Visualization of Grand Slam results

Maran Many*
University Claude Bernard Lyon 1

Alexis Rabilloud†
Polytech Lyon

ABSTRACT

The main idea in this paper is to visualize the performances of men tennis players in the Grand Slams since 2000. In general, what we observe is that the top 32 players tend to perform very well in Grand Slams compared to the other players. Thus we have tried to understand if there is any particular reason why it is easier for a top player to win these tournaments compared to lesser ranked players. The obvious reason for top players to perform well is the fact that they are more talented, that's why they are actually highly ranked, but we can also think of the polling systems in these tournaments for the rounds, where top players are favoured, in the sense that the top players can't play against each other in the first rounds. That's why we have analysed the performances of the players depending on their ranking. Furthermore, for each player, we visualize the rankings of their opponents at each round and consequently the difference in the ranking between these 2 players. This visualization shows that players ranked above 32 are in danger from the first round itself, because they play against players with the same level, whereas top players play against players which are easily in their reach in the 2 to 3 first rounds, which stems from the polling system. The added value of this visualization is actually to show the imaginary path that a player ranked above 32 would have done to win the tournament, i.e. the players he would have had to beat to win the tournament. Moreover, with this visualization we can also campare the performances of two players in a given Grand Slam, for example compare the rankings of the opponents of the winner and the finalist.

1 INTRODUCTION

Grand Slams are the biggest tournaments in ATP (Association of Tennis Players) World Tour, and a player should win at least 7 matches consequentively to win the title, and there are 4 Grand Slams per year. The specific study on players performances in Grand Slams came from the observation that for the last 12-13 years, almost all the Grand Slams are won by the so called "Big Four" which comprises of four players, namely Roger Federer, Rafael Nadal, Novak Djokovic and Andy Murray. When winning a single Grand Slam is already a huge feat for any player, winning more than 10 Grand Slams is really a big achievement, and this is which is actually achieved by 3 of the 4 members of the "Big Four" (17 for RF, 14 for RN and 12 for ND as December 2016). Of course the main reason of these results is the immense talent of these players, but we also can't deny the fact that at least a little part is due the ATP polling system with "seeds" (tete de serie). In effect, this polling system ensures that the best players don't play against each other in the first rounds, and they rather play with relatively low ranked players in these rounds. This system has a double benefit. First it allows lowly ranked players to compete with world top players, and then, if the hierarchical ranking is respected, in the last rounds, we only have the best players competing against each other. This coming from the fact that the public is more attracted by matches between World number 1 and 2, rather than 54 and 55 for example. However, in spite of all these advantages of this polling system, there are also obvious drawbacks which is the fact that with this system, for a player ranked above 32, if there is no exceptional circumstances, it becomes extremely difficult to win a Grand Slam, because in theory, he will have to win consecutively against players ranked much higher than him. The simple statement here is that statistically it is easier for top ranked player to win a grand slam compared to a relatively lowly ranked player.

To show that top players are a little privileged in the beginning rounds, we have decided to visualize the path taken by all the players in a given tournament, the path of a player being the opponents he encountered in the tournament, or more precisely the rankings of the opponents. With this, we can see how difficult or easy was the path for a given player, and consequently compare the paths of different players.

We have also decided to construct the imaginary path for all the players, the imaginary path being the path of the player if he had actually won the tournament. Of course there is no imaginary path for the winner because he actually won the tournament, but for all the other players we have an imaginary path. The idea behind this path is to show how difficult it is for a relatively low ranked player to win the tournament.

2 RELATED WORK

There exists actually two main visualization methods of the players' results in a tennis tournament, Grand Slams included.

The first one is the simplest, and it consists in creating a table corresponding to each round in a tournament.This means that we have a table for the first round for example, and each row contains the details of a first round match with the name of the two players, their rankings and the final score of the match. Then we have a table for the second round, third round... till the finals with the name of the winner at the end. This visualization is still widely used in a lot of tennis related websites, including the ATP website (see figure 1). The drawbacks of this visualization is that it is not obvious to see which stage a given player has reached in the tournament, apart from the winner and the finalist. Moreover, it does take some time to reconstite the different opponents of the winner for example. That's why a second type of visualization, which is much more intuitive has emerged.

This second visualization is a kind of dendogram. Indeed, after the draws for the first round matches, we can separate all the players into 2 groups, in such a way that the best players in the two groups will play the finals against each other. Knowing this match system, we can create a graph where in the first level we have all the first round matches separated into two groups, the players of group 1 being at the top and the players of the group 2 being at the bottom for example. Then in the second level, we have all the matches of the second round, and each second round match has two roots corresponding to the matches won by the players of this match, etc... till the finals. Figure 2 is an excerpt of this graph of Wimbledon 2012.

Although this visualization is definitely better than the first visualization, because it is easier to follow the path of a player, is still has some drawbacks. In effect, for a Grand Slam, we have 128 players for the first round, so it becomes really difficult to have a global vision of the graph from the beginning. Furthermore, it is not really easy to compare the paths of two given players, as we first have to

*maran.many@gmail.com
†arabilloud@hotmail.fr

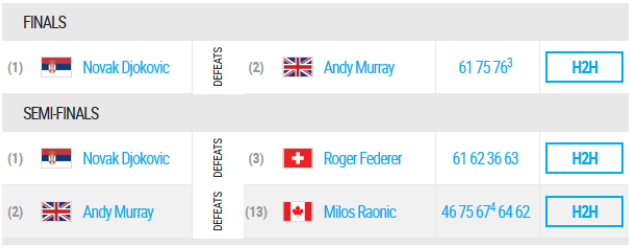

Figure 1: A partial visualization of the results of the Australian Open 2016 with tables. The image is from the ATP website [1]

.

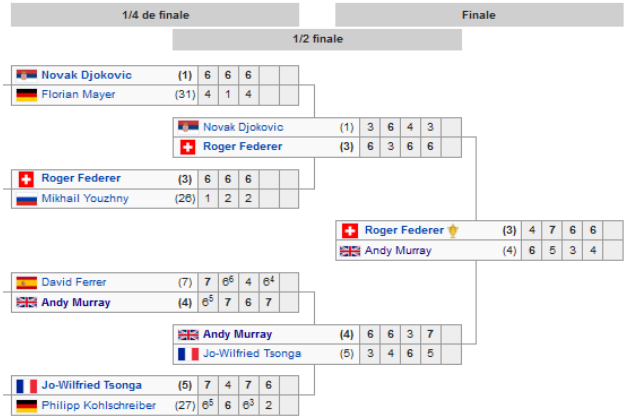

Figure 2: A partial visualization of the results of Wimbledon 2012. The image is from Wikipedia and is in the public domain.

find the players in the tree and then note their opponents, etc...

In this paper we propose a new way to visualize Grand Slam results, in which we can compare the paths of several players, and even compare the results of groups of players, regrouped by their ranking. We even go further by building the imaginary path of a player who would have won the tournament.

3 Project Description

Throughout this project, our aim has been to prove that, while it seems normal for top ranked player to perform well in tournaments, the organisators of said tournament are a part of the reasons why the first-rate players are monopolizing the places of honor. The visualizations we realized as part of this project serves directly this objective.

3.1 Visualization 1

3.1.1 Observations

This visualization shows in the form of a line chart the rank difference the tennis players we are studying are facing during the tournament considered. To realize this chart, we used the data all the results of the matches of the four Grand Slam from 2000 to 2016. For the players not reaching the final, we also predicted their route had they won all their matches using the results of the player they lost against, and so on if that player ended up losing as well. This allows us to compare the players with a similar point of view, there wouldnt be much to analyse if the players had highly varied number of matches. It is indeed relevant to note which players the losers would have been opposed to as it is among the possibles scenarii.

We can note that for each player we compute the following formula : rankdiff = rankopponent - rankplayer which means that if the

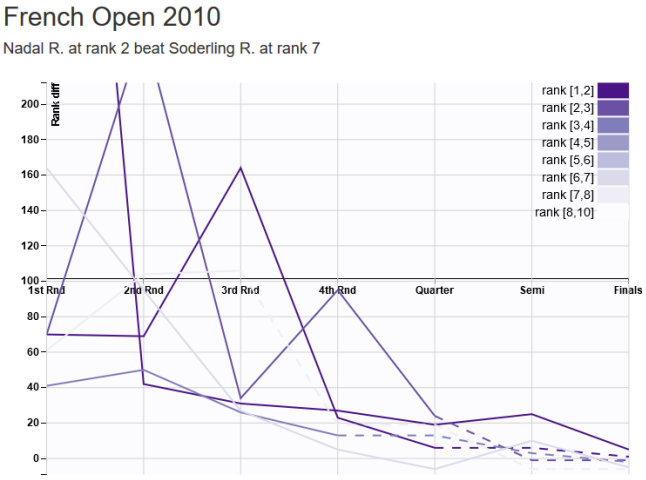

Figure 3: Interactions for the first visualization (1)

.

player plays against a lower ranked opponent, this difference will be positive and otherwise if the opponent is better than the player. It also has a meaning because if the player has to play against a better opponent, it means that he will have to display more of his talent and energy to win and it is a loss for the following rounds of the tournament.

What we can discern from this visualization is that a large majority of the top ranked players goes through the first few rounds of any tournament among the four Grand Slam without much troubles, given that most of them end up passing. They are opposed to highly inferior players in terms of ranking as can be seen from the large difference in rank. This difference in rank grows sharply smaller as the tournament goes on, to most of the time ending with a difference in rank less that five.

3.1.2 Interactions

While our objective was clearly to study the route of first-rate players, we chose to allow a high level of interaction with the user, leaving him the freedom to choose the ranking of the players he wish to display as well as choosing the tournament he wish to study.

The user can also accede to more accurate information about the displayed player if he overflies the corresponding line with the pointer.

3.1.3 Use Case

As an example of an utilization of our visualization, we display the rendering given by the application for the French Tour 2010, considering the players ranked in the top 8 of the ATP ranking.

On this visualization, we can see that all the lines representing each player seem to be converging toward a rank difference of 0, which means that all the players participating in the late stages of this tournament are players whose rank are close to the players selected. Since the selected players are first-rate players, we can conclude that all the players participating in the final rounds of the tournament are also top ranked players.

We can also observe that before reaching the final stages of the tournament, all the rank differences are positive or even extremely positive. For example, the winner of this tournament, Rafael Nadal, never encountered any players of the ATP top 20 before the final round. What can be emphasized from this is that Rafael Nadal in

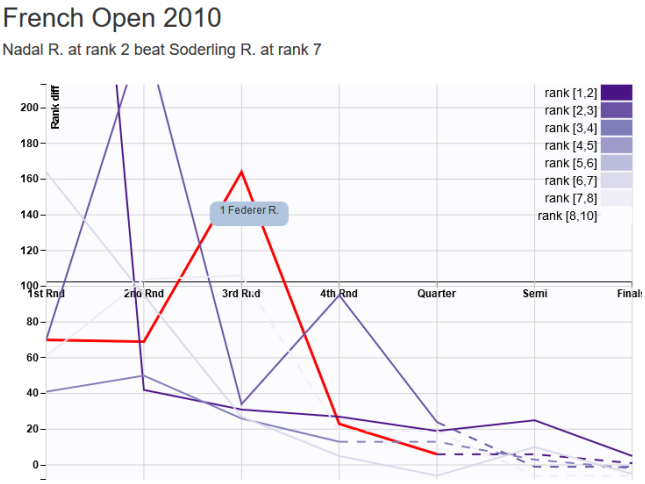

Figure 4: Interactions for the first visualization (2)

.

this tournament was greatly helped by this arrangement because he had less trouble winning his matches et could save his energy since the matches were less challenging. As for the other finalist, Robin Soderling, he met opponents from the ATP top 20 since the fourth round, twice among those better ranked than him, and had to deplete a lot of energy to dispose of his opponents.

3.2 Visualization 2

3.2.1 Observations

As a support to our first visualization, we decided to provide another representation of the available data. In the form of a barchart, this visualization allows us to represent directly the ever growing proportion occupied by top ranked players as a tournament progress.

Our aim with this visualization is to show the proportion represented by the players whose rank are included in the domain the user specified, from the minimum ranking to the maximum ranking. The results of this visualization emphasizes the fact that as the domain covers the top ranked players, the proportion occupied by that domain increases in the late stages of the tournament.

Our first hypothesis was that the omnipresence of the top ranked players in the late stages of a tournament can not simply be explained by the superiority of said players in terms of talent.

3.2.2 Interactions

We implemented the same level of interaction as our first visualization to let the user control the knowledge displayed by the visualization and link it to the first visualization. Thus, the user is able to choose which range of ranks he wish to display and which tournament in which year he want to study. The user can also use its pointer to display additional information on any of the bar of the chart.

3.2.3 Use Case

The case of the French Tour 2010:
On this visualization, we can see immediately that the proportion of the available places occupied by the players we selected is increasing in an exponential way. When we consider that the player selected are the players with a rank among the ATP top 8, it means that most of these players are not eliminated early from the tournament. The proportion occupied by these players even reaches 100% in the final round, which means that both players are from the selected lot. On

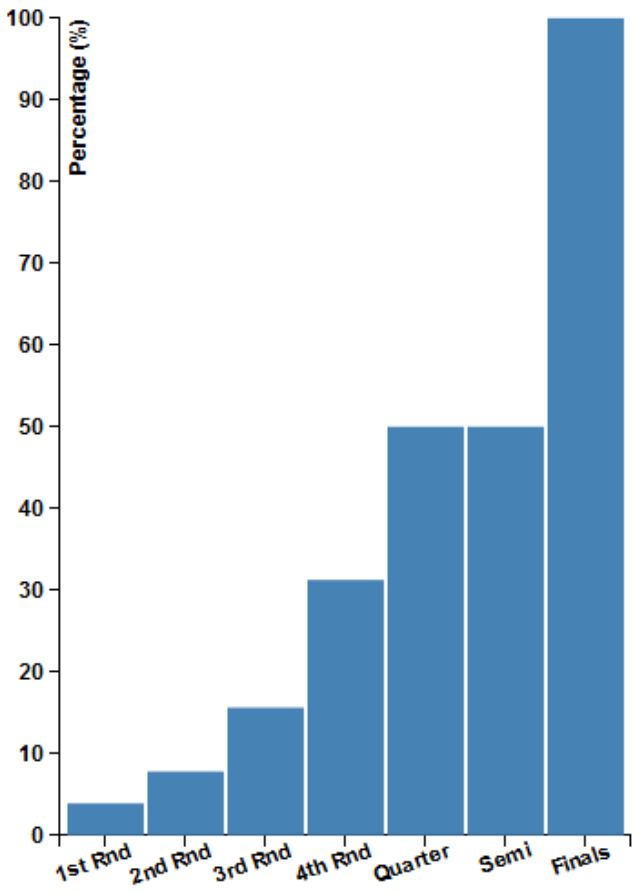

Figure 5: Interactions for the second visualization (1)

.

the other hand, if players with lower ranks, for example from rank 30 to 60, are selected, this proportion doesnt grow with the progress of the tournament and tends to stagnate or even decrease with each passing round.

3.3 Interesting results

If we analyse the performance of the players ranked 1 to 4 in all the 68 Grand Slams that we have analysed, at least one of these players reach the semi finals, with only 4 exceptions that are : US Open 2000, Australian Open 2001, Australian Open 2002 and Wimbledon 2003. Australian Open 2002 expecially represents the Grand Slam where there was the most of counter-performances, as the world number one and two lost in the first round itself and the world number four lost at the second round. This means that from 2004, the top 4 players are very steady and they regularly reach the semi finals and the finals.

Now if we analyse the performances of players ranked between 33 to 64, we can see that it is very rare for these players to reach the finals (only in 4 tournaments) or even semi-finals. Among the best performances of a player in this range of ranking is that of Gaudio G. ranked 44 at the time of the French Open 2004 where he consecutively beat 3 players who had a rank at least 20 better than him to win the title.

For players ranked above the ranking 64, we only have 3 Grand Slams out of the 68 where a player in this range reach at least the semi-finals. However among these 3 Grand Slams, in Wimbledon

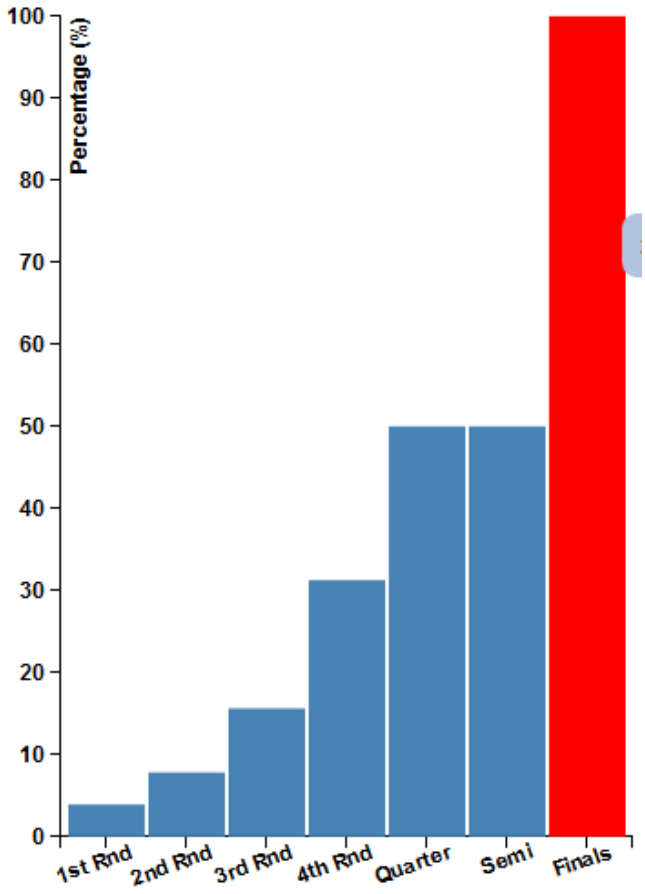

Figure 6: Interactions for the second visualization (1)

.

2001 we have a great surprise as the winner is Goran Ivanisevic ranked 125 at that time, who won the title by beating consecutively 3 players who where among the 20 players.

4 Discussion

The database we used [2] had some issues, for example some players were marked as winners of a match when they actually lost it. Consequently we had to manually correct these values for a number of matches. Therefore it is possible that some minor errors went unnoticed, as we can't check all the matches.

As expected, highly ranked players can reach the last rounds more easily than lowly ranked players with the current polling system. But after these "easy" rounds, they have 2-3 difficult rounds to score to win the title. Which means that assuming the eight best players reach the quarter finals, if at least three out of the four best reach the semi finals regularly, this clearly indicates that these players dominate the other players, so this means that these four players will surely reach the last rounds if they do not meet each other.

One of the main limitation of this visualization is that we only visualize men players' results, and not womens'. The justification for this choice, as explained in the introduction, is that a small number of players are dominating their sports for over a decade, which is not really the case for women, where it is more balanced, even if there was a period of domination of Serena Williams over the other players.

Another limitation here is that we only visualize the Grand Slam results, and we could have added other tournaments such as Masters 1000. However, we strongly believe that these tournaments would have given the same results, because these tournaments are also rarely won by anyone who does not belong to the "Big Four".

The initial idea was actually to have another graph, where we will visualize the time spent at each round by each player, because we expect that in the first rounds, top players don't spend too much time on court, compared to other players. Which means that they will have more energy for the next rounds compared to other players, and consequently they are more likely to win the next round. However the database we had does not contain this information. Therefore, this visualization will surely get better if we could see the opponent and the time spent by the player at each round.

Another thing that we can't do here, is that we can't select a player by his name to see his path. To do this, we actually need to know his ranking, as we can select a player by his ranking. It is something that is quite feasible, but due to time constraits, we didn't do it.

5 Conclusion

In this paper, we have explored a new way to visualize the progress of a tennis player, that consists in analyzing the rank difference between this player and his opponent at each round. We have also visualized the performances of a given group of players, delimited by a minimum and maximum ranking, which has enabled us to have interesting results. Especially these visualizations participate in explaining why top ranked players are more likely to win big tournaments like Grand Slams, as they play the first rounds against players who they can easily beat normally.

This is a kind of vicious circle because with the current polling system, as top ranked players are more likely to go deep in the tournaments, they will win more ATP points than other players, so they will be highly ranked, and so on. As changing the round systems will be difficult, one solution might be to change the allocation of ATP points of a player depending on the rank of his opponent, for example giving more points if the opponent has a rank between 1 and 8.

References

[1] ATP. Official site of men's professional tennis, January 2017.
[2] http://www.tennis data.co.uk/alldata.php. Tennis-data.