A Graph-Based Approach to Better Sports Rankings

Abigail R. Roane, Chaiwat Ekkaewnumchai, Connor W. McNamara, Kyle Richards, Gabor Sarkozy and Craig E. Wills*

Computer Science Department
Worcester Polytechnic Institute
Worcester, MA 01609

WPI-CS-TR-19-03

June 2019

Abstract

Rankings are used in sports to reflect the relative strength of teams in comparison with each other. In this work we take a new approach to sports rankings that focuses on the actual games played where the primary goal of the ranking is to minimize "misranked" game results in which a lower-ranked team defeats a higher-ranked team. This approach is based on the premise that the fairest rankings system adheres as closely as possible to the actual games that have been played by minimizing the percentage of these misranked games.

In addition to defining this new misrank percentage metric, we have developed a graph-based ranking algorithm that determines the optimal ranking for a set of teams that best matches actual game results. By default all games are weighted equally in the rankings, but we can adjust this weight based upon game recency and point margin parameters. Applying our graph-based ranking algorithm to six sports leagues and twelve existing rankings, we learn the values of these parameters that best match existing rankings. For example, we find that College Football rankings value early season game results as much as those later in the season.

We find that our new graph-based algorithm yields significantly lower misrank percentages than existing sports rankings. For example for the 2018 National Football League regular season, the misrank percentage for the NFL.com ranking is 40% higher than for the comparable graph-based ranking using best-matched game recency and margin parameters. Even worse, the misrank percentage for the 2018 College Football Playoff Poll is twice as much as the graph-based ranking with the same parameters. The misrank percentage for the 2018 AP Poll College Basketball rankings is more than twice as much as the comparable graph-based ranking. These results show that our graph-based rankings are better than existing rankings in remaining true to actual game results and that minimization of games between misranked teams can and should be used as a basis for sports rankings.

*Corresponding Author: cew@cs.wpi.edu

1 Introduction

Rankings are used in sports to reflect the relative strength of teams in comparison with each other. Many approaches are used to rank teams—based upon a rating [14], polls by sports writers [1] or coaches [6], consensus power rankings from an expert panel [17] or computerized algorithms [19] that reflect various weights. These rankings account for (either explicitly or implicitly) factors such as wins/losses, when games are played, where games are played, margin of victory, strength of schedule, injuries to key players, performance in past years and name recognition.

In this work we take a new approach to sports rankings that focuses on the actual games played where the primary goal of the ranking is to minimize misranked game results in which a lower-ranked team defeats a higher-ranked team. This approach is based on the premise that the fairest rankings system adheres as closely as possible to the actual games that have been played by minimizing the percentage of "misranked" game results in which a lower-ranked team has defeated a higher-ranked team. In a perfect rank ordering there would not be any misranked games in which the lower-ranked team won, but with enough games played in any sports league season it is unlikely to have a perfect ordering. In the simplest such case, if Team A beats Team B, Team B beats Team C and then Team C beats Team A then there is no ordering between these three teams that will avoid such a scenario.

In addition to defining this new misrank percentage metric, we design and implement an algorithm that is able to determine the ordering of teams based on using it as the optimizing metric. We apply the algorithm to game results for a variety of sports league seasons. We use an approach that parameterizes the value of games based upon recency and point spread as well as doing so in a manner that disambiguates amongst multiple equivalent orderings to determine the best ordering.

Our work makes a number of contributions:

1. definition of a misrank percentage metric that measures how closely a ranking adheres to actual game results,

2. a graph-based ranking algorithm that can be applied to any sports league allowing rankings for the teams in the league and comparisons across sports leagues,

3. algorithmic approximations that make it computational feasible to determine rankings with minimized misrank percentage for sports leagues with a varied number of teams,

4. adjustable parameters to handle recency and point differential of games,

5. the ability to learn the inherent importance of game recency and margin in existing rankings, and

6. an evaluation of existing rankings showing that they do not perform as well as our graph-based algorithm in adherence to actual game results.

In the remainder of this report we describe the graph-based ranking algorithm and how we resolved various issues in its design and implementation. We describe the data and external rankings to which we apply the algorithm and compare with the subsequent graph-based ranking results. We go on to show sample rankings from actual data and how the misrank percentage metric varies across six sports leagues. We use the parameterized algorithm to learn the relative importance of game recency and margins for the various sports rankings and evaluate the misrank percentage for a variety of rankings. We conclude with a summary of results and directions for future work.

2 Graph-Based Ranking Algorithm

In computer science, a *graph* is a convenient data structure consisting of *nodes* and *edges* between those nodes. An example graph is shown in Figure 1, where A, B, C, and D are the four nodes (also called *vertices*) and the lines between them are the five edges of the graph. This graph is *directed* because the edges have a direction as indicated by the arrows. One might use a graph like this to represent flights between cities—nodes represent cities and edges represent existing flights between those cities.

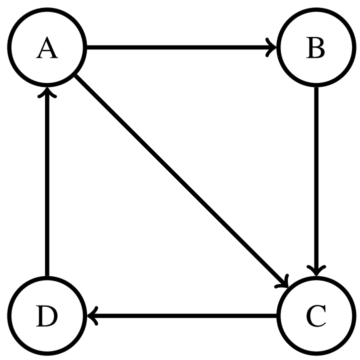

Figure 1: A Directed Graph with Four Nodes and Five Edges

Figure 1 is *cyclic* because there is a cycle from node A to C to D and back to A. A *feedback arc set* is a set of edges that can be removed from a graph to turn it into a directed *acyclic* graph. The *minimum feedback arc set* is a feedback arc set with the least possible number of edges [24]. There can be multiple minimum feedback arc sets for each graph. For example, removing the edge from node C to D in Figure 1 results in a directed acyclic graph. Similarly, removing the edge from node D to A results in a directed acyclic graph, which is shown in Figure 2. This single "backedge" (shown in red in the figure) is a minimum feedback arc set for the graph and the remaining edges are rank ordered with no cycles.

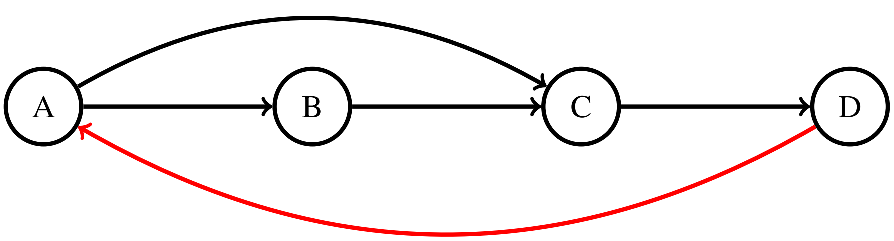

Figure 2: Ordered Graph with a Single Backedge

In this work we use a graph-based approach to represent games played between teams in a sports league, such as the National Football League, where nodes are teams and a directed edge from node A to node B in the graph indicates team A defeated team B in a game. We define a ranking for the set of teams based on the games played between them by finding a minimum feedback arc set (as done in Figure 2) and using the resulting node ordering as the ranking. We define the metric "misrank percentage" as the percentage of backedges in an ordered graph out of the total number of edges. This metric is so named because it is the percentage of games in which a lower-ranked team defeats a higher-ranked team. In Figure 2, the misrank percentage is 20% as there is one backedge out of a total of five edges.

While this is a simple and elegant approach to ranking a set of teams based upon the games played between them, there are a number of issues to resolve in doing so. These issues include:

1. Finding a minimum feedback arc set for any graph is an *NP-hard problem* (see [12], p. 192), which means that there is no known algorithm that can solve this problem in polynomial time [23]. From a practical standpoint, this problem means that the computation time grows exponentially as the number of nodes (teams) increases. In our work we see a dramatic increase in computation time as the number of teams grows to more than 30.

2. There can be multiple minimum feedback arc sets for a graph resulting in multiple ordered graphs with the same misrank percentage. Determining the best ordering from multiple equivalent (based on the misrank percentage metric) orderings is an issue we need to resolve.

3. As described thus far, all edges have the same weight meaning that all games count the same. However, factors such as the score margin in the game (was it close or was it a blowout) or the recency of the game (was the game played early or late in the season) may mean that some edges are weighted more than other edges. We need to resolve how to consider and account for different edge weights in our algorithm.

In the remainder of this section we summarize how we resolve each of these issues in our approach. See [18] for more details on how each of these aspects are handled, variants of the algorithm that we explored, and how the algorithm was implemented.

2.1 Handling Computational Cost of Algorithm

Our approach needs to determine which order of nodes gives the smallest percentage (or weight) of backedges. The straightforward approach to solve this problem is to simply use "brute force" and consider all possible orderings of nodes and select an ordering that provides the minimal weight. However, this approach does not scale. Instead we use a dynamic programming algorithm, which finds an optimal order with a significantly reduced time complexity compared to the brute-force algorithm.

We designed a dynamic programming algorithm based on subproblems. For each subproblem, we find the minimum feedback arc set for a smaller set of nodes. This dynamic programming algorithm is bottom-up, which keeps only an optimal solution for each set of nodes.

In addition to dynamic programming, we use a number of other techniques to increase the computational efficiency and improve the resulting order. These include:

1. *pruning* to remove subrankings with a high total backedge weight;

2. *parallelization* of subproblems among separate threads of execution;

3. *preording* of teams based on net edge weight in and out of nodes, which by default is simply the team's win/loss record; and

4. the use of smaller *sliding windows* (or ranges) of teams where our dynamic programming algorithm is applied to a smaller range of teams at a time.

2.2 Postprocessing to Choose From Multiple Orderings

Once a ranking is created by our algorithm, there is typically flexibility to move teams without changing the backedge weight. Such a common situation is when two teams are next to each other in a ranking, but they have not played each other. Since there is no edge between the two nodes in the graph, no backedge will be introduced by switching the team's places in the ranking. As a consequence, once a ranking is determined by the algorithm we next postprocess the results in two ways.

First we determine what we call the *Range of Correctness (RoC)* for each team in the ranking. This postprocessing step indicates the certainty of each team's placing within a ranking. The Range of Correctness does not choose the best ranking given equivalent backedges. Rather, it is intended to allow a viewer to understand the "play" of each team within a given ranking. For a given team, the top of the range is one below the next lowest ranked team to which the team has lost, while the bottom of range is one above the next highest rank team over which the team has won. Note: an undefeated team will always have an upper range of 1 since any undefeated team could be the top-ranked team. Similarly any winless team will always have a lower range of the number of teams since any such team could be the lowest ranked team.

The second postprocessing step reconsiders each team's ranking (within the limits of the Range of Correctness) based on a secondary metric that considers the relative strengths and weaknesses of a team's wins and losses. This step is where the strength of a team's schedule is utilized. This secondary metric gives a higher score to teams who defeat teams that are higher in the ranking, and a lower score to teams who defeat teams that are lower in the ranking. It penalizes teams less when they are beaten by higher-ranked teams and penalizes teams more when they are beaten by lower-ranked teams. The weighted score is defined in Equation 1.

$$score(v_i) = \sum_{j=0}^{n-1} (n + 1 - rank(v_j)) \times w(v_i, v_j) - \sum_{j=0}^{n-1} rank(v_j) \times w(v_j, v_i) \tag{1}$$

where $rank(v_i)$ is the rank (starting from 1) of node (vertex) v_i, and n is the number of teams. When this score is maximized, we have the best possible ranking according to our secondary metric. In the default case, the weight of all edges ($w(v_i, v_j)$) is 1.0, meaning all games count equally regardless of the score or when the game was played.

2.3 Accounting for Recency and Margin of Games

A third issue to resolve with our approach is how to weight the value of each game played between two teams. As indicated in Equation 1 for our secondary weighted score metric, there is a weight associated with each edge with a default value of 1.0. However, depending on the sport this weight may not be appropriate to use for all games in generating the best ranking. We introduce two parameters, where each can take on a value between 0.0 and 1.0, to account for game recency and margin in our rankings.

The recency parameter considers the timeframe from the beginning of a sport season until its current point. Its value specifies the minimum weight of a win at the beginning of a season. For example, if the recency parameter is 1.0 then a win at the beginning of season counts the same as a win at the current point of the season—all wins are the same weight based on recency. If the recency parameter is 0.0 then a win at the beginning of season counts for zero weight. We use a linear distribution between the beginning and current point of the season. Thus if we are computing a ranking at the end of the season with a recency parameter of 0.5, then a win at the beginning of the season counts 0.5, a win half way through the season counts 0.75 and win at end of the season counts 1.0.

Similarly, we use the margin parameter to weight the value of the game based upon the point spread of the win. This parameter specifies the minimum weight of a close win in contrast to a "big" win that counts for a weight of 1.0. For example, if the parameter has a value of 1.0 then a close win counts the same as a big win—all wins are the same weight based on margin. If the parameter has a value of 0.0 then a close win counts for zero weight. We use linear distribution between a close and a big win. Thus if we are computing a ranking with a margin parameter of 0.5, then a close win counts 0.5, a moderate-sized win half way between a close and big win counts 0.75, and big win counts 1.0.

A key decision with the margin parameter is what constitutes a "big" win. Obviously point differential for such a win varies across different sports. After examination of distributions of point differences, we chose to define the big win point differential as the 75th-percentile of all point differentials for a given sport. Examples of this 75th percentile point differential for various sports are: National Football League, 17 points; National Hockey League, 3 goals; Major League Baseball, 5 runs; and NCAA Basketball, 18 points. A big win with full margin value is assigned to any game in which a team wins by at least this point differential and the value for smaller wins is linearly scaled from the margin parameter value for a one-point win.

The end result of these considerations is that the weight for each edge is calculated as the product of the recency and margin weights for the edge. For example, if the recency and margin parameters are each 0.5 then a close win at the beginning of the season will translate into an edge weight of 0.25 in comparison to a big win at the end of the season, which will translate into an edge weight of 1.0.

3 Methodology

A feature of our approach is that it can be applied to compute rankings for any sports league. We chose to compute rankings for the following six leagues: Major League Baseball (MLB), the National Basketball Association (NBA), the National Football League (NFL), the National Hockey League (NHL), NCAA College Basketball (CBB) and NCAA College Football (CFP). These sports leagues vary significantly in number of teams, number of games, and average points per game. We chose the first four leagues because they are popular professional sports in the United States. We also wanted to include NCAA College Basketball and Football because they present interesting ranking challenges due to the large number of teams. Additionally, rankings for these leagues are important because championship tournament members in NCAA College Football and

Basketball are determined by a ranking.

The MLB, NFL, NBA, and NHL leagues all result in relatively dense graphs because teams play each other many times. In the NBA and NHL, all nodes are connected (have an edge between them) since every team plays every other team at least twice. Additionally, MLB teams each play 162 games, meaning that two teams may play each other many times.

NCAA Basketball and Football provide different graphs from those of professional leagues. NCAA Basketball and Football graphs have many more nodes and are sparser in the set of edges (games) than their professional league counterparts. These graphs often have small, dense subgraphs (due to conference play) within the larger, sparser graphs.

Overall, the differences between each sport analyzed in this project significantly affect the ranking algorithm design. It must account for both sparse and dense graphs, as well as varying numbers of nodes. It was important to consider these factors when creating the algorithm to translate graphs into rankings.

Table 1 summarizes the sports leagues that are analyzed along with the specific sports season that is used in the reported results. Data for each each league were obtained from websites referenced in the table. The table also shows other rankings that were available for each league. These external rankings allow us to compare the rankings that we generate with known rankings. In all cases, the set of data and the rankings were captured at the end of the regular season games, but before any post season games were played.

Table 1: Analyzed Sports Leagues

Sports League	Abbrev.	Season	Comparable External Rankings
Major League Baseball [3]	MLB	2018	ESPN [7], NBC Sports [9]
National Basketball Association [4]	NBA	2017-18	ESPN [8], NBA.com [15]
National Football League [16]	NFL	2018	ESPN [10], NFL.com [17]
National Hockey League [13]	NHL	2017-18	ESPN [11], SI [20]
NCAA College Basketball [21]	CBB	2017-18	AP Poll [2], Coaches Poll [6]
NCAA College Football [22]	CFB	2018	AP Poll [1], CFP [5]

4 Sample Ranking

As an illustration of our approach, Table 2 shows the application of our graph-based ranking algorithm to the 2018 NFL season where the recency and margin parameters are both set to 1.0 so that all games played are weighted equally regardless of when played or the point differential of the game.

Above the table is shown that the misrank pct. for this ranking and parameters is 19%. This percentage means that despite being an optimal ranking there are 19% of game results teams where a lower-ranked team defeated a higher-ranked team. The misrank pct. is a reflection of the sport and weighting of the games. In subsequent results we show how the misrank pct. varies as the sport and parameters change.

The results in Table 2 show that using these parameters the best ranking has New Orleans as the best team with New England ranked number two. Again keep in mind that these rankings are based

Table 2: 2018 NFL Regular Season Ranking with All Games Weighted Equally

Recency = 1.0 Margin = 1.0 Misrank Pct. = 19%

Rank	Team	RoC
1.	New Orleans	[1, 4]
2.	New England	[1, 2]
3.	Houston	[3, 8]
4.	Chicago	[3, 4]
5.	L.A. Rams	[5, 5]
6.	Kansas City	[6, 6]
7.	Baltimore	[7, 7]
8.	L.A. Chargers	[8, 8]
9.	Cleveland	[9, 9]
10.	Denver	[10, 10]
11.	Pittsburgh	[11, 12]
12.	Seattle	[11, 16]
13.	Cincinnati	[12, 13]
14.	Indianapolis	[14, 14]
15.	Buffalo	[15, 15]
16.	Tennessee	[16, 16]
17.	Dallas	[17, 18]
18.	Minnesota	[16, 18]
19.	Philadelphia	[19, 19]
20.	Atlanta	[20, 20]
21.	Washington	[21, 21]
22.	Jacksonville	[22, 24]
23.	Detroit	[19, 23]
24.	Carolina	[24, 24]
25.	N.Y. Giants	[25, 25]
26.	Tampa Bay	[26, 30]
27.	Arizona	[24, 27]
28.	Green Bay	[28, 28]
29.	Miami	[29, 29]
30.	N.Y. Jets	[30, 32]
31.	San Francisco	[29, 31]
32.	Oakland	[32, 32]

only on the regular season results. The Range of Correctness (RoC) shows that the rank range for New Orleans is between 1 and 4 (they beat the 5th-ranked L.A. Rams) while it is between 1 and 2 for New England (they beat 3rd-ranked Houston). There is no explicit ordering between the two teams because they did not play each other in the regular season. New Orleans is ranked higher because it has a higher secondary weight metric value (See Section 2.2) than New England.

The RoC results in the table show that many teams are fixed into a particular ranking because they have lost to the team ranked immediately above them and beat the team ranked immediately below. Such an example are the L.A. Rams, ranked 5th, who lost to 4th-ranked Chicago and beat 6th-ranked Kansas City. Changing the ordering would increase the overall misrank percentage.

5 Comparison of Sports Rankings Using Misrank Percentage

Now that we have seen an example of actual rankings for one sports league, the availability of a ranking system that can be applied to any sport affords a number of opportunities for further analysis. The first of these analyses compares sports leagues by using a common set of ranking parameters to understand how the misrank percentage varies as we examine different leagues. For this comparison we again use a recency and margin parameter value of 1.0 as done for results in Table 2.

Figure 3 shows the relative misrank pct. for the six sports leagues using the parameters that weight all games equally. The results are rank ordered with CFB having the lowest misrank pct. at 11% with the NFL second best at 19% (as also shown in Table 2). MLB has the highest misrank pct. for these parameters at 38%.

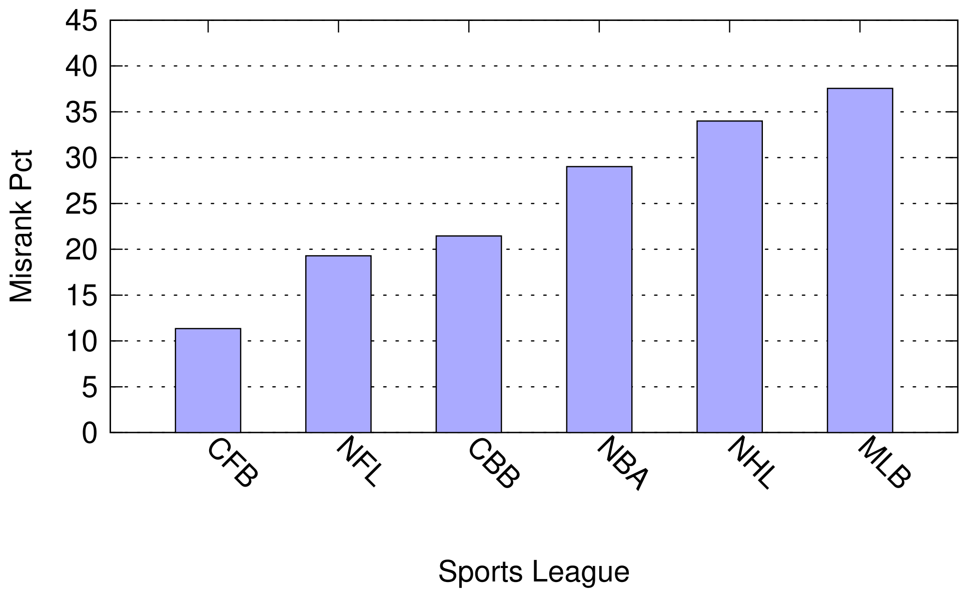

Figure 3: Relative Misrank Percentage for Best Ranking of Six Sports Leagues with All Games Weighted Equally (Regardless of Margin and Recency)

These results are influenced by the graph density of games played as well as the many number of games between the same teams within a season. MLB teams do not play all other teams, but the schedule has many games between the same two teams and it is unlikely that one team always beats the other.

6 Learning What Factors Are Important in Sports Rankings

The flexibility of our ranking systems allows us to learn about the importance of game recency and margins in the existing rankings used for a variety of sports. In some rankings these factors may be explicitly used while in other rankings these factors may implicitly be considered as part of the ranking.

As shown in Table 1 we gathered information on two well-known external rankings for each of the six leagues. These rankings are from organizations such as ESPN, NBC Sports, the Associated Press, NFL.com, and NBA.com. These are respected rankings that provide context for the rankings that we generate. Additionally, we develop a simple means to compare rankings that allows us to look at these external rankings and learn the influence of game recency and point differential in these rankings.

Our method of learning about these external rankings is to generate our rankings for a variety of parameter combinations. We test with the recency parameter set to the five values of 0.0, 0.25, 0.5, 0.75 and 1.0. In combination we test with the margin parameter set to the same five values for a total of 25 parameter combinations. For each combination we find the graph-based ranking that results in the lowest misrank pct. and compare the ranking to determine its closeness to an external ranking.

As an example, we return to the NFL where we find the ranking that most closely matches the end-of-the-regular-season ESPN ranking [10] is generated with a recency parameter of 0.75 (not much variance in weight based on recency) and a margin parameter of 0.25 (more variance in weight based on margin). The resulting graph-based ranking is shown along side the ESPN ranking in Table 3. The table shows New Orleans and New England are the two top ranked teams as shown for different parameters in Table 2, but there is variation in the remainder of the rankings with Chicago ranked third with these new parameters.

Table 3 shows that in the ESPN rankings the L.A. Rams are the top-ranked team followed by New Orleans, Kansas City and New England. The last column in the table shows the relative difference in the ranking place between our graph-based ranking and the ESPN ranking. The Rams have rank difference of +3 while New Orleans has a rank difference of -1. Buffalo has the largest rank difference (-13) with a ranking of 28 in ESPN and 15 in our graph-based approach.

We compute the *average rank difference* between our generated ranking and an external ranking by accumulating the absolute value of the rank difference for each team and then dividing by the number of teams to get an average. This average of 4.0 between the graph-based and ESPN rankings is shown at the top of Table 3. This average rank difference is the smallest of the 25 graph-based rankings generated with the various parameter combinations for College Football. The top of the table also shows that this graph-based ranking has a misrank pct. of 16%.

As another example of comparing a generated ranking with an external ranking, Table 4 shows the graph-based ranking of College Football that is closest to the end-of-regular-season College Football Playoff rankings [5]. As shown, this is the ranking generated with a recency parameter of 1.0 and a margin parameter of 0.5. The recency parameter of 1.0 means that games are weighted equally regardless of when played in the season with the margin of each game having a moderate effect on the game weight.

The table has a number of interesting results. Most notably the first four teams are exactly the same as were chosen for the past year's College Football Playoff System. A difference between the two rankings does not occur until the fifth rank where the CFP ranked Georgia at that spot while the graph-based approach ranked them 9th. The biggest difference between the two rankings was for Iowa State, which is ranked 24th in the CFP, but is 56th in the graph-based ranking. Note that while only the top 25 teams are shown for the graph-based ranking all 130 teams are ranked in order by the algorithm with the full ranking available in [18].

Once again the average rank difference is shown at the top of Table 4 with a value of 5.6. Again this Average Rank Difference is the smallest of the 25 graph-based rankings generated with

Table 3: NFL Comparative Rankings with Smallest Average Rank Difference between Graph-Based and ESPN Rankings

Recency = 0.75 Margin = 0.25 Avg Rank Diff = 4.0 Misrank Pct. = 16%

Rank	Graph-Based	ESPN	Rank Difference
1.	New Orleans	L.A. Rams	[+3]
2.	New England	New Orleans	[-1]
3.	Chicago	Kansas City	[+2]
4.	L.A. Rams	New England	[-2]
5.	Kansas City	L.A. Chargers	[+2]
6.	Baltimore	Houston	[+2]
7.	L.A. Chargers	Pittsburgh	[+3]
8.	Houston	Chicago	[-5]
9.	Denver	Seattle	[+2]
10.	Pittsburgh	Baltimore	[-4]
11.	Seattle	Dallas	[+7]
12.	Cleveland	Minnesota	[+5]
13.	Cincinnati	Indianapolis	[+1]
14.	Indianapolis	Carolina	[+9]
15.	Buffalo	Denver	[-6]
16.	Tennessee	Tennessee	[0]
17.	Minnesota	Philadelphia	[+2]
18.	Dallas	Washington	[+4]
19.	Philadelphia	Green Bay	[+9]
20.	Atlanta	Miami	[+9]
21.	Detroit	Atlanta	[-1]
22.	Washington	Cleveland	[-10]
23.	Carolina	Tampa Bay	[+3]
24.	Jacksonville	Jacksonville	[0]
25.	N.Y. Giants	Cincinnati	[-12]
26.	Tampa Bay	N.Y. Giants	[-1]
27.	Arizona	Detroit	[-6]
28.	Green Bay	Buffalo	[-13]
29.	Miami	N.Y. Jets	[+1]
30.	N.Y. Jets	Arizona	[-3]
31.	San Francisco	San Francisco	[0]
32.	Oakland	Oakland	[0]

the various parameter combinations. The top of the table also shows that this graph-based ranking has a misrank pct. of 10%.

While these two specific ranking comparisons are interesting in themselves, a contribution of our graph-based ranking approach is using the same approach for all twelve external rankings to learn the relative importance of game recency and margin in each of these rankings.

The results of this analysis are shown in Figure 4 where the game margin and recency parameters that best match each of the twelve sport league rankings are shown (along with the average rank difference). Results shown in Table 3 locate the point on the grid for NFL:ESPN at recency=0.75, margin=0.25 while the results shown in Table 4 locate the point on the grid for CFB:CFP at recency=1.0, margin=0.5.

Figure 4 shows a number of interesting results regarding the importance of game recency and margin in sports league rankings.

Table 4: College Football Comparative Rankings with Smallest Average Rank Difference between Graph-Based and CFP Rankings

Recency = 1.0	Margin = 0.5	Avg Rank Diff = 5.6	Misrank Pct. = 10%
Rank	Graph-Based	College Football Playoff	Rank Difference
1.	Alabama	Alabama	[0]
2.	Clemson	Clemson	[0]
3.	Notre Dame	Notre Dame	[0]
4.	Oklahoma	Oklahoma	[0]
5.	Central Florida	Georgia	[+4]
6.	Ohio State	Ohio State	[0]
7.	Michigan	Michigan	[0]
8.	Louisiana State	Central Florida	[-3]
9.	Georgia	Washington	[+7]
10.	Kentucky	Florida	[+10]
11.	West Virginia	Louisiana State	[-3]
12.	Boise State	Penn State	[+2]
13.	Fresno State	Washington State	[+4]
14.	Penn State	Kentucky	[-4]
15.	Texas	Texas	[0]
16.	Washington	West Virginia	[-5]
17.	Washington State	Utah	[+8]
18.	South Carolina	Mississippi State	[+3]
19.	Missouri	Texas A&M	[+3]
20.	Florida	Syracuse	[+16]
21.	Mississippi State	Fresno State	[-8]
22.	Texas A&M	Northwestern	[+11]
23.	Vanderbilt	Missouri	[-4]
24.	North Carolina State	Iowa State	[+32]
25.	Utah	Boise State	[-13]

1. All sports rankings have at least a moderate amount of consideration for games regardless of when they are played (with a recency parameter of 0.5 or greater).

2. That result is particularly true for College Football where rankings match the closest with a recency parameter of 1.0 meaning early season games matter as much as those in the late season.

3. There is much variation in terms of importance of margin. One ranking (NHL:ESPN) was best matched with a parameter of 1.0 meaning point differential did not make a difference, but the other NHL ranking had the parameter at 0.0 so there are no clear conclusions for the sport.

4. The two NBA rankings are consistent with a moderate value of 0.5 for each parameter.

5. The two College Basketball rankings are also consistent with a moderate value for recency and a value of 0.0 for margin parameter indicating maximum sensitivity to point margin for this sports league.

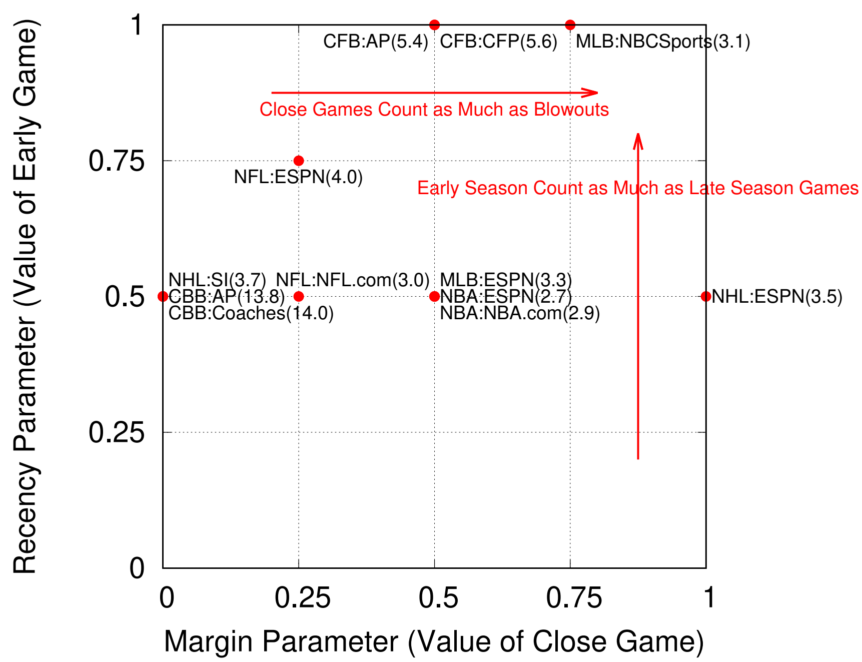

Figure 4: Game Margin and Recency Parameters That Best Match Each Sport:Ranking (Average Rank Difference)

7 Misrank Percentage Comparison

In the final part of our work we examine how the graph-based rankings compares with the relevant external rankings for each of the six sports leagues. For this portion of the analysis we focus on the misrank percentage for each ranking across a variety of game recency and margin parameters. In each case, we compare with parameter value pairs of (0.0, 0.0), (0.5, 0.5) and (1.0, 1.0) for each ranking and sports leagues. We also compute misrank percentage results for additional parameter value pairs shown in Figure 4 that result in the closest matches between the graph-based and an external rankings.

The first set of misrank pct. results are shown for the NFL in Figure 5. They include results for the rankings shown in Tables 2 and 3. The results show that the misrank pct. for the NFL.com rankings are a bit better than the ESPN ranking for all parameter value pairs. More importantly, the figure shows that the misrank pct. for the graph-based ranking is much lower than the two relevant external rankings for all parameter value combinations. For example, as shown in Figure 4, the closest match for the NFL.com rankings is the (0.5,0.25) value pair. Figure 5 shows that the graph-based Ranking has a misrank pct. of 16% while the NFL.com has a misrank pct. of 23%, roughly a 40% increase in the number of actual game results that do not match the rankings order of the teams involved.

Similarly, Figure 6 shows College Football misrank pct. results for various parameter value pairs, including (1.0,0.5) used for the closest graph-based ranking in Table 4. The results show the the CFP rankings produced slightly lower misrank pct. results, but the graph-based rankings are much better. For example, the (1.0,0.5) recency,margin parameter value pair results in a 10%

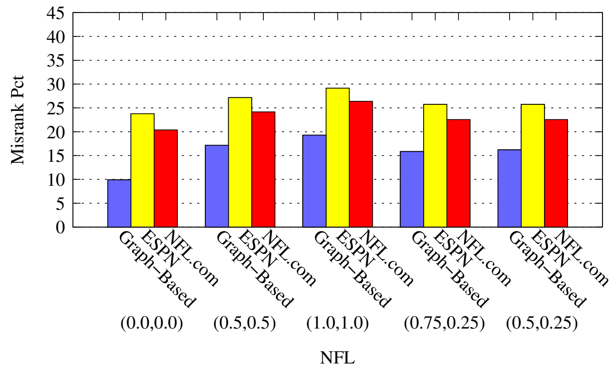

Figure 5: NFL Misrank Percentages for Three Rankings and Across (Recency, Margin) Parameters Value Pairs

misrank pct. for the 25 ranked teams while the CFP has a misrank pct. that is twice as much. Again, the graph-based approach is more accurate in ordering the teams to reflect actual game results.

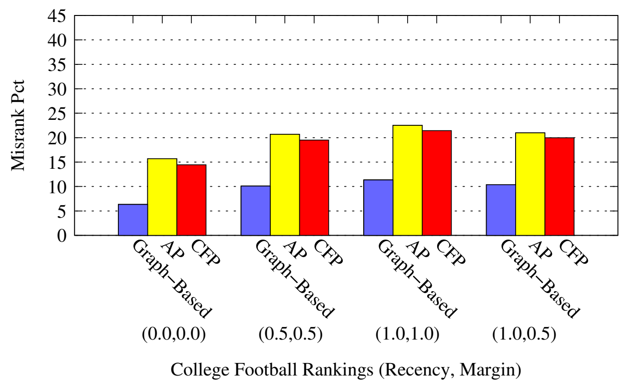

Figure 6: College Football Misrank Percentages for Three Rankings and Across (Recency, Margin) Parameters Value Pairs

Differences also exist between misrank pct. results for the graph-based rankings and relevant existing rankings for MLB, the NBA and the NHL as shown in Figures 7-9. However, the misrank percentage for all of these rankings is higher and the relative difference between the graph-based ranking and the existing rankings is less.

The misrank pct. results differences are even more pronounced for College Basketball rankings as shown in Figure 10. Using the closest (recency, margin) value pair of (0.5,0.0) from Figure 4, the misrank pct for the AP Poll ranking results in a misrank pct. that is more than twice as much as the graph-based ranking.

8 Summary and Future Work

In this work we have taken a new approach to sports rankings that focuses on the actual games played where the primary goal of the ranking is to minimize "misranked" game results in which a lower-ranked team defeats a higher-ranked team. This approach is based on the premise that the

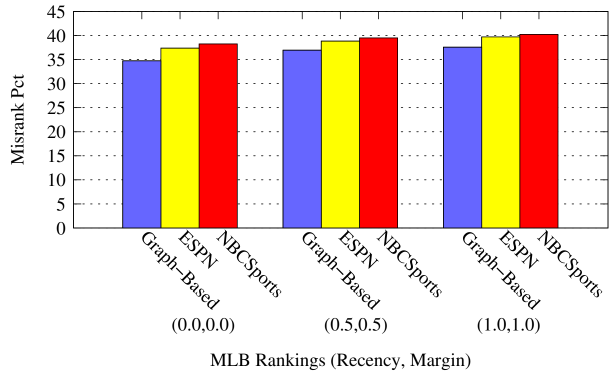

Figure 7: MLB Misrank Percentages for Three Rankings and Across (Recency, Margin) Parameters Value Pairs

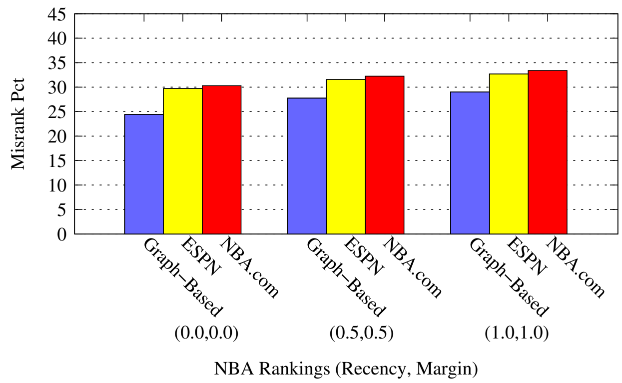

Figure 8: NBA Misrank Percentages for Three Rankings and Across (Recency, Margin) Parameters Value Pairs

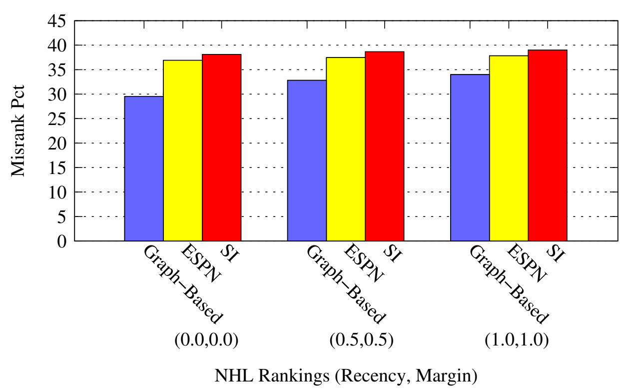

Figure 9: NHL Misrank Percentages for Three Rankings and Across (Recency, Margin) Parameters Value Pairs

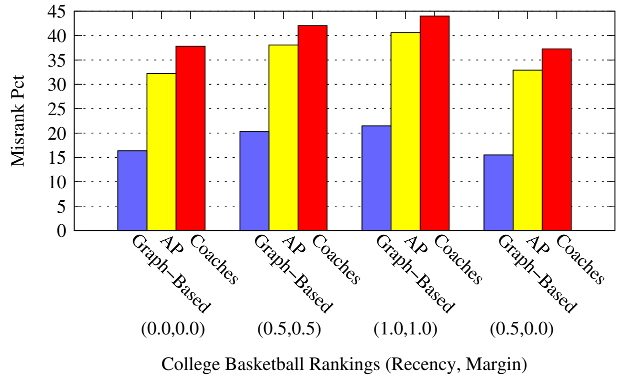

Figure 10: College Basketball Misrank Percentages for Three Rankings and Across (Recency, Margin) Parameters Value Pairs

fairest rankings system adheres as closely as possible to the actual games that have been played by minimizing the percentage of these games.

In addition to defining this new misrank percentage metric, we have developed a graph-based ranking algorithm that determines the optimal ranking for a set of teams based on the actual game results. By default all games are weighted equally in the rankings, but we can adjust this weight based upon game recency and point margin.

We apply the graph-based algorithm to games from six sports leagues and find that we are able to generate a ranking for College Football that results in the lowest misrank percentage with the optimal National Football League ranking having the second-lowest percentage. The ranking for Major League Baseball teams results in the highest misrank percentage amongst the six leagues because of the volume and density of games played between teams.

We also examined twelve existing rankings for these sports leagues in our work. The flexibility of our graph-based ranking systems allows us to learn about the importance of game recency and margins in each of these rankings. The results show that all sports rankings have at least a moderate amount of consideration for each game regardless of when it was played. This result is particularly true for College Football where early season games matter as much as those later in the season. There is more variation amongst rankings for the importance of game margin.

Our final set of results show that our new graph-based algorithm yields significantly lower misrank percentages than existing sports rankings. For example in the NFL, the misrank percentage for the NFL.com ranking is 40% higher than for the graph-based ranking using best-matched game recency and margin parameters. Even worse, the misrank percentage for the College Football Playoff Poll is twice as much as for the graph-based ranking with the same parameters. The misrank percentage for the AP Poll College Basketball rankings is more than twice as much as the comparable graph-based ranking. These results show that our graph-based rankings are better than existing rankings in remaining true to actual game results and that consideration of misrank percentage should be included in sports rankings.

Despite these promising results, there are a number of direction for future work. These include to:

1. make algorithm improvements for better orderings as the number of teams grows,

2. investigate additional approaches for the selection of the best ranking from amongst the set

that yield the minimal misrank percentage,

3. consider other factors about a game such as the home team or injuries to key players as well as include results from the post season,

4. apply our graph-based ranking algorithm to other sports and sports seasons, and

5. work to not just generate an ordering of teams, but to also consider a rating for each team that could be used as a predictive model for future games.

References

[1] AP Top 25 Poll - Week 15. https://collegefootball.ap.org/poll/2018/15, 2018. [Online; accessed 7 Feb. 2019].

[2] 2017-18 College Basketball Polls. https://www.sports-reference.com/cbb/seasons/2018-polls-old.html, 2018. [Online; accessed 7 Feb. 2019].

[3] Basebal Reference. https://www.baseball-reference.com, 2018. [Online; accessed 27 Feb. 2019].

[4] Basketball Referernce. https://www.basketball-reference.com, 2018. [Online; accessed 27 Feb. 2019].

[5] FBS Football Rankings - College Football Playoff. https://www.ncaa.com/rankings/football/fbs/college-football-playoff, 2018. [Online; accessed 7 Feb. 2019].

[6] Men's College Basketball Rankings - Week 18. http://www.espn.com/mens-college-basketball/rankings/_/week/18/year/2018/seasontype/2, 2018. [Online; accessed 24 Feb. 2019].

[7] Power Rankings: Astros fight off their superteam rivals. http://www.espn.com/mlb/story/_/id/24811499/ranking-mlb-teams-week-26, 2018. [Online; accessed 7 Feb. 2019].

[8] NBA Power Rankings: Ready for a wild West finish? http://www.espn.com/nba/story/_/id/23083860/nba-power-rankings-our-expert-panel-unveils-rankings-week-26, 2018. [Online; accessed 7 Feb. 2019].

[9] Final MLB Power Rankings. http://www.rotoworld.com/articles/mlb/82577/522/final-mlb-power-rankings?pg=2, 2018. [Online; accessed 7 Feb. 2019].

[10] NFL Power Rankings: New Year's resolutions for every team. http://www.espn.com/nfl/story/_/id/25656646/2018-nfl-week-18-power-rankings-new-year-resolutions-every-team, 2018. [Online; accessed 7 Feb. 2019].

[11] Power Rankings: The Farewell Tour edition. http://www.espn.com/nhl/story/_/id/22759910/nhl-2017-18-power-rankings-greg-wyshynski-farewell-tour-edi 2018. [Online; accessed 7 Feb. 2019].

[12] Michael R. Garey. *Computers and intractability : a guide to the theory of NP-completeness*. A Series of books in the mathematical sciences. W. H. Freeman, San Francisco, c1979.

[13] Hockey Reference. https://https://www.hockey-reference.com, 2018. [Online; accessed 27 Feb. 2019].

[14] metinmediamath.wordpress.com. How to Calculate the Elo-Rating (Including Examples). https://metinmediamath.wordpress.com/2013/11/27/how-to-calculate-the-elo-rating-including-example/, 2013. [Online; accessed 21 Jan. 2019].

[15] Week 26 Power Rankings: Cleveland Cavaliers and Golden State Warriors look vulnerable as regular season closes. http://www.nba.com/powerrankings/2017-18-week-26, 2018. [Online; accessed 7 Feb. 2019].

[16] NFL.com. https://www.nfl.com/, 2019. [Online; accessed 27 Feb. 2019].

[17] NFL Power Rankings: Pats up, Seahawks down entering playoffs. http://www.nfl.com/news/story/0ap3000001006802/article/nfl-power-rankings-pats-up-seahawks-down-entering-playoffs, 2019. [Online; accessed 7 Feb. 2019].

[18] Abigail R. Roane, Chaiwat Ekkaewnumchai, Connor W. McNamara, and Kyle Richards. Graph-based sports rankings. Technical Report MQP-CEW-1902, Worcester Polytechnic Institute, March 2019. https://digitalcommons.wpi.edu/mqp-all/6724/.

[19] Jeff Sagarin Ratings NFL. https://www.usatoday.com/sports/nfl/sagarin/, 2019. [Online; accessed 27 Feb. 2019].

[20] Power Rankings: Teams on the Bubble Battling to Make the Playoffs. https://www.si.com/nhl/2018/04/02/power-rankings-playoff-bubble-teams-stars-blues-panthers, 2018. [Online; accessed 7 Feb. 2019].

[21] Sports Reference College Basketball. https://www.sports-reference.com/cbb/boxscores/, 2018. [Online; accessed 27 Feb. 2019].

[22] Sports Reference College Football. https://www.sports-reference.com/cfb/boxscores/, 2018. [Online; accessed 27 Feb. 2019].

[23] Vreda Pieterse and Paul E. Black. Algorithms and Theory of Computation Handbook. https://www.nist.gov/dads/HTML/nphard.html, September 2013. [Online; accessed 2 Feb. 2019].

[24] D. Younger. Minimum feedback arc sets for a directed graph. *IEEE Transactions on Circuit Theory*, 10(2):238–245, June 1963.