Spring 2014

Reducing the Effects of Unequal Number of Games on Rankings

T. P. Chartier

J. Harris

K. R. Hutson

See next page for additional authors

Chartier, T.P., et al. "Reducing the Effects of Unequal Number of Games on Rankings." IMAGE - The Bulletin of the International Linear Algebra Society 52 (Spring 2014), 15-23.

Reducing the Effects of Unequal Number of Games on Rankings

Abstract
Ranking is an important mathematical process in a variety of contexts such as information retrieval, sports and business. Sports ranking methods can be applied both in and beyond the context of athletics. In both settings, once the concept of a game has been defined, teams (or individuals) accumulate wins, losses, and ties, which are then factored into the ranking computation. Many settings involve an unequal number of games between competitors. This paper demonstrates how to adapt two sports rankings methods, the Colley and Massey ranking methods, to settings where an unequal number of games are played between the teams. In such settings, the standard derivations of the methods can produce nonsensical rankings. This paper introduces the idea of including a super-user into the rankings and considers the effect of this fictitious player on the ratings. We apply such techniques to rank batters and pitchers in Major League baseball, professional tennis players, and participants in a free online social game. The ideas introduced in this paper can further the scope that such methods are applied and the depth of insight they offer.

Keywords
rank, competitors

Disciplines
Mathematics

Authors
T. P. Chartier, J. Harris, K. R. Hutson, A. N. Langville, D. Martin, and Charles D. Wessell

Reducing the Effects of Unequal Number of Games on Rankings

Timothy P. Chartier,[1] John Harris,[2] Kevin R. Hutson,[2]
Amy N. Langville,[3] Daniel Martin, and Charles D. Wessell[4]

Abstract. Ranking is an important mathematical process in a variety of contexts such as information retrieval, sports and business. Sports ranking methods can be applied both in and beyond the context of athletics. In both settings, once the concept of a game has been defined, teams (or individuals) accumulate wins, losses, and ties, which are then factored into the ranking computation. Many settings involve an unequal number of games between competitors. This paper demonstrates how to adapt two sports rankings methods, the Colley and Massey ranking methods, to settings where an unequal number of games are played between the teams. In such settings, the standard derivations of the methods can produce nonsensical rankings. This paper introduces the idea of including a super-user into the rankings and considers the effect of this fictitious player on the ratings. We apply such techniques to rank batters and pitchers in Major League baseball, professional tennis players, and participants in a free online social game. The ideas introduced in this paper can further the scope that such methods are applied and the depth of insight they offer.

1. Introduction. Ranking is an important mathematical process that informs decision-makers, whether they be consumers finding webpages returned from queries to search engines, the Major League Baseball (MLB) determining who will play in its playoffs, the Bowl Championship Series (BCS) selecting college football teams for the holiday bowl games, or the Association of Tennis Professionals (ATP) deciding on tournament invitations and seeding. Underneath such rankings are mathematical algorithms. Google uses variations of the classic PageRank algorithm, MLB leans on winning percentage, the BCS aggregates rankings obtained from human opinion and mathematical calculation, and the ATP awards ranking points in a way that rewards players who compete in a lot of tournaments in addition to playing well in them. This paper will demonstrate one way to adapt two specific sports ranking methods to settings in which teams or individuals play an unequal number of games.

What contexts arise where teams play unequal numbers of games? Most team sports design seasons with relative uniformity in the number of games played. For example, a season of NCAA Football Bowl Subdivision games is typically 12 to 14 games. Contrast this with the 2011 ATP Tour season which featured 304 players competing in 2566 completed matches.[5] Figure 1 displays a graph of the player by player matrix of the 2011 ATP Tour. In this matrix the players are listed in order of their ATP ranks (at the end of 2011) and entry a_{ij} equals the number of times player i played player j. In the figure, nonzero elements (nz) are colored blue. As we see, players differ widely in the number of matches played.

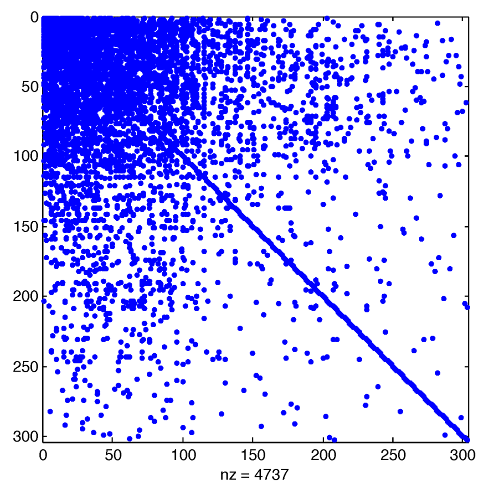


Figure 1: *A player by player matrix of the 2011 ATP Tour where a_{ij} contains a 1 if player i played player j and 0 otherwise. In the color-coded matrix above, nonzero elements are colored blue.*

Weaker players, for example, will compete less (and therefore have fewer blue pixels in their row) as they either do not qualify for many top-level tournaments without a special invitation from the tournament organizers, or when they do qualify, they are often defeated in the first round. Under these circumstances, 73 of the 304 competitors on the 2011 ATP Tour played only one match while the median number of matches per player was 6. This inherent disparity in the number of matches directly impacts the Colley and Massey ranking of professional tennis players.

To demonstrate this effect consider Jose Acasuso who in 2011 was nearing the end of a solid, if unspectacular, professional tennis career. The Argentine had an ATP ranking as high as 20$^{\text{th}}$ in 2006, but in the first few months of 2011 his rank had fallen into the 200s, and he was competing in tournaments on the second-tier Challenger Tour or trying to qualify for major tournaments he would have been invited to just a few years earlier. The one exception came in February, when the local organizers of the Buenos Aires ATP tournament awarded him a wild card entry. Acasuso, playing in his home country and on his favorite surface (clay), defeated Alexandr Dolgopolov and Pablo Cuevas in the first two rounds before losing in the quarterfinals to Nicolas Almagro. All three of his opponents were good players,

[1]Department of Mathematics and Computer Science, Davidson College, Davidson, NC, 28035, USA, tichartier@davidson.edu. This research was supported in part by a research fellowship from the Alfred P. Sloan Foundation.

[2]Department of Mathematics, Furman University, Greenville, SC 29613, USA, john.harris@furman.edu kevin.hutson@furman.edu

[3]Department of Mathematics, College of Charleston, Charleston, SC 29401, USA, langvillea@cofc.edu. This research was supported in part by NSF grant CAREER-CCF-0546622.

[4]Department of Mathematics, Gettysburg College, Gettysburg, PA 17325, USA, cwessell@gettysburg.edu

[5]These figures include only completed matches. Matches won by walkover (i.e. forfeit) or because a player retired during a match were not counted. Also, matches at team competitions (Davis Cup and World Team Cup) were not counted.

ranked 29^{th}, 64^{th}, and 13^{th} by the ATP's ranking system at the time of the tournament. Those were Acasuso's only ATP tournament matches of the year.

Both the Colley and Massey methods are designed to consider the strength of the opposition when calculating a player or team rank, and Acasuso's three matches provide a "perfect storm" for both algorithms. By winning twice as many matches as he lost (2-1) against solid opposition, Acasuso ended the year ranked 33^{rd} by Colley and 14^{th} by Massey.

This dynamic is not new. In fact, it is often quite common in the statistics that we use to rank athletes. Let us now consider MLB. Also in 2011, Joey Gathright made only one plate appearance for the Boston Red Sox. He made the most of it with a walk, a stolen base, and an eventual scored run. While Gathright was clearly proficient with his one plate appearance; it is difficult, if not impossible, to claim he was the best batter in the major leagues during 2011. Likewise, pitcher Jarrod Parker had a 0.00 earned run average during the 2011 season for the Arizona Diamondbacks. While this is impressive, he only faced 22 batters to achieve this ERA. There are likely dozens of major league pitchers who faced 22 batters in a row without allowing an earned run to cross the plate. Parker had an ERA below any Cy Young award winning pitcher. Yet, this ERA statistic, which is often used to compare pitchers, is not as meaningful in this case. Neither example (Parker nor Gathright) gives sufficient information in which to judge whether or not either athlete can sustain his performance. Most would agree that a batting average of 0.350 is more impressive if it can be maintained over 400 at-bats than over 40 at-bats. In fact, Rule 10.22 of the MLB Charter [1] states "The individual batting, slugging or on-base percentage champion shall be the player with the highest batting average, slugging percentage or on-base percentage, as the case may be, provided the player is credited with as many or more total appearances at the plate in league championship games as the number of games scheduled for each club in his club's league that season, multiplied by 3.1 in the case of a Major League player." If 162 games are scheduled in a given year, to win one to these batting titles a player would need $162 \cdot 3.1 = 502$ plate appearances. If no such cut-off is made for the Colley method Joey Gathright is ranked as the top batter and Jarrod Parker as the second-best pitcher in 2011.

This paper introduces adaptations to the Colley and Massey methods that aid in contexts where players play unequal numbers of games. Section 2 reviews the derivation of both the Colley and Massey methods and demonstrates how playing a small number of games can be an advantage. Section 3 introduces the idea of including a super-user into the rankings. Section 4 applies the adaptation of the Colley and Massey methods to ranking batters and pitchers in Major League Baseball. Sections 5 and 6 look at the effect of the super-user and how to choose the minimum cutoff at which point the super-user is not included for a player's ranking. Section 7 applies the super-user method to professional tennis. Finally, in Section 8 we look at the need for such methods in applications beyond sports. The concluding remarks summarize the method and its implications.

2. Ranking with the Colley and Massey Methods. This paper will focus on adapting two common sports ranking methods - the Colley method and Massey method. Both rely on linear systems to create their ratings. Both methods are also used by the Bowl Championship Series to aid in its rankings of NCAA FBS teams.

The Colley Method modifies winning percentage to create a linear system

$$C\mathbf{r} = \mathbf{b}, \tag{5}$$

which produces a rating for each team. For a full description and derivation of the Colley Method, see [12]. The linear system can be formed by the following definition of the rating r_i of team i

$$r_i = \frac{1 + (w_i - l_i)/2 + \sum_{j \in O_i} r_j}{2 + t_i}, \tag{6}$$

where w_i (l_i) represents the number of winning (losing) interactions for team i, t_i represents the total number of games involving team i, and O_i represents the set of opponents of team i. Loosely interpreted, this formula computes a team's rating as its winning percentage plus the average rating of its opponents. Equation (6) gives a row-wise description of the linear system in (5), where $\mathbf{r}$ is the vector of team ratings, $\mathbf{b} = [b_i]$ is a vector such that $b_i = 1 + \frac{1}{2}(w_i - l_i)$ and $C = T - A$, where $T = [t_{ij}]$ is a diagonal matrix in which the diagonal entries $t_{ii} = 2 + t_i$ and A is a matrix in which the (j, k)th entry is the number of times that the j and kth teams play. Solving this system of equations provides a rating for each team encoded in the ratings vector $\mathbf{r}$. This rating can be sorted to provide a ranking (or relative standing) for each team.

Another ranking approach, the Massey Method, is similar to the Colley Method in that it sets up a system of linear equations whose solution provides a rating for players. This linear system

$$M\mathbf{r} = \mathbf{p}, \tag{7}$$

is derived from the assumption that the teams' ratings will describe the point differential in their competitions. For example, if teams i and j compete with team i winning by p^{ij} points, then

$$r_i - r_j = p^{ij}.$$

However, this system is usually inconsistent and the method of least squares is employed to find a "best fit" solution. For a full derivation of the Massey Method, see [12]. Similar to the Colley Method, a team's rating in the Massey Method can be loosely interpreted as equal to the team's average point differential plus the average of the team's opponents' ratings. Again, solving this system of equations gives a rating vector that can be sorted to provide a ranking of participants.

3. Limiting the Effects of Unequal Numbers of Games. One common solution to this problem of disparate numbers of games is to simply drop athletes (or teams or items) with fewer than some minimum number of games. For example, the FIDE chess federation has required that players have at least 30 matches before being rated. While the minimum cutoffs are usually carefully chosen for the particular application, this approach is a bit arbitrary. Further, a ranking can be sensitive to the chosen cutoff. That is, a cutoff of 20, as opposed to 30, may produce quite a different ranking. Further, excluding players from rankings in this manner diminishes their contributions and reduces the visibility of the player. We seek to penalize these players without excluding them.

In this paper, we propose an alternative to the minimum cutoff approach for dealing with the disparate number of games issue for the Colley and Massey Methods. Our approach utilizes a super-user, a dummy team (athlete or item) that plays and beats every actual team that it plays. We force teams to lose to the super-user in every case, rather than a combination of wins and losses, in order to penalize teams that are perhaps artificially inflated in rank due to good performance in only a few games. Teams below the minimum cutoff face the super-user multiple times so that, with the addition of these artificial games, there is much more parity in the number of games that teams play. Because the super-user's ranking is artificially generated and has no practical meaning, it is removed from consideration so that each actual team's rank is incremented accordingly.

4. Ranking Baseball Players. In order to test the super-user idea in a scenario with a multitude of unequal numbers of games, we apply it to ranking pitchers and batters in the MLB. Note that the point of this exercise is not to find the best way to rank pitchers and batters but rather to explore the effects of adding a super-user to the Colley and Massey ranking methods. The Colley method can be used to rank individual players, pitchers and batters, in a similar way that it is used to rank teams. We formulate the interactions of batters and pitchers as a multi-edge bipartite (two-mode) network disregarding the interactions that National League Pitchers have with one another. This approach is similar to the mutually-antagonistic network formulation in Saavedra, et al. [14]. Bipartite networks are formulated using two disjoint sets of nodes, P (pitchers) and B (batters), and an edge connects a node x in P with a node y in B for each instance for which batter y had a plate appearance against pitcher x during the season. This network can be encoded in the Colley Matrix C where element C_{ij} equals the negative of the number of times pitcher i faced batter j during the season. Each batter-pitcher interaction results in either a hit, a walk, or an out. We ignore interactions that involve runners getting thrown out stealing to end innings, sacrifices, or that result in hit-by-pitches, and we strictly look at interactions that result in a hit, walk, or out. If the result is a hit or a walk, the hitter is awarded a "win" in that interaction and the pitcher a "loss." If the result is an out, the pitcher is given a "win" and the batter a "loss." From this, we can form the appropriate $\mathbf{b}$ vector, and the Colley Method, as described in the introduction, can be used to rank pitchers and batters. In using the Massey Method, the notion of a point differential p^{ij} in the interaction between pitcher i and batter j needs some explanation. There may be several ways to assign points to a pitcher-batter interaction. Here, we use the Bill James-defined RUE value [11], or runs to end of inning, to define a score for each possible plate appearance. The possible events and their corresponding score value are: generic out (0.240), strikeout (0.207), walk (0.845), single(1.025), double (1.132), triple (1.616), and home run (1.942). The batter receives the corresponding score for each walk, single, double, triple, and home run that he achieves or otherwise, he receives a score of 0. The pitcher receives a score of 0 if the batter gets a hit or a walk. Otherwise, for the pitcher, we calculate the average number of runs per inning for each year (2002-2011), call this value E and subtract the plate appearance outcome value (either 0.240 for an out or 0.207 for a strikeout) from the value E. The value of E varies from year to year. In 2011, $E = 0.478$. So, if a pitcher wins an interaction with a strikeout, he is given a value of $E - 0.027 = 0.478 - 0.207 = 0.271$. In this way the pitcher gets a score indicating how many runs were "saved" from the interaction. Using these values we define the values p^{ij} for the Massey Method.

As noted above, the number of plate appearances (or innings pitched) for baseball batters (or pitchers) can vary to a great degree, which can cause the Colley and Massey results to be skewed. For example, pinch hitters are usually called upon in later innings, if at all, to perform in certain situations. Players get injured or are called up from or sent down to the minors. Some teams assign catchers to only certain pitchers, and thus those play only when his assigned pitcher starts the game or he is called to pinch hit. Pitching rotations fluctuate throughout the year and relief pitchers have quite a bit of variability as to when they get into the game. All of these issues contribute to some players getting more plate appearances or innings pitched than others. To alleviate this, we set up two super-user nodes, the SuperBatter node and the SuperPitcher node. If a batter has $x < 300$ plate appearances during the year, he is forced to lose to the SuperPitcher $300 - x$ times. Here the cutoff of 300 was chosen to be roughly the mean/median numbers of plate appearances for batters. In 2011, the mean was 274 and the median was 291. Later, we let this vary. Likewise, if a pitcher has $y < 400$ batter interactions during the year, he is forced to lose to the SuperBatter $400 - y$ number of times. In each interaction with the SuperPitcher, the batter loses the interaction with an out, and the SuperPitcher is credited with $E - 0.240$ points. In each interaction with the SuperBatter, the SuperBatter wins the interaction with a single and is credited 1.025 points. These outcomes can change as the user sees fit though. Tables 1 and 2 show the Colley rankings of batters and pitchers both using and not using the super-user nodes for the 2011 season.

Without super-user				**With super-user**			
Rank	Batter	OBP	Interactions	Rank	Batter	OBP	Interactions
1	Joey Gathright	1.00	1	1	Miguel Cabrera	0.448	680
2	Esteban German	0.462	12	2	Jose Bautista	0.447	645
3	Gil Velazquez	0.429	6	3	Mike Napoli	0.414	427
4	Logan Schafer	0.500	4	4	Adrian Gonzalez	0.410	704
5	Antoan Richardson	0.5	4	5	Joey Votto	0.416	709
6	Russ Canzler	0.400	4	6	Lance Berkman	0.412	580
7	Miguel Cabrera	0.448	680	7	David Ortiz	0.398	603
8	Jose Bautista	0.447	645	8	Prince Fielder	0.415	676
9	Chris Parmelee	0.443	88	9	Matt Kemp	0.399	676
10	Mike Napoli	0.414	427	10	Dustin Pedroia	0.387	721
11	Adrian Gonzalez	0.410	704	11	Alex Avila	0.389	537
12	Jesus Montero	0.406	68	12	Paul Konerko	0.388	620
13	Joey Votto	0.416	709	13	Todd Helton	0.385	480
14	Cedric Hunter	0.400	5	14	Michael Young	0.380	678
15	Lance Berkman	0.412	580	15	Ryan Braun	0.397	621
16	David Ortiz	0.398	603	16	Jose Reyes	0.384	580
17	Alejandro De Aza	0.400	169	17	Victor Martinez	0.380	586
18	Leonys Martin	0.375	8	18	Carlos Beltran	0.385	591
19	Yonder Alonso	0.398	98	19	Jacoby Ellsbury	0.376	712
20	Hector Gomez	0.429	7	20	Nick Swisher	0.374	621
21	Prince Fielder	0.415	676	21	Alex Gordon	0.376	678
22	Matt Kemp	0.399	676	22	Matt Holliday	0.388	506
23	Dustin Pedroia	0.387	721	23	Chase Headley	0.374	433
24	Alex Avila	0.398	537	24	Casey Kotchman	0.378	548
25	Cole Gillespie	0.429	7	25	Yunel Escobar	0.369	574

Table 1: *Top* 25 *Colley batters with and without super-user.*

An alternative way of dealing with this issue, which we refer to as the Minimum-Game Method, is to allow every batter and pitcher to contribute to the network and to the ratings of each player, but only rank those pitchers achieving at least 400 plate appearances and those batters achieving at least 300 plate appearances. It should be noted that the results from this method are similar to the results using the super-user method. In fact, the first major discrepancy between these two Colley methods occurs at rank 40 in the Super-User Method, where Jesus Guzman appears despite having only 271 plate appearances. In 2011, Guzman appeared in 76 games, mostly as a pinch hitter. Thus we see that the Super-User Method ranks pinch hitters, players who suffered lengthy injuries, and catchers who alternate with pitchers, while the Minimum-Game Method does not. Pitchers Josh Johnson (Super-User Colley Rank 40) and Stephen Strasburg (Super-User Colly Rank 109) were having banner years until season-ending injuries. The Super-User Method allows these contributions to be noted whereas the Minimum-Game Method does not.

Without super-user				**With super-user**			
Rank	Pitcher	OBPA	Interactions	Rank	Pitcher	OBPA	Interactions
1	Justin Verlander	0.238	964	1	Justin Verlander	0.238	964
2	Jarrod Parker	0.227	20	2	Jered Weaver	0.257	913
3	Stephen Strasburg	0.193	85	3	Dan Haren	0.256	932
4	Jered Weaver	0.257	913	4	Josh Beckett	0.258	746
5	Dan Haren	0.256	932	5	Josh Tomlin	0.269	656
6	Josh Beckett	0.258	746	6	Josh Collmenter	0.266	609
7	Josh Tomlin	0.269	656	7	Guillermo Moscoso	0.266	521
8	Chris Young	0.242	96	8	Ian Kennedy	0.268	877
9	Josh Johnson	0.252	238	9	James Shields	0.267	968
10	Guillermo Moscoso	0.266	521	10	Clayton Kershaw	0.250	913
11	Josh Collmenter	0.266	609	11	Doug Fister	0.263	853
12	Ian Kennedy	0.268	877	12	Michael Pineda	0.270	685
13	Clayton Kershaw	0.250	913	13	Jeremy Hellickson	0.282	767
14	James Shields	0.267	968	14	David Price	0.278	899
15	Doug Fister	0.263	853	15	Alexi Ogando	0.277	680
16	Michael Pineda	0.270	685	16	Brandon McCarthy	0.280	677
17	Randall Delgado	0.293	146	17	Tommy Hanson	0.281	535
18	Luis Mendoza	0.267	57	18	Johnny Cueto	0.269	612
19	David Price	0.278	899	19	Tim Hudson	0.277	867
20	Alexi Ogando	0.277	680	20	Scott Baker	0.288	542
21	Jeremy Hellickson	0.282	767	21	Ricky Romero	0.279	897
22	Brandon McCarthy	0.280	677	22	Cole Hamels	0.251	848
23	Tommy Hanson	0.281	535	23	Brandon Beachy	0.289	578
24	Johnny Cueto	0.269	612	24	Daniel Hudson	0.290	911
25	Tim Hudson	0.277	867	25	Gavin Floyd	0.282	775

Table 2: *Top* 25 *Colley pitchers with and without super-user.*

These discrepancies also occur with the Massey Method. Tables 3 and 4 show the Massey rankings, the Massey Minimum Plate Appearance (dentoed minPA), and the Massey Super-User Rankings (denoted SU) for batters and pitchers in 2011.

Rank	Massey	Massey (MinPA)	Massey (SU)
1	Joey Gathright	Jose Bautista	Jose Bautista
2	Esteban German	Mike Napoli	Mike Napoli
3	Cole Gillespie	Miguel Cabrera	Miguel Cabrera
4	Russ Canzler	Matt Kemp	Matt Kemp
5	Jesus Montero	Adrian Gonzalez	Adrian Gonzalez
6	Jose Bautista	Alex Avila	Alex Avila
7	Mike Napoli	Joey Votto	Daivd Ortiz
8	Chris Parmelee	David Ortiz	Joey Votto
9	Miguel Cabrera	Curtis Granderson	Curtis Granderson
10	Jason Giambi	Giancarlo Stanton	Lance Berkman

Table 3: *Top* 10 *Massey batters with the three methods.*

Rank	Massey	Massey (MinPA)	Massey (SU)
1	Brad Peacock	Matt Cain	Matt Cain
2	Stephen Strasburg	Guillermo Moscoso	Guillermo Moscoso
3	Luis Mendoza	Roy Halladay	Roy Halladay
4	Jarrod Parker	Jordan Zimmerman	Jordan Zimmerman
5	Matt Cain	Justin Verlander	Justin Verlander
6	Josh Johnson	Doug Fister	Doug Fister
7	Chien-Ming Wang	Cole Hamels	Cole Hamels
8	Guillermo Moscoso	Ryan Vogelsong	Kyle Lohse
9	Roy Halladay	Kyle Lohse	Ryan Vogelsong
10	Chris Young	Cliff Lee	Cliff Lee

Table 4: *Top* 10 *Massey pitchers with the three methods.*

Note that the Minimum Plate Appearance approach and the super-user approach have very similar rankings with only a few transpositions of players. As with the Colley Method, we see that batters and pitchers who made the most of their limited game exposure show up high in the rankings.

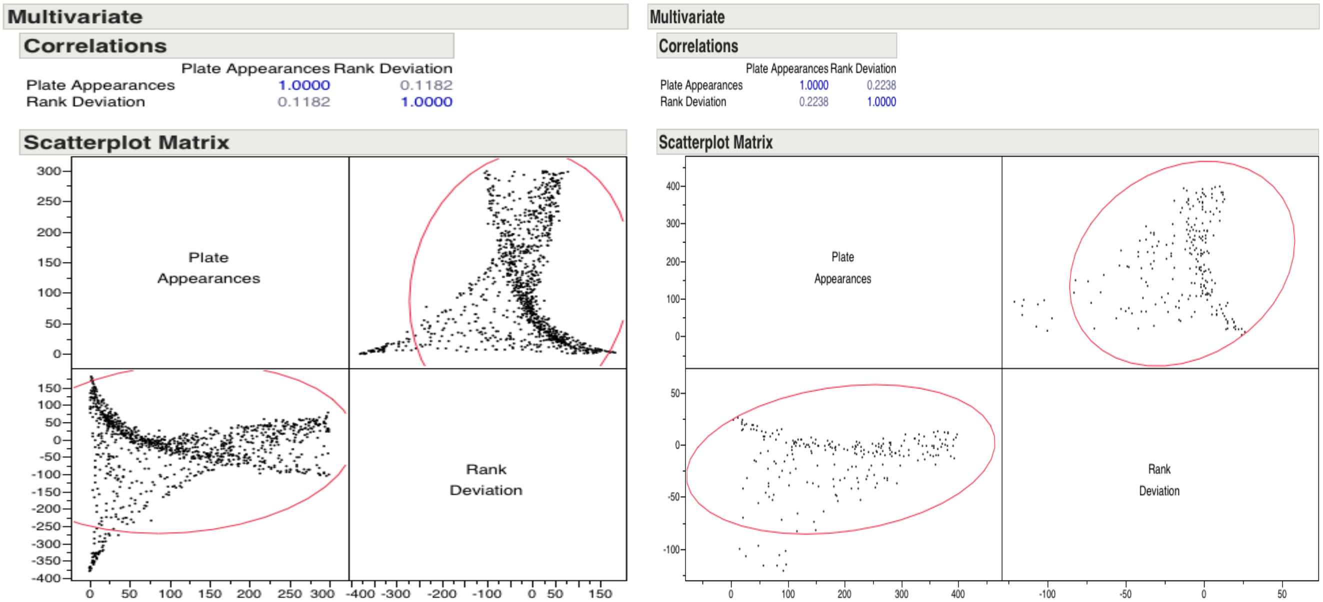
Rank	Super-User	Min Plate Appearances
20	Andruw Jones	Mark Reynolds
21	**Alejandro De Aza**	Carlos Beltran
22	Mark Reynolds	Justin Upton
23	Paul Konerko	Troy Tulowitzki
24	Carlos Gonzalez	Wilson Betemit
25	Carlos Beltran	Josh Hamilton
26	Justin Upton	Hunter Pence
27	Wilson Betemit	Matt Holliday
28	Troy Tulowitzki	Josh Willingham
29	**Allen Craig**	Carlos Pena
30	Josh Hamilton	Pablo Sandoval
31	Matt Holliday	Ryan Howard
32	Hunter Pence	Dustin Pedroia
33	Carlos Pena	Todd Helton
34	**Chris Parmelee**	Nick Swisher
35	Josh Willingham	Mike Carp
36	Pablo Sandoval	Lucas Duda
37	**Ike Davis**	Kevin Youkilis
38	Ryan Howard	Robinson Cano
39	**Jesus Montero**	Matthew Joyce
40	Dustin Pedroia	Corey Hart

Table 5: *Snapshot of Massey Super-User and MinPA Rankings.*

Table 5 gives a snapshot of the Massey Super-User and Massey Minimum Plate Appearance Methods from rank 20 to 40. The boldfaced names are those players who appear in the Massey Super-User rankings but not in the Massey Minimum Plate Appearance rankings. There are a couple of items to note. From Table 3 we see that Chris Parmelee is ranked 8^{th} in the overall Massey rankings. He had 88 plate appearances in 2011 and a batting average of 0.355 and was called up from the minors during the season. The super-user penalty that Parmelee incurred drops him in the rankings to 34, but he remains in the rankings. Alejandro De Aza is 15^{th} in the overall Massey Rankings but only drops 6 places to 21^{st} in the Massey Super-User Rankings. In 2011, De Aza appeared in only 54 games with 179 plate appearances and batted 0.329. He was an outfielder who rotated with other outfielders during the year and was used as a pinch hitter. Like Parmelee, he incurs a super-user penalty, but it is not quite as severe since he has more than twice the plate appearances as Parmelee. Finally, Allen Craig, who was instrumental in the St. Louis Cardinals winning the World Series in 2011, shows up at 29 in the Massey Super-User Rankings while not appearing at all in the Massey Minimum Plate Appearance method. In 2011, Craig had 219 plate appearances with a batting average of 0.315. He was used mostly as a utility player by the Cardinals and played 6 different positions for them throughout the season. Ike Davis was having a good year until injury struck, and Jesus Montero was a rotating designated hitter throughout the year. Contributions from these players are noted in the super-user approach but are omitted from the minimum plate appearance approach.

5. Effect of Playing the Super-User. One potential issue that could arise in the Super-User Method is whether players who are closer to the minimum cutoff appear higher in the rankings despite a poorer performance relative to other players. Could this be why players like Chris Parmelee and Alejandro De Aza moved in the rankings when a super-user node is added to produce Tables 1 and 5? To aid in answering this question, consider Figures 2 (a) and (b) of plots compiled of players from the 2009-2011 major league baseball seasons that failed to meet minimum cutoff for interactions. The number of interactions against the super-user is plotted against the overall rank change from ranking with and without the super-user nodes. The output from the statistical software package JMP is shown in the figures. In both pitchers and batters, the deviation in rank has a low correlation ($r = 0.1182$ for batters and $r = 0.2238$ for pitchers) with the number of interactions these players have against their corresponding super-user, and thus the variation in rank deviations is primarily due to other factors. This suggests that performance in the Super-User Method has more to do with where players are ranked than the number of plate appearances.

6. Choosing the Minimum Cutoff. How does one choose the minimum cutoff for games played by the super-user? Further, how sensitive is the method to such choices? Are there patterns to the perturbations seen in the rankings as evidenced by the super-user? To aid in answering these questions, let's continue our study of Major League Baseball. To simplify our discussion, we will concentrate solely on the ranking of batters.

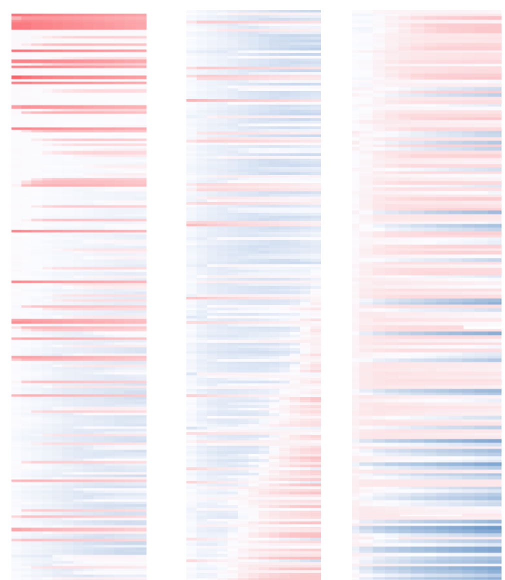


(a) 2009-2011 Batters below cutoff (b) 2009-2011 Pitchers below cutoff

Figure 2: *Number of super-user interactions against rank change for batters (a) and pitchers (b).*

The sensitivity of rankings to changing parameters associated with these linear algebra based ranking systems has been studied in the literature. Notably, the famous PageRank method for ranking webpages results in a linear system with a scalar parameter known as the teleportation constant [8, 12]. The sensitivity of PageRank to changes in this parameter is studied in two main ways: (1) through an analysis of the derivative of the ranking with respect to this parameter [2, 7] and (2) with computational studies that vary the parameter over its domain [2, 8, 9, 10]. Because the Colley and Massey super-user ranking system also result in a linear system with one scalar parameter, namely the number of times the batter/pitcher faces their respective super-users, we consider similar sensitivity analysis. However, the first approach, a derivative analysis, will not work for the super-user application because the parameter changes depending on the number of plate appearances of each player and is not the same for each player. Thus, we apply the second type of traditional sensitivity analysis to our problem and run computational studies varying our cutoff parameter over its domain.

Figure 3 shows rank variations of batters as the minimum cutoff varies from 0 to 700 incrementing by 50. We use 700 since the maximum number of plate appearances was 721 in 2011. Figure 3 is a heat map showing decreases (light red indicates a slight decrease and dark red, a large decrease) and increases (blue) in players' ranks as the minimum cutoff increases. Note that Figure 3 (a), (b), and (c) are the upper third, middle third, and bottom third of the rankings, respectively. The first column in (a), (b), and (c) shows the ranks of the players when the minimum cutoff is 0 and the last column, when the cutoff is 700. Each row in the map corresponds to an individual player and his rank changes.



(a) (b) (c)

Figure 3: *Heat map showing decreases (light red indicates a slight decrease and dark red, a large decrease) and increases (blue) in player's ranks as the minimum cutoff increases where (a), (b), and (c) are the upper third, middle third, and bottom third of the rankings.*

Three patterns are apparent in Figure 3.

- Red dominates toward the top of the heat map (Figure 3(a)). Players with few plate appearances see an immediate penalty (severe drop in rank indicated by dark red) when the minimum is raised above their number of plate appearances. This is working as intended as these players had undeservedly high rankings due to good performances in just a few games and now with our method must face the dominant super-user multiple times.
- Conversely, blue dominates the bottom of the heat map (Figure 3(c)). This indicates that low-ranked players receive some increase in rank as the parameter increases, i.e., as they play more games against the dominant undefeated

super-user. Of course, these low-ranked players lose each time they face the super-user, so their subsequent increase in rank seems contradictory. However, this increase in rank is a result of the "strength of schedule" influence in the ranking method. Because the super-user is the #1 team, these low-ranked players receive a boost by playing the dominant team so many times.

- The middle section of the heat map (Figure 3(b)) is very interesting, showing a triangular section of decrease in red with a symmetric triangular section of increase in blue. This section justifies setting the minimum cutoff parameter somewhere between 300 and 500. In this range, players have a slight decrease in rank, followed by a stable rank period, followed by a slight increase in rank. Within this range, the penalty movement down is balanced by the strength of schedule movement up.

Finding the best minimum cutoff is difficult since the problem does not lend itself to the aforementioned derivative-based approach and must be analyzed by empirical means. Further, there may be several different objectives governing the quality of the rankings produced by the methods. Who is to say which objective ranks batters and pitchers best? For these reasons we choose to empirically find the cutoff that minimizes the squared differences in current ranking and previous ranking as the minimum cutoff is incrementally increased. For each of the deviations in the minimum cutoff we can calculate a squared deviation in ranks (or ratings) that occurs for batters. As the minimum cutoff is varied from 0 to 700 (this time by 5), the minimum squared deviation in rank occurs at a cutoff of 425. Note that this value is larger than the 300 that we prescribed but smaller than the 502 that MLB prescribes.

One can also vary how the super-user is applied. In some scenarios it might be deemed more fair for every team to play the super-user a fixed number of times (rather than ensuring everyone has a minimum number of games). If we apply this so-called uniform game method to MLB, the effects are similar to the method of minimum plate appearances. We see an immediate drop in the rankings of players with few, but successful plate appearances in both methods. In the uniform games method, these players drop into the middle third of the rankings. In the super-user method, the immediate drop is more severe, sending these players into the bottom fourth of the rankings. What is significantly different between the two applications of the super-user is that in the uniform games application those players with few, but unsuccessful plate appearances also jump into the middle third of the rankings whereas in the minimum plate appearance method these players stay near the bottom of the rankings. This difference is due to the fact that the super-user plays everyone (instead of just those with few attempts) and has thus garnered a better rating for playing a tougher strength of schedule.

7. Ranking Tennis Players. As mentioned earlier, there is a great disparity between the number of matches played by the very best professional tennis players and those players whose ability puts them on the border between the ATP Tour tournaments and lesser competitions on the Challenger Tour and Futures Series.

Unlike the ranking of baseball players, the tennis ranking problem cannot be modeled as a bipartite network. Another difference is that in tennis the super-user does not end up as the #1 player in either the Colley or Massey rankings. Evidently, in tennis the top players are so dominant, that the super-user's 2560 wins and 0 losses compiled almost exclusively against sub-par competition was only good enough for 10^{th} place in the Colley rankings and 22^{nd} in the Massey rankings. As the minimum cutoff is increased, however, the super-user's ranking increases as it is allowed to play competition with higher ratings and thus improve the average ratings of its opponents. For instance, as the minimum cutoff is increased to 40, the super-user's Colley ranking is 5^{th}. Once the minimum cutoff is increased to 80, the super-user is playing the top-ranked players and is ranked 2^{nd}. The super-user takes the top spot in the Colley ranking when the minimum cutoff is 95.

Player	W	L	Colley	Colley (SU)	Massey	Massey (SU)	Avg. Change
Acasuso, J.	2	1	33	96	14	110	79.5
Gonzalez, M.	3	11	224	288	193	281	76.0
De Bakker, T.	2	11	240	294	197	294	75.5
Ramirez-Hidalgo, R.	5	15	239	296	198	290	74.5
Brands, D.	4	11	202	268	195	258	64.5
Delbonis, F.	3	2	47	95	20	99	63.5
Serra, F.	3	12	257	297	217	300	61.5
Van Der Merwe, I.	3	1	51	102	53	123	60.5
Bozoljac, I.	1	1	101	137	71	148	56.5
Cervenak, P.	2	1	81	121	82	154	56.0

Table 6: *The ten players whose rankings declined the most after introduction of the super-user. Based on largest average rank decline over both the Colley and Massey methods.*

Nevertheless, the issue of players who played comparatively few matches being ranked higher than expected still arises. Though these unjustly high rankings are not as dramatic as they are in the baseball example, where players with only two or three matches appear at the very top of the rankings. In addition to Jose Acasuso, Federico Del Bonis, Izak Van Der Merwe, Ilija Bozoljac, and Pavol Cervenak, players unknown to all but the most serious tennis fans, were all able to achieve a top 100 ranking from Colley and Massey by playing in five or fewer matches and winning at least as many as they lost. Table 6 shows the decline of these players after implementation of a super-user approach with a cut-off value of 17. (The mean number of matches played per player on the 2011 ATP tour was 16.88.)

Since the Massey method requires a margin of victory for each match and tennis's scoring system does not provide an obvious one, the formula

$$\text{Margin of Victory} = 5 + 5(\text{Sets Won Differential}) + \text{Games Won Differential}$$

was used with the Massey algorithm. This formula assures that each match winner has a positive margin of victory. The winner of a five-set match by the closest score possible (0-6,0-6,7-6,7-6,7-6), gets a margin of victory equal to one. For 2011, the mean margin of victory so defined was 19.21, so the value of 19 was used as the margin of victory in each of the super-user's matches.[6]

Table 7 shows the ten players whose rankings improved the most once the super-user was introduced. These 10 players can be divided into two groups, with each group populated by a similar type of player. The three players in the first group, Andreas Vinciguerra, Alexander Sadecky, and Aljaz Bedene, lost their only completed ATP match to players whose ATP rank at the time of the match was 727th, 401st and 204th respectively. This performance was so abysmal that the addition of sixteen losses to the super-user slightly improved their Colley and Massey ranking. Nevertheless these three still remain in the bottom third of ranked players in 2011.

Player	W	L	Colley	Colley (SU)	Massey	Massey (SU)	Avg. Change
Vinciguerra, A.	0	1	287	260	299	246	40.0
Haas, T.	7	11	141	105	132	89	39.5
Paire, B.	5	10	157	123	159	115	39.0
Falla, A.	7	14	153	116	139	100	38.0
Istomin, D.	10	20	152	117	140	103	36.0
Sadecky, A.	0	1	280	254	282	236	36.0
Gabashvili, T.	10	19	144	108	137	104	34.5
Volandri, F.	12	19	131	101	144	106	34.0
Souza, J.	7	8	126	100	147	107	33.0
Bedene, Aljaz	0	1	261	231	251	217	32.0

Table 7: *The ten players whose rankings improved the most after introduction of the super-user. Based on largest average rank improvement over both the Colley and Massey methods.*

The other group of seven players is much more interesting. Each of them played between 15 and 31 matches and had 2011 Colley/Massey rankings between 126 and 159 before the introduction of the super-user. Notice that few of these players ended up playing in any matches against the super-user, but their ranking benefited since other players originally ranked above them had their rankings fall after repeated losses to the super-user. These seven players played 70 of their combined 159 real-world matches against players in the ATP's top 50, and despite a combined record of 12-58, these "quality" losses are rewarded by Colley and Massey, especially once super-user matches were included.

Of course, it is possible that these players are merely being rewarded for having drawn tough first-round opponents, though it is interesting to note that Benoit Paire, Alejandro Falla, and Denis Istomin have all recorded their career-high ATP ranking since the 2011 season. Tommy Haas is an interesting case. The oldest player on the list, he was the ATP's #2 player in May 2002. His career has been interrupted twice by injury breaks that lasted over a year. He returned from one of these in mid-2011 and has since returned to the ATP's top 20. It appears that an increase in the Super-User Ranking may indicate future success and would be a promising area for future study.

As in the baseball example, the issue of where to place the minimum cutoff arises. If we again try to find a cutoff that minimizes the sum of the squared deviations from current rank to previous rank, one (local) minimum occurs at 17, which is where we set this cutoff. A slightly better minimum occurs when the cutoff is 30, although with this cutoff, the ranking for the top 40 players changes very little compared to when the cutoff is set at 17.

8. Games Beyond Sports. Sports ranking methods can be applied outside the context of sports. One such application is Prediculous, a free online social game where users compete with other users by predicting the future of events in sports, politics, entertainment, business, and the world. An example question regarding the prediction of a sporting event is shown in Figure 4. Players compete for points and leaderboard status. In all, approximately 18,600 unique users have played the game.

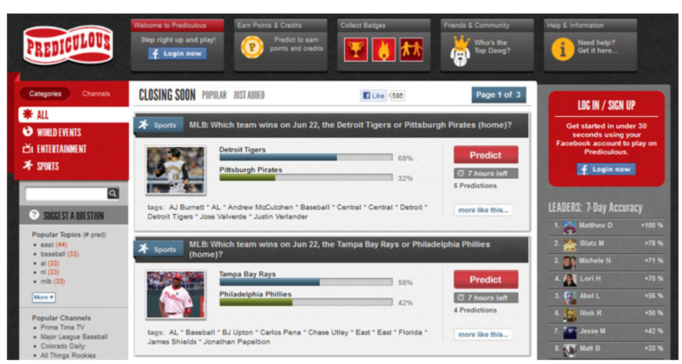


Figure 4: *A sample question contained on* `http://www.prediculous.com`.

In 2011, several authors of this paper applied the Colley method to this context with the intent of employing a more advanced ranking method in the website's leader board. The first step was defining a game. In this context, a game occurs over a prediction. Note, both players can be

[6]The Margin of Victory formula is the creation of Gettysburg College student Michael McLauglin, who studied tennis ranking for his 2012 senior capstone project.

correct, both incorrect, or one is correct in a prediction. Further, games occur between each pair of users participating in a prediction question. As such, ties must be integrated into the Colley method.

Like tennis, players (which are Prediculous users in this case) vary in the number of games played. New users begin with one prediction. Other users have over 100 predictions. The need for a super-user quickly arises. If one removes users who do not meet some minimum cut-off, new users will not receive a rank or will simply be put at the bottom until they surpass the cut-off. For Prediculous, this is undesirable since encouraging new users' play is important to growing participation on the site. Consequently, the super-user approach has business implications and is very helpful. The cut-off was chosen in consultation with Prediculous with their business goals in mind.

Earlier in the paper, other applications beyond sports were listed. If one ranks Netflix movies from user ratings, then again, movies vary in the number of ratings they receive which can elevate lesser known movies. A similar effect happens with Amazon products when a sports ranking method is applied to the user ratings of products.

9. Concluding Remarks. From sports to the Internet to businesses looking at their products, ranking is an important mathematical process. As demonstrated in this paper, the Colley and Massey methods can offer valuable insight in the ranking of batters and pitchers in Major League Baseball, of tennis players, and of participants in an online social game. These representative examples underscore how easily contexts arise in which an unequal number of games can arise. As discussed, removing data of less active players or participants removes valuable information. However, the presence of unequal numbers of games can decrease the value of the resulting rankings. To make sports ranking methods adaptable and valuable in such settings, this paper demonstrated how to adapt both the Colley and Massey ranking methods with the introduction of a super-user. This fictitious player aids in identifying strong players who have not played a significant number of games. At the same time, the effect of simply playing (and losing) a few games against strong opponents is not an immediate advantage. Further, the paper showed that such a method does not favor players approaching the cut-off and is adaptable but also not highly sensitive to the cut-off used with the super-user. The ideas introduced in this paper can further the scope that such sports ranking methods are applied and the depth of insight they offer.

References

[1] `http://mlb.mlb.com/mlb/official_info/official_rules/`. Online; accessed 31–July–2013.

[2] P. Boldi, M. Santini, and S. Vigna, PageRank as a function of the damping factor, *The* 14^{th} *International Conference on the World Wide Web*, New York, 2005, ACM Press.

[3] T. Chartier, A. Langville, and P. Simov, March madness to movies, *Math Horizons*, (2010) 16–19.

[4] T. P. Chartier, E. Kreutzer, A. N. Langville, and K. E. Pedings, *Mathematics and Sports*, Mathematical Association of America, 2010, 55–70.

[5] T. P. Chartier, E. Kreutzer, A. N. Langville, and K. E. Pedings, Sensitivity of ranking vectors, *SIAM J. Sci. Comput.*, 13 (2011) 1077–1102.

[6] W. N. Colley, Colley's bias free college football ranking method: The Colley Matrix explained, 2002. `http://www.colleyrankings.com/matrate.pdf`.

[7] D. Gleich, P. Glynn, G. Golub, and C. Greif, Three results on the pagerank vector: eigenstructure, sensitivity, and the derivative, in A. Frommer, M. W. Mahoney, and D. B. Szyld, eds., *Web Information Retrieval and Linear Algebra Algorithms*, No. 07071 in Dagstuhl Seminar Proceedings, Internationales Begegnungsund Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany, 2007.

[8] D. F. Gleich, *Models and Algorithms for PageRank Sensitivity*, Ph.D. Thesis, Stanford University, September 2009.

[9] D. F. Gleich, P. G. Constantine, A. Flaxman, and A. Gunawardana, Tracking the random surfer: empirically measured teleportation parameters in PageRank, *WWW'10: Proceedings of the* 19^{th} *International Conference on the World Wide Web*, April 2010, 381–390.

[10] D. F. Gleich, L. Zhukov, and P. Berkhin, Fast parallel PageRank: A linear system approach, *Tech. Report YRL*-2004-038, Yahoo! Research Labs, 2004.

[11] B. James and J. Henzler, *Win Shares*, STATS, Inc., 2002.

[12] A. N. Langville and C. D. Meyer, *Who's #1?: The Science of Rating and Ranking*, Princeton University Press, Princeton, 2012.

[13] K. Massey, *Statistical models applied to the rating of sports teams*, Honors Thesis, Bluefield College, 1997. `http://www.masseyratings.com/theory/massey97.pdf`.

[14] S. Saavedra, S. Powers, T. McCotter, M. A. Porter, and P. J. Mucha, Mutually-antagonistic interactions in baseball networks, *Physica A*, 389 (2010) 1131–1141.