*$ (binary and fractional\(^5\)) creating the following four outcomes.

0. $\delta^*$ is non-integer and $X^*$ is binary.

---

\(^5\)If $X^*$ contains at least one fractional value, we say it is fractional.
1. $\delta^*$ is non-integer and $X^*$ is fractional.
2. $\delta^*$ is integer and $X^*$ is binary.
3. $\delta^*$ is integer and $X^*$ is fractional.

Case 0 is actually not possible and therefore not an outcome because since $C$ being a sum of counts is integer and $X^*$ is binary, then the objective value $\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}^*$ must be integer. Case 1 means that the LP solution is not optimal for the IP. Our experiments show that Case 1, though possible, is unlikely. This is also supported by Anderson et al. [4] and Reinelt et al. [28, 31]. Case 2 means that $\rho = 1$, there is a unique optimal solution, and the LP solution is optimal for the IP. Case 3 is the most interesting to us as we can use its solution matrix $X$ to get information about the set $P$ of optimal solutions.

**Theorem 6.2.** If the Interior Point solver of the LP relaxed weighted rankability problem of Model (2) ends in Case 3 ($\delta^*$ is integer and $X^*$ is fractional) and the LP’s optimal face is the IP’s optimal face, then

1. $\delta^*$ is the optimal objective value for the integer program,
2. $X^*$ is on the interior of the optimal face (i.e., the convex hull of all optimal solutions) of the integer program, and
3. fractional entry $(i, j)$ in $X^*$ means that there exists at least one optimal ranking in $P$ with $x_{ij}^* = 1$ (thus, $i > j$) and at least one with $x_{ij}^* = 0$ (thus, $i < j$).

Being a relaxation, the LP optimal face can actually be larger than the IP’s optimal face. As a result, the second hypothesis of Theorem 6.2 (that the LP’s optimal face is equal to the IP’s optimal face) may not always hold. This condition is difficult to check. Fortunately, in practice, the centroid of the LP optimal face is near the centroid of the IP’s optimal face so that Theorem 6.2 is still useful.

When $X^*$, the interior point solution of LP relaxation of Model (2), is binary, $X^*(r, r)$ is a strictly upper triangular matrix and $r$ is an optimal ranking. In fact, $r$ is the lone member of $P$. On the other hand, when $X^*$ is fractional, the rowsum ranking $r$ may or may not be a member of $P$. Nevertheless, the reordering of $X^*$ induced by $r$ is very helpful because $X^*(r, r)$ is an approximately upper triangular matrix with deviations from the upper triangular structure that are noticeable and helpful as shown by the example of next section.

**6.1. Illustrative example.** We return to Example 1, weighted data from the Big 12 conference of college football. Due to space, we only show the Hillside rankability results; the LOP rankability results are similar. Figure 3 shows the optimal solution matrix $X^*$ output by the Interior Point solver of the linear programming relaxation of the weighted rankability problem. Notice this is a very good approximation to the exact $X^f$ shown a few pages earlier. The presence, location, and value of fractional elements of the approximate $X^*$ are a good match to those of the ideal exact $X^f$.

As mentioned above, in practice, the set of all optimal solutions is not often of interest; however, building the full $P$ set for Example 1 from $X^*(r, r)$ demonstrates the theory presented in this section. For small examples or examples for which the $X^*$ matrix contains a small number of fractional entries, the full $P$ set can often be obtained by examination of $X^*(r, r)$ as follows. The first row and column of $X^*(r, r)$ are binary, which means that the first item, item 10, belongs in the first rank position. There are no other candidates for this position. Similarly, the last row and column are binary, meaning that item 9 is the only choice for the final rank position. In addition, there is another binary structure in the matrix; notice
the binary cross near the center of the matrix, covering the bands corresponding to the rows and columns for items 6, 7, 11, and 4. This means that these items must appear in the sixth through ninth rank positions in that order. The remaining rank positions in $X^*(r,r)$ contain fractional values, which represent potential alternatives for the corresponding rank positions (see Theorem 6.2). For example, in the second and third rank positions, items can be ordered either 8 then 12 or 12 then 8. In the fourth and fifth rank positions, items 3 and 2 can be ordered in any of the $2!$ ways. Finally, the same thing happens in the tenth and eleventh rank positions with items 1 and 5. This creates a set of $2 \times 2 \times 2 = 8$ rankings that must be evaluated for optimality. In this case, all 8 rankings built from $X^*(r,r)$ are indeed optimal with an objective value of $\delta^* = 255$. And these are same 8 rankings found by SCIP and presented with Example 1 at the start of this section.

Section 9 uses a larger example, the March Madness basketball tournament, to show additional ways to use the summary information in $X^*$.

### 6.2. Lowerbound on $\rho$

In this section, we provide a lowerbound and thus, estimate, on $\rho$, the number of rankings in the set $P$ of all optimal rankings. This bound may be helpful for a large example that has a complicated highly fractional $X^*$ matrix. It also provides another piece of summary information about the rankability of a dataset.

![Figure 3. Approximate fractional matrix $X^*(r,r)$ for Example 1 obtained by the Interior Point solver of the linear programming relaxation.](image-url)
Theorem 6.3. If $X^*$ is the exact centroid of all optimal rankings for a weighted rankability problem, then
\[ \rho \geq \left\lceil \frac{1}{m} \right\rceil, \]
where $m$ is the smallest fractional element in $X^*$.

When the Interior Point LP solver concludes in Case 3 ($\delta^*$ integer, $X^*$ fractional) and the LP optimal face is the IP optimal face, then Theorem 6.2 showed that $X^*$ is a convex combination of all optimal rankings. When an Interior Point solver such as Mehrotra and Ye [29] is used even if the LP optimal face is larger than the IP optimal face, $X^*$ of the LP is near the centroid of the IP. While this is not the exact centroid required by the hypothesis of Theorem 6.3, it is close enough to give an estimate of a lowerbound. We found this to be the case when we compared the lowerbound estimate to the exact value of $\rho$ for several real datasets from the Big 12 conference of college football. For Example 1, $m = 0.47$ and $\rho \geq \left\lceil \frac{1}{0.47} \right\rceil = 3$ is a lowerbound of the true $\rho = 8$.

Corollary 2. If $X^*$ is the exact centroid of all optimal rankings for a weighted rankability problem, then fractional entry $(i,j)$ is the percentage of rankings in $P$ that have $i > j$ (and, hence, $X^* = X_f$ of Section 6).

For Case 3 (the LP concludes with an integer $\delta^*$ and fractional $X^*$), interior point methods conclude near the exact centroid and thus a fractional entry in the optimal solution is an approximation to the percentage of rankings in $P$ that have $i > j$.

7. Revisiting the unweighted problem. Anderson et al. designed rankability methods for unweighted graphs [4]. Yet the weighted methods from this paper can be applied to unweighted data, creating three rankability methods for unweighted data: the original method of [4], the LOP of Model (1) and the Hillside of Model (2). See reference [6] for more on this.

8. Diameter of $P$. This section presents another example and optimization model that argue for examining more information about the set $P$. The example comes from the unweighted data from the 1999 season of the ACC conference of college football. We run the original rankability method of Anderson et al., using the LP relaxation of the alternative formulation of [6] so that Theorem 6.2 and the ideas from Section 6 apply.

Example 2. The interior point solver produces an integer $\delta = 12$ and the following highly fractional matrix $X^*(r,r)$, which usually portends a large $\rho$ value, yet in this case $\rho$ is small, namely $\rho = 4$.

$$X^*(r,r) = \begin{pmatrix} 3 & 1 & 4 & 8 & 2 & 6 & 9 & 5 & 7 \\ 3 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & .36 & .73 & 1 & .62 & 1 & 1 \\ 4 & 0 & .64 & 0 & .36 & 1 & 1 & .64 & 1 \\ 8 & 0 & .28 & .64 & 0 & .64 & .40 & 1 & 1 \\ 2 & 0 & 0 & 0 & .36 & 0 & .26 & .64 & 1 \\ 6 & 0 & .38 & 0 & .10 & .74 & 0 & .38 & .74 \\ 9 & 0 & 0 & .36 & 0 & .36 & .62 & 0 & .36 \\ 5 & 0 & 0 & 0 & 0 & .26 & .64 & 0 & .64 \\ 7 & 0 & 0 & 0 & .36 & .62 & 0 & .36 & 0 \end{pmatrix}.$$
Even though the set $P$ shown below contains just 4 optimal rankings, it is very diverse. Items vary greatly in their rank positions. For instance, item 6 ranges from third place to the ninth and last place.

\[
P = \left\{ \begin{array}{ccc}
3 & 3 & 3 \\
8 & 4 & 4 \\
4 & 6 & 1 \\
6 & 1 & 2 \\
1 & 2 & 8 \\
2 & 8 & 5 \\
5 & 5 & 9 \\
9 & 9 & 7 \\
7 & 7 & 6
\end{array} \right\}.
\]

Figure 4 compares the $P$ sets of two examples, Example 1 and Example 2. Example 2 has 8 rankings in its $P$ set while Example 2 has just 4. The spaghetti plots show differences in neighboring rankings.\(^6\)

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{Spaghetti plots and summary of diversity of $P$ sets for Examples 1 and 2.}
\end{figure}

For Example 1, these differences are less dramatic and occur only between neighboring items in the rankings, e.g., items 8 and 12 swap as do items 1 and 5, and 2

\(^6\)A complete spaghetti plot would establish lines between all $\binom{10}{2}$ pairs of rankings. Since this is too messy as it requires 3-D plots, our point is made by using the incomplete 2-D spaghetti plots shown in Figure 4.
and 3. The relative position of items in the rankings appears rather definite. On the other hand, Example 2 has messier spaghetti plots. Notice also the average Kendall rank correlation between the two examples. Example 1’s rankings have a high rank correlation whereas Example 2’s rankings do not. This numerical indicator of the diversity of the two \( P \) sets corroborates the visual indicator. Example 2 also has a much higher percentage of fractional entries than Example 1. A high percentage of fractional entries in the optimal solution matrix can indicate a large \( p \) and/or a very diverse \( P \). In either case, the rankability should be low.

For both weighted and unweighted data, rankability considers two values, \( \delta \) and \( \rho \). Example 2 makes the case for consideration of another value, a value associated with the diversity of \( P \). Once again, we want to approximate such a measure without computing the full, or even a partial, set \( P \). This leads to the following optimization Model (3).

\[
\begin{align*}
\max & \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (u_{ij} + v_{ij}) \\
& x_{ij} + x_{jk} - x_{ik} \leq 1 \quad \forall i < j < k \quad \text{(transitivity)} \\
& y_{ij} + y_{jk} - y_{ik} \leq 1 \quad \forall i < j < k \quad \text{(transitivity)} \\
& (d_{ij} - d_{ji})x_{ij} + d_{ji} = \delta^* \quad \text{(optimal hyperplane)} \\
& (d_{ij} - d_{ji})y_{ij} + d_{ji} = \delta^* \quad \text{(optimal hyperplane)} \\
& u_{ij} - v_{ij} = x_{ij} - y_{ij} \quad \forall i < j \quad \text{(absolute deviation)} \\
& x_{ij} \in \{0, 1\} \quad \forall i < j \quad \text{(binary)} \\
& y_{ij} \in \{0, 1\} \quad \forall i < j \quad \text{(binary)} \\
& u_{ij} \in \{0, 1\} \quad \forall i < j \quad \text{(binary)} \\
& v_{ij} \in \{0, 1\} \quad \forall i < j \quad \text{(binary)}
\end{align*}
\]

This model finds two optimal rankings, two members of \( P \), that are farthest apart. Before solving the model, the optimal objective value \( \delta^* \) of the LOP or Hillside model must be input as the righthand side of two constraints. Decision variables \( x_{ij} \) create a solution matrix for one optimal ranking while decision variables \( y_{ij} \) create a solution matrix for the other optimal ranking in this most distant pair. We denote the optimal objective value as \( n_{max}^{opt} \), which is the number of discordant pairs between the maximally distant pair of optimal rankings in \( P \). The maximal Kendall tau distance between an optimal pair of rankings can then be computed with the standard definition \( \tau = 1 - \frac{4n_{max}^{opt}}{n(n-1)} \). We refer to \( \tau \) as the diameter of the set \( P \).

Several pairs of distant optimal rankings can be produced by running Model (3) multiple times in sequence, each time adding a constraint that forces the objective function value to be different from the previous run. For example, if the first run of Model (3) finds a maximally distant pair with an objective value of 34. Then a constraint is added that \( \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (u_{ij} + v_{ij}) \leq 33 \), which forces the model to
find a next most distant pair of optimal rankings. The user can continue in this fashion until any desired number of distant optimal pairs is produced.

Analyzing Examples 1 and 2 with Hillside Count, the diameter of $P$ as measured by the maximum Kendall rank correlation between any two optimal solutions is found to be approximately 0.95 and 0.67, respectively. The two most maximum discordant pairs for examples 1 and 2 are shown in Figure 5. The number of discordant pairs for Examples 1 and 2 are 3 and 7, respectively.

![Figure 5. Two maximally discordant optimal solutions for Examples 1 and 2.](image)

9. March Madness example application. In this section, we conduct a rankability analysis of the well-known NCAA March Madness basketball tournament. March Madness provides a nice playground for ranking and rankability. Many researchers and fans use ranking methods to complete tournament brackets [23]. While there are many aspects of the March Madness tournament that are of interest, predicting and anticipating the upsets that occur each year is of broad appeal [17, 27, 33]. For this reason, we use rankability to predict the degree to which a year will exhibit upsets.

9.1. Upset measure. One method for computing the degree of upsets in a given year is to compare the expected number of wins for each team to their actual number of wins [8, 21]. For example, the expected number of wins for seeds 9-16 is 0 (ignoring play-in games). Seeds 5-8 should win only 1 game as they will play a higher seed in the next round. Seeds 3 and 4 should win 2 games. Seed 2 should win 3 games. Teams seeded 1 should advance to the Final Four where two 1-seeded teams will win 4 games, one 1-seeded team will win 5 games, and one 1-seeded team will win 6 games. On average, 1-seeded teams will win $4.75 = ((4 + 4 + 5 + 6)/4)$ games. An upset score can then be computed for each team by subtracting the seed-expected number of wins from the actual number of wins. For example, if
a 1-seeded team loses in the second round after winning only one game, then that
team’s upset value is $1 - 4.25 = -3.25$. Teams outperforming their seed expectation
have a positive upset score while teams underperforming their seed value have a
negative upset score. Because this is a zero-sum game, the sum of the upset scores
over all teams is 0. Therefore, we define the upset measure for a particular year as
the sum of the positive upset values, i.e., the sum of the upset scores for all teams
that outperformed their seed.

9.2. Modeling dominance relationships, creating $D$ for March Madness.
In the context of March Madness, we will compare hillside and LOP optimization
formulations with respect to two different methods for modeling dominance
relationships (direct and direct+indirect, as defined below). We study how the
combination of these formulations affect a predictive model of the upset measure
using rankability of weighted data features.

For March Madness, the amount of direct dominance information is sparse. There
are over 350 teams in Division 1 NCAA college basketball, yet most teams play only
about 17 different opponents during the season. Thus, of the $350 \times 349 = 122,150$
dominance relations in college basketball, only up to $(350)(34)=10,500$ or 10% are
direct matchups. In other words, over 90% of the relations have no direction domi-
nance information. Fortunately, indirect information can be used to create a fuller
picture. We created indirect information using the ideas described earlier in Section
2.

For the 2018 NCAA men’s basketball season (excluding March Madness and
the last week which consists primarily of conference tournaments), there are 6,390
direct matchups and 65,701 indirect matchups. Including only direct matchups and
generating a matrix for only those teams who participated in March Madness results
in 418 nonzero entries in $D$ with a corresponding density of approximately 10%.
When indirect data is added, the dominance matrix $D$ becomes 72% dense. Our
experiments show the positive effect that indirect relations have on the performance
of rankability algorithms on the application of March Madness.

For this section only the 68 teams in the March Madness tournament are of
interest. Rather than creating a $350 \times 350$ matrix $D$, we instead create a $68 \times 68$
matrix of the 68 teams in the March Madness tournament. Game data from the
remaining 280+ non-tournament teams is incorporated in the $D$ matrix through
the use of indirect relations. In particular, any non-tournament team who played
two tournament teams may cast a vote that becomes part of the $68 \times 68$ matrix.

9.3. Predicting the upset measure. The purpose of our experiments is to study
modeling and formulation alternatives in the context of minimizing the mean abso-
lute error of a predicted upset measure for the 2002-2018 March Madness seasons.
For this study, we selected support vector machines (SVM) as they are considered
well-suited for small sample classification and regression [32]. The mean absolute
error was calculated using 5 fold cross-validation, and all models were compared
for statistical significance to a baseline model predicting the mean of the training
dataset. Statistical significance was evaluated using the $5 \times 2$ method described by
Dietterich [16]. Hyperparameters were tuned for the support vector machine (ker-
nel, $C$, $\epsilon$). Three hyperparameters were tuned for computing the direct+indirect
dominance matrix (direct threshold, spread threshold, and weight of indirect sup-
port) and the single hyperparameter of direct threshold was tuned for the direct
only dominance matrix.
Table 4. 5 fold cross-validation results for predicting the upset measure for NCAA Men’s March Madness 2002-2018

| Dominance Method | Parameters | Method | MAE     |
|------------------|------------|--------|---------|
| 1. **Direct+Indirect** | $dt=0, st=1, wi=1$ | Hillside | 3.148221 |
| 2. **Direct+Indirect** | $dt=0, st=0, wi=1$ | Hillside | 3.148221 |
| 3. **Direct+Indirect** | $dt=1, st=0, wi=1$ | LOP    | 3.172793 |
| 4. **Direct+Indirect** | $dt=1, st=1, wi=1$ | LOP    | 3.172793 |
| 5. **Direct+Indirect** | $dt=0, st=0, wi=0.25$ | Hillside | 3.213978 |
| 6. **Direct+Indirect** | $dt=0, st=1, wi=0.25$ | Hillside | 3.213978 |
| 7. **Direct**          | $dt=2$     | Hillside | 3.309455 |
| 8. **Direct+Indirect** | $dt=2, st=1, wi=1$ | LOP    | 3.311588 |
| 9. **Direct+Indirect** | $dt=2, st=0, wi=1$ | LOP    | 3.311588 |
| 10. **Direct+Indirect** | $dt=2, st=2, wi=0.5$ | Hillside | 3.331698 |

9.3.1. Results. Here we present the summary of the experiments focused on the top performing results and present a feature importance analysis. The top 10 results sorted by the mean absolute error are shown in Table 4. A baseline MAE score of 3.65625 is found by predicting the average upset measure. The best results were seen using the hillside formulation with both direct and indirect information to construct the dominance matrix $D$ using the direct threshold $dt = 0$, spread threshold $st = 1$, and weight of indirect support $wi = 1.0$. Six out of the top ten models were achieved using the hillside formulation. The remaining top four models were found using the LOP formulation. Nine out of the top ten models incorporated indirect information directly in the dominance matrix.

The top performing models for both the LOP and hillside formulations were built using a linear kernel (hyperparameters found via grid search). The features included in the model include the number of fractional elements in $X^*$, the number of fractional top 20 elements in the reordered $X^*(r,r)$, and the distance from perfection ($\delta$). All features were normalized to each have a mean of 0 and standard deviation of 1. The SVM coefficients for the distance from perfect ($\delta$) for both the LOP and the hillside formulation are 0.346 and 0.482, respectively. This indicates that the upset score increases as $\delta$ the distance from perfection, increases. For the LOP formulation, the coefficients for number of fractional elements in $X^*$ and the number of fractional top 20 elements in the reordered $X^*(r,r)$ are 0.378 and 0.333, respectively. This indicates that more fractional elements are predictive of a higher degree of upsets. This relationship does not hold for the hillside formulation because the hillside formulation uses a much stricter definition for distance from perfection, and therefore, in general, has fewer fractional elements in $X^*$. For both formulations, each year may be visualized by rank ordering the $X^*$ matrix and color coding fractional elements according to whether they indicate movement up the rankings (green) or movement down the rankings (red). Color-coded visualizations for all years are available in the Supplemental Section. We highlight three years in particular (2009, 2013, and 2016) for both the LOP (Figure 6) and hillside (Figure 7 formulations).
An inspection of $X^*$ shows optimal and near optimal alternative solutions. A few additional uses of $X^*$ for March Madness are: (1) one can identify teams with many fractional elements, particularly green pixels in the $X^*$ plot, as cinderella teams with the potential to upset opponents in the tournament, (2) one can, prior to the first game of the tournament, predict whether this year’s tournament will either be straightforward because the data has high rankability or wild and likely to have many upsets because the data has low rankability, and (3) one can use the fractional components of $X^*$ to determine which items are ranked with more certainty and which should be viewed with caution.

For example, look at the $X^*$ plots for years 2013 and 2016 in Figure 6. These clearly have considerable fractional elements as indicated by the amount of red and green in the plots. In 2013, the Final Four had a 9 seed (Wichita), two 4 seeds (Syracuse and Michigan) and 1 seed Louisville. In 2016, a 15 seed (Middle Tennessee) upset a 2 seed (Michigan State). This was just the eighth time a 15 seed beat a 2 seed in March Madness. Further, at least one 9, 10, 11, 12, 13, 14, and 15 seed won a first-round game for the third time in tournament history, with 2013 being another such time.
In contrast, consider the $X^*$ plot for 2009 in Figure 6. Note the fractional elements, again indicated by the red and green portions of the plot in the bottom right of the graph. These fractional elements indicates that the rankings vary with respect to the lower ranked teams. In March Madness, these weaker teams will play much stronger teams in the first round. As such, we expect a more predictable tournament, which was exactly how the 2009 tournament played out. In 2009, for the first time since seeding began in 1979, all 1, 2 and 3 seeds reached the Sweet 16 (i.e., the round containing the 16 teams who won their first two games), and for the third consecutive time, all 1 seeds reached the Elite Eight (winning their first three games).

10. **Future work.** In [4], Anderson et. al proposed an overall rankability measure, $r$, after obtaining an entire set $P$. This paper’s introduction to the diameter of $P$, $\tau$, does not require obtaining the entire set $P$. Because $\tau$ is computationally less expensive than $P$, it opens a pathway to compute a variation of the $r$ measure that uses $\delta$, a maximum possible $\delta$ for a matrix of a given size, $\rho$, and $\tau$. This is one avenue of future work.

While the methods of this paper involve a ranking with rankability information, future work will explore the rankability measure’s connection to ranking methods outside the scope of those motivated from minimum violations. For example, at what point is a rankability measure so poor that it indicates that essentially any ranking method will fail? Similarly, given $X^*$, can we determine which parts of the data are suitable for ranking and which may require more data, domain expert insight, or simply be viewed suspiciously in any ranking?

The hillside formulation uses a rather strict definition of distance from perfection and thus produces fewer multiple optimal solutions. For the March Madness experiments we found that, for the hillside model, the $\delta$ value of rankability was more helpful in predicting upsets than the $\rho$ value. Thus, another area of future work broadens the set of solutions of interest to include both optimal and near-optimal solutions to see how this affect $\rho$ and rankability.

11. **Conclusions.** Establishing a linear ordering is inherent in many problems from ranking colleges or sports teams to the recommendation of products. This paper extends the breadth and impact of rankability relative to the work of Anderson et. al [4]. First, this paper introduces the inclusion of weighted data into a measure of rankability. Binary relationships exist in many forms from thumbs up or down ratings to simply looking at wins and losses. The inclusion of weighted data allows for important nuances to enter our understanding of rankability. This paper looks at various applications that are varied in both in size and context. In particular, the March Madness example demonstrates that datasets containing 300 or more items can now be explored. This paper also introduces methods to summarize information about all optimal solutions, enabling us to not only know the rankability of data but see where differences occur and what portions of a ranking are less certain. Rankability can serve as a critical tool in determining the quality of a dataset relative to ranking and give a sense of the underlying variability one could expect in a linear ordering of the items.
Acknowledgments. This work was supported by Cal Poly Digital Transformation Hub powered by Amazon Web Services (https://dxhub.calpoly.edu/).

REFERENCES

[1] T. Achterberg, Scip: Solving constraint integer programs, *Mathematical Programming Computation*, 1 (2009), 1–41.
[2] N. Ailon, M. Charikar and A. Newman, Aggregating inconsistent information: Ranking and clustering, *Journal of the ACM*, 55 (2008), 27 pp.
[3] I. Ali, W. D. Cook and M. Kress, On the minimum violations ranking of a tournament, *Management Science*, 32 (1986), 660–672.
[4] P. Anderson, T. Chartier and A. Langville, The rankability of data, *SIAM Journal on the Mathematics of Data Science*, (2019), 121–143.
[5] P. Anderson, T. Chartier, A. Langville and K. Pedings-Behling, IGARDS technical report, 2019. Available from: https://igards.github.io/research/IGARDS_Technical_Report_November_2019.pdf.
[6] P. Anderson, T. Chartier, A. Langville and K. Pedings-Behling, Revisiting rankability of unweighted data technical report, 2020. Available from: https://igards.github.io/research/Revisiting_Rankability_Unweighted_Data.pdf.
[7] R. D. Armstrong, W. D. Cook and L. M. Seiford, Priority ranking and consensus formation: The case of ties, *Management Science*, 28 (1982), 638–645.
[8] J. Brenner and P. Keating, Chaos kills, *ESPN, The Magazine*, (2016), 20–23.
[9] S. Brin and L. Page, The anatomy of a large-scale hypertextual web search engine, *Computer Networks and ISDN Systems*, 33 (1998), 107–117.
[10] S. Brin, L. Page, R. Motwami and T. Winograd, The PageRank citation ranking: Bringing order to the web, *Tech. Report 1999-0120*, Computer Science Department, Stanford University, 1999.
[11] T. R. Cameron, A. N. Langville and H. C. Smith, On the graph Laplacian and the rankability of data, *Linear Algebra and its Applications*, 588 (2020), 81–100.
[12] C. R. Cassady, L. M. Maillart and S. Salman, Ranking sports teams: A customizable quadratic assignment approach, *INFORMS: Interfaces*, 35 (2005), 497–510.
[13] T. P. Chartier, E. Kreutzer, A. N. Langville and K. E. Pedings, Sensitivity of ranking vectors, *SIAM Journal on Scientific Computing*, 33 (2011), 1077–1102.
[14] B. J. Coleman, Minimizing game score violations in college football rankings, *INFORMS: Interfaces*, 35 (2005), 483–496.
[15] W. D. Cook and L. M. Seiford, Priority ranking and consensus formation, *Management Science*, 24 (1978), 1721–1732.
[16] T. G. Dietterich, Approximate statistical tests for comparing supervised classification learning algorithms, *Neural Computation*, 10 (1998), 1895–1923.
[17] S. Dutta, S. H. Jacobson and J. J. Sauppe, Identifying ncaa tournament upsets using balance optimization subset selection, *Journal of Quantitative Analysis in Sports*, 13 (2017), 79–93.
[18] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, 1979.
[19] F. M. Harper and J. A. Konstan, The movielens datasets: History and context, *Acm Transactions on Interactive Intelligent Systems (TIIS)*, 5 (2015), 1–19.
[20] X. Jiang, L.-H. Lim, Y. Yao and Y. Ye, Statistical ranking and combinatorial Hodge theory, *Mathematical Programming*, 127 (2011), 203–244.
[21] P. Keating, personal communication.
[22] Y. Kondo, Triangulation of input-output tables based on mixed integer programs for intertemporal and inter-regional comparison of production structures, *Journal of Economic Structures*, 3 (2014), 1–19.
[23] A. N. Langville and C. D. Meyer, *Who’s #1? The Science of Rating and Ranking Items*, Princeton University Press, Princeton, 2012.
[24] A. N. Langville, K. Pedings and Y. Yamamoto, A minimum violations ranking method, *Optimization and Engineering*, (2011), 1–22.
[25] A. S. Lee and D. R. Shier, A method for ranking teams based on simple paths, *UMAP Journal*, 39 (2018), 353–371.
[26] W. W. Leontief, *Input-Output Economics*, 2nd edition, Oxford University Press, 1986.
12. SUPPLEMENT.

Theorem 12.1 (Theorem 6.1). Every ranking of a weighted rankability problem corresponds to a binary extreme point of the LP weighted rankability polytope.

Proof. Every ranking \( r \) has a corresponding binary strictly upper triangular matrix \( X(r, r) \) which denotes \( X \) after it has been symmetrically reordered according to \( r \). The matrix \( X \) is binary and clearly feasible since anti-symmetry and transitivity are easy to verify from the upper triangular form of \( X(r, r) \). It remains to show that \( X \) is an extreme point, i.e., that \( X \) cannot be written as a convex combination of other extreme points. We do this by contradiction. Suppose that there exists a scalar \( 0 < \alpha < 1 \) and, without loss of generality, exactly two binary feasible matrices \( Y \neq Z \) such that \( X = \alpha Y + (1 - \alpha)Z \). Since \( Y \neq Z \), there exists at least one element, say \((i, j)\) such that \( y_{ij} \neq z_{ij} \). Suppose, without loss of generality, that \( y_{ij} = 1 \) and \( z_{ij} = 0 \). Then \( x_{ij} = \alpha y_{ij} + (1 - \alpha)z_{ij} = \alpha \), which means that \( X \) is fractional, which contradicts the statement that \( X \) is binary. Therefore, the assumption that \( X \) is a convex combination of \( Y \) and \( Z \) is false and rather it is that \( X \) is an extreme point. \( \square \)

Theorem 12.2 (Theorem 6.2). If the Interior Point solver of the LP relaxed weighted rankability problem of Model (2) ends in Case 3 (\( \delta^* \) is integer and \( X^* \) is fractional) and the LP’s optimal face is the IP’s optimal face, then

1. \( \delta^* \) is the optimal objective value for the integer program,
2. \( X^* \) is on the interior of the optimal face (i.e., the convex hull of all optimal solutions) of the integer program, and
3. fractional entry \((i, j)\) in \( X^* \) means that there exists at least one optimal ranking in \( P \) with \( x^*_{ij} = 1 \) (thus, \( i > j \)) and at least one with \( x^*_{ij} = 0 \) (thus, \( i < j \)).

Proof. (1) (By Contradiction.) Assume otherwise. That is, assume \( \delta^* \), the optimal objective value of the linear program, is not the optimal objective value of the integer program. Then \( \delta^* \) is suboptimal for the integer program and the integer program’s optimal objective value must be an integer superior to \( \delta^* \) such as \( \delta^* - 1, \delta^* - 2, \ldots \). However, this is impossible because the linear program, being a relaxation to the integer program, must have an objective value equal to or superior to the objective value of the integer program. In other words, the only possible superior objective value for the linear program is a non-integer value yet this contradicts the fact that we are in Case 3 with an integer objective value.
(2) That $X^*$ of the LP is in the interior of the IP’s optimal face follows trivially from the assumption that the LP’s optimal face is the IP’s optimal face.

(3) By (2) above, we know that $X^*$ is in the interior of the optimal face of the integer program, which means that $X^*$ is a convex combination of the $\rho$ binary optimal extreme points of the integer program, each of which, by Theorem 6.1, corresponds to a ranking $h$ denoted by the binary matrix $X^h$. Thus,

$$X^* = \alpha_1 X^1 + \alpha_2 X^2 + \ldots + \alpha_\rho X^\rho,$$

where $0 < \alpha_i < 1$, $\sum_{i=1}^\rho \alpha_i = 1$, and $X^h$ is the binary matrix corresponding to optimal ranking $h$. If the $(i,j)$ entry of $X^*$, $x^*_{ij}$, is 1, then all rankings in $P$ agree that $i > j$ because $x^*_{ij}$ can only be 1 if all $x^h_{ij} = 1$.

$$x^*_{ij} = \alpha_1 x^1_{ij} + \alpha_2 x^2_{ij} + \ldots + \alpha_\rho x^\rho_{ij} = \alpha_1(1) + \alpha_2(1) + \ldots + \alpha_\rho(1) = \alpha_1 + \alpha_2 + \ldots + \alpha_\rho = 1.$$

Similarly, at the other extreme, the only way that $x^*_{ij} = 0$ is if all rankings in $P$ agree that $i < j$, i.e., $x^h_{ij} = 0$ for all $h$. The remaining option for $x^*_{ij}$ is a fractional value, which can happen only if some $x^h_{ij} = 1$ (meaning $i > j$) and some $x^h_{ij} = 0$ (meaning $i < j$). Thus, a fractional value in the $(i,j)$ entry of $X^*$ represents disagreement among the members of $P$ about the pairwise ranking of items $i$ and $j$.

**Theorem 12.3** (Theorem 6.3). If $X^*$ is the exact centroid of all optimal rankings for a weighted rankability problem, then

$$\rho \geq \left\lceil \frac{1}{m} \right\rceil,$$

where $m$ is the smallest fractional element in $X^*$.

**Proof.** Assume it is the $(i,j)$ entry of $X^*$ that holds the smallest fractional value $m$. The only way this entry can have a nonzero value is if at least one of the $\rho$ binary optimal rankings $X^h$ for $h = 1, 2, \ldots, n$ has $i > j$, which means there exists at least one $x^h_{ij} = 1$ for $h = 1, 2, \ldots, n$. Suppose that exactly one of the optimal rankings, say $X^1$, has $i > j$ so that $x^1_{ij} = 1$. $X^*$ is the centroid of all binary optimal rankings $X^1, X^2, \ldots, X^\rho$ and can be written as the following convex combination

$$X^* = \frac{1}{\rho} X^1 + \frac{1}{\rho} X^2 + \cdots + \frac{1}{\rho} X^\rho.$$

Thus, $m = x^*_{ij} = \frac{1}{\rho}(1) = \frac{1}{\rho}$ and $\rho = \frac{1}{m}$. Now suppose exactly two of the $\rho$ binary optimal rankings have $i > j$, then $m = x^*_{ij} = \frac{1}{\rho}(1) + \frac{1}{\rho}(1) = \frac{2}{\rho}$ and $\rho = \frac{2}{m} > \frac{1}{m}$. Continuing in this fashion, it follows that $\rho \geq \frac{1}{m}$, regardless of the number of binary optimal rankings that contribute to the fractional $m$. Since $\rho$ is an integer, $\frac{1}{m}$ can be rounded up to the nearest integer. □

Received August 2020; revised December 2020.

E-mail address: pander14@calpoly.edu
E-mail address: tichartier@davidson.edu
E-mail address: langvillea@cofc.edu
E-mail address: kathryn@behling.org
Figure 8. $X^*$ color-coded visualization for each year of NCAA March Madness using the top performing LOP formulation.
Figure 9. $X^*$ color-coded visualization for each year of NCAA March Madness using the top performing hillside formulation.