A Network-Based Approach to Ranking College Football Teams

Ilan Goodman (igoodman), Kat Gregory (katg), Sunil Pai (sunilpai)

December 8, 2015

Abstract

Using win probabilities, centrality measures, Bradley-Terry rankings, and loop detection and removal, we construct an optionally weighted directed graph representing a season of college football and derive from it a ranking of the college football teams for that season. We use win probability to define "garbage time" and use this garbage time detection to produce an analogous weighted graph, which we can use for more powerful rankings. We use weekly prediction to evaluate our ten implemented ranking algorithms and compare performance with the AP Poll. We find that our modified BeatPaths algorithm outperforms all the other algorithms we tried, weighted algorithms outperform their unweighted counterparts, and our garbage time weighting scheme slightly outperforms a margin of victory weighting scheme. Notably, many of our network-based algorithms outperform the AP poll, which is traditionally considered to be the go-to metric for comparing college football teams. All our algorithms aim to be allowed for consideration by the College Football Playoff Selection Committee.

1 Problem and Context

College football has enjoyed a special place in the hearts, minds, and pocketbooks of Americans since the late nineteenth century [14]. However, the sport faces a unique challenge: although many football fans would vehemently argue that their team is the best, there is no objective way to determine the national champion team. Because most games are played within conferences, each team will only directly play a fraction (on the order of $\frac{1}{10}$) of the other teams. Instead of a round robin or a knockout elimination tournament between all conference champions, the College Football Playoff (CFP) Ranking Committee incorporates a sheet of statistics and the votes of expert judges to choose four teams to compete for the championship [16]. Their choice is often met with grumbling and general discontent.

Thus, there remains a standing need for a more robust system for ranking college football teams. There is considerable debate over what factors such an algorithm should incorporate. According to their FAQs, the Selection Committee is allowed to consider "conference championships won, strength of schedule, head-to-head competition, comparative outcomes of common opponents (without incenting margin of victory) and other relevant factors that may have affected a team's performance during the season or likely will affect its postseason performance," so any ranking system we create should be in line with these guidelines so it may be used as a tool by the Selection Committee [3]. Simply observing the difference between the number of wins and losses (win-loss differential) is not enough to definitively rank teams: aside from the fact that this would lead to numerous ties, not all wins and losses are created equal. An ideal method might also factor in the magnitude of each victory and each team's strength of schedule. Margin of victory, in addition to not being permitted, does not suffice since teams should not be encouraged to run up the score at the end of a game. Traditionally strength of schedule only accounts for opponents' record and opponents' opponents' record, i.e., only two steps away on the graph of college football teams [4]. This misses a lot of information encoded in the graph (namely higher order effects) that can easily be considered by using the entire graph in the algorithm. In our paper, we leverage the latent network structure of a season of college football in order to model each team's strength of schedule, and with this information we evaluate several ranking algorithms against the human-generated AP Poll. Our goal is to illustrate which teams have built the best season résumé and are most deserving to reach the playoffs.

While many ranking systems exist, most are either completely subjective (such as the AP Poll) or are entirely stats-driven [18]. The statistics that the CFP Selection Committee uses to rank teams do not tend to correlate well with team strength or ability and do not reference margin of victory, even though margin of victory predicts winning percentage better than nearly any other statistic [8]. Most existing ranking systems do not directly consider the complete network of NCAA Football Bowl Subdivision (FBS) teams when establishing their ranking, and none manage to take score or statistics into account without encouraging a team to run up the score during "garbage time"[1] (a facet of a ranking system that would not be allowed for discussion in the CFP Selection Committee).

Our project endeavors to objectively rank college football teams (based exclusively on games from the current season) using a network-based approach. As discussed in the literature, analyzing the directed graph formed at the end of college football seasons (for example, through feedback arc removal and cycle detection) can prove to be a powerful technique in ranking sports teams [19]. By using win probability plots to determine when garbage time begins in each game, we can

[1]The time at which the outcome of the game is effectively determined, which often occurs well before the game ends, see section 4.2.1

transform any season of college football into a weighted directed graph that accounts for scores and statistics (and hence should be more descriptive than any system that solely looks at wins, losses, and matchups), yet should be valid for the CFP Selection Committee to use when they rank teams. Since we are the only algorithm that can automatically identify garbage time and exclude it in our calculations, and since the network of football teams matters so greatly in college football, we expect that we should be able to outperform any other system that conforms with the CFP Committee's standards for inclusion in consideration. Furthermore, our algorithm is objective and starts fresh every year, so it is fair to teams that are traditionally overlooked for making the CFP. Using the evaluation method in section 5, we find that our modified BeatPower algorithm on the unweighted graph outperforms all other variants (including weighted variants) and the human polls.

2 Related Work

Outside of the AP Poll, there have been several studies performed on college football ranking algorithm analysis. In "A network-based ranking system for US college football," J. Park and J Newman define the "total win score" of each team as the sum of this team's direct and (discounted) indirect wins at all distances, the "total loss score" as the sum of direct and (discounted) indirect losses at all distances, and the "total score" of each team as the difference between this team's "total win score" and "total loss score" [16]. Team A has an indirect win over team B if team A beats a team C which has (indirectly) beaten team B. This factors in strength of schedule but assumes transitivity of wins [7]. In "Offense-Defense approach to ranking team sports," A. Govan, A. Langville, and C. Meyer characterize a team's strength through distinct offensive and a defensive ratings instead of a single metric [9]. This model has the potential to incorporate other performance metrics, like total yardage, but it relies on assumptions about game data that are not guaranteed to hold true given the sparseness of college football match-ups. In "A Markov method for ranking college football conferences," R. Mattingly and A. Murphy use a Markov model similar to Google PageRank to simulate fans changing allegiances between teams according to transition probabilities that bias fans to support winning teams [12]. This is efficient but does not take into account home field advantage.

3 Baselines

We approach the challenge of ranking college football teams by performing network analysis on a directed graph that represents a football season. Our graph is built from all the games that occur during the season. Each node is a team and each directed edge e_{ab} represents a game where team a beat team b (or for certain algorithms, such as PageRank where we want strong teams to have high in-degree, the edges are in the opposite direction). Multiple edges can exist between the same pair of teams because a pair of teams can play each other multiple times, although this situation is not common in college football.

3.1 Our Baseline Algorithms

As our baseline ranking algorithms we assign each team the node score calculated when respectively running Katz centrality, PageRank, eigenvector centrality, and Laplacian centrality on the unweighted directed graph described above. We describe the results of these algorithms along with others in section 6.

3.1.1 Katz Centrality

Katz centrality designates as important those nodes that are either highly linked themselves or are linked to other important nodes. For our purposes, it captures strength of schedule by ranking highly those teams that have defeated many teams or have defeated highly ranked teams.

The Katz centrality of a team i is given by

$$x_i = \alpha \sum_j A_{ij} x_j + \beta,$$

where A is the adjacency matrix of our graph G, α is less than $\frac{1}{\gamma_{\max}}$ if γ are the eigenvalues of A, and β controls the initial centrality [10]. For our calculations, we used $\alpha = 0.1$ and $\beta = 1.0$.

3.1.2 PageRank

PageRank adjusts Katz centrality to account for the fact that being linked to an important node is less meaningful if that node also links to many other nodes. For our purposes, it captures the nuance that it is not as catastrophic to be beaten by a well ranked team (that has beaten many other teams) as it is to be beaten by a mediocre team (that has not beaten many other teams).

The PageRank centrality of a team j is given by

$$r_j = \alpha \sum_i A_{ij} \frac{r_i}{d_i} + \beta_j$$

where d_i is the out-degree of team i or 1 if team i has out-degree of 0, A is the adjacency matrix of our graph G, α is the damping factor, and β_j represents the "personalization value" for node j [15]. In our calculations, we used $\alpha = 0.85$ and a uniform distribution over β.

3.1.3 Eigenvector Centrality

Eigenvector centrality is identical to Katz centrality, using $\alpha = \frac{1}{\gamma_{\max}}$ and $\beta = 0$. The eigenvector centrality score of a node is proportional to the sum of the scores of its neighbors [13].

3.1.4 Laplacian Centrality

Laplacian centrality is a measure that gives "an intermediate centrality score between global and local characterization of the position of a vertex in weighted networks" [17]. Let $L(G)$ be the Laplacian matrix of a graph $G = (V, E)$, as defined in [17]. Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of L, and define the Laplacian energy of G to be $E_L(G) = \sum_{i=1}^{n} \lambda_i^2$. Let $V = \{v_1, \dots, v_n\}$. Then if G_i is the subgraph of G obtained by removing v_i from G, we define the Laplacian centrality $C_L(v_i, G)$ of vertex v_i to be

$$C_L(v_i, G) = \frac{E_L(G) - E_L(G_i)}{E_L(G)}.$$

As proven in [17], this is a nonnegative quantity, and it performs strongly when one wants to achieve a balance between global and local characterization of nodes in a graph.

3.2 AP Poll

As a further baseline, we also evaluate the performance of the AP Poll using the same metric we use to evaluate our own rankings. The AP Poll is a weekly vote by sports writers and broadcasters who have extensive experience in covering college football [1]. Prior to 2014, this poll was considered the definitive ranking system, and it continues to be the default used in reporting team rankings when the official CFP rankings are unavailable.

4 Our Approach

We develop and run several algorithms on the college football season network to compare their performance. We explore algorithms on an unweighted graph, which correspond to strength of record metrics, and also algorithms on a weighted graph, which additionally encapsulate quality of victory.

4.1 Unweighted Algorithms

4.1.1 Modified BeatPower

BeatPower is a graphical power ranking system based solely off wins and losses developed by Brad Siffert [18]. We define a BeatLoop as a cycle within season graph G. BeatLoops represent ambiguity because we cannot infer a relative ranking for the teams that make up that loop (i.e., we cannot impose a strict ordering on or topologically sort a graph with cycles). We handle BeatLoops by removing[2] from the graph the games that create them based on the assumption that these games do not give us actionable information about the relative strength of the teams. To do so, we perform the following BeatPath algorithm on the graph G:

- At each iteration i, we repeat the following steps:
  - First, if there are no BeatLoops left in the graph, the algorithm is complete.
  - Second, we find all BeatLoops of length i. If there are no more BeatLoops of length i, we move on to the next iteration.
  - Third, for each BeatLoop of length i, we remove the edges associated with that cycle. If the same edge is part of multiple BeatLoops of length i, we remove only a single copy of that edge in this step regardless of whether or not there are duplicate copies of that edge [18].

From the original graph G and the acyclic graph G' that results from the BeatPaths algorithm, we can calculate each team's BeatPower ranking [18]. To determine the BeatPower ranking of a team a, define $\mathsf{BeatWins}(a)$ as the number of

[2] in order of smallest BeatLoop to largest remaining BeatLoop

descendents of a in G', $\mathsf{BeatLosses}(a)$ as the number of ancestors of a in G', $\mathsf{TotalWins}(a)$ as the number of descendents of a in G, and $\mathsf{TotalLosses}(a)$ as the number of ancestors of a in G. Then the BeatPower of a can be calculated by

$$\mathsf{BeatPower}(a) = \frac{\mathsf{BeatWins}(a) - \mathsf{BeatLosses}(a)}{\mathsf{TotalWins}(a) - \mathsf{TotalLosses}(a)}.$$

BeatPower ranges from -1 to 1; better teams have higher BeatPower and are hence ranked above teams with lower BeatPower.

Because the original BeatPaths algorithm was designed to evaluate the structure of graphs representing NFL football seasons, we had to make several modifications to run the algorithm on graphs representing college football seasons. There are more teams in college football than in the NFL[3] and, as a result, many more BeatLoops. Our basic BeatLoops algorithm discovered tens of millions of BeatLoops with lengths on the order of $40-50$ teams before we halted it for efficiency concerns. Still, we need to remove all BeatLoops, otherwise our BeatPower algorithm will not give us any useful data. To work around this limitation, we utilize the fact that the conference structure of college football creates near-cliques of teams. By adding a pre-proccessing step where we run BeatPaths first on individual conferences to remove inner-conference loops, we ensure that we have sparser graph to work with later on. This also has intuitive justification: in some sense, conference games matter more in college football than non-conference games and the in-conference loops should be treated with higher priority. For some seasons with exceptionally long and large counts of BeatLoops, notably 2008, we also run BeatPaths on each pairwise combination of conferences after running on each conference individually but before running on the rest of the graph.

4.1.2 Modified Bradley-Terry

The Bradley-Terry model for ranking teams is a standard paired comparison model for NFL data analysis that we apply, with a few modifications, to college football [11]. The analysis takes into account home field advantage through the indicator h_{ijr}, which is 1 if the game was on i's home field, -1 if the game was on j's home field, and 0 if the game was on a neutral field, and which is given a weight τ. A team's abilities are given by the vector $\boldsymbol{\mu}$. The algorithm aims to estimate this abilities vector $\boldsymbol{\mu}$ and use it to rank the teams. The standard algorithm does so via maximization of the log-likelihood estimate:

$$\ell(\boldsymbol{\mu}, \tau) = \sum_{i<j}^{k} \sum_{r=1}^{n_{ij}} y_{ijr}(\tau h_{ijr} + \mu_i - \mu_j) - \log(1 + \exp(\tau h_{ijr} + \mu_i - \mu_j)).$$

Here, we consider the indicator variable Y_{ijr} which is who won the game (the data for this is y_{ijr}). The actual probability that we are optimizing is very much like a sigmoid function:

$$p(Y_{ijr}) = \frac{1}{1 + \exp(\tau h_{ijr} + \mu_i - \mu_j)}$$

We modified the problem by performing gradient descent on the following optimization problem:

$$\max_{\boldsymbol{\mu}, \tau} \ \ell(\boldsymbol{\mu}, \tau)$$
$$\text{subject to } \|\boldsymbol{\mu}\| = 1$$

The main modification was to maintain $\|\boldsymbol{\mu}\| = 1$ on each iteration by projecting onto the unit sphere after each gradient update. Interestingly, we noticed that convergence for this algorithm with $\alpha = 0.05$ occurred in relatively few (< 30) iterations for all of the seasons as can be seen in Figure 1.

Since there is a large computational cost to calculating Bradley-Terry rankings and it did not outperform our BeatPower metric on our milestone experiments, we opted to not evaluate Bradley-Terry further, instead spending our time and computational resources on more successful algorithms.

4.2 Weighted Algorithms

As discussed earlier, the college football playoff selection committee is not allowed to take margin of victory into account when establishing their rankings [3]. Nonetheless, it is important to identify how commanding each victory or loss is, since this data has tremendous predictive and descriptive power [8]. It reflects better on a team to lose a close game than to be blown out; inversely, a dominating win illustrates more potential for a team than a win where the team barely squeaks on by. To this end, we seek a measure that illustrates how decisive each game was to weight the edges on our graph.

4.2.1 Garbage Time

Garbage time refers to the time toward the end of the game when one team has effectively already won the game, and the other team has next to no chance of coming back. The scores and statistics racked up in this portion of the game are

[3] 128 FBS (Football Bowl Subdivision) teams compared to 32 NFL teams

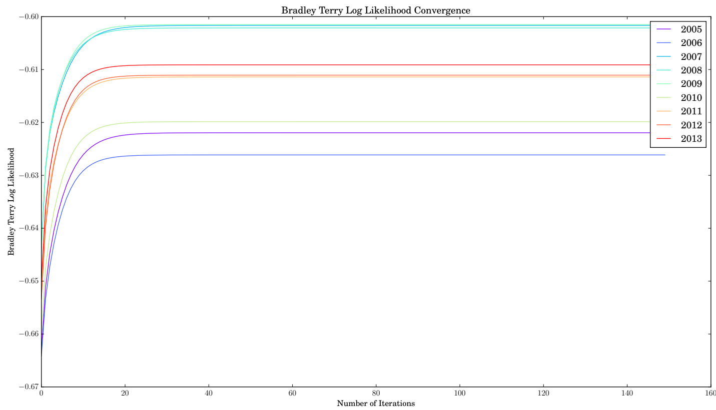

Figure 1: Convergence behavior of Bradley-Terry Gradient Descent

effectively meaningless: teams put reserves in and play calling is substantially different from the rest of the game, so no respectable ranking algorithm should use these numbers [2]. Prior to this paper, no one has systematically found a way to identify and remove garbage time in the game analysis that cannot be manipulated by the teams (e.g., delaying substitutions to rack up more stats). Using win probability, we formally define *garbage time* and use our garbage time detection to build a weighted graph and improve our ranking algorithm.

Using the data from the $2008 - 2013$ seasons, we train a model based on score, time remaining in the game, down, distance, yard line, and similar features to predict win probability for each team on each play. With this model in hand, we can plot the probability that the winning team will win the game over the course of the game. Comparing these plots to the play-by-play data, including player substitutions and postgame recaps, we define the start of garbage time to be *the earliest time at which the winning team has at least a* 95% *chance of winning for the entire rest of the game*, and garbage time lasts from its start until the end of the game (by definition). If the winning team does not have at least a 95% chance of winning immediately prior to the final play of the game or if the game goes to overtime, we say the game has no garbage time.

To build our weighted graph, we have a node for each team. We use the same edges as in the unweighted graph. The weight of each edge is equal to a small constant (for simply winning the game) plus a quantity proportional to the amount of time remaining at the start of garbage time (higher weight corresponds to a more dominating victory). The relative influence of these two terms depends on the algorithm and can be relearned every week to produce the optimal descriptive power. With this weighted graph in hand, we can perform more detailed graphical analysis of college football than any other algorithms are capable of, without providing any benefit for running up the score after the game is in hand.

4.2.2 Weighted Katz Centrality, PageRank, Eigenvector Centrality, and Laplacian Centrality

Weighted Katz centrality, PageRank, eigenvector centrality, and Laplacian centrality are each the same as their respective unweighted counterparts, but they use the weighted adjacency matrix instead of the adjacency matrix, thereby taking into account edge weights. The unweighted versions of all these algorithms are identical to setting the weight of each edge equal to 1 and performing the weighted algorithm.

4.2.3 Goodman Centrality

Goodman centrality is a variant of current flow centrality that works on directed graphs [6]. It is a novel graph metric designed to favor nodes with many, high-capacity outgoing edges and few, low-capacity incoming edges. In our graph, this means that teams with a high Goodman centrality score have many, quality victories and few, close losses.

Formally, if the max flow from source node u to sink node v is given by $\mathsf{maxFlow}(u,v)$, then the Goodman centrality score of a node u in a directed weighted graph $G=(V,E)$ is given by

$$\mathsf{GoodScore}(u) = \sum_{v \in V} \left(\mathsf{maxFlow}(u,v) - \mathsf{maxFlow}(v,u)\right).$$

Computing the $\mathsf{GoodScore}$ for all nodes is computationally equivalent to the all pairs max flow problem on a weighted graph. It is an open problem if there is an algorithm to solve all pairs max flow faster than computing max flow for each of the $\Theta(n^2)$ pairs of nodes individually [5]. Accordingly, Goodman centrality takes a long time to calculate, so we only were able to calculate rankings for the years $2011 - 2013$.

5 Evaluation

We have implemented ten distinct ranking systems thus far, plus we have data on the AP Poll every week after week 10 (the week the CFP Selection Committee publishes their first rankings of the season) of every year from $2008-2013$. To this end, it is crucial for us to have an evaluation method so we can tell which of our ranking algorithms perform the best and how they stack up against other rankings.

Given a ranking of the top 25 teams (such as the AP Poll), we can predict the binary outcome of a match where at least one team is ranked by picking the better ranked team to win. We evaluated this way only for the AP Poll as an approximation to the performance of the human ranking system in predicting Top 25 games whereas for evaluating our ranking performance (e.g. BeatPower, PageRank, or other centralities) we use all of the games of a given week to evaluate ranking performance. Thus, in order to evaluate each ranking system, we use the following procedure: for each week starting at week 10 of each season, calculate the ranking for that week given all the games played until that point in the season. Use this ranking to predict the binary outcomes of the games for the upcoming week, as described above. After computing this for every week of every year, we calculate the fraction of games we predicted correctly. We strive to predict the greatest fraction of games correctly.

6 Results and Analysis

Ranking algorithms for unweighted graphs demonstrate similar predictive power. The accuracy of predictions from each ranking algorithm increases throughout the season as the graphs that underlie the rankings acquire more edges and, as a result, more information about the relative strength of teams. However, notice that the last column of each figure is dark: rankings for bowl games break this trend of increasing predictive power because bowl games tend to match teams which are very similar in skill, so the games are appreciably harder to predict than a typical season game. This pattern of increasing accuracy up until the bowl games persists throughout most of the rankings (see Figure 2).

Ranking algorithms on weighted graphs have significantly greater predictive power than those on unweighted graphs. This confirms our intuition that game statistics are fundamental to our understanding of team performance. Of the different rankings algorithms for weighted graphs, Katz centrality performs best (Table 1).

The predictive power of ratings derived from graphs with our garbage time weights is very similar to that of ratings derived from graphs with weights corresponding to margin of victory; indeed, our garbage time weights result in slightly better performance. This suggests that we have successfully derived a way to incorporate the predictive power associated with margin of victory in a manner that can legally be used by the College Football Playoff Selection Committee (Figures 3 and 4).

We also evaluated the predictive power of the AP Poll, Goodman centrality (labeled max flow), and BeatPower rankings. We find BeatPower to be the most powerful; the ratings it produces outperform even the AP rankings (Figure 5).

Final tabulated results taking into account the entire seasons for all of our ranking metrics are shown in Table 1.

Overall, we find that the BeatPower ranking algorithm most accurately predicts the outcomes of future games. This is intriguing given that it is one of the simplest of the algorithms we evaluated. The removal of BeatLoops makes intuitive sense given the ambiguity cycles introduce, and while we observed that rankings on weighted graphs usually outperform those on unweighted graphs, BeatPower derives a robust ordering even without information about garbage time or margin of victory.

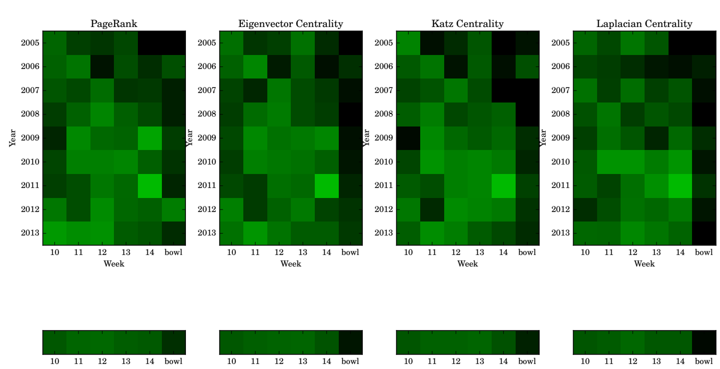

Figure 2: Predictive accuracy of rankings on unweighted graphs. Brighter green corresponds to greater predictive accuracy. Plots at the bottom represent the average evaluation accuracy over all years.

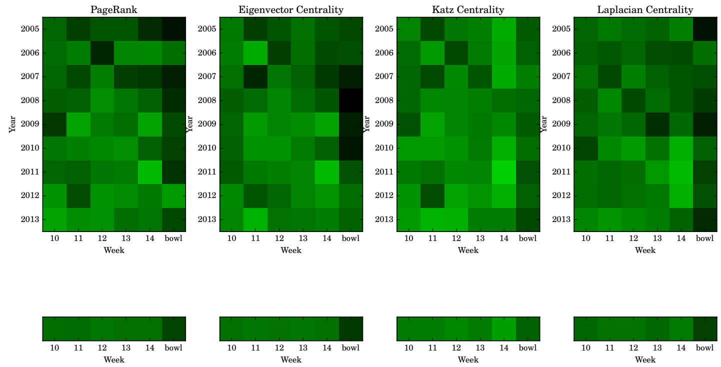

Figure 3: Predictive accuracy of graphs weighted by garbage time. Brighter green corresponds to greater predictive accuracy. Plots at the bottom represent the average evaluation accuracy over all years.

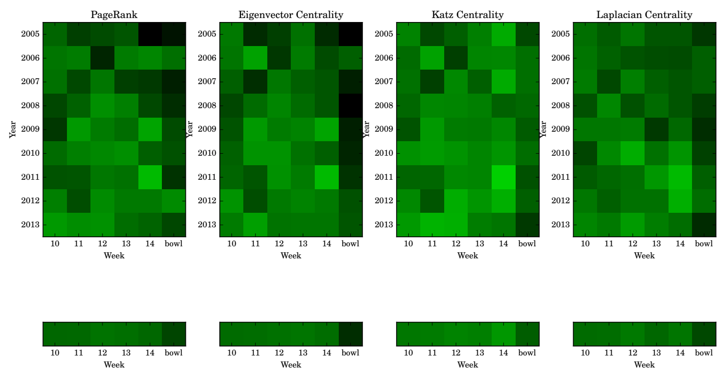

Figure 4: Predictive accuracy of graphs weighted by margin of victory (MOV). Brighter green corresponds to greater predictive accuracy. Plots at the bottom represent the average evaluation accuracy over all years.

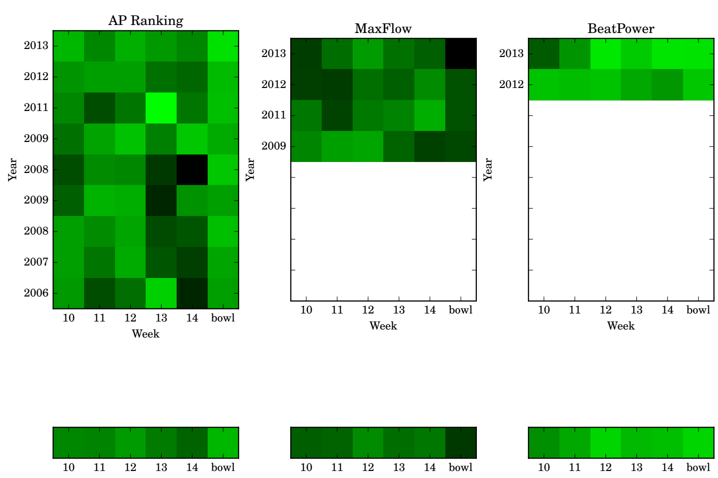

Figure 5: Predictive accuracy of other metrics. Note that due to time constraints, we only analyzed 2 years for BeatLoops graph and 4 years of MaxFlow (Goodman) centrality. Brighter green corresponds to greater predictive accuracy. Plots at the bottom represent the average evaluation accuracy over all years. Plots at the bottom represent the average evaluation accuracy over all years.

Ranking Algorithm	Mean Predictive Accuracy
Unweighted Katz Centrality	0.660 ± 0.046
Unweighted PageRank	0.670 ± 0.037
Unweighted Eigenvector Centrality	0.658 ± 0.053
Unweighted Laplacian Centrality	0.650 ± 0.061
BeatPower	$\mathbf{0.863 \pm 0.048}$
Garbage-Weighted Katz Centrality	$\mathbf{0.748 \pm 0.037}$
MOV-Weighted Katz Centrality	$\mathbf{0.743 \pm 0.034}$
Garbage-Weighted PageRank	0.705 ± 0.032
MOV-Weighted PageRank	0.698 ± 0.029
Garbage-Weighted Eigenvector Centrality	0.707 ± 0.043
MOV-Weighted Eigenvector Centrality	0.695 ± 0.049
Garbage-Weighted Laplacian Centrality	0.703 ± 0.037
MOV-Weighted Laplacian Centrality	0.705 ± 0.030
Garbage-Weighted Goodman Centrality	0.700 ± 0.049
AP Poll	0.770 ± 0.052

Table 1: As seen in the weekly predictive plots, the BeatPower rating system performs the best by far, followed by Katz centrality and other centrality measures (both weighted and unweighted), showing the power of direct, unweighted win-loss data and a network perspective in predicting outcomes.

7 Future Work

Since our modified BeatPower rankings significantly outperformed all other rankings and our weighted metrics outperformed their unweighted counterparts, we would like to create a weighted analog to BeatPower and try it out on our garbage-weighted graph. Furthermore, we would like to see the performance of Katz Centrality of the directed acyclic graph output by our weighted BeatLoops-free graph since Katz centrality is typically applied to such graphs.

Given the inherent flaws and biases that characterize any human ranking system, we hope that our work will bring a greater level of objectivity to the ranking of college football teams. We believe that a weighted graph holds far more predictive power than its unweighted analog and accordingly we should be able to tease far more information from it than we can from the unweighted counterpart.

References

[1] The AP Top 25 Poll.

[2] Defining 'Garbage Time'.

[3] Selection Committee FAQs.

[4] Strength Of Schedule Explained.

[5] All Pairs Maximum Flow, 2012.

[6] Ulrik Brandes and Daniel Fleischer. *Centrality measures based on current flow.* Springer, 2005.

[7] Shuo Chen and T Joachims. Modeling Intransitivity in Matchup and Comparison Data. February 2016.

[8] Bill Connelly. College Football Playoff Committee is BCS-esque at Avoiding Helpful Stats, November 2015.

[9] Anjela Y Govan, Amy N Langville, and Carl D Meyer. Offense-defense approach to ranking team sports. *Journal of Quantitative Analysis in Sports*, 5(1), 2009.

[10] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

[11] Guido Masarotto and Cristiano Varin. The ranking lasso and its application to sport tournaments. *The Annals of Applied Statistics*, 6(4):1949–1970, 2012.

[12] R Mattingly and A Murphy. A Markov method for ranking college football conferences. *SUNY Cortland*, 2010.

[13] Mark Newman. *Networks: an introduction.* Oxford University Press, 2010.

[14] R Ours. Introduction: A Brief History of College Football.

[15] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. *The PageRank citation ranking: bringing order to the Web.* PhD thesis, Stanford University, 1999.

[16] Juyong Park and Mark EJ Newman. A network-based ranking system for US college football. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(10):P10014, 2005.

[17] Xingqin Qi, Eddie Fuller, Qin Wu, Yezhou Wu, and Cun-Quan Zhang. Laplacian centrality: A new centrality measure for weighted networks. *Information Sciences*, 194:240–253, 2012.

[18] Curt Siffert. BeatPaths: The Winning Ways of Winners, 2005.

[19] Sean J. Taylor. Optimal Descriptive NFL Rankings.