ReputationRank: A Novel Weighting System for Retrodictive College Football Ranking

Journal Title
XX(X):1–10
©The Author(s) 0000
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Alex Wang

Abstract

Since the era of the Bowl Championship Series, computer ranking systems have been used extensively to rank college football teams. Although most of these methods are black-box models with methodology unavailable publicly, the publicly available models indicate a wide variety of philosophies as to the quality of a single game, including but not restricted to pure margin of victory, pure wins and losses, or some combined "power" metric utilized in tandem with wins and losses. This article seeks to apply ranked-choice voting concepts to existing graph-theory-based ranking frameworks by utilizing a norm-referenced scoring mechanism. Furthermore, this article also incorporates additional hyperparameters to better represent team merit and strength of schedule.

Introduction

In this section, we discuss the basis behind computer rankings for college football, their different philosophies, and an introduction of a norm-weighted philosophy based on the concept of ranked-choice voting.

Why Computer Rankings?

In the earlier days of college football, sportswriters did not actively rank teams and, as the sport was still in its infancy, there was a distinct lack of head-to-head competition. Today, with the implementation of the Bowl system, conference championships, the College Football Playoff, and fixed-length regular seasons, there is, in general, a much better idea of how teams compare to the rest of the field, even without head-to-head competition.

Rankings, both computer-generated and human-generated, continue to dominate college football discourse today, and up until 2013, both were used to determine the seeding for the national championship game and other high-quality bowl games (Low C, 2023). Despite computer rankings not being used to seed bowls and playoff games today, the methodology behind ratings still serves as a good baseline in identifying how humans perceive teams differently from mathematical models.

Different Philosophies in Computer Rankings

Computer Rankings often reflect the philosophy of the creator. This section discusses the individual philosophies of several NCAA-designated "major selectors," and is not meant to be a comprehensive mathematical guide to the framework of these models.

The Colley Matrix Method The Colley Matrix, developed by Wes Colley, is notable in the fact that it only considers wins and losses (Colley W, 2002). It is a Laplacian matrix model which solves a system of equations whose right-hand side is a function of a team's wins and losses, discounting margin of victory. This model, being a matrix model, does not consider recency of games.

Massey's Method Massey's Method, developed by Kenneth Massey, is notable in the fact that it only considers margin of victory, game recency, and the venue of the game (home, away, neutral) (Massey K, no date). In its current iteration, it utilizes least-squares regression, treating every individual game as one equation. Massey's original 1997 paper used pure margin of victory; however, this is no longer the case.

The Billingsley Report Richard Billingsley's model considers recency of games, wins, losses, and home- and away-games. It contains various post-hoc adjustments which consider head-to-head competition, and as such, agrees more with human polls (Billingsley R, no date).

Through the fact that the NCAA still considers various major computer systems as "major selectors," it can be seen that there is no empirical consensus on what metric to rank college football teams on (NCAA, 2025).

The ReputationRank System

Ranked-Choice Voting The core of the ReputationRank methodology is that it considers each team as a ”voter” within a network. That is to say that each team's information is limited: its knowledge is restricted to the teams that it has played. It ranks the teams that it has played on a subjective voting scale normalized between 0 and 1, an extension of ranked-choice voting as a proxy for a team's ”opinions” for the others on its ”ballot.” In practice, the ”mean opponent” of team A receives a vote of 0.5 from team A. A team's rating is a weighted average of the votes they receive. The iterative ReputationRank formula is

$$r_i^k = \frac{1}{n_i} \sum_{j \in o_i} v_{j,i} \cdot f(r_j^{k-1})$$

Where r_i^j is the rating of team i on iteration k, n_i is the number of games played by team i, $v_{j,i}$ is the vote value j casts upon i, and $f(r_j^{k-1})$ is a function of the rating of team j on iteration number $k-1$.

Recursive Vote-Weighting The base ReputationRank algorithm relies on recursively rating teams. At each iteration, a team's votes are weighted by a function of their own rating. We opt for an affine transformation, as simply multiplying a team's rating by their vote to create a weighted vote would indicate that a 0.95 vote from a team with rating 0.2 should be worth five times less than a 0.95 vote from a team with rating 1.0. Furthermore, when normalizing based on the number of interactions, this penalizes parties for merely interacting with worse parties, regardless of the worse party's view of the better. As such, a logical choice is an affine transformation of the form

$$f(r_j^{k-1}) = a + b \cdot r_j^{k-1}$$

with a and b as hyperparameters tuned at the discretion of the election coordinator. The affine transformation in this case assumes that every vote must be worth a “base” value and prevents the vast vote value disparity that comes from lesser-regarded parties. With that defined, the final algorithm is:

$$r_i^k = \frac{1}{n_i} \sum_{j \in o_i} v_{j,i} \cdot (a + b \cdot r_j^{k-1})$$

which must be iterated until convergence.

Rewarding Wins Without explicit win-loss weighting, teams that consistently lose but perform above opponent-average are overvalued. The prior framework simulates ranked-choice voting in a non-competitive environment, in which teams have ”votes” toward other teams. It does not, however, consider wins and losses. In order for the ReputationRank framework to be applied to college football, wins and losses must be rewarded explicitly; otherwise, teams are only rewarded for performance relative to the opponents of their opponents. Teams that perform relatively well by this metric but still tend to lose games will have their ratings inflated despite poor performance on the field. In section 3, we will discuss an additional parameter to reward wins.

Background

In this section, we explore the mathematical basis for ReputationRank and discuss its graph-theoretical inspiration.

On the Motivation Behind this Model

This model does not claim to be an entirely unique class of ranking algorithms. It borrows from the framework in existing graph-theory models.

Keener's Method Keener's method is an eigenvector-based ranking algorithm (Langville and Meyer, 2012). It stems from the philosophy that a team's rating should be proportional to a function of the ratings of its opponents. Keener's method solves the following equation:

$$\mathbf{A}\mathbf{r} = \lambda \mathbf{r}$$

where $\mathbf{r}$ is the principal eigenvector of the matrix $\mathbf{A}$. $\mathbf{A}$ is defined such that

$$\mathbf{A}_{ij} = \frac{p_{ij} + 1}{p_{ij} + p_{ji} + 2}$$

where p_{ij} is defined as the number of points i scored on j. ReputationRank borrows from the philosophy of Keener's method in the fact that a team's rating should be proportional to a function of the ratings of its opponents; however, it contains two major differences that distinguish it from other methods in its class in two major ways:

- Firstly, that the affine transformation applied to a team's ”votes” indicate that ReputationRank cannot be solved as the principal eigenvector of an unmodified voting matrix;
- secondly, that the individual values in the voting matrix are norm-referenced. Keener's method uses Laplace's Rule of Succession to assign values within the matrix; this is, then, independent of the other opponents of a certain team. Practically, a 20-6 win will be treated the same regardless of what teams were playing in it. ReputationRank is a norm-referenced algorithm which provides game-scores, or votes, based on how well a team performed against an opponent's opponents.

PageRank ReputationRank belongs in the same family of ranking algorithms as Google's PageRank due to the fact that in their base form, they are iterative forms of vote aggregation. Furthermore, in a weighted PageRank model, PageRank can simulate a norm-referenced system, as the weights of all outbound links on a page must add to 1. However, there are two key distinctions that justifies the novelty of ReputationRank:

- Firstly, that PageRank calculates the sum of a team's inbound links, and ReputationRank calculates the average of a team's inbound links. This makes ReputationRank more aligned with the initial ranked-choice voting philosophy, ensuring that every team's "ballot" is weighted evenly regardless of the number of wins or games they have;
- secondly, that PageRank typically calculates one-way influence. In college football, the concept of a "quality loss" indicates that a team can be positively affected by a loss to a strong team; this motivates the fact that ReputationRank generates bi-directional links for each game instead of a one-directional "link" or endorsement.

Positioning ReputationRank

ReputationRank, then, belongs in the broad graph-theoretic family of ranking approaches such as Keener's method and PageRank, but departs from these methods through norm-referenced scoring, an affine transformation of vote weights to preserve the vote importance of weaker entities, and an explicit win-loss modifier. These changes position ReputationRank as a member of this graph-based family of retrodictive ranking algorithms, but provide a novel weighting philosophy designed to interpret college football's schedule disparity, large number of teams, and lack of head-to-head competition.

Methodology

Interpreting the College Football Season

To apply ReputationRank to college football, we first have to place college football in a graph-theory standpoint. Each season may be represented as a graph, with each "node" being an individual team. Each "edge" represents a game played, with votes being propagated along said edges. With this defined, we begin creating the voting scheme.

Defining the Subjective Voting Scale

The first step in defining the ReputationRank methodology is the definition of the subjective voting scale. Teams' vote values should be between 0 and 1.We opt for the following for $v_{i,j}$, the vote i casts on j for a game played between the two:

$$v_{i,j} = \frac{1}{2} \cdot [\tanh(\tau \cdot z) + 1]$$

with z being the z-score of that game's point margin against all point margins attained against team i. τ is a hyperparameter applied to scale the hyperbolic tangent function such that information about margin of victory is relevant, but not overly rewarding wins with high z-score. By doing so, this achieves minimal impact for the unsportsmanlike practice of "running up the score."

An analysis of the 2024 college football season indicates that a vast majority of games have a z-score between -2 and 2, and only 30 out of more than 1300 games had a z-score greater than 2.5. Setting $\tau = 0.5$ places this typical z-score range squarely in the nonlinear region of $\tanh$, yielding vote values between approximately 0.12 and 0.88. This ensures that most games contribute meaningful gradations of information, while preventing blowouts from saturating the scale.

Affine Vote Weighting

Returning to the prior ReputationRank formula

$$r_i^k = \frac{1}{n_i} \sum_{j \in o_i} v_{j,i} \cdot (a + b \cdot r_j^{k-1})$$

indicates that when votes should be weighted based on an affine transformation of the voting team's rating at the current iteration. This is so that teams with good performances against weaker schedules do not suffer a collapse in the value of their votes, and furthermore so that teams do not get penalized harshly for simply interacting with a worse team. Each team's vote should be worth a base amount, indicating a base "trust" in each team's subjective ballot.

We now demonstrate why this is necessary. Running ReputationRank with $f(r_j^{k-1}) = r_j^{k-1}$ on the 2024 college football season demonstrates that teams with weaker schedules are penalized for playing the weaker schedule, rather than being rewarded for good quality of play against those teams. We take note of two specific case studies of the lack of an affine transformation harming or helping a team's rating.

Boise State In 2024, Boise State had a record of 12-2, with close losses to #3 Oregon, and a playoff loss against #5 Penn State. This strong win-loss record positioned them at #8 in the AP Poll. Boise State played in the Mountain West Conference, a weaker mid-major conference; however, their exemplary performance allowed voters to rank them higher. Within our initial non-affine transformed ReputationRank test, Boise State ranks at #57, with a rating of 0.534; this indicates that Boise State's schedule strength outweighs its performance.

Boise State's highest-rated opponent, #23 UNLV, had a rating of 0.492 in the ReputationRank test; this indicates that the entire Mountain West Conference network became a rating sink due to the lack of an affine transformation.

Kentucky In 2024, Kentucky had a record of 4-8, with a three-point win over #11 Ole Miss. However, despite the losing record, Kentucky's end rating in ReputationRank was 0.719, at #25. Kentucky plays in the Southeastern conference, a conference well-known for its strength of schedule; Kentucky, then, demonstrates another example of a team's schedule strength outweighing its performance.

Adding Back the Affine Adding back a heuristic affine transformation with $a = \frac{1}{3}$ and $b = \frac{2}{3}$, Boise State once again ranks at #25. The fact that this weakly tuned affine transformation is able to fix most of the discrepancies with mid-major teams demonstrates both its effectiveness and its necessity.

Win-Loss Modifier

Notably, the subjective voting scale

$$v_{i,j} = \frac{1}{2} \cdot [\tanh(\tau \cdot z) + 1]$$

does not consider whether a team won or lost. This is inapplicable to the world of college football, due to the fact that win-loss record is still important outside of the context in which these wins and losses occur. A team which plays stronger teams very closely, but still loses the game, should not be greatly rewarded for playing above the mean opponent.

To combat this, we add the δ hyperparameter, which rewards wins and penalizes losses, while still maintaining the context in which these wins and losses occur:

$$v_{i,j} = \frac{1}{2} \cdot [\tanh(\tau \cdot z \pm \delta) + 1]$$

For the remainder of this article, δ will be set at a heuristic value of 0.5. Additional tests for local stability in δ will be discussed further in the "Parameters" section.

Per-Step Normalization

A key step that we must take after every iteration is to ensure that the ratings are normalized to a mean of 0.5, while preserving the rating ratio between teams. The affine transformation

$$f(r_j^{k-1}) = a + b \cdot r_j^{k-1}$$

assumes a stable rating scale, the scale of the ratings must be identical at every step. In practice, without per-step normalization, the vector $\vec{r}^k$ tends to shrink without normalization; as the magnitude of $\vec{r}^k$ decreases, the term $b \cdot r_j^{k-1}$ becomes negligible and the affine transformation is reduced to the a term. Rescaling at every step is also mathematically necessary to maintain the ratio between the importance of a and the relative importance of a voter's current rating.

We normalize the ratings at every iteration such that their mean is 0.5. This is not an arbitrary choice; the fact that the vote values are bounded in $(0, 1)$ indicates that rescaling to a mean of 0.5 keeps the affine map interpretable without changing magnitudes and leading to instability.

Mathematically, the iterative scheme operates in a projective space, where magnitudes of team ratings do not matter and only the direction of the ratings vector does. By Perron-Frobenius theory, the process has one unique stable direction, being the principal eigenvector of the update matrix; this eigenvector is the final ratings vector.

Handling the FCS

As stated previously, Division 1 college football is divided into two subdivisions, the FBS and the FCS. The FCS is a division consisting of smaller and less-funded schools, and thus has a noticeably lower standard of play. FBS teams can pay FCS teams money in order to have a near-guaranteed win in their season (Trahan K, 2015). These "money games" are, a vast majority of the time, wins by exceedingly large margins and have little value in seasonal rankings. Thus, we incorporate FBS-FCS games by treating all FCS teams as one "super-team," and ranking the teams as normal.

This implementation of FCS ratings leads to loss in granularity, as variation of skill within the FCS is not captured. However, having the model include all FBS teams greatly increases computational cost at minimal benefit to the ranking of FBS teams.

Future extensions may include a more robust FCS grouping systems, or weighting FCS games based on an FCS team's performance within the FCS season.

Iterative Convergence Criteria

In practice, when iterating this process multiple times, the rankings tend to converge toward a single fixed value. When working with a college football season, the ratings after 15 iterations tend to only shift by a magnitude of approximately 10^{-6} per iteration. We define the point at which the order of rankings have remained static for 10 iterations as the point to stop iteration.

Parameters

Subjective Voting Scale

The subjective voting scale is defined as

$$v_{i,j} = \frac{1}{2} \cdot [\tanh(\tau \cdot z \pm \delta) + 1]$$

in which τ and δ are subjective hyperparameters used to control how the scale views outliers in margin of victory and how the scale views the importance of a win. In the previous section, we defined $\tau = 0.5$, and for the rest of this article $\delta = 0.5$ as well. δ and τ are chosen heuristically; a discussion of the role of δ will be included later in this section.

The Affine Coefficients

Recall the affine transformation

$$f(r_j^{k-1}) = a + b \cdot r_j^{k-1}$$

which prevents low-rated teams from having their votes collapse. We seek to tune the coefficients a and b. Firstly, we define the relationship

$$b = 1 - a$$

to not only preserve all possible ratios of a and b, but also to reduce the possible space of the grid-search; by tuning a only, we can sufficiently tune both values.

With that, we define a metric to measure the quality of a set of coefficients. Define r_i as the final, converged rating of team i.

Define

$$w_{i,j} = v_{i,j} \cdot (a + b \cdot r_i)$$

where $w_{i,j}$ is the *weighted vote* from team i to team j, weighted by the affine transformation of team i's stable rating.

Define

$$CV_i = \frac{1}{r_i} \sqrt{\frac{\sum_{j \in o_i} |w_{j,i} - r_i|^2}{N_i}}$$

as the coefficient of variation of the weighted votes that team i received. Recall once again that by the definition of ReputationRank, $r_i = \bar{w}_i$, the average of the weighted votes i received.

The philosophy behind ReputationRank assumes that the best a and b will cause each team's votes to vary the *least*, indicating that the transformation has succeeded in weighting votes. As such, it is possible to tune the affine parameters.

We choose our dataset as all Division I FBS seasons from 2014 to 2024, more specifically, the era of the College Football Playoff; we exclude the 2020 season as a statistical outlier due to the relative isolation of the graph nodes from COVID-19-related issues. We perform a grid-search with interval 0.01 on $a \in (0, 1)$. The value of a with the lowest value of $\bar{CV}$, the average CV, is $a = 0.37$. We utilize this value for the remainder of the article.

The Delta Parameter

Recall that the δ parameter controls the degree to which the subjective voting scale values the act of winning. We contend that due to the subjective nature of the voting scale, δ cannot be tuned empirically. However, we can verify the robustness of the parameter within a local interval.

Delta is Locally Stable We analyze, for the years 2014-2024 without 2020, the stability of the rankings in the local interval $\delta \in [0.45, 0.55]$. We (i) take the Spearman correlation ρ relative to the rankings at $\delta = 0.5$, and (ii) find the number of instances in which a top-25 team's ranking shifted by more than three places relative to the rankings at $\delta = 0.5$.

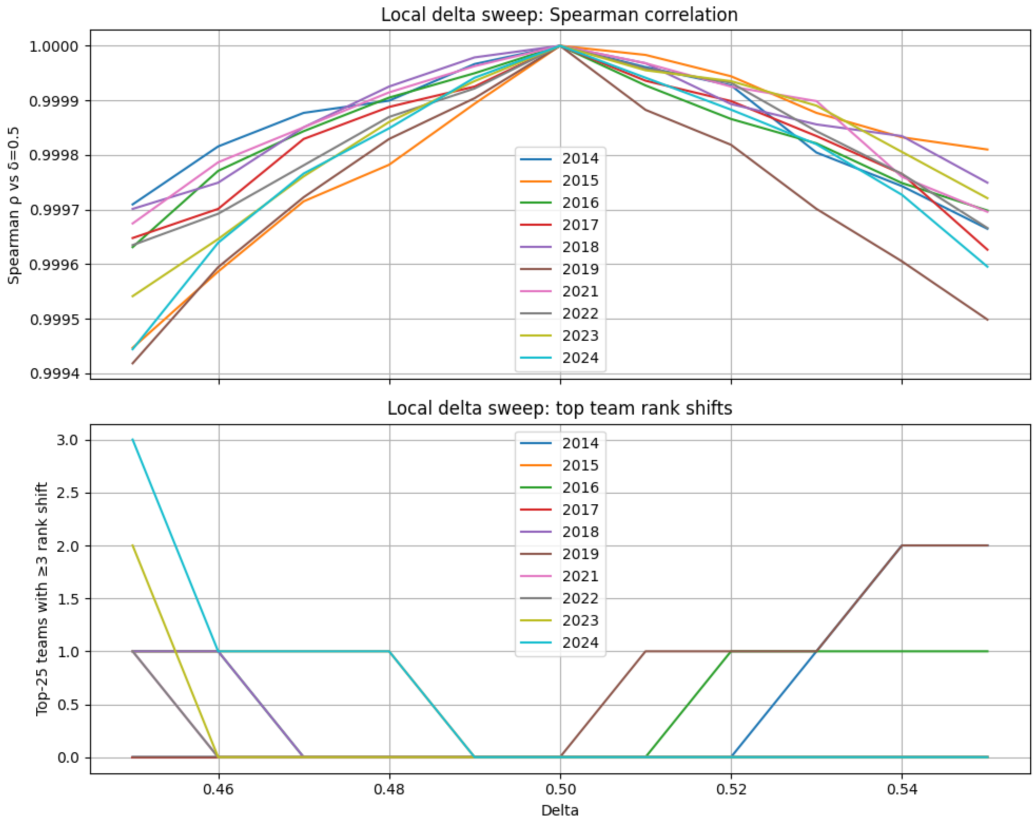

Figure 1. Spearman ρ and top-25 rank shifts of 3 or greater relative to $\delta = 0.5$

Spearman's correlation coefficient remains above 0.999 relative to $\delta = 0.5$ for all $\delta \in [0.45, 0.55]$, and top-25 rank shifts of considerable magnitude remain relatively rare. We conclude that the behavior of rankings for a range of δ is locally stable.

Framing the Parameters

All ranking systems, whether they claim objectivity or not, have author-specific biases within them. Colley, for example, emphasizes the act of winning. The above selections for τ, δ, and a do not represent empirical truths about the structure of the game of college football. They instead encode a subjective, normative choice. We have demonstrated that the values chosen for the three hyperparameters are reasonable values stable within their respective local ranges. These parameters, then, should not be viewed as objective truths but instead as tunable values that allow coordinators to emphasize different philosophies.

Results

In this section, we seek to demonstrate ReputationRank's comparison with various human and computer polls in order to show its efficacy.

The 25 teams being compared within the following tables will be the top 25 from the AP Poll. This is due to the fact that certain teams in the top 25 in the various computer polls may not have received votes in the AP.

We will utilize two metrics to track correlation of ReputationRank (abbreviated Rep.) against other polls: the Spearman correlation and the Kendall τ.

2024

We analyze the 2024 season due to its recency, and the creation of the first 12-team college football playoff.

Team	**AP**	**Rep.**	**Colley**	**Massey**
Ohio State	1	1	1	1
Notre Dame	2	2	3	2
Oregon	3	3	2	3
Texas	4	4	4	5
Penn State	5	5	5	6
Georgia	6	6	6	4
Arizona St.	7	9	12	16
Boise St.	8	14	10	28
Tennessee	9	21	22	8
Indiana	10	8	7	10
Ole Miss	11	7	18	9
SMU	12	11	13	21
BYU	13	12	8	13
Clemson	14	21	24	19
Iowa St.	15	18	11	18
Illinois	16	25	9	23
Alabama	17	10	15	7
Miami	18	15	21	25
S. Carolina	19	13	19	14
Syracuse	20	17	16	38
Army	21	24	23	40
Missouri	22	22	20	15
UNLV	23	27	30	46
Memphis	24	33	27	42
Colorado	25	34	32	32
Spearman vs Rep.	**0.874**	**1**	**0.845**	**0.853**
Kendall vs Rep.	**0.751**	**1**	**0.711**	**0.704**
Spearman vs AP	**1**	**0.874**	**0.847**	**0.808**
Kendall vs AP	**1**	**0.751**	**0.707**	**0.660**

Table 1. Various Rankings for the 2024 Season

2017

We analyze the 2017 season due to the controversy surrounding UCF's strength of schedule.

Team	AP	Rep.	Colley	Massey
Alabama	1	1	2	1
Georgia	2	2	4	3
Oklahoma	3	8	9	5
Clemson	4	3	6	4
Ohio State	5	5	5	2
UCF	6	6	1	9
Wisconsin	7	4	3	6
Penn State	8	7	7	7
TCU	9	12	15	10
Auburn	10	10	13	11
Notre Dame	11	9	8	8
USC	12	14	12	13
Miami	13	11	11	16
Oklahoma St.	14	13	21	12
Michigan St.	15	16	10	15
Washington	16	15	20	14
Northwestern	17	18	14	19
LSU	18	24	24	17
Mississippi St.	19	23	25	18
Stanford	20	25	28	21
South Florida	21	29	26	34
Boise St.	22	20	17	33
NC St.	23	22	16	20
Virginia Tech	24	30	29	22
Memphis	25	17	19	39
Spearman vs Rep.	**0.930**	**1**	**0.921**	**0.925**
Kendall vs Rep.	**0.793**	**1**	**0.767**	**0.780**
Spearman vs AP	**1**	**0.930**	**0.851**	**0.968**
Kendall vs AP	**1**	**0.793**	**0.653**	**0.853**

Table 2. Various Rankings for the 2017 Season

2016

We analyze the 2016 season due to all three computer models selecting a non-consensus national champion.

Team	AP	Rep.	Colley	Massey
Clemson	1	2	2	2
Alabama	2	1	1	1
USC	3	11	9	9
Washington	4	7	4	4
Oklahoma	5	6	6	5
Ohio State	6	3	3	3
Penn State	7	9	8	12
Florida St.	8	8	5	7
Wisconsin	9	4	7	10
Michigan	10	5	10	6
Oklahoma St.	11	13	16	14
Stanford	12	17	12	13
LSU	13	12	19	8
Florida	14	25	14	17
Western Mich.	15	10	11	29
Virginia Tech	16	23	20	16
Colorado	17	19	18	24
West Virginia	18	24	17	19
South Florida	19	22	15	36
Miami	20	14	27	15
Louisville	21	20	24	18
Tennessee	22	21	13	21
Utah	23	38	30	28
Auburn	24	18	25	13
San Diego St.	25	30	39	45
Spearman vs Rep.	**0.837**	**1**	**0.838**	**0.856**
Kendall vs Rep.	**0.620**	**1**	**0.653**	**0.698**
Spearman vs AP	**1**	**0.837**	**0.896**	**0.846**
Kendall vs AP	**1**	**0.620**	**0.727**	**0.664**

Table 3. Various Rankings for the 2016 Season

Analysis

Across multiple CFP-era seasons (2024, 2017, 2016), the human and computer polls displayed a relatively strong correlation with ReputationRank, indicating its consistency with the consensus notion of the quality of a football team. All three models also have a comparable correlation to the AP Poll. Furthermore, correlation values in both Spearman ρ and Kendall τ are relatively high for ReputationRank against the AP Poll, indicating that ReputationRank agrees well with sportswriters. Furthermore, the values indicated do not represent an absolute consensus amongst the ranking methods, suggesting that ReputationRank still contains its nuances from its subjective voting scale and hyperparameter tuning.

Conclusion and Acknowledgements

ReputationRank is a Framework

ReputationRank is a tunable and flexible retrodictive framework and the hyperparameters defined earlier do not represent an empirical or rigid description of college football merit. The subjective voting scale can be tuned for each separate ranking purpose and can be used to control the effects of outlier wins and losses, blowouts, and the value of a win as opposed to performing against the mean. Furthermore, as the college football landscape continues to evolve, especially regarding issues of parity, the affine hyperparameters will likely shift to represent that.

Limitations

The most notable drawback of the ReputationRank system as described prior is the FCS ”super-team” system. Individual variety in FCS teams are not thoroughly demonstrated through this grouping. Future work regarding ReputationRank may include a better FCS grouping system, or combining all of Division I football to best and most accurately represent the relationship between FBS and FCS teams. Furthermore, additional analysis on margin of victory could be implemented into the subjective voting scale to best represent a team’s performance within a single game.

Additional Rankings

Additional rankings from 1982 onward are housed at smalliftrue.github.io.

Acknowledgements

I would like to thank Bill Radjewski of CollegeFootballData.com for his easily accessible database of college football data and the Python API used to generate these rankings.

References

Billingsley R (no date) *In Search of a National Champion*. College Football Research Center.
Colley W (2002) *Colley’s Bias Free College Football Ranking Method: The Colley Matrix Explained*. Colley Matrix.
Langville A and Meyer D (2012) *Who’s #1? The Science of Rating and Ranking*, 1st edn. Princeton University Press.
Low C (2023) *How the creation of the BCS set the stage for the current playoff format*. ESPN.
Massey K (no date) *Massey Ratings Description*. MasseyRatings.com.
NCAA (2025) *Football Bowl Subdivision Records*. NCAA.
Trahan K (2015) *Pay-for-Flay: Why FBS-FCS Beatdowns Will Never Go Away*. Vice.