

WHO'S THE GOAT? SPORTS RANKINGS AND DATA-DRIVEN RANDOM WALKS ON THE SYMMETRIC GROUP

GIAN-GABRIEL P. GARCIA AND J. CARLOS MARTÍNEZ MORI

ABSTRACT. Given a collection of historical sports rankings, can one tell which player is the greatest of all time (i.e., the GOAT)? In this work, we design a data-driven random walk on the symmetric group to obtain a stationary distribution over player rankings, spanning across different time periods in sports history. We combine this distribution with a notion of stochastic dominance to obtain a partial order over the players. We implement our methods using publicly available data from the Association of Tennis Professionals (ATP) and the Women's Tennis Association (WTA) to find the GOATs in the respective categories.

1. INTRODUCTION

The question of which player is the *greatest of all time* (i.e., the GOAT) is one that brings about strong opinions from sports fanatics. Within individual sports in particular, a wide variety of individual performance criteria can be used to argue in favor or against a given player being the GOAT. These criteria are often quantitative at face value, taking the form of individual performance statistics. However, these statistics need not be consistent with one another, and thus they ultimately prompt subjective value judgments on their relative importance. To add to this, other qualitative aspects about how a specific sport develops over time might alter their significance across generations.

For instance, consider Steffi Graf and Serena Williams; two widely successful players in women's professional tennis. Between the two, Williams leads in the number of *major* titles, holding 23 compared to the 22 held by Graf. However, Graf leads in the number of titles, holding 107 compared to the 73 held by Williams. Which of these statistics—the number of titles or the number of major titles—is more important? How does 1 additional major title fare against 34 additional titles? Note that these players had little overlap in competition: Graf played professionally from 1982 to 1999, whereas Williams turned professional in 1995 and only retired recently in 2022. How do the numerous differences in professional tennis from the 1980s to the 2020s (e.g., changes in equipment, training, court surfaces, and point systems; consolidation of prestige around certain tournaments; increased competition/access to the sport; a global pandemic that suspended multiple tournaments in 2020) relate to the significance of these and other performance statistics?

1.1. Contributions. The Graf-Williams example highlights an inherent difficulty with the problem of identifying the GOAT: individual performance statistics can only offer snapshots of history, and it is challenging to objectively contextualize them over time. Fortunately, most individual sports maintain regularly-updated *rankings* that reflect the moment in time

Date: September 19, 2024.

2020 Mathematics Subject Classification. 60J20.

Key words and phrases. sports analytics, rank aggregation, Markov chains.

in which they were curated. In professional tennis specifically, rankings are curated on a weekly basis by both the Association of Tennis Professionals (ATP) [2] and the Women’s Tennis Association (WTA) [21].

In this work, we propose a mathematical framework that leverages historical ranking data to *compare* players, and in particular players spanning across different time periods in sports history. In broad strokes, our method is as follows:

- (1) First, we design a data-driven random walk on the symmetric group to obtain a stationary distribution over player rankings.
- (2) Next, we combine this distribution with a notion of stochastic dominance to obtain a partially ordered set (poset) of players.

This approach to player comparison represents a departure from existing methods in two principal ways. First, compared to the use of individual performance statistics or match-level data, the use of historical rankings more directly reflects the evolution of circumstances and/or value systems. For example, players might be strategic around tournament participation in response to the ranking system in place—what one had to do to be top-ranked in the 1980s might be different from what one had to do in the 2010s—and this might not be as evident in other datasets. Moreover, it opens the opportunity for longitudinal comparisons in ranked systems for which match-level data is not available/does not exist (e.g., individual players within team sports, music charts, college rankings). Second, the adoption of a notion of stochastic dominance helps come to terms with the possibility that there is no single GOAT, instead limiting pairwise comparisons to cases in which these can be established conclusively.

To illustrate the practical merits of our approach, we perform a numerical implementation using singles ranking data from the “Open Era” of the ATP and the WTA. We experiment with a concept of GOAT that captures individual performance at the top of ranking cut-offs—performance as a “top 3,” “top 5,” “top 10,” or “top 20” player (refer to Section 3.1 for details)—to obtain data-driven posets and find the GOATs in the respective categories.

Tying back to the introduction, we conclude that Graf and Williams are *both* consistently the GOATs. In particular, they are distinct maximal poset elements in all corresponding experiments. We summarize these and other main findings about the GOATs in the ATP and the WTA in Tables 1 and 2, respectively. In addition, we observe that prominent young

Cutoff	GOATs of the ATP	Notes
Top 3	Djokovic, Federer, Nadal, and Sampras	- All better than Connors and Lendl
Top 5	Djokovic, Federer, and Nadal	- Federer and Nadal better than Sampras - Djokovic and Sampras incomparable
Top 10	Djokovic, Federer, and Nadal	- All better than Connors, Lendl, and Sampras - Federer and Nadal better than Becker - Djokovic and Becker incomparable
Top 20	Federer and Nadal	- Federer better than Djokovic - Nadal and Djokovic incomparable

TABLE 1. Summary of GOATs of the ATP. The surnames mentioned correspond to Boris Becker, Jimmy Connors, Novak Djokovic, Roger Federer, Ivan Lendl, Rafael Nadal, and Pete Sampras (in alphabetical order). Refer to Section 4.2 for more details.

Cutoff	GOATs of the WTA	Notes
Top 3	Graf and S. Williams	- Graf better than Navratilova - S. Williams and Navratilova incomparable
Top 5	Graf and S. Williams	- Graf better than Navratilova - Graf and S. Williams better than Seles - S. Williams and Navratilova incomparable
Top 10	Graf, Navratilova, and S. Williams	- Navratilova better than Seles - Graf and Seles incomparable - S. Williams and Seles incomparable
Top 20	Graf, Navratilova, Seles, S. Williams, V. Williams, and Wozniacki	- Numerous players close to maximal

TABLE 2. Summary of GOATs of the WTA. The surnames mentioned correspond to Steffi Graf, Martina Navratilova, Monica Seles, Serena Williams, Venus Williams, and Caroline Wozniacki (in alphabetical order). Refer to Section 4.3 for more details.

players such as Carlos Alcaraz in the ATP and Aryna Sabalenka in the WTA perform well as top 3 players, but that they are at a disadvantage as top 5, 10, or 20 players due to their short career lengths at the time of writing.

Lastly, we analyze the obtained posets and note qualitative differences between the rankings of the two associations. We find that, in the ATP, being a “good top 20 player” is correlated with being a good top 10, 5, and 3 player (with decreasing correlation in that order). However, in the WTA, *being a “good top 20 player” is only mildly correlated with being a good top 10 player, and it is nearly uncorrelated with being a good top 5 or top 3 player*. This highlights a measurable difference in terms of career and dominance longevity for top players across the ATP and the WTA.

1.2. Related Work. Radicchi [14] addressed the question of who is the GOAT in men’s professional tennis using historical match-level data and a network diffusion algorithm similar to PageRank [7]. Baker and Hale [4] addressed the same question using time-varying *paired comparison* models and the results of matches in major tournaments. Temesi, Szádóczki, and Bozóki [20] addressed the question of who is the GOAT in women’s professional tennis using pairwise comparison matrices and the results of matches between players who were ever ranked first. Still, from a philosophical point of view, sports rankings have been criticized for their inability to capture comparisons that are “too close to call” [6].

In contrast to existing methods, we propose the aggregation of historical rankings into a single poset as an alternative to the stringent requirements of a linear order. That is, some pairs of players can be conclusively compared, others cannot, and there might be various, pairwise incomparable GOATs. Thus, our conceptualization of the problem is most closely related to that of *rank aggregation* [10, 18] (refer also to Lin [13] for a survey).

However, our methods are actually most closely related to those from the theory of card shuffling—the design and analysis of methods to sample uniform random permutations. The set $[n] := \{1, 2, \dots, n\}$ corresponds to a deck of $n \in \mathbb{N} := \{1, 2, \dots\}$ cards, and different card shuffling procedures (e.g., riffle shuffling, shuffling by transposing two random cards, shuffling by transposing two random adjacent cards) correspond to different sampling methods.

The central questions in this area relate to convergence rates: how many times does a given shuffling procedure need to be repeated so that a given deck is sufficiently close to uniformly distributed? Bayer and Diaconis [5] showed that $(3/2)\log n + \epsilon$ riffle shuffles are necessary and sufficient for a deck of n cards to start becoming well-mixed, exhibiting a cutoff phenomenon [1] (so that a standard deck of 52 cards requires about 6 repetitions). Diaconis and Shahshahani [9] showed that the random transposition shuffle has a cutoff at $(1/2)n \log n$, whereas Lacoïn [11] showed that the random adjacent transposition shuffle has a cutoff at $(1/\pi^2)n^3 \log n$. Refer to Diaconis and Fulman [8] for more on card shuffling.

Our method in Section 3 is most similar to shuffling by random (not necessarily adjacent) transpositions. However, our goal in this work is different from card shuffling in that we do not seek to sample uniform random permutations. Rather, we seek to sample permutations from a data-driven distribution that is somehow in agreement with historical ranking data. Given the large amount of existing work for the case of uniform random permutations, we anticipate analytical questions around the data-driven sampling of permutations to be challenging and of independent mathematical interest.

1.3. Organization. The remainder of this work is organized as follows. For the purpose of self-containment, in Section 2 we summarize some necessary technical background. In Section 3 we design our methods. In Section 3 we implement our methods using data from the ATP and the WTA, and analyze the results of our numerical implementation. Lastly, in Section 5 we make concluding remarks.

2. BACKGROUND

Readers familiar with the symmetric group may choose to skip Section 2.1, whereas readers familiar with Markov chains may choose to skip Section 2.2. In Section 2.3 we formalize our problem input along with other necessary technical concepts and notation.

2.1. Symmetric Group. Let $n \in \mathbb{N} := \{1, 2, 3, \dots\}$. A permutation of $[n] := \{1, 2, \dots, n\}$ is a bijection $\pi : [n] \rightarrow [n]$. The *symmetric group* of degree n , denoted \mathfrak{S}_n , is the set of all permutations of $[n]$ forming a group under function composition.

Fact 2.1. For $n \geq 1$, $|\mathfrak{S}_n| = n!$

The following example illustrates the different ways in which we express permutations.

Example 2.2. Let $\pi : [5] \rightarrow [5]$ be the permutation expressed explicitly as $\pi(1) = 5$, $\pi(2) = 4$, $\pi(3) = 1$, $\pi(4) = 2$, and $\pi(5) = 3$. Equivalently, we express π as follows:

- In one-line notation, and write $\pi = 54123$. Here, the i th entry of the word is $\pi(i)$.
- In cycle notation, and write $\pi = (153)(24)$. Here, the number following number i in a cycle is $\pi(i)$. For example, 5 follows 1 in (153) so that $\pi(1) = 5$. Similarly, 1 follows 3 in (153) so that $\pi(3) = 1$. Note that function composition is performed from right to left, so that $(153) = (13)(15)$ and equivalently $\pi = (13)(15)(24)$.

Cycles of length 2 are known as *transpositions*, since the effect of a cycle (i, j) for distinct $i, j \in [n]$ is to transpose the i th and j th entries. By convention, a transposition (i, j) is written as (i, j) if $i < j$ and as (j, i) if $j < i$.

Fact 2.3. For $n > 1$, every $\pi \in \mathfrak{S}_n$ is a product of at most $n - 1$ transpositions.

Moreover, every transposition can be expressed as a product of *adjacent transpositions*. For ease of notation, let $s_{ij} = (i, j)$ denote a transposition and $s_i = (i, i + 1)$ for $i \in [n - 1]$ denote an adjacent transposition. Consider the following example.

Example 2.4. *Note that*

- $s_{13} = (13) = (12)(23)(12) = s_1 s_2 s_1$,
- $s_{15} = (15) = (12)(23)(34)(45)(34)(23)(12) = s_1 s_2 s_3 s_4 s_3 s_2 s_1$, and
- $s_{24} = (24) = (23)(34)(23) = s_2 s_3 s_2$.

Therefore,

$$\begin{aligned}
\pi &= (153)(24) \\
&= (13)(15)(24) \\
&= (12)(23)(12)(12)(23)(34)(45)(34)(23)(12)(23)(34)(23) \\
&= (12)(34)(45)(34)(23)(12)(23)(34)(23) \\
&= s_1 s_3 s_4 s_3 s_2 s_1 s_2 s_3 s_2.
\end{aligned}$$

The last example illustrates the following.

Fact 2.5. *For $n > 1$, every $\pi \in \mathfrak{S}_n$ is a product of at most $n(n-1)/2$ adjacent transpositions.*

In this work, we associate with \mathfrak{S}_n two bi-directed graphs over $n!$ nodes (the number of nodes follows from Fact 2.1):

- (1) The first one, depicted in Figure 1a for $n = 4$, contains arcs (π, σ) and (σ, π) if and only if $\pi = \sigma \cdot s_{ij}$ for some $1 \leq i < j \leq n$. For example, if $\pi = 4213 \in \mathfrak{S}_4$ and $\sigma = 4312 \in \mathfrak{S}_4$, then the graph has the arc (π, σ) since $4213 \cdot s_{24} = 4312$ and s_{24} is a transposition. The undirected version of this bi-directed graph is known as the Bruhat graph.
- (2) The second one, depicted in Figure 1b for $n = 4$, contains arcs (π, σ) and (σ, π) if and only if $\pi = \sigma \cdot s_i$ for some $i \in [n - 1]$. For example, if $\pi = 4213 \in \mathfrak{S}_4$ and $\sigma = 4231 \in \mathfrak{S}_4$, then the graph has the edge (π, σ) since $4213 \cdot s_3 = 4231$ and s_3 is an adjacent transposition. The undirected version of this bi-directed graph is realized as the 1-skeleton graph of the permutahedron.

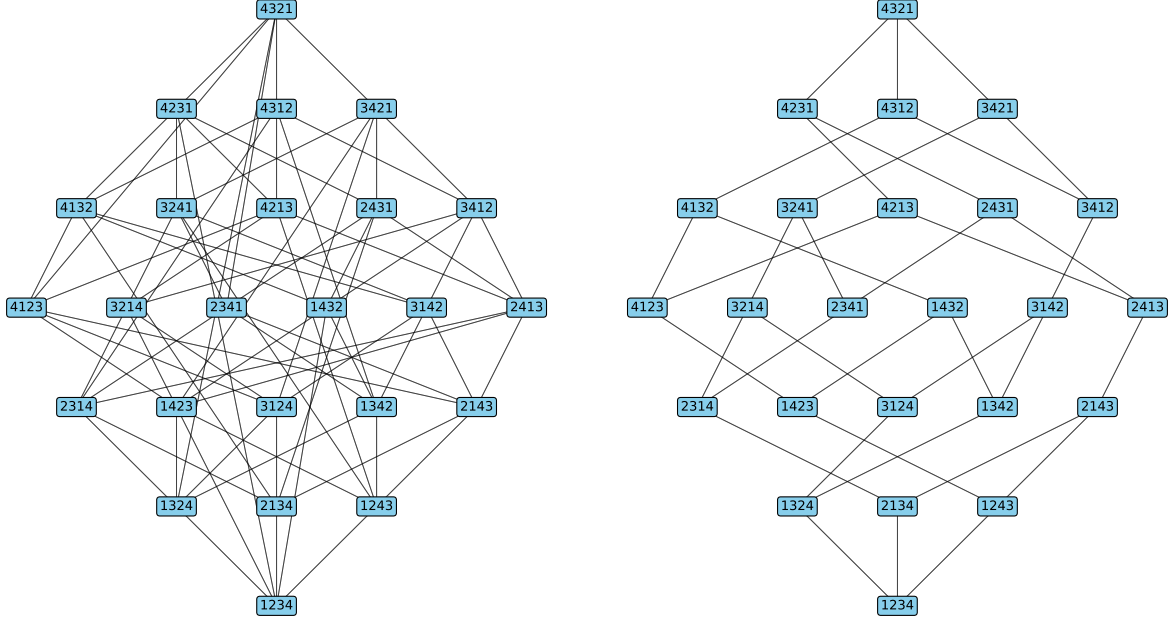
The bi-directed graph in Figure 1b illustrates the commonly-used fact that adjacent transpositions suffice to generate \mathfrak{S}_n . However, as we explain in Section 3, the more edge-dense class of graphs illustrated in Figure 1a is better suited for the purposes of this work. The main properties of this class of graphs that we leverage in this work follow from Fact 2.3 and are summarized below.

Remark 2.6. *For $n > 1$, the bi-directed version of the Bruhat graph on \mathfrak{S}_n is strongly-connected and has diameter $n - 1$.*

Refer to any abstract algebra textbook for more on the symmetric group.

2.2. Markov Chains. This section is based on Levin and Peres [12]. A discrete-time Markov chain is a sequence X_1, X_2, X_3, \dots of random variables with the Markov property. That is,

$$\Pr[X_{t+1} = x | X_1 = x_1, X_2 = x_2, \dots, X_t = x_t] = \Pr[X_{t+1} = x | X_t = x_t]$$



(A) Generated by transpositions.

(B) Generated by adjacent transpositions.

FIGURE 1. Bi-directed graphs on \mathfrak{S}_4 , with nodes as permutations in one-line notation. (Bi-directed arcs are depicted as undirected edges.)

for all $t \in \mathbb{N}$ whenever the conditional probabilities are well-defined. If the random variables share a finite co-domain, then the chain has a finite state space—say, indexed by $[m]$ for some $m \in \mathbb{N}$. The chain is time-homogeneous if

$$\Pr[X_{t+1} = x | X_t = y] = \Pr[X_t = x | X_{t-1} = y]$$

for all $t \in \mathbb{N}$, so that the transition probabilities are independent of t .

In the finite and time-homogeneous case, the transition probabilities can be represented by a row-stochastic matrix $P \in \mathbb{R}^{m \times m}$ where

$$P_{ij} = \Pr[X_{t+1} = j | X_t = i]$$

for all $i, j \in [m]$. In effect, P encodes the transition probabilities of a random walk on a directed graph $G_m = ([m], A)$, where $(i, j) \in A$ if and only if $P_{ij} > 0$. Specifically, $X_t \in [m]$ corresponds to the node visited during the t th step and, conditioned on $X_t = i$, the i th row of P governs the random transition to the $(t + 1)$ th step.

The chain P is *irreducible* if, for any $i, j \in [m]$, there exists some $\tau \in \mathbb{N}$ such that $P_{ij}^\tau > 0$. For any $i \in [m]$, its *period* is defined as the greatest common divisor of $\{t \in \mathbb{N} : P_{ii}^t > 0\}$. The chain is *aperiodic* if $\gcd(\{t \in \mathbb{N} : P_{ii}^t > 0\}) = 1$ for all $i \in [m]$. Note that $P_{ii} > 0$ for all $i \in [m]$ is a sufficient condition for aperiodicity.

The following are classical results in the theory of Markov chains.

Theorem 2.7. *If P is an irreducible finite Markov chain, then there exists a unique probability distribution $\mu \in \mathbb{R}^m$ over $[m]$ with $\mu = \mu P$. The distribution μ is referred to as the stationary distribution of P .*

Theorem 2.8 (Convergence Theorem). *If P is an irreducible and aperiodic finite Markov chain with stationary distribution μ , then there exist constants $\alpha \in (0, 1)$ and $C > 0$ such that*

$$\max_{i \in [m]} \|(P^t)_i - \mu\|_{\text{TV}} \leq C\alpha^t,$$

where $(P^t)_i$ is the i th row of P^t and $\|\cdot\|_{\text{TV}}$ is the total variation distance.

Theorem 2.8 gives analytical grounds for the repeated application of an irreducible and aperiodic finite Markov chain to sample from its stationary distribution.

2.3. Sports Rankings and Partially Ordered Sets. A *partially ordered set* (poset) is a pair $\pi = ([n], \preceq_\pi)$ consisting of a ground set $[n]$ and a partial order \preceq_π satisfying:

- (1) *Reflexivity:* For any $i \in [n]$ we have that $i \preceq_\pi i$.
- (2) *Antisymmetry:* For any $i, j \in [n]$ we have that if $i \preceq_\pi j$ and $j \preceq_\pi i$, then $i = j$.
- (3) *Transitivity:* For any $i, j, k \in [n]$ we have that if $i \preceq_\pi j$ and $j \preceq_\pi k$, then $i \preceq_\pi k$.

For $i, j \in [n]$, let $i \prec_\pi j$ if $i \preceq_\pi j$ and $i \neq j$. An element $i \in [n]$ is *maximal* if there is no $j \neq i \in [n]$ with $i \prec_\pi j$. The definition of *minimal* is analogous.

A poset may be represented pictorially by a Hasse diagram, which is a drawing of its transitive reduction. For example, if the edges in the graphs in Figure 1 were all oriented upwards, we would obtain Hasse diagrams of two different posets on the symmetric group of order 4, both with unique minimal element 1234 and unique maximal element 4321.

A totally ordered set is a poset in which every two elements are comparable, and a chain is a subset of a poset that is totally ordered for the induced order (note that the use of the word “chain” in the context of a poset differs from its use in Section 2.2). For example, $\{1234, 1243, 2143, 2413, 2431, 4231, 4321\}$ forms a chain in Figure 1b. An *anti-chain* is a subset of elements such that no pair of elements are comparable. For example, $\{1324, 2134, 1243\}$ forms an anti-chain in Figure 1b. A poset $\pi^* = ([n], \preceq_{\pi^*})$ is an extension of $\pi = ([n], \preceq_\pi)$ if, for any $i, j \in [n]$ with $i \preceq_\pi j$, we have that $i \preceq_{\pi^*} j$ as well. An extension is a linear extension if it is totally ordered. Refer to Stanley [19, Chapter 3] for more on posets.

Let $[n]$ represent a set of players. A *ranking* is a totally ordered set $\pi = ([n], \preceq_\pi)$, which we treat as a permutation $\pi \in \mathfrak{S}_n$ (in explicit notation) given the natural ordering \leq of \mathbb{N} . In particular, we have that

$$j \preceq_\pi i \iff \pi(i) \leq \pi(j)$$

for all $i, j \in [n]$. We more generally extend the notion of ranking to proper subsets of players, so that a ranking is a totally ordered set $r = (S_r, \preceq_r)$ for some $S_r \subseteq [n]$, which we again treat as a bijection $r : S_r \rightarrow [|S_r|]$ given the natural ordering \leq of \mathbb{N} . In this way, our problem input is a collection

$$\mathcal{R} = \{r_1, r_2, \dots, r_\ell\}$$

of historical sports rankings on subsets of players. In subsequent sections we will consider partial orders on the full set of players $[n]$ together with their corresponding linear extensions as part of our problem output.

Lastly, note that we adopt the following notation. For $i \in [n]$, let $\mathcal{R}_i = \{r \in \mathcal{R} : i \in S_r\}$ denote the sub-collection of rankings containing player i . Similarly, for distinct $i, j \in [n]$, let

$\mathcal{R}_{ij} = \{r \in \mathcal{R} : i, j \in S_r\}$ denote the sub-collection of rankings containing both player i and player j . By convention, let $\mathcal{R}_{ii} = \emptyset$.

3. DATA-DRIVEN RANDOM WALKS ON THE SYMMETRIC GROUP

In Section 3.1 we discuss the meaning of GOAT in relation to assumptions (in particular, ranking cutoffs) on the problem input. In Section 3.2 we design a data-driven transition rule for random walks on \mathfrak{S}_n . In Section 3.3 we adopt a notion of stochastic dominance to build a poset from the stationary distribution of this random walk. In Section 3.4 we discuss sampling procedures for this random walk.

3.1. Cutoffs and the meaning of GOAT. Recall that $[n]$ represents the set of players and that the problem input is a collection

$$\mathcal{R} = \{r_1, r_2, \dots, r_\ell\}$$

of historical sports rankings on subsets of the players, so that $\cup_{r \in \mathcal{R}} S_r = [n]$. Now, consider any player $i \in [n]$ and any ranking $r \in \mathcal{R}_i$. Certainly, player i being “top ranked” in r is consistent with them being the GOAT, but how “top” is top ranked? A full sports ranking might involve thousands of players. However, player i being among the top 10 in r is much more indicative than them being among the top 100, which is itself even much more indicative than them being among the top 1000. In some sense, we must adopt a definition for the concept of being at the top.

In the experimental part of this work, we restrict the rankings in \mathcal{R} to involve the top 3, 5, 10, and 20 players as cutoffs from the full ranking data that might be available. We do this for two reasons:

- (1) The first reason is conceptual. As argued previously, the difference between a player being ranked 100 or 101 (let alone 1000 or 1001) is not particularly relevant in relation to them being the GOAT, whereas the difference between them being ranked 1 or 2 very much is. By experimenting with different cutoffs we gain insights into the consequences of this subjective choice.
- (2) The second reason is computational. Indeed, if each ranking $r \in \mathcal{R}$ involved thousands of players, the set $[n]$ would be very large and $n!$ would be astronomically large. In our experiments, the different cutoffs always keep the size of the set under 300 players. Note that 300! is still an astronomically large number, with 615 decimal digits.

In short, we assume that being the GOAT is consistent with being a “good top κ ” player, and we experiment with the choice of $\kappa \in \mathbb{N}$. In this way, having ever made it to the top κ is a necessary condition for being the GOAT, which restricts the set of contenders.

3.2. Transition Rule. Let $W \in \mathbb{R}_{\geq 0}^{n \times n}$ be a matrix with entries

$$W_{ij} = 1 + \underbrace{|\mathcal{R}_i| / |\mathcal{R}_j|}_{\text{prevalence ratio}} \cdot \underbrace{|\{r \in \mathcal{R}_{ij} : j \prec_r i\}|}_{\text{\# of times } i \text{ beats } j} \quad (1)$$

for all distinct $i, j \in [n]$. The second term in the product, $|\{r \in \mathcal{R}_{ij} : j \prec_r i\}|$, encodes the number of (cut off) historical rankings in which player i is ranked better than player j . We scale this quantity by the first term in the product, $|\mathcal{R}_i| / |\mathcal{R}_j|$, which is the players’ relative prevalence above the cutoff. The motivation for this is as follows.

Remark 3.1. Suppose players i and j have short and long prevalence in the (cut off) historical rankings, respectively. In particular, it might be that player i made it to the top of the ranking for a brief period of time, consequently beating player j during this period. However, since the prevalence of player i is short, there might be significant periods of time during which player i fell below the cutoff while player j did not—by definition, player j beat player i during these periods. This artifact of cutoffs is not captured by the term $|\{r \in \mathcal{R}_{ij} : j \prec_r i\}|$, which by itself favors player i over player j . We correct for this in the model using their prevalence ratio $|\mathcal{R}_i| / |\mathcal{R}_j|$.

Intuitively, the term $|\mathcal{R}_i| / |\mathcal{R}_j|$ ensures that if the players have similar prevalence above the cutoff, the data points in which player i beats player j are taken at face value. Conversely, if the prevalence of player i is much smaller than that of player j , the data points in which player i beats player j are proportionally discounted. Finally, the additive unit is included to later on obtain an irreducible random walk. Note that (1) can be pre-computed efficiently.

Next, for each $\pi \in \mathfrak{S}_n$ and each $1 \leq i < j \leq n$, let

$$\mathcal{W}(s_{ij}|\pi) \propto \begin{cases} W_{ji}, & \text{if } j \prec_\pi i \\ W_{ij}, & \text{otherwise} \end{cases} \quad (2)$$

form a probability distribution over (not necessarily adjacent) transpositions. The interpretation of $\mathcal{W}(s_{ij}|\pi)$ is as follows:

- $j \prec_\pi i \implies i \prec_{s_{ij} \cdot \pi} j$, whereas
- $i \prec_\pi j \implies j \prec_{s_{ij} \cdot \pi} i$.

In other words, the probability of selecting transposition s_{ij} given π is proportional to the agreement between the resulting permutation $s_{ij} \cdot \pi$ and the historical data, particularly with respect to the relative order between players i and j .

Remark 3.2. Note that (1)-(2) highlight the need to consider the set of all transpositions, not just those that are adjacent (recall the difference in Figure 1). Indeed, depending on the structure of \mathcal{R} and the indexing of the players, it is possible that

$$|\mathcal{R}_{ij}| = 0$$

for a large fraction of the adjacent choices of $1 \leq i < j(=i+1) \leq n$, in turn rendering (1) uninformative. For example, this could occur if the players are indexed in such a way that adjacent players generally cannot be compared pairwise because of their prevalence during non-overlapping time periods.

Next, we use (2) to compactly define a row-stochastic matrix $P \in \mathbb{R}_{\geq 0}^{n! \times n!}$ such that, for any given $\pi \in \mathfrak{S}_n$, we have the transition probabilities

$$P_{\pi, \sigma} = \begin{cases} \frac{1}{2} \cdot \mathcal{W}(s_{ij}|\pi), & \text{if } \sigma = s_{ij} \cdot \pi \text{ for some } 1 \leq i < j \leq n \\ 1/2, & \text{if } \sigma = \pi \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

The matrix P governs a data-driven random walk on \mathfrak{S}_n . Note that its support along any given row is of size $O(n^2)$ and can be evaluated efficiently on the fly.

Lemma 3.3. Given (3), P is irreducible and aperiodic.

Proof. The additive 1 term in (1) implies the support graph of P is the bi-directed Bruhat graph, with the addition of self-loops. Fact 2.3 implies this graph is strongly connected, so that P is irreducible. Moreover, the non-zero probability of a lazy step in the second case of (2) implies $P_{\pi,\pi} > 0$ for all $\pi \in \mathfrak{S}_n$, so that P is aperiodic. \square

Lemma 3.3 and Theorem 2.7 directly imply the following.

Corollary 3.4. P has a unique stationary distribution μ over \mathfrak{S}_n .

3.3. Partial Ordering. In light of Corollary 3.4, consider the stationary distribution μ of P . We combine this distribution with a suitable notion of stochastic dominance to obtain a partial order $([n], \preceq_\mu)$ over the players. In particular, for each $i, j \in [n]$, let

$$j \preceq_\mu i \iff \Pr_{X \sim \mu} [X(i) \leq k] \geq \Pr_{X \sim \mu} [X(j) \leq k], \forall 1 \leq k \leq n. \quad (4)$$

Recall that μ is a distribution over \mathfrak{S}_n , so that X is a random permutation (in explicit notation) and $X(i) \in [n]$ denotes the ranking of player $i \in [n]$. Therefore, at an intuitive level, (4) states that player i is conclusively “better” than player j in \preceq_μ if and only if, given that $X \sim \mu$, player i is across the board more likely to rank better than player j .

Let $\mathcal{L}(\preceq_\mu)$ denote the set of linear extensions of \preceq_μ (refer to Section 2.3). In the experimental part of this work, we will consider the average rank of a player with respect to a random linear extension as a summary statistic of the player’s career performance. In particular, for each $i \in [n]$ we will consider

$$\bar{X}(i) = \frac{1}{N} \sum_{t=1}^T X_t(i), \quad (5)$$

for some $T \in \mathbb{N}$ and where X_1, X_2, \dots, X_T are sampled uniformly at random from $\mathcal{L}(\preceq_\mu)$.

3.4. Sampling. Our next goal is to obtain independent samples from the stationary distribution μ induced by (3) to form an empirical poset according to (4). Based on Theorem 2.8, this can be achieved through the repeated execution of (3). At this point, the question of speed of convergence comes to the forefront. If we knew the mixing time $\tau \in \mathbb{N}$ of this Markov chain for a desired tolerance to error, we could execute the random walk collecting every τ th sample. However, based on the discussion in Section 1.2, we anticipate this is a very challenging theoretical question. Neither is it easy to obtain numerical upper bounds based on the eigenvalues of the transition matrix P due to its $n! \times n!$ size.

Given the sports analytics scope of this work, we do not attempt to answer these questions and simply acknowledge that analytical questions around data-driven random walks on the symmetric group are of independent mathematical interest. In the experimental part of this work, we execute this random walk for a large number of steps collecting every $2(n-1)$ th sample. Given the $n-1$ diameter noted in Remark 2.6 and the $1/2$ probability of taking a lazy step in (3), this is the expected number of steps needed for the entire state space to be reachable from any given state.

4. NUMERICAL IMPLEMENTATION

In this section, we implement our methods using singles rankings from the “Open Era” of the ATP and the WTA. As discussed in Section 3.1, we experiment with taking the top 3, 5, 10, and 20 players as different cutoffs from the full ranking data that is available.

For each experiment, we execute a random walk as described in Section 3.4 until we collect 100,000 samples. We then use these samples to obtain data-driven posets over the players, as described in Section 3.3. We illustrate these posets as horizontal Hasse diagrams with maximal elements on the right-hand side (the full diagrams can be found in the Appendix). To estimate the average rank of the players in a given poset, we first sample 100,000 linear extensions uniformly at random using SageMath [17] and then follow [5].

The remainder of this section is organized as follows. In Section 4.1 we further describe the datasets we use for our experiments. In Section 4.2 we present our results using rankings from the ATP, whereas in Section 4.3 we present our results using rankings from the WTA. Refer to Tables 1 and 2 for a summary of some of the results. Lastly, in Section 4.4 we compare the results obtained for the ATP and the WTA to note qualitative differences between the two associations. All mentioned statistics that are external to our numerical implementation can be verified online [3, 22].

4.1. Ranking Data. Both the ATP and the WTA singles rankings are based on rolling 52-week period systems that award points to individual players according to their placement in different tournament categories. Rankings are typically published on Mondays, and points are dropped 52 weeks after being awarded. Note that the point systems have changed over the years: details on the current systems are available online in the official rulebooks [2, 21].

We retrieved historical rankings from the repositories compiled by Sackmann [15, 16] for both the ATP and the WTA. In the case of the ATP, this corresponds to 2230 weekly rankings from 1973 to 2023 (Sackmann notes that this data is complete from 1985, but that it is intermittent from 1973 to 1984 and that it is missing 1982). In the case of the WTA, this corresponds to 2052 weekly rankings from 1984 to 2023.

Naturally, the limitations of the data we use induce limitations on our experimental results. Aside from missing data points, we in particular do not consider the “Amateur Era” and only partially capture players that were active during the the onset of the “Open Era” rankings. As a concrete example, we do not consider Margaret Court; a widely successful player in women’s professional tennis that was active from 1960 to 1977. This data availability issue speaks to the difficulty of capturing the Amateur Era due to the lack of a centralized ranking system.

4.2. Rankings from the ATP. In Figure 2 we plot the top portion of the Hasse diagram of the poset over “top 3 players,” where Djokovic, Federer, Nadal, and Sampras are incomparable maximal elements—no one of these players is conclusively better than another. Note also that they all have a similar average rank. Conversely, we conclude that each of Djokovic, Federer, Nadal, and Sampras is better than Connors and Lendl as a top 3 player. Similarly, as top 3 players, Connors is better than Andre Agassi, who is better than Björn Borg, who is better than Andy Murray, who is better than Carlos Alcaraz. In other words, these players form a chain in the poset. Formally, this is

$$\text{Djokovic} \succ_{\mu} \text{Connors} \succ_{\mu} \text{Agassi} \succ_{\mu} \text{Borg} \succ_{\mu} \text{Murray} \succ_{\mu} \text{Alcaraz}.$$

This is also reflected by significant jumps in the average rank from one player to the next. Interestingly, Alcaraz is near the top of the Hasse diagram despite turning professional most recently, in 2018.

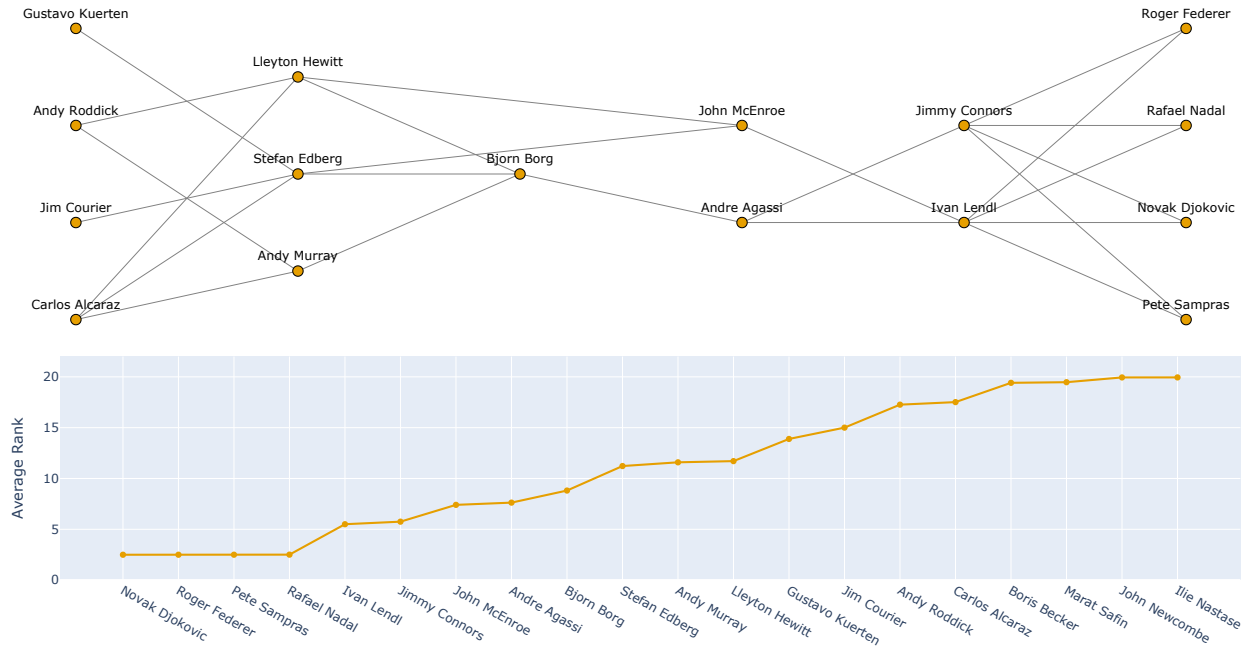


FIGURE 2. Cutoff at top 3 of full ATP rankings.

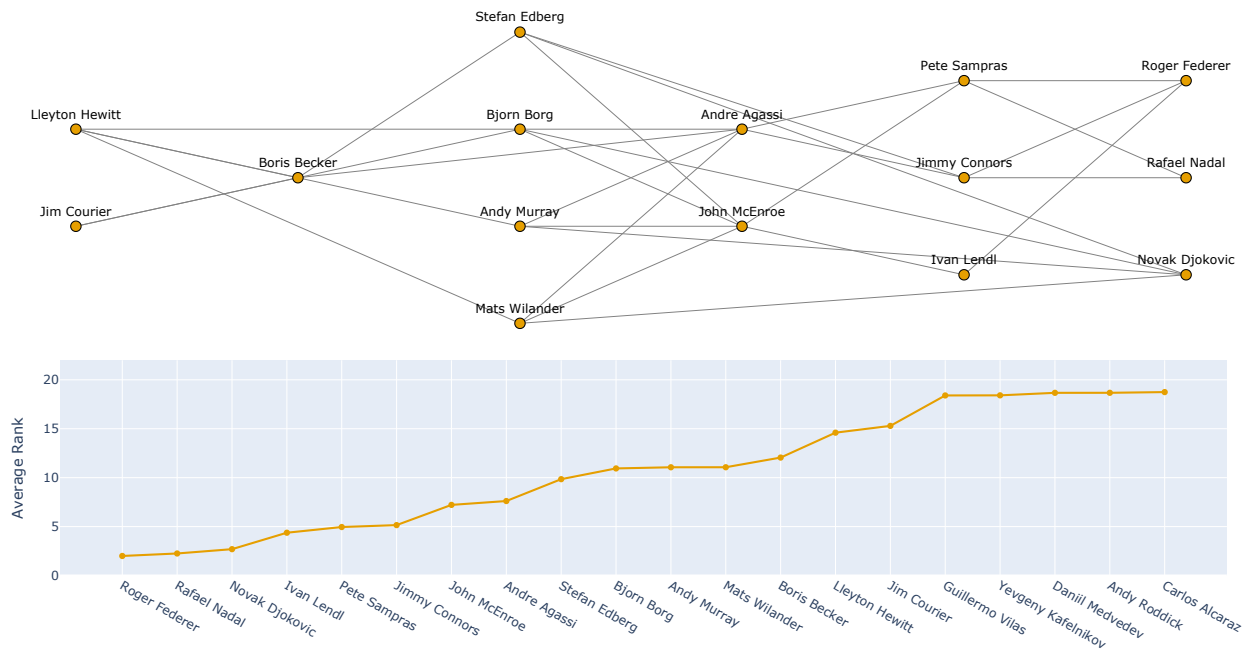


FIGURE 3. Cutoff at top 5 of full ATP rankings.

In Figure 3 we plot the top part of the Hasse diagram of the poset over “top 5 players,” where Djokovic, Federer and Nadal remain incomparable maximal elements, although Federer fares slightly better than the other two in terms of average rank. However, Sampras is

no longer a maximal element: as a top 5 player, he is worse than both Federer and Nadal, and he is incomparable to Djokovic. Sampras is moreover incomparable to Connors and Lendl (both of whom he was better than as a top 3 player). Note that Djokovic and Connors are incomparable as top 5 players: their difference in average rank favors Djokovic, but it is not sufficiently conclusive. Moreover, note that Alcaraz no longer appears near the top of the Hasse diagram, having been displaced by other players who are better top 5 players than they are top 3 players, such as Guillermo Vilas (this is reflected in the average rank plot). First, note that Alcaraz and Vilas did not overlap in competition: Vilas played professionally from 1968 to 1992. Second, note that Alcaraz transitioned very quickly from below the top 20 to within the top 3 from February to September of 2022, and has consistently prevailed within the top 3 until the time of writing. His prevalence within the top 3 is significant independently of his short career at present. However, when the cutoff is relaxed to top 5, his short career-to-date puts him at a disadvantage compared to retired players with significant career-long prevalence in the top 5 (but perhaps not in the top 3), such as Vilas.

In Figure 4 we plot the top part of the Hasse diagram of the poset over “top 10 players,” where Djokovic, Federer, and Nadal once again remain incomparable maximal elements. However, unlike the previous cutoff, we conclude that Djokovic is better than Connors as a top 10 player. Qualitatively, we observe the inclusion of more players near the top of this Hasse diagram compared to the previous cases, indicating that there are more “good top 10 players” than there are good top 3 or top 5 players. For example, Becker (who was far from maximal in Figures 2 and 3) is now close to maximal: only Federer and Nadal are better than him, and he is incomparable to Djokovic.

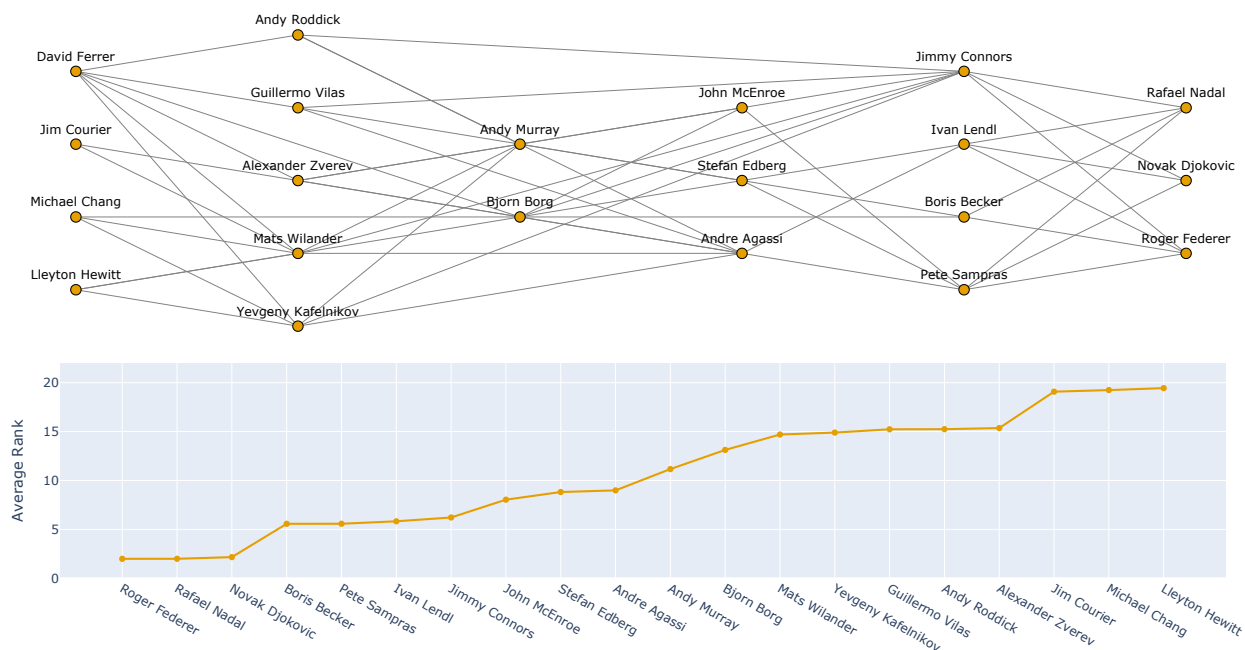


FIGURE 4. Cutoff at top 10 of full ATP rankings.

Lastly, in Figure 5 we plot the top part of the Hasse diagram of the poset over “top 20 players.” Interestingly, only Federer and Nadal remain incomparable maximal elements.

Despite this, only Federer is conclusively better than Djokovic as a top 20 player—Djokovic and Nadal are incomparable as top 20 players. Qualitatively, we again observe the inclusion of even more players near the top of this Hasse diagram compared to the previous cases. For instance, the Djokovic \succ_μ Connors \succ_μ Agassi chain from Figure 2 is broken in Figure 5, where all three of these players are equally close to the top of the Hasse diagram and form an anti-chain.

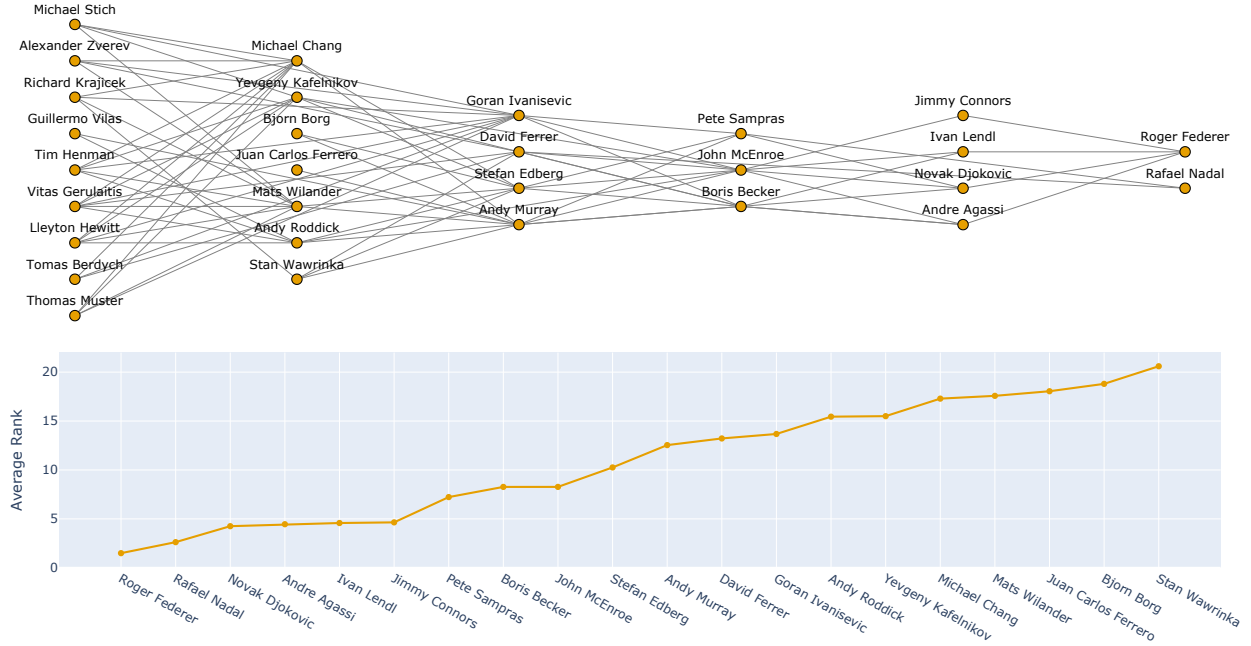


FIGURE 5. Cutoff at top 20 of full ATP rankings.

4.3. Rankings from the WTA. In Figure 6 we plot the top portion of the Hasse diagram of the poset over “top 3 players,” where Graf and S. Williams are incomparable maximal elements—no one of these players is conclusively better than the other. Note also that they both have a similar average rank, albeit with a slight advantage for Graf. Note that S. Williams and Martina Navratilova are incomparable. On the other hand, we conclude that, as top 3 players, Graf is better than Navratilova, who is better than Justine Henin, who is better than Aryna Sabalenka. In other words, these players form a chain in the poset. Formally, this is

$$\text{Graf} \succ_\mu \text{Navratilova} \succ_\mu \text{Henin} \succ_\mu \text{Sabalenka}.$$

We also conclude that, as a top 3 player, Sabalenka, Kim Clijsters, and Maria Sharapova are pairwise incomparable (i.e., they form an anti-chain).

In Figure 7 we plot the top portion of the Hasse diagram of the poset over “top 5 players,” where Graf and S. Williams are again incomparable maximal elements. Note that while Graf is better than Navratilova, S. Williams and Navratilova are incomparable. Moreover, note that unlike the case in Figure 6, as top 5 players, both Clijsters and Sharapova are better than Sabalenka. This is a similar effect to that seen with Alcaraz in Section 4.2: Sabalenka turned professional in 2015 and had little overlap in competition with Sharapova,

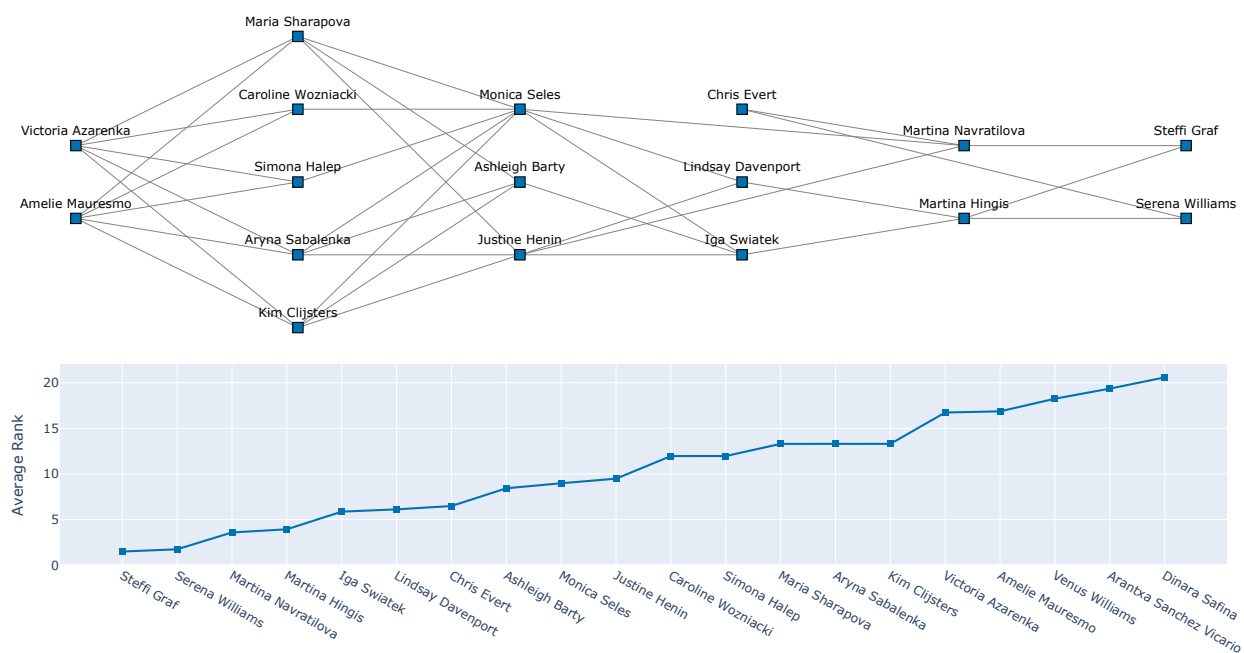


FIGURE 6. Cutoff at top 3 of full WTA rankings.

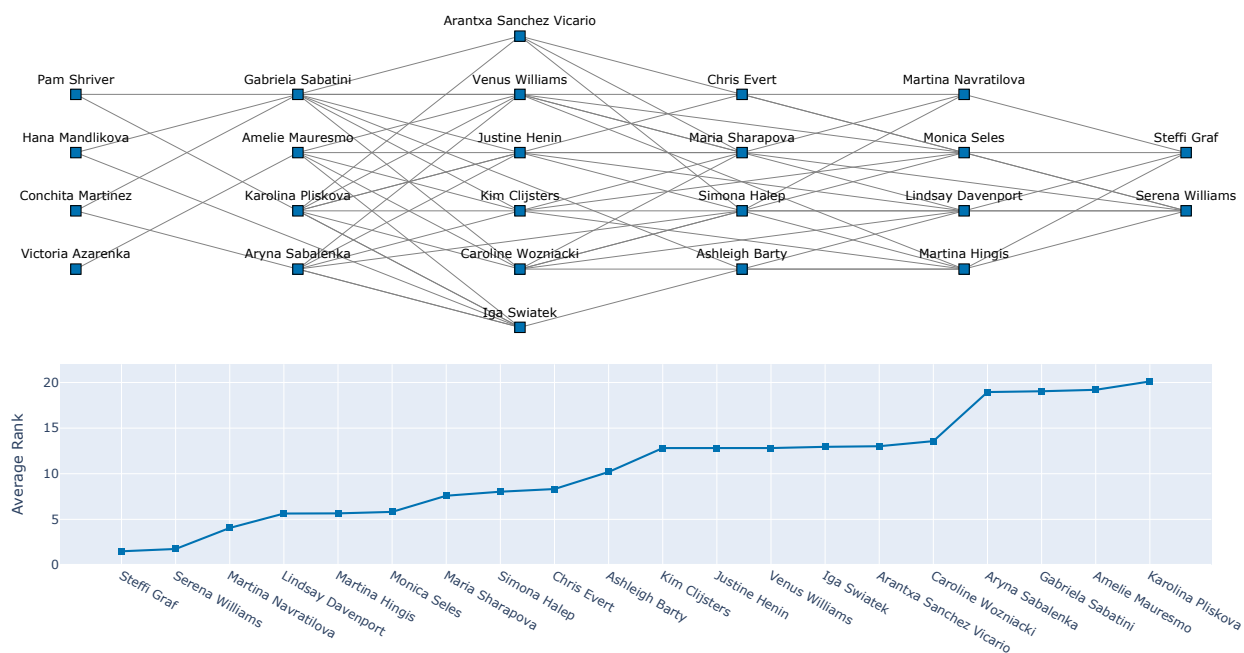


FIGURE 7. Cutoff at top 5 of full WTA rankings.

who retired in 2020, and Clijsters, who retired in 2012 except for a brief comeback between 2020 and 2022. In particular, her prevalence within the top 3 is significant independently of her short career at the time of writing. However, when the cutoff is relaxed to top 5, her

short career-to-date puts her at a disadvantage compared to long-established players with significant career prevalence in the top 5, such as Clijsters and Sharapova. Qualitatively, we observe the inclusion of more players near the top of this Hasse diagram compared to the previous cases, and that chains of players that were comparable as top 3 players become anti-chains and close to maximal as top 5 players (e.g., Lindsay Davenport and Seles).

In Figure 8 we plot the top portion of the Hasse diagram of the poset over “top 10 players,” where now the three of Graf, Navratilova, and S. Williams are incomparable maximal elements. In other words, while Navratilova is close to but not quite the GOAT as a top 3 or top 5 player, she is conclusively one of the GOATs as a top 10 player. As a top 10 player, Navratilova is better than Seles. However, Graf and S. Williams cannot be conclusively compared to Seles (even though they fare slightly better than her in terms of average rank).

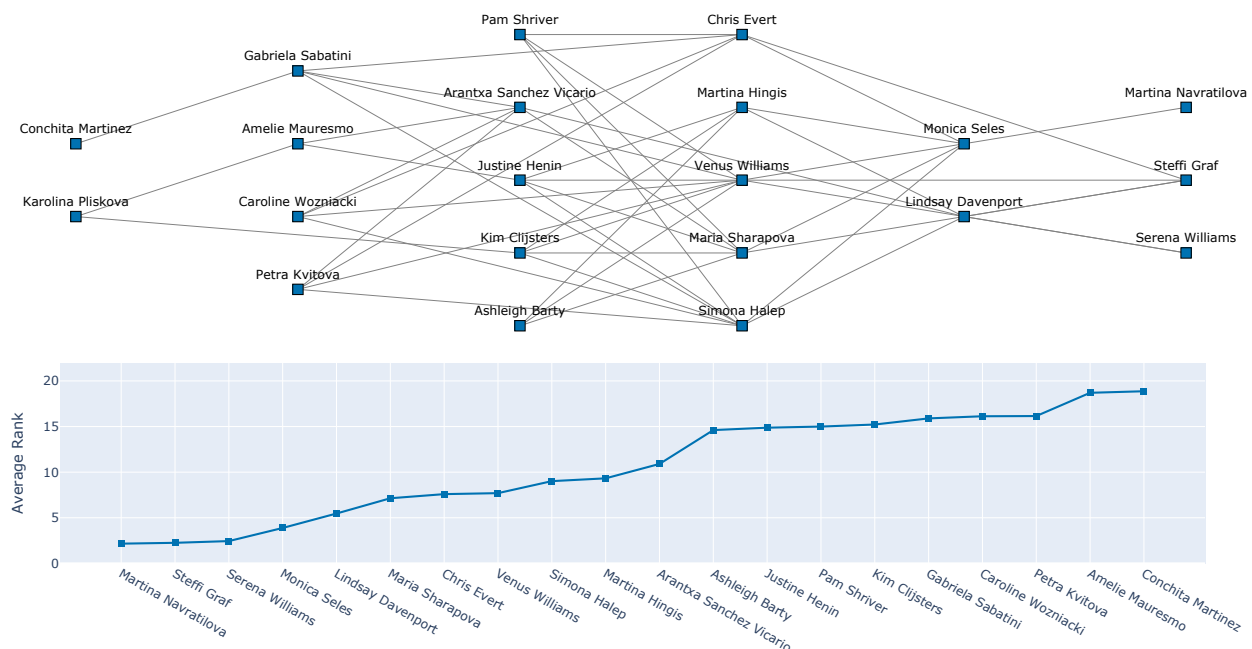


FIGURE 8. Cutoff at top 10 of full WTA rankings.

Lastly, in Figure 9 we plot the top portion of the Hasse diagram of the poset over “top 20 players,” where now all six of Graf, Navratilova, Seles, S. Williams, V. Williams, and Wozniacki are incomparable maximal elements. Qualitatively, we observe a significant and across the board increase in the number of players near the top of the Hasse diagram. This indicates that, in the WTA, there are many more good top 20 players than there are good top 3, 5, or 10 players.

4.4. Comparing the Associations. Figures 5 and 9 are qualitatively different in that, as top 20 players, the number of players close to maximal is much larger for the WTA than it is for the ATP. Comparing the results from Sections 4.2 and 4.3, this difference is not nearly as apparent for the posets over top 3, 5, or 10 players. Here, we further investigate this difference from a more quantitative point of view. In particular, we ask questions of

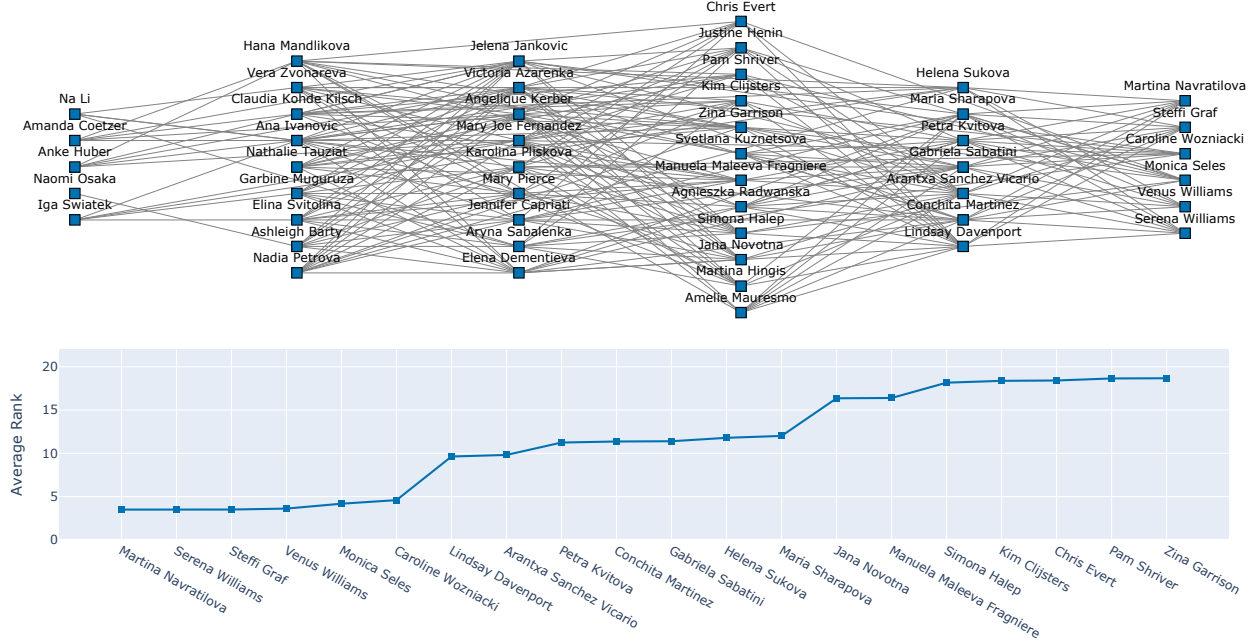


FIGURE 9. Cutoff at top 20 of full WTA rankings.

the form: to what extent is being “a good top κ player” a predictor of being “a good top κ' player,” where $\kappa' \leq \kappa$? We formalize the notion of “a good top κ player” by average rank (from the linear extensions) for the poset cutoff κ . Intuitively, the smaller the difference between κ and κ' , the stronger we expect the correlation to be.

In Figure 10, we consider the set of players that appear among the first 25 by average rank for *all* cutoffs $\kappa = 3, 5, 10, 20$. This amounts to 15 players in both the ATP and in the WTA. Then, for each of these sets of 15 players, we plot a matrix of scatter plots of average rank and varying cutoffs. In this way, each point represents a player, and its coordinates correspond to the player’s average rank for a combination of cutoffs. For each scatter plot, we also fit a linear model and report its coefficient of determination r^2 .

From Figure 10 we note that, in both associations, being a good top 5 player is a reasonable predictor of being a good top 3 player, as is being a good top 10 player a reasonable predictor for being a good top 5 player. In both cases, being a good top 10 player is still correlated with being a good top 3 player, although much less so than with being a good top 5 player. However, when we consider being a good top 20 player as a predictor, we observe a major difference between the two associations. In the ATP, being a good top 20 player as a predictor is correlated with being a good top 10, 5, and 3 player (with decreasing correlation in that order). Conversely, in the WTA, being a good top 20 player is only mildly correlated with being a good top 10 player, and it is nearly uncorrelated with being a good top 5 or top 3 player. This quantitative finding highlights a major difference in terms of career and dominance longevity for top players across the two associations. While we do not delve into causality in this work, we believe it would be practically important to further study the reasons and broader implications of this phenomenon.

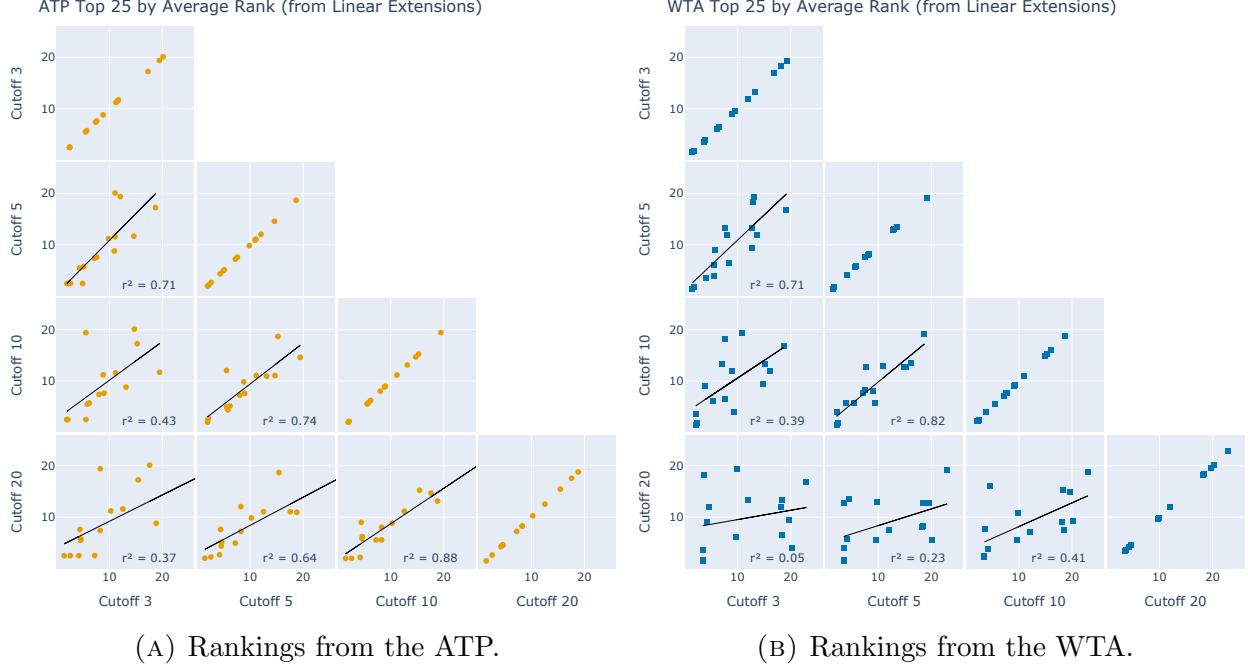


FIGURE 10. Average ranks (from linear extensions) with varying cutoffs.

5. CONCLUSION

In this work, we proposed a mathematical framework to compare players, and in particular players spanning across different time periods in sports history. We aggregated cutoffs of historical ranking data into a single poset over top players by *i*) sampling from a data-driven random walk over the symmetric group and *ii*) adopting a notion of stochastic dominance. Ultimately, some pairs of players can be conclusively compared while others cannot. This approach departs from existing methods in that it relies on ranking data (rather than match-level data, which may or may not exist/be available) and formally comes to terms with the possibility that some comparisons are “too close to call.” The two steps in our framework leave ample room for further modeling choices, namely by adapting the construction of (1) and (4) as needed for the particular application in mind. These applications also motivate the analytical study of data-driven sampling of permutations.

We implemented our methods using data from both the ATP and the WTA to find the GOATs in the respective categories. Our experimental results suggest that Nadal, Federer, and Djokovic are the ones that tend to come ahead as GOATs of the ATP under most circumstances, whereas Graf and S. Williams are the ones that do so in the case of the WTA. Through our experiments, we also noted major differences in terms of career and dominance longevity for top players across the two associations.

ACKNOWLEDGEMENTS

J. C. Martínez Mori would like to thank Patrick Kastner for computational resources and Amanda Priestley for a helpful discussion. Part of this research was performed while J. C. Martínez Mori was visiting the Mathematical Sciences Research Institute (MSRI), now becoming the Simons Laufer Mathematical Sciences Institute (SLMath), which is supported