Using Numerical Methods to Rank eSports Teams

Philip Yoo

Abstract

With the rise of eSports, the competitive form of videogames, what is the best way to rank these teams in the most objective way using numerical methods? This paper focuses on the use of a linear least-squares method called the Massey method to produce rankings based on matches won and point differences and explores how the Massey method behaves with changes to the data with numerical experimentation. The goal is to ultimately produce the best objective rankings for the Team Fortress 2 Invite teams from Season 20. The match data needed to be altered in order to achieve the most accurate rankings. Weights were also applied in order to obtain even more information on how to rank the teams. Also, the Massey matrix is well-conditioned after the so-called "cheat" step while it is ill-conditioned before it. The best ranking calculated by the weighted Massey Method from first to last is froyotech, Ascent, Team SoloUber, Strawberry Mangoes, woodpig, Getawhale Fanclub, Meat Market, MAD LUXURIOUS, and Kawaii Five-0.

1. Introduction

For many years, researchers, mathematicians, and even some diehard fans have used numerical methods to obtain objective rankings of sports teams and players. There is a wide variety of different methods each with their own unique computations and advantages. The Pythagorean expectation method has been used in baseball to rate teams and predict how many games a given team will win based on the number of runs the players score. The Elo rating system is used to rank top chess players through the exchange of skill points when a match concludes. There are also the Colley method and the Massey method that have been used to predict match wins and to fill out extremely accurate March Madness brackets [1].

Recently, there has been an emergence of a new type of "sport" involving videogames with the rise in popularity of playing matches using electronic devices. As such, this new form of sport is called eSports. Some games have grown so popular that competitive teams have received corporate sponsorships from major companies such as LG and Samsung and have participated in tournaments with multi-million dollar prize pools.

These new eSports should not excluded from the various ranking methods that have been used on real sports like baseball, football, and soccer. After all, there are a lot of similarities between the two forms: large fan bases, high-stakes competition, and collections of teams of different skill levels. Therefore, there is a need to rank these teams in the most objectively way possible using data of their performance.

We will be using the aforementioned Colley and Massey methods for this paper. However, the Colley method will only be discussed briefly while the Massey method will be the main focus for reasons explained later. We will be applying the Massey method to the match data from the twentieth season of the Invite division of the E-Sports Entertainment

Association league (abbr. ESEA) for the video game Team Fortress 2 (abbr. TF2). This Invite division is the highest division available for this particular video game where teams can only participate by the league's invitation. A total of nine teams of six players each played a total of 72 games (each team played each other twice) for Season 20 of ESEA Invite.

I chose this particular set of match data because Season 20 offered a lot of great drama, rivalries, close games, and an unexpected conclusion, which I believed would provide interesting results with the Massey method [3]. The goal of this paper is to explore how the Massey method behaves with changes to the data and to ultimately produce the best objective rankings for the TF2 Invite teams.

2. Methods

The two methods we will use to rank the teams of ESEA Invite are the Colley and Massey methods. The nine teams in Invite are Ascent, froyotech, Team SoloUber, woodpig, Strawberry Mangoes, Getawhale Fanclub, Meat Market, MAD LUXURIOUS, and Kawaii Five-0. Each team played 16 matches total with win-loss statistics of (14-2), (14-2), (12-4), (9-7), (8-8), (7-9), (6-10), (1-15), and (1-15) respectively.

The Colley method applies Laplace's rule of succession and uses the number of matches won to produce a rating. It then involves solving a linear system in the form:

$$Cr = b$$

The matrix C is the Colley matrix. In this example, since there are nine teams, the Colley matrix will be a 9×9 matrix with each row and column associated with a specific team. Let the total number of matches played by the i-th team be t_i and the total number of matches played by the i-th and j-th team be n_{ij}. Then, the entries for this matrix are $c_{ij} = 2 + t_i$ when $i = j$ and $c_{ij} = -n_{ij}$ when $i \neq j$. The vector b are the values $b_i = 1 + 0.5(w_i - l_i)$ where w_i is the number of wins by and l_i is the number of losses by the i-th team. When we solve for the vector r we will have the Colley ratings for the individual teams corresponding to their respective indices. For this example, the Colley matrix will have the following entries in the order Ascent, froyotech, Team SoloUber, woodpig, Strawberry Mangoes, Getawhale Fanclub, Meat Market, MAD LUXURIOUS, and Kawaii Five-0.

```
18  -2  -2  -2  -2  -2  -2  -2  -2
-2  18  -2  -2  -2  -2  -2  -2  -2
-2  -2  18  -2  -2  -2  -2  -2  -2
-2  -2  -2  18  -2  -2  -2  -2  -2
-2  -2  -2  -2  18  -2  -2  -2  -2
-2  -2  -2  -2  -2  18  -2  -2  -2
-2  -2  -2  -2  -2  -2  18  -2  -2
-2  -2  -2  -2  -2  -2  -2  18  -2
-2  -2  -2  -2  -2  -2  -2  -2  18
```

The b vector is $(7, 7, 5, 2, 1, 0, -1, -6, -6)^T$. Using MATLAB, we can easily solve this linear system for the vector r [1].

Team	**Colley Rating**	**Colley Rank**	**Round Win-Loss**	**Match Win-Loss**	**ESEA Rank**
Ascent	0.8000	1	61-22	14-2	1
froyotech	0.8000	1	52-14	14-2	2
Team SoloUber	0.7000	3	55-36	12-4	3
woodpig	0.5500	4	42-48	9-7	4
Strawberry Mangoes	0.5000	5	43-40	8-8	5
Getawhale Fanclub	0.4500	6	41-40	7-9	6
Meat Market	0.4000	7	44-52	6-1	7
MAD LUXURIOUS	0.1500	8	14-62	1-15	8
Kawaii Five-0	0.1500	8	7-45	1-15	9

We can see from the table that without tied ratings the Colley rank is basically identical to the ranks that ESEA awarded the teams at the end of the season. ESEA, as a league, only cares about match wins; teams that win more matches will be ranked higher. In the case of ties in match win-loss, ESEA then uses the number of rounds won in a match to rank one team above the other. For example, Ascent and froyotech both had 14 match wins and 2 match losses. However, Ascent had 61 round wins and 22 round losses while froyotech had 52 round wins and 14 losses. The round win-loss difference for Ascent is 39 while froyotech's round win-loss difference is 38. This allows Ascent to narrowly rank above froyotech in ESEA's system.

We see that the Colley method provides very similar ranking results to the one ESEA awarded to the Invite teams. However, that is because both ranking systems uses match wins and losses to produce the ranking, so of course the rankings will be similar. This brings up the question of whether or not the Colley method is the best objective ranking method. By the Colley method alone, there are ties in the ranking namely for Ascent and froyotech (0.8000) and for MAD LUXURIOUS and Kawaii Five-0 (0.1500). Without any additional information, the Colley method cannot determine which of the two teams with the tied rating score is objectively better. That small difference in the round win-loss should be significant enough to be reflected in the rankings. Yet, the Colley method gives them the exact same rating score despite that fact since it only looks at match wins and losses. Therefore, we will use the Massey method, which takes into account both match wins and round wins to compute the rankings. The rest of the paper will now discuss and focus on the Massey method.

The reason why the Colley method was included in this paper was to show that it, as a numerical method for ranking, was flawed when given certain data, specifically with ties in win-loss. Its inclusion will show that the Massey method is the overall better method. It is,

however, important to note that the Colley method is much simpler to compute compared to the Massey method.

The Massey method takes into account both match wins and round wins to calculate ratings. With more data and parameters, it will provide a better ranking result than the Colley method has. It is a method that solves a linear least-squares problem; it aims to minimize the square of the two-norm $\|b - Ax\|_2^2$ for the linear system $Ax = b$. The first step will be to set up a linear system in the form:

$$M_1 r = p_1$$

The matrix M_1 is the matrix that contains information about match wins and losses. In this example, we will have the teams in the order Ascent, froyotech, Team SoloUber, woodpig, Strawberry Mangoes, Getawhale Fanclub, Meat Market, MAD LUXURIOUS, and Kawaii Five-0. The first match of the season was played between Meat Market and MAD LUXURIOUS that ended with a score of 5-1 with Meat Market's victory. Meat Market is the seventh team in the order and MAD LUXURIOUS is the eighth team in the order. Next, the second match was between Getawhale Fanclub and MAD LUXURIOUS with the final 5-0 with Getawhale Fanclub's victory. Getawhale Fanclub is the sixth team in the order and MAD LUXURIOUS is the eighth team in the order. So we have the first two rows of the M_1:

```
0   0   0   0   0   0   1   -1  0
0   0   0   0   0   1   0   -1  0
```

We fill up the rest of the matrix with the total 72 games. We use 1 to indicate the winning team and -1 to indicate the losing team in a given match. Then, the values of p_1 are the absolute values of the point differences of the 72 matches. So, the first two matches have point differences of 4 and 5. Therefore, the vector $p_1 = (4,5, \dots, d_{72})^T$ where d_{72} is the point difference of the 72nd match. We fill up the rest of the vector with the point differences of each match. The vector r is similar to the vector in the Colley method in that it will contain the Massey ratings for the teams. We will be trying to solve for this vector r. However, we have a problem here. M_1 is a 72×9 matrix, r is a 9 element vector, and p_1 is a 72 element vector. The linear system is overdetermined, so it is very likely that it is inconsistent and has no solution. We will then have to use the normal equations by multiplying the transpose of M_1 to the left on both sides of the equation.

$$M_1^T M_1 r = M_1^T p_1$$

Now, we let $M_2 = M_1^T M_1$ which is a 9×9 matrix and $p_2 = M_1^T p_1$. The 9 element vector r^* will contain the Massey ratings for the teams.

$$M_2 r^* = p_2$$

Now, a new problem arises: the matrix M_2 is singular so there can be an infinite number of solutions to the linear system. To remedy this situation, Massey just sets all the entries on the bottom row of M_2 to 1 and the bottom entry of p_2 to 0. We will call this step the "cheat" step. Now, the system will have a unique solution, and we solve for r^* which will have the Massey ratings [2]. This concludes the Methods section, and the Massey method used again in the Results section is the same with a few input differences.

3. Results

This is the result of the Massey method using the 72 matches played in Season 20 of ESEA TF2 Invite.

Table 1

Team	Massey Rating	Massey Rank	Round Win-Loss	Match Win-Loss	ESEA Rank
Ascent	2.1667	1	61-22	14-2	1
froyotech	2.1111	2	52-14	14-2	2
Team SoloUber	1.0556	3	55-36	12-4	3
woodpig	-0.3333	6	42-48	9-7	4
Strawberry Mangoes	0.1667	4	43-40	8-8	5
Getawhale Fanclub	0.0556	5	41-40	7-9	6
Meat Market	-0.4444	7	44-52	6-1	7
MAD LUXURIOUS	-2.6667	9	14-62	1-15	8
Kawaii Five-0	-2.1111	8	7-45	1-15	9

We now see very different rankings for the nine Invite teams compared to both the Colley method results and the official ESEA rankings. We see that the fourth place team, woodpig, is now ranked as sixth place even though it had more wins than Strawberry Mangoes and Getawhale Fanclub. MAD LUXURIOUS now places ninth instead of eighth. These changes are due to the round win-loss differences. Though woodpig won more matches, its round win-loss difference is -6 while Strawberry Mangoes' and Getawhale Fanclub's round win-loss differences are 3 and 2 respectively. Thus, Strawberry Mangoes and Getawhale Fanclub are rewarded for their ability to win more rounds while woodpig is punished for their round losses using the Massey method.

It is also interesting to note a special property of the Massey method ratings. The difference of two teams' rating scores is the expected point difference of a given match between those two teams [1]. For example, Ascent is expected to score 4.8334 more points than MAD LUXURIOUS and 1.1111 more points than Team SoloUber. This means that while Getawhale Fanclub was expected to win matches against woodpig with 0.3889 points, woodpig won one of the two matches it had with Getawhale Fanclub even though they were expected to lose both. This means that woodpig defied expectations when they played against Getawhale Fanclub. These upset victories over Getawhale Fanclub and other teams are what allowed woodpig to rank higher on the ESEA ranking system though they rank lower using the Massey method.

However, there is a major problem with the data set provided by ESEA of the 72 matches played for Season 20. At the start of the season, there were nine total teams willing to play fun and exciting videogame matches. MAD LUXURIOUS and Kawaii Five-0 were the

two new teams that were invited to play in the Invite division after their stellar performance in the Intermediate division. However, after a string of match losses by large point differences against the stronger and more skilled veteran Invite teams, the discouraged Kawaii Five-0 lost the will to play further and disbanded in the middle of the season, specifically after the fifth week of the total nine weeks of the season. This caused automatic forfeiture of Kawaii Five-0's scheduled matches. In ESEA, forfeited matches are given a 2-0 match score to the victory of the team that did not forfeit, one point per half of the match. This can cause inflation of other teams' number of wins since these forfeited matches are assigned as wins for the non-forfeiting team and can create inaccuracies in the round point differences. It is difficult, if not impossible, to predict the final score of a forfeited match if it were played since it never happened. Perhaps, the forfeiting team could have ended up winning the match with a final score of 5-0 instead of losing with a score of 2-0 as automatically assigned by ESEA. Hence, forfeited matches can provide incorrect data and can affect the results of the Massey ratings.

So, for this next use of the Massey ratings, we will remove all Kawaii Five-0 match data and remove all forfeited matches in order to improve accuracy.

Table 2

Team	**Massey Rating**	**Massey Rank**	**Round Win-Loss**	**Match Win-Loss**	**ESEA Rank**
Ascent	2.4403	2	55-22	11-2	1
froyotech	3.1937	1	40-12	8-1	2
Team SoloUber	0.5558	3	39-36	7-4	3
woodpig	-0.6773	5	35-40	7-4	4
Strawberry Mangoes	0.0558	4	35-38	4-7	5
Getawhale Fanclub	-0.8848	6	32-38	4-8	6
Meat Market	-1.0796	7	37-50	4-9	7
MAD LUXURIOUS	-3.6038	8	12-49	0-10	8
Kawaii Five-0					9

From this table, we can see the new Massey ranks calculated without the Kawaii Five-0 matches and without all forfeited matches. We see some significant changes. After all, we made significant changes to the data, so the results should reflect them. We now have froyotech, the previously second place team, as placing first by a large margin. woodpig's drop from its ESEA fourth place is not as dramatic as the Massey method result where it dropped to sixth place and it is now in fifth place. However, it is important to note that in this scenario, teams play different numbers of matches due to the number of forfeited matches. Ascent played 13 matches without forfeitures and froyotech played 9 matches without forfeitures. One can say that because of the difference in the number of matches played by each team, these ratings are not the best or the most objective. But, these results

are definitely improved and more accurate since we removed the inaccurate data used in the first use of the Massey method.

Now that we have the rankings we can arguably consider the "better" results, we will apply weights to the matches to explore the effects weights can have on the Massey ratings. Instead of the 1 point that is awarded to the winning team of a match in the Massey matrix, we arbitrarily choose a value greater than 1 to assign the winners. For this weighting experiment, we will use 1.5 for the weight value. This weight will award teams that more matches by more round wins effectively ranking the teams by "dominance." Teams that rank high are teams that have dominated and are expected to dominate the lower teams by winning many matches by large point margins.

Table 3

Team	**Massey Rating**	**Massey Rank**	**Round Win-Loss**	**Match Win-Loss**	**ESEA Rank**
Ascent	1.9895	2	55-22	11-2	1
froyotech	2.3336	1	40-12	8-1	2
Team SoloUber	0.7734	3	39-36	7-4	3
woodpig	-0.2002	6	35-40	7-4	4
Strawberry Mangoes	0.5963	4	35-38	4-7	5
Getawhale Fanclub	-0.2000	5	32-38	4-8	6
Meat Market	-0.5000	7	37-50	4-9	7
MAD LUXURIOUS	-4.7925	8	12-49	0-10	8
Kawaii Five-0					9

From this table, we see a very interesting result. We see that woodpig, which previously ranked fifth in the non-weighted Massy method, has dropped down to sixth place. This means that Getawhale Fanclub, the team that took woodpig's place, has won many matches by large point margins and have dominated more teams than woodpig. It is also interesting to note that MAD LUXURIOUS has a very low weighted Massey rating which implies that this team lost many matches by large point margins. Using this weighted method is one way of ranking teams by their dominating performance.

Now, we will observe the condition numbers of the Massey matrices. In the steps to calculating the rankings, we have a matrix M_2 which turned out to be singular. Massey then uses a "cheat" step by changing the last row of this matrix to a row of 1's and the last element of p_2 to 0. We will now calculate the condition number of the matrix M_2 before and after the "cheat" step using MATLAB's `cond()` function. We will use the data from the second use of the Massey method when we first removed the Kawaii Five-0 matches and forfeited matches and did not apply weights.

Before the "cheat" step, the condition number of the matrix is 1.6181e+17, and after the "cheat" step, the condition number of the matrix is 5.3818. We can see, as expected, the

matrix before the "cheat" step is ill-conditioned. After the "cheat" step is applied, the matrix condition number is much lower and the matrix is no longer ill-conditioned. We will test the whether or not the matrix is well-conditioned by creating a small change and seeing if that causes a large or small change in the Massey ratings. We will slightly change the data from the second use of the Massey method when we first removed the Kawaii Five-0 matches and forfeited matches and did not apply weights. We will change one of the match results between Ascent and froyotech (specifically the match played in Week 6 of Season 20) by changing the point difference from 3 to 4.

Table 4

Team	**Massey Rating**	**Massey Rank**	**Round Win-Loss**	**Match Win-Loss**	**ESEA Rank**
Ascent	2.3736	2	55-22	11-2	1
froyotech	3.1942	1	40-12	8-1	2
Team SoloUber	0.6336	3	39-36	7-4	3
woodpig	-0.6829	5	35-40	7-4	4
Strawberry Mangoes	0.0503	4	35-38	4-7	5
Getawhale Fanclub	-0.8849	6	32-38	4-8	6
Meat Market	-1.0796	7	37-50	4-9	7
MAD LUXURIOUS	-3.6043	8	12-49	0-10	8
Kawaii Five-0					9

We see that the ratings differ very slightly and none of the rank numbers changed. This shows that a small change caused only just a small change and that the matrix is indeed well-conditioned as indicated from our previous condition number calculation.

4. Discussion

We see from our experiments with the Massey method that the results behave well with certain changes and the Massey matrix is well-conditioned. We have produced the best rankings for the 9 Invite teams in Table 3 with the dominance weights. This ranking from first to last is froyotech, Ascent, Team SoloUber, Strawberry Mangoes, woodpig, Getawhale Fanclub, Meat Market, MAD LUXURIOUS, and Kawaii Five-0. Table 2 is also a good candidate for the best ranking calculated with the Massey method, but the dominance weights provide useful information about the way certain teams performed. We have previously discussed how woodpig defied expectations when they played against Getawhale Fanclub from Table 1's results analysis. This weighted ranking places woodpig below Getawhale Fanclub as if to show that Getawhale Fanclub rightfully ranks above woodpig in terms of dominating performance. This could mean that, though woodpig won more matches than expected, woodpig was only "lucky" and were objectively a "worse"

team than Getawhale Fanclub. Thus, the weighted ratings should be more objective since it removes the aspect of "lucky" wins and awards solid dominating wins.

There also exists a problem with using mathematical methods to predict match results and overall season performance. In sports, there is an aspect of luck in certain circumstances that will allow one team or player to perform better than expected. In a setting devoid of luck, the more skilled team will without a doubt defeat the less skilled team. However, luck and other factors can allow the underdog team to win. The same holds true for eSports. It is possible for low-ranking teams to upset high-ranking teams due to un-measurable factors such as luck, player mood on the day of the match, computer issues, network issues, etc. Typically psychological, emotional, and technical problems can affect the players' performance and therefore the outcome of a match.

In fact, after the end of the season, ESEA hosted a playoff tournament for the top four teams of the division: Ascent, froyotech, Team SoloUber, and woodpig (according to ESEA's ranks). Though froyotech placed second and Team SoloUber placed third during the season, Team SoloUber ended up placing second and froyotech placed third in the tournament. Ascent placed first and woodpig placed fourth in the tournament. This unexpected victory for Team SoloUber over froyotech was caused by an unfortunate situation on froyotech's side. After a heated argument during the lower bracket finals match, a team member from froyotech suddenly resigned and refused to play. froyotech had to bring in a substitute player, and as a result this affected froyotech's performance in the match. Team SoloUber defeated froyotech in the lower bracket finals and moved on to play Ascent in the grand finals. The conclusion of the playoff tournament definitely differed from the results calculated and predicted from the Massey method.

The paper *Bracketology: How can math help?* by Tim Chartier, Erich Kreutzer, Amy Langeville, and Katheryn Pedings mentions that "although methods generally perform well from year to year, because of the inherent randomness of sports, it is difficult for a method to consistently take the title of the best mathematical method" [1]. We have already shown the specific flaw that the Colley method has with certain data. Though the Massey method failed in predicting the winner of the ESEA playoff tournament, this Massey's linear least-squares method is still a powerful tool for accurately predicting match outcomes in any sport, both physical and electronic.

References

[1] Chartier, T., Kreutzer, E., Langeville, A., and Pedings, K. 2010. Bracketology: How can math help? available at: http://www.mathaware.org/mam/2010/essays/ChartierBracketology.pdf

[2] Wessell, C. 2013. Massey's Method available at: http://public.gettysburg.edu/~cwessell/RankingPage/massey.pdf

[3] ESEA Season 20 Invite Team Fortress 2 match data available at: http://comp.tf/wiki/ESEA_Season_20/Invite