RANK AND SCORE AGGREGATION METHODS IN COMPETITIVE CLIMBING*

KIRA PARKER†

ADVISOR: BRAXTON OSTING‡

Abstract. Ranking methods are used in all aspects of life, from Google searches to sports tournaments. Because all ranking methods necessarily have advantages and disadvantages, USA Climbing, the organizer of national climbing competitions in the United States, changed their ranking method three times between 2009 and 2016. The combined rank method employed in 2015 marked a drastic step away from the previous two in that it failed to meet the independence of irrelevant alternatives (IIA) criterion and was almost impossible for spectators to use to calculate ranks on their own. We compare this more recent rank aggregation method with older USA Climbing score aggregation methods as well as other methods from the literature. Three particularly important methods we consider are (i) the combined rank method, (ii) a combination of the previous two USA Climbing score aggregation methods (the merged method), and (iii) a linear programming (LP)-based rank aggregation method from the literature. Using data from the 2016 Bouldering Youth National Championships, we perform leave-one-out cross validation and the Friedman hypothesis test to conclude that at the 99% confidence level, the LP-based rank aggregation method has significantly more predictive power than the other two methods, while there was insufficient evidence to distinguish between the predictive power of the combined rank method and the merged method. However, due to the desirable properties, such as the IIA criterion, satisfied by the merged method, we recommend this method for use in competitive climbing.

1. Introduction.

Ranking methods are used in all aspects of life, from Google searches to sports tournaments. One particular class of ranking methods is those used to combine multiple sets of information about the same alternatives into one overall rank. Both rank and score aggregation methods fall into this class: rank aggregation combines several ranks of alternatives, while score aggregation is used when the alternatives each have multiple scores associated with them. Score aggregation can take into account the magnitude of difference in the preference of one alternative over another, whereas rank aggregation cannot. In this paper, we will consider rank and score aggregation in the context of competitive climbing.

USA Climbing (USAC) is a non-profit organization for competitive climbing in the United States. Although USAC holds a variety of competitions for different groups of people and areas of climbing, we will focus solely on their *youth onsight bouldering competitions*. In each competition, climbers are divided into 10 categories based on age and gender: female junior, female youth A, female youth B, female youth C, female youth D, and their male counterparts. Each category is given between three and six never-before-seen boulder problems to climb with four to five minutes per boulder problem per person. Climbers score points (called *hold points* for the remainder of the paper) for each hold grabbed on each of their three to six boulders. If the climber finishes the problem, it is called a *top*, and he or she scores the maximum number of hold points for the climb. Otherwise, the highest hold reached is recorded. The number of *attempts* a climber takes to reach his or her high hold (or the top) is also recorded. If a climber finishes the problem on his or her first attempt, it is called a *flash*. Figure 1 illustrates some of this terminology.

Climbers are ranked on each problem based on their highest hold reached with number of attempts as a tie breaker. The problem we face is how to best combine these scores or rankings on each problem into an overall rank of climbers in the competition. For more information on USAC and competitive climbing, see [3].

Rank aggregation methods have been studied extensively in the context of social choice theory, with Arrow playing a large role in the 1950's. His "Impossibility Theorem," detailed in [4], states that no non-dictatorship rank aggregation method with unrestricted domain can satisfy both independence of irrelevant alternatives (IIA) and Pareto efficiency, which suggests that no method is ideal (see section 3 for an explanation of these properties). Different methods, therefore, are better suited to different situations.

Score aggregation appears to be a better way of ranking climbers, assuming there is a method of associating scores with them. In [5], Balinski and Laraki discuss the advantages of what they call "grading." They define a *common language* to be a set of strictly ordered grades, and a *method of grading* to be a function that assigns to any input profile (matrix of grades) one final grade in the same common language to each candidate and also satisfies some basic properties, one of which is independence of irrelevant alternatives (a desirable property for methods that rank climbers). Using a common language, then, allows a function to satisfy non-dictatorship, Pareto efficiency, IIA, and unrestricted domain,

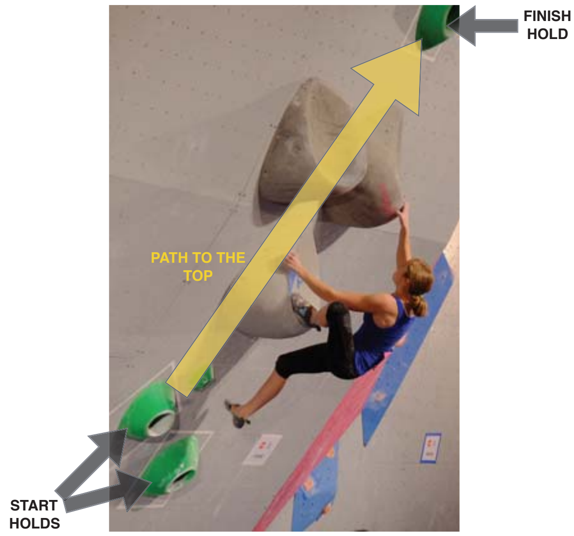
*Submitted to the editors June 2017.

Funding: This work was funded by the University of Utah Math Department

†University of Utah, Salt Lake City, UT 84124 (kira.parker314@gmail.com).

‡Department of Mathematics, University of Utah, Salt Lake City, UT 84112 (osting@math.utah.edu).

Fig. 1: Female youth A boulder problem from the 2015 Bouldering Youth National Championships illustrating the climbing terms discussed above. ©Kira Parker.



meaning that Arrow's Impossibility Theorem does not apply. As we will see, competitive climbing lends itself to score aggregation methods, as climbers have already been given various "grades" on each problem they climb (problems corresponding to the judges in the aforementioned paper).

The rise of American competition climbing in the 1990's brought with it a variety of different rank and score aggregation methods. USAC changed their ranking method for onsight bouldering competitions three times between 2009 and 2016. The first two methods, further explained in subsection 2.4.2 and subsection 2.4.1, respectively, were score aggregation methods. They revolved around how many holds a competitor reached on their climb, where his or her score was the sum of those hold points. In the first method, climbers received a twenty point bonus for finishing a climb on their first try, while in the second method climbers were ranked first by the number of climbs completed and then by their total hold points.

USAC's third method was a rank aggregation method, and will be described in subsection 2.3.1. Drastically different from the first two, its introduction caused an uproar in the competition climbing community. Whereas before climbers and spectators could predict the overall placement of themselves or their favorite climber fairly easily, nothing was certain under the new method. Some obscure formula, with geometric means and "Ranking Points," was going on behind the scenes, and everyone was forced to sit anxiously waiting for the final results to be posted. Even those who understood the formula had trouble predicting climbers' overall placements, and no one was happy. Our primary motivation for this work was the quantity of complaints surrounding this third USAC method, called the combined rank method, and the results it produced. A large portion of these upsetting results were due to lower-ranked climbers finishing and changing the placements of climbers ranked above them, a violation of the IIA criterion. Additionally, competitors would change their score on a problem after the competition, as scoring errors occurred frequently, and the relative ranks of other climbers would be disrupted. Thus methods that satisfy the IIA criterion should be preferred for use in ranking climbers.

In this paper, we examine USAC's combined rank method in comparison with their older score aggregation methods as well as with some other rank aggregation methods found in the literature. We begin with an examination of the datasets produced from climbing competitions and define the methods we use to rank these climbers. We then compare the properties of the various methods, looking specifically at monotonicity, Pareto efficiency (unanimity), and independence. We find that the older USAC methods satisfy the desirable independence of irrelevant alternatives property which all of the rank aggregation methods fail to satisfy. In our analysis of the predictive power of these methods when they are used to rank climbers, we follow the steps laid out in [6]. First, we perform leave-one-out cross validation on the methods. To determine if some methods have consistently lower cross validation scores (and thus have

Fig. 2: Sample dataset of results from the female youth A final round of the 2016 Bouldering Youth National Championships ([3]). It contains information about high point, number of attempts, rank, and ranking points for each climber on each problem: F1, F2, and F3.

FEMALE YOUTH A

		F1				F2				F3				F TOTALS	
	NAME	HIGH HOLD	ATTEMPTS	RANK	POINTS	HIGH HOLD	ATTEMPTS	RANK	POINTS	HIGH HOLD	ATTEMPTS	RANK	POINTS	TOPS	POINTS
1	Maya Madere	14	1	1	5	14	1	1	1	14	1	1	1.5	3	1.96
2	Georgia Bank	14	1	1	5	9	1	2	2	14	1	1	1.5	2	2.47
3	Sidney Trinidad	14	1	1	5	9	2	3	3	14	3	5	6	2	4.48
4	Audrey Miller	14	1	1	5	7	1	6	7	14	2	3	3.5	2	4.97
5	Julia Talbot	14	1	1	5	8	4	5	5	14	3	5	6	2	5.31
6	Laili Couper	14	1	1	5	7	2	9	9	14	3	5	6	2	6.46
7	Miriam Timson	14	2	10	10	7	3	10	10	14	2	3	3.5	2	7.05
8	Melina Costanza	14	1	1	5	8	1	4	4	13	5	8	8	1	5.43
9	Bimini Horstmann	14	1	1	5	7	1	6	7	11	1	9	9	1	6.8
10	Kara Herson	14	1	1	5	7	1	6	7	11	3	10	10	1	7.05
11	Katie Malinowski	11	4	11	11	6	2	11	11	8	1	11	11		11

higher predictive power), we use the nonparametric Friedman statistical test. Our results from this test were highly significant, with a p-value of 1.26×10^{-4}, and we were able to use the post-hoc Nemenyi test to find the best-performing methods. Unsurprisingly, the combined rank method outpredicted the two previously used methods, although not by a large margin. A linear program introduced in [8] had the best predictive accuracy (see subsection 2.3.4 for a description of the linear program). However, a combination of the two old score aggregation methods, a method we call the *merged method*, was the best-predicting independent method, leading us to recommend the use of this method to USA Climbing.

This paper is organized as follows: Section 2 explains the data and methods used to rank climbers. The properties of the methods are analyzed in section 3. Section 4 comprises the evaluation of the predictive power of the methods, and section 5 discusses the results and concludes the paper.

2. Data and Methods.

2.1. Data. As illustrated in Figure 2, the datasets analyzed contain information for each of M climbers on N problems. This information includes high hold, or the number of hold points obtained, on the problem (an integer between 0 and the total number of holds on the problem), number of attempts the climber required to reach that high hold, the climber's rank on the problem, and the ranking points he or she earned for that problem. The climbers are ranked based on their hold points, where a higher point value is better, and then by number of attempts if two climbers have the same hold points for a problem, where fewer attempts is preferable. Ranking points are calculated from the rank, where a climber's points are equal to either his or her rank for the problem or, if multiple climbers have the same rank, the average ranking of the tied climbers. From each dataset, we calculate the following:

1. A matrix of hold points, $H \in \mathbb{R}^{M \times N}$, where H_{ij} is the number of hold points the ith climber earned on the jth problem.
2. A vector of total hold points, $h \in \mathbb{R}^M$, where $h_i = \sum_{j=1}^{N} H_{ij}$.
3. A matrix of attempts, $A \in \mathbb{R}^{M \times N}$, where A_{ij} is the number of attempts the ith climber had on the jth problem to reach his or her high hold.
4. A vector of total attempts, $a \in \mathbb{R}^M$, where $a_i = \sum_{j=1}^{N} A_{ij}$.
5. A matrix of ranks, $R \in \mathbb{R}^{M \times N}$, where R_{ij} is the rank of the ith climber on the jth problem.
6. A matrix of ranking points, $P \in \mathbb{R}^{M \times N}$, where P_{ij} is the number of ranking points the ith climber received on the jth problem.
7. A vector of tops, $t \in \mathbb{R}^M$, where the ith component is the total number of tops for climber i.

The data used in our analysis comes from the 2016 Bouldering Youth National Championships, and was downloaded from the USAC webpage immediately following the competition ([3]). This data and the code implementing the methods used to analyze it is available at [9].

Example from the 2016 Bouldering Youth National Championships. For example, using the final results from the female youth A category of the 2016 Bouldering Youth National Championships (see dataset in Figure 2), we get the following data:

$$H = \begin{bmatrix} 14 & 14 & 14 \\ 14 & 9 & 14 \\ 14 & 9 & 14 \\ 14 & 7 & 14 \\ 14 & 8 & 14 \\ 14 & 7 & 14 \\ 14 & 7 & 14 \\ 14 & 8 & 13 \\ 14 & 7 & 11 \\ 14 & 7 & 11 \\ 11 & 6 & 8 \end{bmatrix} \quad h = \begin{bmatrix} 42 \\ 37 \\ 37 \\ 35 \\ 36 \\ 35 \\ 35 \\ 35 \\ 32 \\ 32 \\ 25 \end{bmatrix} \quad A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 2 \\ 1 & 4 & 3 \\ 1 & 2 & 3 \\ 2 & 3 & 2 \\ 1 & 1 & 5 \\ 1 & 1 & 1 \\ 1 & 1 & 3 \\ 4 & 2 & 1 \end{bmatrix} \quad a = \begin{bmatrix} 3 \\ 3 \\ 6 \\ 4 \\ 8 \\ 6 \\ 7 \\ 7 \\ 3 \\ 5 \\ 7 \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 3 & 5 \\ 1 & 6 & 3 \\ 1 & 5 & 5 \\ 1 & 9 & 5 \\ 10 & 10 & 3 \\ 1 & 4 & 8 \\ 1 & 6 & 9 \\ 1 & 6 & 10 \\ 11 & 11 & 11 \end{bmatrix} \quad P = \begin{bmatrix} 5 & 1 & 1.5 \\ 5 & 2 & 1.5 \\ 5 & 3 & 6 \\ 5 & 7 & 3.5 \\ 5 & 5 & 6 \\ 5 & 9 & 6 \\ 10 & 10 & 3.5 \\ 5 & 4 & 8 \\ 5 & 7 & 9 \\ 5 & 7 & 10 \\ 11 & 11 & 11 \end{bmatrix} \quad t = \begin{bmatrix} 3 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

2.2. Data Cleaning. USAC only posts information in their datasets that is relevant to their current ranking method. Consequently, we lacked the information about which climbers finished which problems that we use for the top score method, ABS10 method, and merged method (see subsection 2.4.1, subsection 2.4.2, and subsection 2.4.3). In order to obtain this information, we wrote a program that considered the total number of tops for each climber, as well as the maximum number of holds points scored on each problem, and tested different sets of top hold points for each climb to see if it resulted in climbers having the correct number of tops. If only one set of top hold points worked, we could then uniquely determine which climbers finished each problem and which did not. For a few datasets, multiple lists of top hold points worked, and we could not determine this information. We were forced to leave these datasets out of our analysis.

For example, consider the following dataset:

	Hold Points P1	Hold Points P2	Hold Points P3	Tops
Climber 1	15	20	10	2
Climber 2	12	20	17	2
Climber 3	11	20	5	1

The maximum number of hold points scored on problems one, two, and three (P1, P2, and P3) are 15, 20, and 17, respectively. If we assume these are also the top hold points for each problem, we find that each climber has the expected number of tops. However, is this solution unique? If the top hold points for problem one were greater than 15, Climber 1 would only have one top, so this set of top hold points does not work. Similar issues rule out increasing the top hold points for problems two and three. Since we cannot decrease the top hold points for any problem (a climber cannot score more points than the problem is worth), we know that the only possible top hold points are 15, 20, and 17 for problems one, two, and three, respectively. Thus we can see that Climber 1 finished problems one and two, Climber 2 finished problems two and three, and Climber 3 finished only problem two.

2.3. Rank Aggregation Methods. The following methods are the rank aggregation methods we implemented and analyzed. They use the ranked lists of M climbers on each of N problems to produce a final ordering of the climbers. As is standard in competitive climbing, climbers are ranked first by number of tops and then by their placement in the ordered list that the given method produces. To see if this tradition of ranking climbers first by number of tops is justifiable, we also use these methods without taking number of tops into account, where a climber's final rank is just that produced by the method.

2.3.1. Combined Rank Method. The combined rank method used by USA Climbing, as detailed in [2], takes the ranking points matrix P and calculates the climbers' geometric mean vector m, where

$$m_i = \left(\prod_{j=1}^{N} P_{ij}\right)^{1/N} = \arg\min_{m_i \in \mathbb{R}} \sum_{j=1}^{N} (\log P_{ij} - \log m_i)^2.$$

The equivalence follows from setting $\partial_{m_i} \sum_{j=1}^{N} (\log P_{ij} - \log m_i)^2 = 0$ to get

$$0 = \partial_{m_i} \sum_{j=1}^{N} (\log P_{ij} - \log m_i)^2 = 2 \sum_{j=1}^{N} (\log P_{ij} - \log m_i)(-\frac{1}{m_i}),$$

which then implies $N \log m_i = \sum_{j=1}^{N} \log P_{ij}$ and thus $m_i = \left(\prod_{j=1}^{N} P_{ij}\right)^{\frac{1}{N}}$. Climbers are ranked by their geometric mean, where a lower value is better.

2.3.2. Geometric Mean Method. The geometric mean method is identical to the combined rank method except that it uses the ranks matrix R to calculate the rating vector instead of the ranking points matrix P. This means that the final rank from a competition where no climbers tied would be the same under either method, but would potentially differ if there were ties. Here

$$m_i = \left(\prod_{j=1}^{N} R_{ij}\right)^{1/N} = \arg\min_{m_i \in \mathbb{R}} \sum_{j=1}^{N} (\log R_{ij} - \log m_i)^2.$$

2.3.3. Borda Method. The Borda method, described in [10], is a standard method used in social choice theory that assigns climber $i = 1, ..., M$ Borda points equivalent to $\sum_{j=1}^{N} (M - P_{ij})$. However, we were interested in a ranking method where climbers with the same ranking on a problem are given Borda points equal to the position of the highest ranked climber, as given in R_{ij}, instead of the average of the positions that P_{ij} gives. Thus we assign Borda points for climber i with

$$b_i = \sum_{j=1}^{N} (M - R_{ij}) = \arg\min_{b_i \in \mathbb{R}} \sum_{j=1}^{N} \left(M - R_{ij} - \frac{b_i}{N}\right)^2.$$

Climbers are ranked by these Borda points, where higher point values are better.

2.3.4. Linear Programming Method. A linear programming approach to aggregating ranks of sports teams was developed in [8]. This method can be applied to the problem of ranking M climbers on N problems as follows. First the matrix C is calculated, where C is a skew-symmetric $M \times M$ matrix with $C_{ij} = \{\#$ of problems with climber i above climber $j\}$ - $\{\#$ of problems with i below $j\}$. We then calculate the $M \times M$ matrix X, which corresponds to the aggregated rank with $X_{ij} = 1$ if climber i beat climber j and 0 otherwise. The matrix X is the solution to the following linear program, which we solve using the `scipy.optimize.linprog` function in Python:

$$\max \sum_{i=1}^{M} \sum_{j=1}^{M} C_{ij} X_{ij}$$

$$X_{ij} \geq 0$$

$$X_{ij} + X_{ji} = 1 \text{ for all distinct pairs } (i, j)$$

$$X_{ij} + X_{jk} + X_{ki} \leq 2 \text{ for all distinct triples } (i, j, k).$$

The first and second constraints together restrict the possible entries in X to numbers between 0 and 1. This is a relaxation of the integer constraint that each $X_{ij} \in \{0, 1\}$. The second constraint alone ensures that two climbers cannot both beat each other. The third constraint maintains transitivity in the overall ranking: if climber i is ranked above climber j and climber j is ranked above climber k, then climber i must be ranked above climber k.

In our problem, we divided the climbers into disjoint sets based on number of tops, where climbers with the same number of tops were in the same set. We then solved the linear program for each set of climbers and combined the ranks at the end, so climbers with higher numbers of tops were ranked above climbers with lower numbers of tops. This was the only rank aggregation method lacking a "no-tops" version.

2.3.5. Geometric Median Methods. The geometric median methods use Weiszfeld's algorithm to approximate the geometric median of the ranks in $\mathbb{R}^M$, where each point corresponds to the ranks of all M climbers on a given problem. The geometric median is defined to be

$$\arg\min_{y \in \mathbb{R}^M} f(y) \coloneqq \sum_{j=1}^{N} \|R_{:j} - y\|_2,$$

and cannot be determined with an explicit formula. Weiszfeld's algorithm is a form of iteratively reweighted least squares that starts with the centroid of the points (its coordinates are the averages of the coordinates of the points) and creates new, improved estimates of the geometric median of the points. For an estimate y_i, the next estimate is

$$y_{i+1} = \left(\sum_{j=1}^{N} \frac{R_{:j}}{\|R_{:j} - y_i\|} \right) \Bigg/ \left(\sum_{j=1}^{N} \frac{1}{\|R_{:j} - y_i\|} \right),$$

where $R_{:j}$ indicates the jth column of R. We iteratively estimate y_{i+1} until $\|y_{i+1} - y_i\| \leq 10^{-6}$. For more information about Weiszfeld's algorithm and its convergence, see [11].

The *integer geometric median method* calculates a final rank by using the output vector y from Weiszfeld's algorithm as a rating and ranking climbers based first on number of tops and then on their rating y_i, where smaller values are better.

The *no-tops geometric median method* uses almost the same final rank as the integer geometric median method except that climbers are ranked solely by their rating.

The *optimal integer geometric median method* uses the integer geometric median method to calculate the final rank for *tie thresholds* t from 0 to .5 (incrementing by .1). If two climbers p and q have the same number of tops and $|y_p - y_q| \leq t$ then they tie in the final rank. The integer geometric median rank for the tie threshold that produces the minimum value of the objective function $f(y)$ is then the rank given by the optimal integer geometric median method.

2.4. Score Aggregation Methods. The following methods are the score aggregation methods we implemented and analyzed. Given that climbers have scores associated with high points, tops, and attempts on each problem, it is not immediately obvious which scores should be used in such a method. Thus these three methods below each use different aspects of the climbers' performance on each climb to rank the climbers overall.

2.4.1. Top Score Method. The top score method was used by USA Climbing before they implemented their combined rank method [7] . First we extract the information about which climbers finished which problems, a process described in subsection 2.2, and store it in a matrix T, where T_{ij} equals 1 if climber i finished problem j and 0 otherwise. Next, we find the score vectors $s_f, s_t, s_h \in \mathbb{R}^M$, which are defined as follows:

s_{f_i} is the total number of flashes climber i had (problems finished in one attempt), where more flashes is better.

$s_{t_i} = \sum_{j=1}^{N} T_{ij} A_{ij}$ is climber i's total number of attempts to finish the problems that he or she finished, where fewer attempts is better.

$s_{h_i} = \sum_{j=1}^{N} (1 - T_{ij}) A_{ij}$ is climber i's total number of attempts to reach his or her high points on problems he or she didn't finish, where fewer attempts is better.

Climbers are then ranked first by number of tops, second by the sum of their hold points h_i, third by their number of flashes s_{f_i}, and forth and fifth by their attempt scores s_{t_i} and s_{h_i}.

2.4.2. ABS10 Method. The ABS10 method is essentially the method used by USA Climbing in 2009, as detailed in [1], before the top score and combined rank methods. As in the top score method, we first find the tops matrix T. The maximum number of holds on each problem is then calculated and stored in a vector $m \in \mathbb{R}^N$, where if no climber finished problem j, m_j is equal to the average number of holds on national-level problems for the given category (this is where the discrepancy with the actual method lies - USAC knew the number of holds on each climb but the datasets do not provide that information). Next, we find the score vector v, where $v_i = \sum_{j=1}^{N} u_{ij}$ and u_{ij} is equal to $H_{ij}\frac{1000}{m_j} + 20$ if climber i finished the problem on his or her first attempt, or $H_{ij}\frac{1000}{m_j} - 5(A_{ij} - 1)$ otherwise. Climbers are then ranked exclusively by their score (with no regard to tops, unlike in the other methods), where a larger v_i is better.

2.4.3. Merged Method. The merged method is a combination of the ABS10 method and the top score method. Again, we find the tops matrix T, and then calculate the maximum number of holds on each problem and store it in a vector $m \in \mathbb{R}^N$, where m_j is equal to the average number of holds on national-level problems for the given category if no climber finished problem j. Next, we find three score vectors s_p, s_t, and s_h as follows:

$s_{p_i} = \sum_{j=1}^{N} H_{ij}\frac{1000}{m_j}$ is the sum of climber i's normalized hold points, where more points is better.

$s_{t_i} = \sum_{j=1}^{N} T_{ij}A_{ij}$ is climber i's total number of attempts to finish the problems that he or she finished, where fewer attempts is better.

$s_{h_i} = \sum_{j=1}^{N} -(T_{ij} - 1)A_{ij}$ is climber i's total number of attempts to reach his or her highpoints on problems he or she didn't finish, where fewer attempts is better.

Climbers are then ranked first by number of tops, second by their normalized overall hold points s_{p_i}, and third and fourth by their attempt scores s_{t_i} and s_{h_i}.

3. Properties of Ranking Methods. In this section we determine properties of the methods described in subsection 2.3 and subsection 2.4, specifically looking at the independence of irrelevant alternatives (IIA) criterion and Pareto efficiency. For both rank and score aggregation, the IIA criterion can be stated, as in [10], as follows:

The relative position of two climbers A and B in the final rank is solely determined by the performance of A and B on the problems and is not affected by the performance of other climbers.

A Pareto efficient method is also defined in [10] as a method that satisfies the following:

If climber A performs better than climber B on all problems in the competition (either has more tops, achieves a higher high point on at least one problem, or has fewer attempts on at least one problem), then B cannot be ranked above A in the final rank.

We consider it beneficial to use a ranking method that satisfies the IIA criterion, even if it has slightly worse predictive power than another non-independent method. This is because competitors' scores are often either entered into the computer incorrectly or scored incorrectly by the volunteer judges. When using an independent method, a climber's score can be changed without impacting the relative placements of other climbers. In addition, a climber's overall placement should not be affected by how another, lower-ranked climber climbed, which had occurred repeatedly under the combined rank method.

Using a method that is Pareto efficient is also a priority. It makes no sense for a climber who performed worse than another climber on every single problem to be ranked ahead in the competition, because clearly the second climber performed better.

We also verify that all the methods are monotone, which is defined in [10] as follows:

If the performance of some climber A is improved on a problem, his or her overall rank cannot become worse.

This encourages climbers to try as hard as they can to finish all their problems in a competition and ensures that improving their score after the competition will not harm their overall placement.

Finally, the Condorcet property is often looked at when analyzing social choice methods and consists of the following ([10]):

Whenever a profile (set of results) has a Condorcet candidate, the method must choose this climber to be the unique winner of the competition. A Condorect candidate is the climber who can beat every other climber in a head-to-head runoff.

However, this property is irrelevant when considering score aggregation methods because in order to calculate the Condorcet candidate, scores must be disregarded (only relative ranks are used). Thus we do not include this property in our analysis and instead look only at independence, Pareto efficiency, and monotonicity.

3.1. Borda Method. The Borda method is not independent. Consider the following example from [10]:

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Tops	Borda Points
Climber 1	1	3	3	3	3	0	2
Climber 2	2	1	1	2	2	0	7
Climber 3	3	2	2	1	1	0	6

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Tops	Borda Points
Climber 1	1	3	3	2	2	0	4
Climber 2	2	1	1	3	3	0	5
Climber 3	3	2	2	1	1	0	6

The performances of Climber 2 and Climber 3 do not change between the two sets of results. In the second table, however, Climber 1 improves his or her performance on problems 4 and 5, meaning he or she goes from being ranked last to being ranked between Climber 2 and Climber 3. When we calculate the Borda points for each climber in the first set of results, we get $b = (2, 7, 6)$ so Climber 2 wins. In the second set of results, however, $b = (4, 5, 6)$ so Climber 3 wins despite the performances of Climber 2 and Climber 3 on each problem remaining exactly the same.

Since all three climbers in this example had the same number of tops, the no-top version of the Borda method also fails to satisfy the IIA criterion.

The Borda method is, however, Pareto efficient. If Climber 1 outperforms Climber 2 on every climb, Climber 1 will be ranked higher on each climb and thus receive more Borda points than Climber 2. As more Borda points are better, Climber 1 must be ranked ahead of Climber 2.

Monotonicity is also satisfied. Moving a climber up in a ranking for a problem can never decrease his or her points, nor can it increase any other climber's points. For more explanation, see [10].

3.2. Combined Rank Method and Geometric Mean Method. The combined rank method and the geometric mean method also fail to satisfy the IIA criterion, as demonstrated by the following sets of results:

	Rank 1	Rank 2	Rank 3	Tops	Geometric Mean
Climber 1	2	1	2	0	1.59
Climber 2	1	2	1	0	1.26
Climber 3	3	3	3	0	3

	Rank 1	Rank 2	Rank 3	Tops	Geometric Mean
Climber 1	3	1	3	0	2.08
Climber 2	2	3	2	0	2.29
Climber 3	1	2	1	0	1.26

Both the combined rank method and the geometric mean method will produce the same result because no climbers tie in the rank for a given problem. In the first set of results, Climber 1 receives a score of 1.59 while Climber 2 receives a score of 1.26, so Climber 2 wins. In the second set of results, however, the performance of only the third climber is altered, but Climber 1 earns 2.08 points and Climber 2 earns 2.29 points, so Climber 1 wins. As none of the climbers finished any problems, this example also illustrates the failure of the no-top versions of the two methods to meet the IIA criterion.

Both methods are, however, Pareto efficient. If Climber 1 performs better than Climber 2 on all problems, he or she will have a strictly better rank on each problem and thus a lower geometric mean of his or her ranking points or ranks (depending on the method used). Therefore, Climber 1 must beat Climber 2.

These methods are both monotone. Improving the rank of a climber on a given problem will decrease his or her rank and ranking points, and thus also decrease his or her geometric mean. In addition, the ranks of the other climbers could only have dropped, so the climber could only increase their rank, ranking points, and geometric mean.

3.3. Geometric Median Methods. The three geometric median methods are not independent. Consider the following result sets:

	Rank 1	Rank 2	Rank 3	Rank 4	Tops	Output Coordinate
Climber 1	4	2	4	1	0	2.71
Climber 2	2	2	3	5	0	2.95
Climber 3	3	4	1	3	0	2.79
Climber 4	4	1	1	3	0	2.13
Climber 5	1	4	4	2	0	2.87

	Rank 1	Rank 2	Rank 3	Rank 4	Tops	Output Coordinate
Climber 1	4	2	3	2	0	2.63
Climber 2	1	2	2	4	0	2.25
Climber 3	1	1	5	4	0	2.70
Climber 4	5	2	3	1	0	2.60
Climber 5	1	2	1	3	0	1.77

Between the two result sets, the relative ranks of Climber 1 and Climber 2 are unaltered. However, in the first result set, the integer geometric median method produces the point $(2.71, 2.95, 2.79, 2.13, 2.87)$ so Climber 1 beats Climber 2. However, in the second result set, $(2.63, 2.25, 2.70, 2.60, 1.77)$ is the produced point, and Climber 2 beats Climber 1, despite Climber 1's performance on each problem remaining exactly the same.

Since none of the climbers finished any problems, this also signifies that the no-tops geometric median method is not independent.

In addition, this example reveals that the optimal integer geometric median method is not independent. For the first set of results, the optimal rank is $(2, 2, 2, 1, 2)$, so Climber 1 and Climber 2 tie. The second set of results, however, produces an optimal rank of $(3, 2, 3, 3, 1)$, and Climber 2 beats Climber 1.

The three methods are, however, Pareto efficient, as the following theorem and discussion shows.

THEOREM 3.1. *The no-tops geometric median method is Pareto efficient.*

Proof. We first claim that if a set of points is contained inside a convex polyhedron, so is the geometric median of those points. This is a routine calculation, as the geometric median is the point that minimizes the sums of the Euclidean distances between that point and every point in the set. For a point y outside the polyhedron containing the set, the closest point on the polyhedron, y', has a smaller sum of Euclidean distances than does y. To show this, we do the following construction:

Let l be the line through y and y' and p be a point inside the convex polyhedron. Then let h_1 be the hyperplane through y' perpendicular to the line l and h_2 be the hyperplane through p perpendicular to the line l. Let q be the point where l intersects h_2 and $d(x, y)$ denote the Euclidean distance between any two points x and y. Then

$$d^2(y, p) = d^2(p, q) + (d(q, y') + d(y', y))^2 > d^2(p, q) + d^2(q, y') = d^2(y', p). \tag{1}$$

Thus the distance from y' to any point p is always less than the distance from y to p, so the geometric median of the points contained in the convex polyhedron must also be either inside or on the boundary of the polyhedron.

Now, consider a set of results with M climbers where some climber k is ranked above another climber j on all of N problems. Let $V = \{v_1, ..., v_N\}$ be the set of the points in $\mathbb{R}^M$ corresponding to each rank of the climbers on a problem. Then $(v_i)_k < (v_i)_j$ for every point v_i, so the set of points V lies on one side of the hyperplane $H = \{x \in \mathbb{R}^M : x_k = x_j\}$. Since there are finitely many points, they can be contained inside a hypercube with one face on H. Thus the geometric median g is on or above this hyperplane, so $g_{i_k} \le g_{i_j}$ and thus climber j cannot beat climber k. ∎

Since taking the number of tops into account when ranking climbers does not affect the Pareto efficiency of the method (the climber who consistently performed worse could not have more tops than the better-performing climber), this theorem implies that the integer geometric median method is also Pareto efficient. Additionally, since the optimal integer geometric median method only differs from the integer geometric median method in that it introduces more ties, it must be Pareto efficient as well, since our only criterion is that the worse-performing climber does not beat the better-performing climber (they could tie).

Finally, none of the geometric median methods are monotone. Consider the following two sets of results:

	Rank 1	Rank 2	Rank 3	Rank 4	Tops	Output Coordinate
Climber 1	2	5	3	5	0	3.23
Climber 2	1	1	1	3	0	1.29
Climber 3	5	1	4	4	0	3.96
Climber 4	3	4	1	1	0	1.57
Climber 5	4	1	4	1	0	3.26

	Rank 1	Rank 2	Rank 3	Rank 4	Tops	Output Coordinates
Climber 1	1	5	3	5	0	3.27
Climber 2	1	1	1	3	0	1.40
Climber 3	5	1	4	4	0	3.77
Climber 4	3	4	1	1	0	1.86
Climber 5	4	1	4	1	0	2.96

The only difference between the two tables is that Climber 1's first rank improves. However, using the integer geometric median method, the first table produces a final rank where Climber 1 is third, and the second table produces a final rank where Climber 1 is fourth. Since every climber in these two tables has zero tops, this also shows that the no-tops geometric median method also fails to satisfy monotonicity.

Although it produces a monotone result for the above set of results, the optimal integer geometric median method is not monotone either. Using the following set of results, Climber 1 is ranked fourth from the first set, but fifth from the second set, despite his or her performance on the first problem improving.

	Rank 1	Rank 2	Rank 3	Rank 4	Tops	Output Coordinate	Final Rank
Climber 1	2	4	4	5	0	4.03	4
Climber 2	1	2	3	1	0	1.88	1
Climber 3	5	1	4	4	0	3.79	4
Climber 4	4	2	1	1	0	1.59	1
Climber 5	2	5	1	1	0	1.65	1

	Rank 1	Rank 2	Rank 3	Rank 4	Tops	Output Coordinate	Final Rank
Climber 1	1	4	4	5	0	3.93	5
Climber 2	1	2	3	1	0	1.92	2
Climber 3	5	1	4	4	0	3.72	4
Climber 4	4	2	1	1	0	1.55	1
Climber 5	3	5	1	1	0	1.83	2

3.4. Merged Method, ABS10 Method, Top Score Method. The merged method, ABS10 method, and top score methods do satisfy the IIA criterion, making them methods of grading as defined in [5]. This is because none of the information used to rank the climbers is changed by the addition of a third climber. If Climber 1 has more tops than Climber 2, he or she will still have more tops than Climber 2 regardless of how many tops Climber 3 has. This is the same for the high hold reached on the climb, the number of flashes a climber has, and the number of attempts a climber has, which determine the various score vectors.

These methods are also monotone. Giving a climber more tops, increasing his or her high hold on a problem, or decreasing his or her number of attempts on a problem cannot harm them overall.

Finally, all three methods are Pareto efficient. If Climber 1 performs better than Climber 2 on every problem, Climber 1 will either have more tops, more hold points, or fewer attempts, all of which imply that Climber 2 cannot be the winner because Climber 1 will be ranked ahead of Climber 2.

3.5. Linear Programming Method. The linear program does not satisfy the independence of irrelevant alternatives criterion, as shown by the following example result sets:

	Rank 1	Rank 2	Rank 3	Rank 4	Tops	Overall Rank
Climber 1	3	2	2	3	0	2
Climber 2	2	4	4	4	0	4
Climber 3	1	2	1	2	0	1
Climber 4	4	1	3	1	0	3
Climber 5	5	5	5	5	0	5

	Rank 1	Rank 2	Rank 3	Rank 4	Tops	Overall Rank
Climber 1	1	2	5	1	0	2
Climber 2	3	4	3	4	0	4
Climber 3	2	3	1	3	0	3
Climber 4	4	1	2	1	0	1
Climber 5	5	5	4	5	0	5

The relative ranks of Climber 3 and Climber 4 do not change; only Climber 1's performance on each problem is altered. However, the first set of ranks produces a final rank where Climber 3 wins and Climber 4 is third, yet with the second set of ranks Climber 4 wins and Climber 3 is third. Although occasionally there are multiple optimal ranks and the linear program only returns one, in this case the overall rank for the first result set is not optimal for the second result set, and vice versa.

We are not aware of a result proving whether or not the linear programming method is Pareto efficient. The following theorem, however, shows that if a certain climber performs better than average in a competition (meaning that the climber has, averaged across all the problems, beaten more competitors than he or she has lost to), he or she must beat at least one climber in the final rank.

Theorem 3.2. *Let $X^\star$ denote the LP ranking. If $\sum_j C_{Aj} > 0$ for some climber A, then there exists a climber k such that $X^\star_{Ak} = 1$ and, equivalently, $X^\star_{kA} = 0$.*

Proof. Assume $X^\star$ is an optimal solution and consider a perturbation of the form $X = X^\star + \epsilon Y$ where $Y_{ij} = \phi_i - \phi_j$ for some $\phi \in \mathbb{R}^N$ that we'll specify later. Since Y is skew-symmetric, the constraint $X_{ij} + X_{ji} = 1$ is satisfied. Additionally we have that

$$Y_{ij} + Y_{jk} + Y_{ki} = \phi_i - \phi_j + \phi_j - \phi_k + \phi_k - \phi_i = 0,$$

so that the constraint $X_{ij} + X_{jk} + X_{ki} \leq 2$ is satisfied.

Let

$$\phi_i = \begin{cases} 1 & i = A \\ 0 & \text{otherwise} \end{cases}.$$

We compute

$$\sum_{ij} C_{ij} X_{ij} - \sum_{ij} C_{ij} X^\star_{ij} = \epsilon \sum_{ij} C_{ij}(\phi_i - \phi_j) = \epsilon \left(\sum_j C_{Aj} - \sum_i C_{iA} \right) = 2\epsilon \sum_j C_{Aj} > 0$$

where we used the skew-symmetry of C. Since $X^\star$ is assumed be optimal, this implies that X must violate the final constraint that $X_{ij} \geq 0$. (Otherwise X would be a feasible solution with a higher objective value, contradicting the optimality of $X^\star$.) This final constraint reads

$$0 \leq X^\star_{ij} + \epsilon Y_{ij} = \begin{cases} X^\star_{ij} + \epsilon & i = A,\ j \neq A \\ X^\star_{ij} - \epsilon & j = A,\ i \neq A \\ X^\star_{ij} & \text{otherwise} \end{cases}.$$

We conclude that there exists a climber k such that $X^\star_{Ak} = 1$, and thus climber A is ranked ahead of climber k in the final rank. ∎

The implementation of the LP algorithm we used is not monotone. In particular, when there are multiple optimal solutions, the solution chosen occasionally produces a non-monotone result, as seen with these result sets:

	Rank 1	Rank 2	Rank 3	Rank 4	Tops	Overall Rank
Climber 1	2	3	4	3	0	3
Climber 2	1	2	5	4	0	5
Climber 3	4	1	1	4	0	4
Climber 4	4	3	1	1	0	2
Climber 5	2	3	1	1	0	1

	Rank 1	Rank 2	Rank 3	Rank 4	Tops	Overall Rank
Climber 1	1	3	4	3	0	4
Climber 2	1	2	5	4	0	5
Climber 3	4	1	1	4	0	2
Climber 4	4	3	1	1	0	3
Climber 5	3	3	1	1	0	1

Climber 1's performance on the first problem is improved, yet he or she drops from third to fourth in the overall rank. However, the overall rank for the second set of results is an optimal rank for the first set of results, and vice versa.

4. Analysis of Ranking Methods. To compare the predictive power of the methods in section 2, we take a similar approach to that of D. Barrow et al. in their analysis of sports ranking methods [6]. First, we perform cross validation on each method for all competition categories. The non-parametric Friedman statistical test is then performed on these cross validation scores to determine if one of the methods performs significantly better or worse than the others, and we finish up with the post hoc Nemenyi statistical test to discern which ranking methods perform significantly better than others.

4.1. Cross Validation. Cross validation determines how effectively an aggregated rank of climbers produced by a given method on some training set of climbs predicts the ranking of the climbers on another climb, the test set. We used data combined from semifinals and qualifiers of each of eight categories of the 2016 Bouldering Youth National Championships, where only the qualifying information from the climbers who climbed in semifinals was considered. Essentially, we had eight datasets containing information on seven boulder problems for twenty or twenty one climbers.

We then used leave-one-out cross validation, splitting the data into training sets and test sets where the training set contained six problems and the test set contained one problem. We applied a ranking method ϕ to the training set and then calculated the *prediction error*, $E(t)$, for the test set t, which Barrow et al. [6] defines to be

$$E(t) = \#\{t_i > t_j, \phi_i < \phi_j\} + \#\{t_i = t_j, \phi_i \neq \phi_j\},$$

where $t_i > t_j$ indicates that climber i is ranked above climber j in the test set and $\phi_i < \phi_j$ indicates that climber i is ranked below climber j in the rank produced by the given method. The lower the predictive error for a ranking method ϕ, the better the method.

The final cross validation score c_g^ϕ for a given method ϕ on a category g was calculated by summing all of the predictive errors and dividing by the number of different test sets (seven). In other words,

$$c_g^\phi = \frac{1}{7}\sum_{t=1}^{7} E(t).$$

By repeating the cross validation testing across each category, we ended up with eight evaluations of each method to be used for the Friedman test.

4.2. Friedman and Nemenyi Tests. We performed a Friedman test on the cross validation scores using `friedman` in R, which also gave us the results from a Nemenyi test if the Friedman test result proved to be significant.

We first ran a Friedman test on the cross validation scores from all the methods and, with a p-value of 2.75×10^{-8}, were able to reject the null hypothesis that all the methods have equal predictive power. As can be seen in Figure 3, the post hoc Nemenyi test revealed that the no-tops version of each method performed significantly worse than the tops version, justifying the use of taking number of tops into account when ranking climbers. Clearly, the linear program is the best-predicting method in this case, with significantly better predictive power than every method except the optimal integer geometric median method.

Fig. 3: Nemenyi test results for all methods. Green spaces (positive numbers) indicate that the method in the column performed significantly better than the method in the row, and yellow spaces (negative numbers) indicate the opposite. A 3 designates a p-value of less than .001, a 2 designates a p-value of less than .01 and a 1 designates p-value of less than .05. Blank squares indicate that the methods were not significantly different in their cross validation scores. NT means that the method is the no-tops version, GM stands for geometric median, and CR stands for combined rank.

Analysis of Ranking Methods

	ABS10	Borda	Borda NT	Geo Mean	Geo Mean NT	Lin. Prog.	Merged	Top Score	CR	CR NT	GM Integer	GM NT	GM Optimal
ABS10	NA	0	-1	2	0	3	1	0	2	0	3	2	3
Borda	0	NA	-2	0	-1	3	0	0	1	0	2	1	2
Borda NT	1	2	NA	3	0	3	3	3	3	1	3	3	3
Geo Mean	-2	0	-3	NA	-3	3	0	0	0	-2	0	0	0
Geo Mean NT	0	1	0	3	NA	3	2	2	3	0	3	3	3
Lin. Prog.	-3	-3	-3	-3	-3	NA	-3	-3	-2	-3	-1	-2	0
Merged	-1	0	-3	0	-2	3	NA	0	0	-1	0	0	1
Top Score	0	0	-3	0	-2	3	0	NA	0	0	1	0	2
Combined	-2	-1	-3	0	-3	2	0	0	NA	-2	0	0	0
Combined NT	0	-1	-1	2	0	3	1	0	2	NA	3	3	3
GM Integer	-3	-1	-3	0	-3	1	0	-1	0	-3	NA	0	0
GM NT	-2	-1	-3	0	-3	2	0	0	0	-3	0	NA	0
GM Optimal	-3	-1	-3	0	-3	0	-1	-2	0	-3	0	0	NA

We then ran a Friedman test considering only methods that did take number of tops into account, since these were the more predictive methods. Again, the p-value of 1.26×10^{-4} allowed us to reject the null hypothesis and perform the Nemenyi test. Figure 4 illustrates the results of the Nemenyi test, which were quite similar to the above results. Unsurprisingly, the linear program remained the method with the best performance, with the geometric median methods a close second. The combined rank method also had good predictive strength, as it was only significantly worse than the linear program. Of the score aggregation methods, the merged method performed the best, since it was only significantly worse than the linear program and the optimal integer geometric median method. However, all three of the ABS10, top score, and merged methods had similar predictive power, as none were significantly better than the others. Interestingly, the Borda method, a common method in social choice theory, had relatively poor predictions, worse than that of the obscure geometric mean and combined rank methods.

Fig. 4: Nemenyi test results for the methods that take number of tops into account. Green spaces (positive numbers) indicate that the method in the column performed significantly better than the method in the row, and yellow spaces (negative numbers) indicate the opposite. A 3 designates a p-value of less than .001, a 2 designates a p-value of less than .01 and a 1 designates p-value of less than .05. Blank squares indicate that the methods were not significantly different in their cross validation scores, and GM stands for geometric median.

Analysis of Ranking Methods with Tops

	ABS10	Borda	Geo Mean	Lin. Prog.	Merged	Top Score	Combined	GM Integer	GM Optimal
ABS10	NA	0	1	3	0	0	1	3	3
Borda	0	NA	0	3	0	0	0	2	2
Geo Mean	-1	0	NA	3	0	0	0	0	0
Lin. Prog.	-3	-3	-3	NA	-3	-3	-3	-1	0
Merged	0	0	0	3	NA	0	0	0	1
Top Score	0	0	0	3	0	NA	0	1	1
Combined	-1	0	0	3	0	0	NA	0	0
GM Integer	-3	-2	0	1	0	-1	0	NA	0
GM Optimal	-3	-2	0	0	-1	-1	0	0	NA

5. Conclusion. In this paper, we analyzed the properties and predictive power of nine different rank and score aggregation methods, three of which have been used by USA Climbing. While all methods were found to be Pareto efficient, and all but the linear program and geometric median methods are monotone, only the three score aggregation methods (ABS10, top score, and merged methods) satisfy the independence of irrelevant alternatives criterion. The IIA criterion is arguably quite important for our recommended method to satisfy. A large portion of the upsetting results under the combined rank method were due to lower-ranked climbers finishing and changing the ranks of climbers ranked above them, a violation of the IIA criterion. Additionally, competitors would change their score on a problem after the competition, as scoring errors occurred frequently, and the relative ranks of other climbers would be disrupted. Thus the necessity of an independent method for competitive climbing becomes clear, and as climbing naturally lends itself to score aggregation, such a method can satisfy monotonicity, Pareto efficiency, and independence while still having an unrestricted domain and being non-dictatorship.

A method used to rank climbers should also be easy for the spectators to calculate so they can accurately predict the performance of their favorite climbers. Examples of such methods are the three score aggregation methods. All of the other methods make use of the ranks of the climbers on each problem, and calculating this rank for a climber is difficult because it will change as more competitors climb. In particular, the combined rank method requires spectators to calculate geometric means of ranking points, the geometric median methods require some iterative process to calculate the geometric median, and the linear programming method is a non-trivial optimization problem.

To study the predictive power of the different methods, we used leave-one-out cross validation to produce eight different evaluations of each method, and used the non-parametric Friedman statistical test to determine if the methods had significantly different predictive powers. With a p-value of less than .001, we rejected the null hypothesis that all methods are equivalent and performed the post hoc Nemenyi test to reveal the best-predicting methods.

We found that the methods with the best predictive power were the linear program and the two geometric median methods that take number of tops into account. The combined rank method's predictive power was about average for this group of methods, indicating that it is an effective ranking method. The merged method had the best performance of the score aggregation methods, and was only slightly worse than the combined rank method.

Although the linear program made the best predictions, it does not satisfy the IIA criterion nor does it allow spectators to calculate the rank of their favorite climbers. In addition, there are often multiple solutions to the linear program, and our algorithm for solving the linear program only returns one of these solutions. This means that, depending on the algorithm used to solve the linear program, different solutions could be given and there is no good way to choose between them.

A similar issue arises with the integer geometric median method. Despite its strong predictive power, the method is not independent and not easy for spectators to use. The optimal integer geometric median performed better than the integer version, but the only difference between the two is the creation of more ties, which are generally considered to be undesirable in climbing competitions.

The combined rank method suffers from the same issues as the integer geometric median method, and climbers have made it clear that they did not appreciate the use of this method. Since the merged method performed only slightly worse than the combined rank method and alleviates these issues, we suggest its use over that of the combined rank method

Thus we recommend the merged method as a ranking method for climbing competitions. It has good predictive power, satisfies all desired criterion, and is easy for spectators to calculate as they watch the competition.

REFERENCES

[1] *American Bouldering Series 2008-2009 member rulebook.* USAC Rules Committee, 2008.
[2] *USA Climbing Rulebook.* USAC Rules Committee, 2016.
[3] *USA Climbing.* http://www.usaclimbing.org/, June 2017.
[4] K. J. Arrow, *A difficulty in the concept of social welfare*, The Journal of Political Economy, 58 (1950), pp. 328–346.
[5] M. Balinskia and R. Laraki, *A theory of measuring, electing, and ranking*, Proc. Natl. Acad. Sci. USA, 104 (2007), pp. 8720–8725.
[6] D. Barrow, I. Drayer, P. Elliott, G. Gaut, and B. Osting, *Ranking rankings: an empirical comparison of the predictive power of sports ranking methods*, Journal of Quantitative Analysis in Sports, 9 (2013), pp. 187–202.
[7] C. Danielson. USAC Routesetter. Personal Communication, December 2016.
[8] A. N. Langville and C. D. Meyer, *The Science of Rating and Ranking: Who's #1?*, Princeton University Press, Princeton, NJ, 2012.
[9] K. Parker. Project Code: https://github.com/kiraclimber/Fall2016-Math-REU, December 2016.
[10] D. H. Ullman and J. Robinson, *A Mathematical Look at Politics*, CRC Press, Boca Raton, FL, 2011.
[11] E. Weiszfeld and F. Plastria, *On the point for which the sum of the distances to n given points is minimum*, Ann. Oper. Res., 167 (2009), pp. 7–41.