2017

Tournaments on Tournaments: Using Graph Theory to Analyze Tennis Tournaments

Megan Scala

Meagan Scala

Senior Mathematics Seminar

Research Project

Tournaments on Tournaments: Using Graph Theory to Analyze Tennis Tournaments

Overview

The purpose of the project is to apply graph theory to tennis tournaments, specifically to rank players based off their performance in a given tournament. Four main methods are studied and applied to tennis tournaments to determine which methods best rank the competing players based off of various factors such as number of rounds won, opponents played, individual match scores, and performance throughout the year. These methods include: the Hare Method, the Kendall-Wei Method, percentage of games won, and the ATP ranking system.

Tennis: The Basics

A tennis match is played best of 3 sets. Therefore, a player must win 2 sets in order to win the match. Some tournaments, specifically the four Grand Slams (Australian Open, French Open, Wimbledon, and US Open), play best of 5 sets, but for simplicity sake the focus of this project will be on matches played as best of 3 sets. A set is won by the first player to win 6 games. However, if the score is 6-5 then a player must win 7-5 or if the score is 6-6 then a tiebreak must be played in order for a player to win 7-6.

In professional tennis, most tournaments are single elimination tournaments. In these tournaments, after each round of matches played the losers are eliminated and the winners move on to the next round. This process continues until there is one player left: the winner. However, there are tournaments, namely the ATP Finals, which are structured as round robin tournaments. In these tournaments, each player competes head-to-head against every other player in the tournament (usually just once). The winner is the player with the greatest number of matches won. In single elimination tournaments, it is obvious who the winner is (since there's only one player left), but for round robin tournaments there could be a tie between multiple players for the win. Therefore, for this project the main focus is on round robin tournaments.

Graph Theory: The Basics

Throughout this project, functions of graph theory are used to analyze and visually represent the outcomes of a round robin tournament. To begin, a given graph is composed of *n* vertices and *m* edges that connect the vertices. Two vertices are said to be *adjacent* if they are connected by an edge. The *degree* of a vertex is the number of edges that are connected to a given vertex. To represent a round robin tournament, *n* represents the number of players and *m* represents the total number of matches played. A *complete graph* is a graph where every vertex is connected to all other vertices by an edge. Therefore, for a complete graph, $m = \frac{n(n-1)}{2}$ [7]. A *directed graph* or *digraph* is a graph consisting of "ordered pairs of distinct vertices" [5, p.27]. Two vertices *a* and *b* are ordered if there exists an edge incident with (or connected to) the given vertices that is directed from either *a* to *b* or *b* to *a*, which can be written as (*a*, *b*) or (*b*, *a*), respectively. For any vertex *v*, "the number of vertices to which a vertex *v* is adjacent is the *outdgree* of *v* and is denoted by od(*v*). The number of vertices from which *v* is adjacent is the *indegree* of *v* and is denoted by id(*v*)" [5, p.162]. The sum of the indegree and the outdegree for any vertex equals the total degree of the vertex. When applied to a round robin tournament, a directed edge (*a*, *b*) represents a match where player *a* defeated player *b*. Since every match has a winner, and thus a directed edge, and all players compete head-to-head, this results in a *complete* *digraph* or *tournament*. In a tournament digraph, the indegree of a vertex represents the number of matches a player lost while the outdegree represents the number of matches a player won. Therefore, the total degree of a vertex represents the total number of matches played for any given player.

In addition, an *a-b walk* in any given graph is “a sequence of vertices,” beginning with *a* and ending with *b*, “such that consecutive vertices in the sequence are adjacent” [5, p.11]. A *cycle* is a walk that begins and ends at the same vertex but no edge or vertex (besides the first and last) is repeated within the walk.

The Problem

Suppose we have five players competing in a round robin tennis tournament. After a round robin of matches is played, the results can be graphed on a tournament digraph. The directed edges show which player won and which player lost in any given match. The player with the largest outdegree is the tournament winner since that player won the most matches. However, there is the possibility for a tie for the winner of the tournament if two or more players won the same number of matches. A tie between any of the players could also cause a problem if the players are to be ranked based off their performance in the tournament. For example, in Figure 1 player A is clearly the winner and should be ranked first since he won the most matches, which was 3. We know this since A has an outdegree of 3. Also, player E should be ranked last since he had the least number of matches won, which was 1. But players B, C, and D all won 2 matches so it is unclear as to how to rank them based off their number of wins.

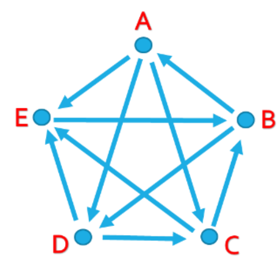

Figure 1

A tournament digraph will clearly show a tournament winner if the graph contains no cycles. However, if a cycle does exist then the results are inconclusive. For example, given three players A, B, and C if the cycle A-B-C-A exists then we know A beat B, B beat C, but C beat A and so there is no clear winner. The reason we could not break the tie between players B, C, and D in Figure 1 was because a cycle existed between the players (B-D-C-B). A cycle shows us that the digraph does not contain a *transitive* relation between the players. A transitive relation is one that implies the results "if $x > y$ and $y > z$, then $x > z$" [8, p.14]. In other words, since player B beat player D and player D beat player C, then player B should in theory also beat player C. But in our example in Figure 1, we had a nontransitive relation since player C beat player B. The nontransitive relation makes it unclear as to how to form a clear ranking for players B, C, and D.

Possible Solutions

A solution is therefore needed to break these ties that appear in tournaments. Through my research and sample problems, I looked at four possible solutions: the Hare Method, the Kendall-Wei Method, the percentage of games won, and the ATP ranking system.

Hare Method: The Hare Method is a process used to select the winner in voting elections through the use of preference ballots. In this method, the candidate with the least amount of votes is eliminated after the first round. The eliminated player's votes are then passed down to the next eligible player on the preference ballots, and the votes are recounted. Again after the second round, the candidate with the least amount of votes is eliminated and that candidate's votes are transferred. The next round of votes are counted, and the process continues until either a candidate has a majority of the votes or until only one candidate remains.

In order to apply this method to a round robin tournament, a few adjustments must be made. First off, the player with the least number of wins is eliminated. But since there aren't any preference ballots, the eliminated player's number of wins will not be passed on to another player. Instead, the number of wins for every other player will be recounted as though the eliminated player never competed. The process is repeated and the next player with the least amount of votes is eliminated. If two players are tied for least number of wins then whoever lost the head-to-head match between those two players is eliminated.

As an example, suppose the adjusted Hare Method is applied to the 5-person round robin tournament in Figure 1. Player E is eliminated in the first round since he had the least number of wins. The number of wins for every other player is recounted (this time excluding player E's matches) and the new totals are represented in the graph in Figure 2.

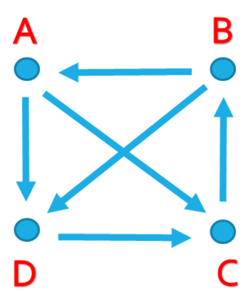

Figure 2

In the second round, players A and B each have 2 wins whereas players D and C each have 1 win. Although players D and C are tied for least number of wins, player C is eliminated next since player D beat player C in their head-to-head match. The wins are recounted and graphed in Figure 3.

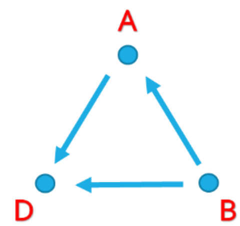

Figure 3

In the third round, player B has 2 wins, player A has 1 win, and player D has 0 wins. At this point, player B is automatically the winner since he won a majority of the matches. If the process is continued one more time and player D is eliminated, it still follows that player B is the winner (Figure 4).

Figure 4

Based on the number of rounds each player survived, the assigned rankings for this 5-player round robin tournament are: B, A, D, C, E.

This adjusted Hare Method seems to give us a clear ranking for our tennis players. Suppose, however, the example is changed slightly so that player D beats player B and player E beats player D. This new tournament is graphed in Figure 5.

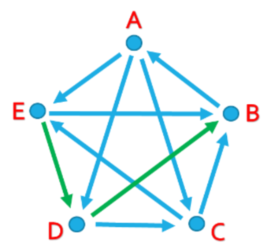

Figure 5

In this example, player A has 3 wins, players E and C both have 2 wins, and players B and D both have 1 win. Player D beat player B in their head-to-head match so player B is eliminated in the first round. The second round wins are then counted and can be seen in Figure 6.

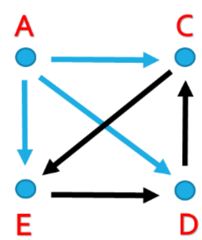

Figure 6

In the second round, player A still has 3 wins, but now players C, D, and E all have 1 win. Looking at the graph we can see there is a cycle (C-E-D-C) between vertices C, D, and E and it is thus unclear as to which player to eliminate next. Therefore, a nontransitive relation exists and a clear ranking cannot be made for these players.

Kendall-Wei Method: The Kendall-Wei Method uses score vectors to "give us an accurate assessment of the relative strengths of the players in the tournament" [2, p.4]. A *score vector* (S_1) lists the number of games each player has won. However, as Bondy and Murty explain, "The drawback here is that this score vector does not distinguish between players 2 and 3 even though player 3 beat players with higher scores than did player 2. We are thus led to the second-level vector…in which each player's second-level score is the sum of the scores of the players he beat" [4, p.186]. We could look at the second- or third-level score vectors to come up with a tournament ranking, but to get the most accurate result we will look at the nth-level score vector also known as the *ultimate strength vector* (S_n). The ultimate strength vector is constructed by finding the largest positive eigenvalue and the corresponding eigenvector for the adjacency matrix of a tournament digraph. We can then rank the numbers in the eigenvector to rank the corresponding players. For example, if the ultimate strength vector is S_n = (1.3, 1.45, 0.75, 0.23, 1.50) then the players would be ranked E, B, A, C, D. An adjacency matrix is an *n* x *n* matrix where *n* represents the number of players in the tournament. For our purposes, each row *i*, column *j* entry in the matrix will either be a 1 or a 0. A 1 indicates that player *i* beat player *j* whereas a 0 indicates player *i* lost to player *j* (since all players are competing head-to-head). A computer program is then used to calculate the largest positive eigenvalue and corresponding eigenvector. In my example of Figure 1, the tournament digraph had an eigenvalue of 1.86040…and its corresponding eigenvector was (2.09712…, 1.86040…, 1.53752…, 1.36396…, 1). Therefore, based on the eigenvector values the ranking for the players would be A, B, C, D, E.

Percentage of games won: As mentioned before, a tennis score is composed of sets that are made up of games. I had originally looked at having the winner be the player with the most number of games won (or least number of games given up). However, this criteria lead to several problems. First of all, if a player beat all of his opponents 6-0, 6-0 then that player should be the overall winner. However, that player only won 12 games in an individual match as opposed to a player who won 14 games in an individual match with a score of 6-4, 2-6, 6-3. This player is then rewarded for going into three sets as opposed to the first player who is punished for winning handily. In a similar way, if the winning criteria is least number of games given up then the losing player in a match is punished for keeping the match close and forcing a third set. For example, a player who loses 0-6, 0-6 only gave up 12 matches. But a player who fought hard and lost 4-6, 7-5, 4-6 gave up a total of 17 games. Therefore, in order to reward players who won by a large margin and players who only lost by a small margin the percentage of games won is the best numerical representation for the overall tournament winner.

For the percentage of games won method, I had to look at all of the games won out of all of the games played for an individual player (for both matches won and matches lost) to determine the player's percentage of games won. This process then leads to a clear winner and can break a given tie. I decided to test this method on an extreme case: a tie between all 5 players where each player won 2 matches (as seen in Figure 7).

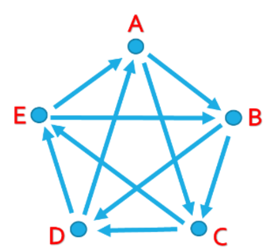

Figure 7

I randomly assigned match scores for each of the ten matches (Figure 7) since we know $m = \frac{n(n-1)}{2} = \frac{5(5-1)}{2} = 10$ and used those scores to determine the percentage of games won for each player (Figure 9).

Match	Score
A vs. B	6-3, 6-4
A vs. C	7-5, 1-6, 6-2
B vs. C	6-4, 6-3
B vs. D	6-3, 3-6, 6-3
C vs. D	2-6, 6-4, 6-4
C vs. E	6-4, 7-5
D vs. A	3-6, 6-2, 6-1
D vs. E	6-1, 6-2
E vs. A	6-1, 6-3
E vs. B	6-2, 4-6, 6-3

Figure 8

A	45.35%
B	48.91%
C	48.96%
D	56.38%
E	50.00%

Figure 9

As seen in Figure 9, this method broke the 5-way tie and provided the following ranking for the players (from highest to lowest percentage of games won): D, E, C, B, A.

Neither the Hare Method nor the Kendall-Wei Method could have been used to break the 5-way tie. The Hare Method cannot go past the first round because no player could be eliminated. In addition, based off the adjacency matrix for Figure 7, Kendall-Wei Method returns an eigenvector of all 1's (1, 1, 1, 1, 1), which simply tells us all the players are tied. Only the percentage of games won method could be used in this extreme case to break the ties between the players.

ATP Ranking System: The Association of Tennis Professionals (ATP) currently uses a points-based system (known as the Emirates ATP Rankings) to rank and seed male professional tennis players. "The year-end Emirates ATP Rankings is based on calculating, for each player, his total points from the four (4) Grand Slams, the eight (8) mandatory ATP World Tour Masters 1000 tournaments and the Barclays ATP World Tour Finals of the ranking period, and his best six (6) results from all ATP World Tour 500, ATP World Tour 250, ATP Challenger Tour and Futures tournaments" [3]. The farther a player continues in any given tournament, the more points that player is eventually rewarded. The tournament itself also has an effect on the total number of points awarded to players. So a player will receive more points for winning a Grand Slam as opposed to winning a Futures tournament. A table of points awarded for the total number of rounds won in each tournament is shown in Figure 10.

	W	F	SF	QF	R16	R32	R64	R128	Q
Grand Slams	2000	1200	720	360	180	90	45	10	25
Barclays ATP World Tour Finals	*1500								
ATP World Tour Masters 1000	1000	600	360	180	90	45	10(25)	(10)	(1)25
ATP 500	500	300	180	90	45	(20)			(2)20
ATP 250	250	150	90	45	20	(5)			(3)12
Challenger 125,000 +H	125	75	45	25	10				5
Challenger 125,000	110	65	40	20	9				5
Challenger 100,000	100	60	35	18	8				5
Challenger 75,000	90	55	33	17	8				5
Challenger 50,000	80	48	29	15	7				3
Challenger 40,000 +H	80	48	29	15	6				3
Futures** 25,000 +H	35	20	10	4	1				
Futures** 25,000	27	15	8	3	1				
Futures** 10,000	18	10	6	2	1				
*Barclays ATP World Tour finals 1500 for undefeated Champion (200 for each round robin match win, plus 400 for a semi-final win, plus 500 for the final win).									
** ATP Doubles Rankings points will be awarded in Futures tournaments beginning with the semi-final round.									
(1) 12 points only if the main draw is larger than 56									
(2) 10 points only if the main draw is larger than 32									
(3) 5 points only if the main draw is larger than 32									

Figure 10

If at the end of the year there are two or more players with the same amount of points, then the ATP implements the following tiebreaker(s): "(1) the most total points from the Grand Slams, ATP World Tour Masters 1000 mandatory tournaments and Barclays ATP World Tour Finals main draws, and if still tied, then (2) the fewest events played, counting all missed Grand Slams, ATP World Tour Masters 1000 tournaments and Barclays ATP World Tour Finals they could have played – as if played, and if still tied, then (3) the highest number of points from one single tournament, then, if needed, the second highest, and so on" [3]. The ATP is most concerned with how far a player advances in a given tournament and thus the total points awarded to each player.

Conclusion

Based on this project there is no clear winner in terms of "best ranking method." Each method has both clear advantages and clear disadvantages. First of all, the adjusted Hare Method turns a round robin tournament into an elimination tournament which is helpful in determining a clear winner since it is the last player standing. A ranking can also be determined based on how many rounds a player survives. However, when a player cannot be eliminated due to a tie the Hare Method is halted and is thus inconclusive. Also, this method does not look at match scores or the opponents a player faces.

On the other hand, the Kendall-Wei Method does incorporate past matches/opponents into a player's ranking through the use of score vectors. Therefore, it is able to break ties (as long as it's not a tie between all players) and provide a clear ranking for the players. But it does not look at the scores of the individual matches so it is unknown if a match was close or not.

The percentage of games won method, however, does look at the scores of each individual match throughout a round robin tournament. It therefore rewards players for fighting hard and losing a close match, for example, 4-6, 5-7 as opposed to losing a blowout match 0-6, 0-6. The winner, however, is not rewarded for winning a match handily. This method also does not take into account an opponent's record such as through score vectors as seen in the Kendall-Wei Method.

Finally, the ATP ranking system rewards players for farther advancement in a given tournament and only looks at a player's best performances throughout the year when determining rankings. However, the ranking system does not take into account how close a player's matches were or who the matches were against.

Therefore, the best ranking method to be used for a given tournament can vary depending on which factor is deemed most important such as number of rounds won (Hare Method or ATP ranking method), opponents played (Kendall-Wei Method), individual match scores (percentage of games won), or performance throughout the year (ATP ranking system).

References

1. Anton, H. (2010). *Elementary Linear Algebra* (10th ed.). Hoboken, NJ: Wiley.
2. Armstrong, D. (2006, September 29). Tournaments. http://wwwusers.math.umn.edu/~reiner/Classes/Tournaments.pdf
3. ATP World Tour. *ATP World Tour*. http://www.atpworldtour.com/
4. Bondy, J. A., & Murty, U. S. R. (1976). *Graph Theory with Applications*. New York, NY: Elsevier Science Publishing. https://www.iro.umontreal.ca/~hahn/IFT3545/GTWA.pdf
5. Chartrand, G., & Zhang, P. (2012). *A First Course in Graph Theory*. Mineola, NY: Dover Publications.
6. ESPN. *ESPN*. ESPN Internet Ventures. http://www.espn.com/
7. Gehrke, M. (2013, May 6). Graph Theory and Sports Scheduling. *Prezi*. https://prezi.com/pssig2j1mj4c/graph-theory-and-sports-scheduling/
8. Heilman, W., & Pasciuto, N. (2017, April). What Nontransitive Dice Exist among Us? *Math Horizons*, 14–17.
9. Malkevitch, J. Who Won! *American Mathematical Society*. http://www.ams.org/samplings/feature-column/fcarc-scores

10. Matrix Eigenvectors Calculator. *Symbolab*. https://www.symbolab.com/solver/matrix-eigenvectors-calculator
11. Random Number Generator. *Random Number Generator*. http://www.randomnumbergenerator.com/
12. Rose, M. L. (2012, July 23). Tie-Breaker Rules for the Round-Robin Format in Tennis. *azcentral*. http://healthyliving.azcentral.com/tiebreaker-rules-roundrobin-format-tennis-5200.html
13. Sackmann, J. (2012, March 30). Three Simple Ways to Improve the ATP Ranking System. *The TennisAbstract blog*. web log. http://www.tennisabstract.com/blog/2012/03/30/three-simple-ways-to-improve-the-atp-ranking-system/
14. Tannenbaum, P. (2018). *Excursions in Modern Mathematics* (9th ed.). Boston: Pearson.
15. Types of Tournaments. *Print Your Brackets*. https://www.printyourbrackets.com/types-of-tournaments.html