Competitive Networks for Individual Sports

Sean Strong, Joseph Taglic, Liyang Sun

https://github.com/LiyangSun/Analysis-of-Individual-Sport-Competition-Networks

Abstract—This paper introduces the concept of individual competitive networks – a unique model for understanding competitive individual sports – and analyzes the properties of these networks in the context of fencing, tennis, and chess.

After a quick review on the relevant mathematical and algorithmic backgrounds, we present our findings and analysis, outline our encountered difficulties, and detail exciting areas for future research.

I. INTRODUCTION

With the rapid development of the internet, social media, and computing infrastructure, social network analysis has become increasingly popular. However, this field has also shown a lot of promising results for a much broader range of subjects, including as biology or criminology. What about sports?

Social network analysis has only been recently introduced to the study of sports, with only a handful of relevant research papers. Of these, all are about team sports rather than individual sports. One obstacle to network analysis in sports seems to be the data collection process. Detailed and specific data about sports can be hard to get, as experts are needed and the data collected for now depend really on the sport type and on the level at which it is played.

However, network analysis in this field has a lot of room for growth: many social network analysis methods are applicable to sport disciplines, and new predictive models can be developed based on competitive network models, leading to a deeper understanding of competitive dynamics across all sports.

Exploring the characteristics of individual sports or competitions poses an interesting challenge in a very visible field. Analyses could provide meaningful insights to various interested parties within the sports industry – competitors, coaches, spectators, and bookies alike.

For instance, can a given sport's competition network be insightful for evaluating its ranking system or level balance? Social network analysis can help us identify competition structures within individual sports, explaining - and hopefully predicting - key phenomena such as parity and variance in both overall and individual results.

In this paper, we present an overview as to how social network analysis can be used to study individual sports' competition results. More specifically, we look at network dynamics within one sport, between different sports, over time, and as a tool for outcome prediction.

We chose to focus on individual sports instead of team sports, as there are more competitors and therefore data points relative to team sports. Additionally, analysis of individual competitors removes the complications of players joining or leaving teams. Moreover, individual competition analysis is of personal significance, as one of our authors is a competitive fencer, himself.

II. RELATED WORK

Social network analysis has already been explored in the context of team sports, namely basketball, football and handball. While Korte and Lames characterized different team sports and their tactical positions in paper [2], Grund (in paper [3]), and Vaz de Melo, Almeida and Loureiro (in paper [4]) tried to assess teams' performance based on the individual performance and interactions of their players.

In paper [2], a player-interaction network was built for each team, based on several matches: nodes represent players and weighted directed edges represent the number of passes from one player to an other player. From this, various centrality metrics were computed, each having a definite meaning for the performance of each player: individual metrics, such as *weighted in-degree* (number of successfully received passes by a player) or *weighted betweenness* (how often a player is on a shortest path between other players), as well as team metrics, such as *weighted in-degree* centralization (indicator for the balance of direct interplay).

By emphasizing strong connections between each tactical positions using *minimum spanning tree* (subset of the edges of the graph that connects all the vertices together, without any cycles and with the minimum possible total edge weight), Korte and Lames were thus able to find the most centralized roles in basketball (point guard), football/soccer (defensive midfielder) or handball (center), and get "network translations" of the nature of different sports.

In paper [3], the same network structure and metrics were used, however for football teams only. The goal of the study was also different, as Grund tried to see how interactions between team members could impact on the team's overall performance. Its main differences with paper [2] were thus the statistical methods used, which will not be discussed here, as we mainly focus on network analysis methods.

Grund managed to support two hypothesis, which are: intense relationships between players (network density) increases team performance, and too much reliance on a small subset of players (high network centrality) decreases performance.

Paper [4]'s goal was to evaluate how individual performance relates to team performance. Using the example of NBA drafting, each NBA player is evaluated according to box score statistics (assists, points, ...), but is this individual performance really representative of his/her influence on the team performance?

The authors built networks for each year and also time cumulative networks, with players and teams as nodes, and edges representing relations between players with teams they played in and players with players they played together with. The metrics used were different and several models were tested. For instance, a *clustering coefficient model* was created, as high clustering coefficient for a team means that this team either has a lot of new players or it frequently makes player transactions. A *degree model* was also tested, as a player with a high degree is probably a player in the end of his career or a player who is traded frequently (in other words not wanted).

These papers were very interesting, as they showed how changes in networks structure or nature have impact on the sports interpretations we can make. A strong common point from all these papers is that they all conducted their research while keeping their knowledge on sports in mind, to get results as relevant and as insightful as possible. In paper [2], the researchers involved had all experience with the studied sports and took role changes when players substitute into account. In paper [4], the historical evolution of the NBA was very useful to explain the evolution of some metrics.

III. MATHEMATICAL AND ALGORITHMIC BACKGROUND

In this section, we give an overview of what methods and concepts we used for our project.

A. Atemporal metrics/scores

1) Clustering Coefficients: Clustering coefficients are measures that attempt to capture how nodes in a graph tend to cluster together. In a directed network, the *local clustering coefficient* of the node i is given by:

$$C_i = \frac{e_i}{k_i(k_i - 1)}$$

with k_i the degree of node i and e_i the number of edges in its neighborhood. Usually if a node is isolated or a leaf ($k_i = 0$ or 1), we set $C_i = 0$.

We can then also compute the average local clustering coefficient of the whole graph by taking the mean of these coefficients.

$$C = \frac{1}{N} \sum_i C_i$$

One flaw of this metric is that if the fraction of isolated nodes and leaves in the network is too large, then the standard clustering coefficient will be penalized a lot and be very small.

In paper [10], the author introduces an alternative clustering coefficient given by:

$$C_a = \frac{1}{1 - \theta} C$$

with θ the fraction of isolated nodes and leaves of the network.

This adjusted metric is more robust to network sparseness, but can also lead to interpretation problems if θ is too large.

2) PageRank: PageRank algorithm was introduced by Google's co-founders Sergey Brin and Lawrence Page (see [1]) and is used to rank sites based on how referenced they are. PageRank is indeed a local metric that measures how each node is being referenced by other nodes.

The PageRank of a node i is recursively defined by:

$$PR(i) = \frac{1 - d}{n} + d \sum_{j \in IN(i)} \frac{PR(j)}{k_j^{out}}$$

with $IN(i)$ the nodes pointing to i, k_j^{out} the out-degree of j, n the number of nodes, d a damping factor between 0 and 1, which is needed in order to treat nodes with no out-links fairly.

There have been different adjustments made to PageRank, which achieve different results. A more common variant is personalized PageRank, which tailors the PageRank results to a certain person's browsing habits.

What interests us in the PageRank, is that it could be used to rank players in a certain sport, instead of the actual ranking system. A player being referenced a lot by other players is indeed a player who won a lot of matches.

3) Authorities and Hubs: Jon Kleinberg developed the Hyperlink-Induced Topic Search (HITS) algorithm in [6] in order to rate Web pages. He defines two local concepts, *hubs* and *authorities*, and their associated scores, inspired by the structure of the Web:

- *Hubs* are directories that are not authoritative in the information that they have, but lead users directly to authoritative pages.
- *Authorities* are pages linked by many different hubs.

To compute them, three steps are needed:

(i) All hub and authority scores are initialized at 1.
(ii) Authority Update Rule:

$$auth(i) = \frac{\sum_{j \in IN(i)} hub(j)}{\sqrt{\sum_{k \in V} auth(k)^2}}$$

with V the nodes of the graph.
(iii) Hub Update Rule:

$$hub(i) = \frac{\sum_{j \in OUT(i)} auth(j)}{\sqrt{\sum_{k \in V} hub(k)^2}}$$

The two update rules can be repeated an unlimited number of times (convergence is assured thanks to normalization).

Similarly as PageRank scores, Hubs and Authorities scores could help us find interesting roles among competitors.

B. Temporal metrics

As we have results of competitions for several years in tennis, it was interesting to study temporal properties of the networks.

In paper [11], a characteristic temporal clustering coefficient is defined, which takes time evolution into account, unlike the standard clustering coefficients.

We consider a sequence of graphs $G_{t_{min}}, \ldots, G_{t_{max}}$, which all have the same nodes. For a node i, we define:

- $N_i(t_{min}, t_{max})$: set of nodes which have been neighbors of i at least in one of the graphs $G_{tmin \leq t \leq tmax}$
- $k_i(t_{min}, t_{max}) = |N_i(t_{min}, t_{max})|$ temporal degree of node i
- $(G_t^{N_i(t_{min}, t_{max})})_{tmin \leq t \leq tmax}$: sequence of subgraphs induced from $(G_t)_{tmin \leq t \leq tmax}$ with nodes $N_i(t_{min}, t_{max})$

The *local temporal clustering coefficient* of node i is thus given by:

$$C_i(t_{min}, t_{max}) = \frac{\sum_{t=t_{min}}^{t_{max}} \text{\# of edges in } G_t^{N_i(t_{min}, t_{max})}}{(t_{max} - t_{min}) \frac{k_i(t_{min}, t_{max})(k_i(t_{min}, t_{max})-1)}{2}}$$

We can then compute the *characteristic temporal clustering coefficient* by taking the mean of the local temporal coefficients.

We also define the alternative temporal clustering coefficient, which takes into account the fraction of isolated nodes and leaves:

$$C_{i_a}(t_{min}, t_{max}) = \frac{\sum_{t=t_{min}}^{t_{max}} \frac{1}{1-\theta_t} \left(\text{\# of edges in } G_t^{N_i(t_{min}, t_{max})} \right)}{(t_{max} - t_{min}) \frac{k_i(t_{min}, t_{max})(k_i(t_{min}, t_{max})-1)}{2}}$$

with θ_t the fraction of isolated nodes and leaves of graph G_t.

C. Network Distance

As one of our main goals is to compare different competition networks, we took a look at which *network distances* were possible in our context.

A first possibility is the *Hamming distance* defined by the sum of the simple differences between the adjacency matrices $A^{(1)}, A^{(2)}$ of two graphs $G^{(1)}, G^{(2)}$:

$$d^H(G_1, G_2) = \sum_{ij} |A_{1_{ij}} - A_{2_{ij}}|$$

However, this distance requires both graphs to have a similar number of nodes (which we can not ensure, as it depends on the number of competitors in each sports) and only focuses on the differences in the number of links, which we do not find relevant here.

In paper [7], the authors define a new distance based on the *Laplacian matrices* of both graphs. As we already saw in lectures, Laplacian matrices can help us infer a lot of structural information on their graphs (e.g. sparsest cut through its second smallest eigenvalue). This new distance thus reflects more structural similarities between graphs than the Hamming distance.

We recall that the Laplacian matrix L of a graph G with adjacency matrix A and degree matrix D is given by"

$$L = A - D$$

In our case, as the graphs are directed, D can be the in- or out-degree matrix with no significant difference.

Let two graphs of node sizes $N^{(1)}, N^{(2)}$, and the eigenvectors sorted from smallest eigenvalue to largest $(\lambda_i^{(1)})_{1 \leq i \leq N^{(1)}}, (\lambda_j^{(2)})_{1 \leq j \leq N^{(2)}}$ of their associated Laplacian matrix.

We first define the *cumulative distribution functions* associated with the rth eigenvectors ($i \in (1,2)$):

$$\rho_r^{(i)}(x) = \frac{1}{N^{(i)}} \sum_{k=1}^{N^{(i)}} H(x - \lambda_{r_k}^{(i)})$$

Then we can define the *Spectral Graph Distance* by:

$$d^{SG}(G^{(1)}, G^{(2)}, d') = \frac{1}{N} \sum_{r=1}^{N} d'(\rho_r^{(1)}, \rho_r^{(2)})$$

with $N = min(N^{(1)}, N^{(2)})$ and d' a function distance.

A few comments on this distance:

- the authors in [7] had some successful results when comparing the performance of this distance to other more common network distances, even when graph sizes were different.
- For d', they chose the distance:

$$d'(\rho_r^{(1)}, \rho_r^{(2)}) = \int_{-\infty}^{\infty} |\rho_r^{(1)}(x) - \rho_r^{(2)}(x)| dx$$

- This distance is generally not well defined in directed graphs: the Laplacian matrix is indeed not symmetric and thus the eigenvectors can be complex vectors.

Concerning the last comment, we could consider competition networks as undirected graphs, and measure their Spectral Graph distance. However, we did not feel satisfied, as it would result in a too big loss of information.

Thus, we thought about a way to generalize the above definition to complex numbers.

The only change is the definition of the cumulative distribution functions.

To be more precise, the cumulative distribution function of a real-valued random variable X is the function given by:

$$F_X(x) = P(X \leq x)$$

The definition given by the authors is simply the discrete version of the above defition. We can then do the same thing

with a complex-valued random variable Z, whose cumulative distribution function would be given by:

$$F_Z(z) = P(Re(Z) \leq Re(z), Im(Z) \leq Im(z))$$

which can be easily made discrete with:

$$\rho_r^{(i)}(x,y) = \frac{1}{N^{(i)}} \sum_{k=1}^{N^{(i)}} H(x - Re(\lambda_{r_k}^{(i)})) H(y - Im(\lambda_{r_k}^{(i)}))$$

IV. DATASETS

A. Data Collection

Our analysis comprises the following datasets:

- US Senior Women's Épée fencing results for 2017-2018[1]
- US Senior Men's Épée fencing results for 2017-2018[1]
- US Senior Men's Saber fencing results for 2017-2018[1]
- US Senior Men's Foil fencing results for 2017-2018[1]
- Tennis ATP Men results from 2000 to 2018[2]
- Tennis WTA Women results from 2007 to 2018[2]
- Chess games results dataset[3], with games on a period of 100 months among 8631 players

For each dataset, the desired information is simply a set of games with a defined winner and loser (except for chess, for which we dropped the draws, but this will be discussed later). While some other information is available, such as margin of victory, we wanted to keep the analysis sufficiently simple that it could be applied across competitions. While margin of victory is well defined for fencing and tennis, results for other competitions such as wrestling or chess might lack this dimension.

B. Network Structure

The most natural idea to explore the properties of these competitive datasets is to load them into directed networks, where:

- Nodes are players' ID (which we assign arbitrarily)
- Edges (p_1, p_2) means "p_1 lost to p_2"

The first tricky decision we had to make was the type(s) of graph we wanted to load the files into. Indeed, during the course of various competitions, one competitor may meet an other competitor multiple times.

The different solutions would be to load them into either a directed unweighted simple graph, a directed unweighted multi-graph or a directed weighted simple graph.

The first solution is too simplistic, erasing significant information about player quality. For two competitors, there is surely a difference in their level of play if one has won 9 of 10 matches rather than 5 of 10, which would be information lost by a simple graph. Our analysis uses thus mostly a directed unweighted multi-graph, which by most measures is

equivalent to a directed weighted graph.

An other observation we can make is that in some competitions (rarely in sports), there can be no winner (for instance in our chess dataset). An idea that we did not try, is to include the number of draws between two players (e.g. by dividing the weight of their edges by the number of draws) instead of ignoring them. In the actual chess ELO ranking system, draws do have significance. As such, this could indeed make the weighted network more adequate, as 44.1% of games in the dataset are drawn!

V. METHODOLOGY

We created the networks as described in Section IV (Datasets) and used a variety of analysis tools to draw conclusions about the networks. Many descriptive statistics were computed with built-in SNAP functions, such as graph size, diameter, and clustering coefficient. Other approaches to analyzing the data were explored on problem sets, such as degree distribution. Some further information was gleaned from more complex functions like PageRank computation and connected component enumeration.

Some experiments were conducted with our implementations of different network analysis tools. For modeling how skill is distributed in the network, we use a plot of the PageRank distribution. We also implement the approaches described in Section III (Mathematical and Algorithmic Background). Ultimately, the combination of traditional metrics and competition-specific concepts allows us to draw interesting conclusions from the data.

[shortlabels]enumitem

VI. RESULTS AND FINDINGS

For our research, we had five key areas of interest: intra-sport analysis, inter-sport analysis, ranking methods, temporal analysis, and predictive analysis.

A. Intra-Sport Analysis: Fencing

For our intra-sport analysis, we looked at modern competitive fencing. More specifically, the three different kinds of fencing: foil, épée and saber. We also looked at the difference between mens and womens épée. One important thing to note is in the US circuit, all people who compete seriously specialize in only one weapon. However, they do all share some important characteristics like footwork, time limits, and score amounts.

In order to make sense of the network characteristics, it is important to provide context. Foil and saber are both fenced with a limited target area, dictated by an electric vest that people wear when they fence. They also both have right of way, which is a standardized set of rules to determine who receives the point after a given action. Épée, like foil, is a point weapon, but it does not have a specific target area - the entire body is the target. Moreover, there is no sense of right of way - the first person who scores, gets the point. If both fencers score within a short time period, they both get a point, which is called a double touch.

As such, people have different preconceptions as to the unique characteristics of each event. As a general rule, épée is viewed as having much greater variability due to the lack of right of way and the existence of the double touch.

Our network data for these four graphs is as follows:

	Saber (M)	Foil (M)
Nodes	226	350
Edges	527	595
Size of SCC	76 (34%)	102 (29%)
Number of WCC	1	2
Clustering Coefficient	0.00408	0.0138
Path Probability	0.344	0.216
Closed triangles	11	7
Effective diameter	11.4	13.7
Full diameter	21	22
Avg shortest path length	6.0	8.0
	Épée (M)	Épée (F)
Nodes	270	233
Edges	630	562
Size of SCC	103 (38%)	91(39%)
Number of WCC	1	1
Clustering Coefficient	0.00310	0.00765
Path Probability	0.39	0.396
Closed triangles	12	10
Effective diameter	6.0	7.27
Full diameter	10	14
Avg shortest path length	4.4	4.9

As we can see, our data actually reflects some of these commonly held beliefs.

Take the probability that a given node is in a triad $\frac{\text{number of triads}}{\text{number of nodes}}$. For mens épée, we find that this probability is around 4.4%, and for womens épée around 4.2%. However, for mens foil we find a probability 2%. Given the explanation above, this makes sense. A triad, in our competitive graph, would be a rock-paper-scissors situation where competitor A beats competitor B, competitor B beats competitor C, but competitor C beats competitor A. Assuming that the better fencer strictly dominates, there should be no existence of triads. However, we see that épée (and saber, to a certain extent) both have noticeably higher rates of triads.

One could also look at the size of the 90th percentile effective diameter. In the context of a competitive graph, the effective diameter would represent roughly the number of matches between two randomly selected players. In a strictly-dominating competition scheme, we would imagine this value to be relatively larger than in a non-strictly-dominating competition scheme, as we would have less short-cuts. Saber especially exhibited this behavior, as we see an effective diameter length of 11.4, whereas for mens and womens épée we see a diameter lengths of 5.9 and 7.2, respectively. (Note: Foil has an effective diameter length of 13.7, but has roughly 50% more nodes in the graph than the other three, so this finding is less significant).

Between mens and womens épée, there are no major differences. We can observe that the clustering coefficient of the mens épée network is less than the one of the womens épée network. However, our alternate clustering coefficient yields

the opposite conclusion which mitigates any conclusions. Both the diameter and the effective diameter of the womens graph are slightly larger than the mens graph, which suggest lesser variability in overall results. In summary, however, we can see that network characteristics are much more dependent on event than on gender.

B. Inter-Sport Analysis

For this portion of our analysis, we looked at the unique characteristics of tennis, chess and fencing competition networks. Importantly, for fencing we only looked at men's épée fencing in order to reasonably scope this portion of our analysis.

For our tennis, fencing and chess networks we computed the following properties for each network:

	Tennis (M)	Tennis (F)
Nodes	1485	963
Edges	52283	29581
Size of SCC	897 (60%)	612 (64%)
Number of WCC	3	1
Alt. Clust. Coefficient	0.467	0.421
Path Probability	0.583	0.624
Closed triangles	12275	9602
Effective diameter	3.5	3.5
Full diameter	7	7
Avg shortest path length	2.9	2.8
	Fencing	Chess
Nodes	270	6832
Edges	630	36387
Size of SCC	103 (38%)	4121 (60%)
Number of WCC	1	94
Alt. Clust. Coefficient	0.0344	0.123
Path Probability	0.39	0.58
Closed triangles	12	2778
Effective diameter	6.0	6.8
Full diameter	10	16
Avg shortest path length	4.4	5.3

From this information, we see that the four networks have significant similarities that are likely shared by other competition networks (these similarities would be caused by the competitive nature of the studied networks), but also some interesting differences, that we will try to explain.

The first important remark is that the fencing network is a lot smaller than the other networks. Thus we have to be careful as to not wrongly over-analyze our results, as less matches' information leads to a higher bias of the data.

We also plotted the out-degree distributions of the different networks on log-log scales. Interestingly, the chess network shows a different distribution than the other three.

We can immediately see that the men and women tennis networks are very similar, compared to the chess network. They both have very few weakly connected components, medium alternative clustering coefficients and short average shortest path length.
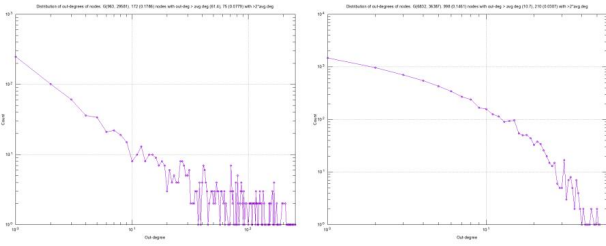
Fig. 1. On the left: out-degree distribution of the tennis women network, which is similar to the tennis men and fencing network. On the right: chess network

The chess network, on the other hand, has many weakly connected components, a much lower alternative clustering coefficient and a longer average shortest path length.

This finding is consistent with the origin of the data for the different sports. While the tennis and fencing networks are based on an elimination-style competition, the chess network likely comes from "Swiss"-style tournaments, in which players each play a fixed number of games, but in each round play games against other competitors with the same or similar records.

We can explain this because the chess network is less grouped than the other networks. Its structure is completely different, and this can be better seen in the distribution plots (Fig. 1). The distributions of the three sports networks follow power laws, but the one from the chess network follows a higher-degree exponential law.

Another interesting property is the number of closed triangles in each network. The one from the fencing network seems however a bit off, we suppose that this is due to its small size. Both tennis networks have similar closed triangles ratios, which are a lot greater than the ratio of the chess network. This already shows that there are some significant difference in the structures of the sports competition networks and of the chess network.

These differences can be explained by how different sports competitions and chess competitions are. Usual sports competitions are represented by complete binary trees. Some chess competitions are also like this, but not always: there are other systems like the Round-robin or Swiss systems.

The same conclusion can be drawn through the connected components analysis. Sports networks typically consist of one giant connected component of highly skilled players and a few weakly connected components, representing less skilled players. This is not true of the chess network, which have a lot weakly connected components. In the chess network, some highly skilled players have no losses, and thus no out-edges, and do not belong to the SCC.

The relative size of strongly connected components to the competitive population is also a valuable metric for measuring the distribution of talent, as SCC's represent some upper echelon of players that are capable of defeating one another. Because the networks have different edge per node ratios, differences in size may be simply due to presence of additional games. To normalize and analyze the standardized SCC size, we remove edges randomly from each graph except the one with minimal edge-to-node ratio until they all have similar edge-to-node ratios. The fencing graph is unchanged, with an SCC making up 38% of the population. Men tennis sees its SCC's size shrink to 27%, women tennis to 30%, and chess to 36%.

The proportion of competitors that have demonstrated an ability to compete with high-caliber players is thus largest for fencing and chess, and smaller for tennis. This interesting twist on the raw size of the SCC of each network help us understand better how the level distribution is among players in each discipline. The higher relative SCC sizes would be due to higher level variance in chess and fencing matches, which would allow weaker players to win matches against decidedly better players. Indeed, in tennis, the rankings are very stable above a certain ranking position (the Big Four and their regular challengers), showing less variance than in fencing.

Our observations are indeed in accordance with the networks' structures, which we visualized in order to get a better idea of key differences between networks. Here are the men's tennis network and the chess network:
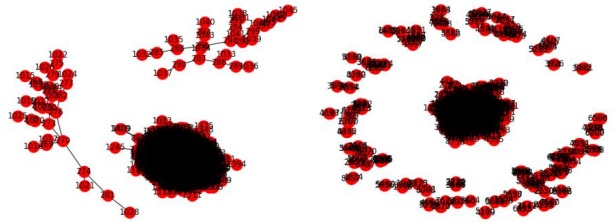

Fig. 2. On the left is the men's tennis network, and on the right is the chess network

As observed above from each network's statistics, the chess network is less clustered and is actually an union of many local competitions. This is unlike the structure of the tennis network, which is largely a concatenation of binary trees that represent direct-elimination-style competitions.

We also tried to apply the network distance defined in section III. However, the computational time was too long and we were not able to get conclusive results to evaluate how relevant the distance is. (For instance, we have found a distance of 0.586 between the men tennis and the women tennis network.)

C. Ranking methods

Looking at each of the networks, we can identify the competitors with the highest PageRank scores, the network hubs, as well as network authorities. This information is as follows:

	Tennis (M)	Tennis (W)
PageRank	Federer R.	Williams S.
	Nadal R.	Wozniacki C.
	Djokovic N.	Radwanska A.
Hubs	Ferrer D.	Radwanska A.
	Berdych T.	Cibulkova D.
	Verdasco F.	Jankovic J.
Authorities	Federer R.	Wozniacki C.
	Nadal R.	Williams S.
	Djokovic N.	Radwanska A.
	Fencing	Chess
PageRank	Kaull J.	#7848
	Ewart S.	#64
	Hoyle J.	#158
Hubs	Thein-Sandler A.	#1594
	White S.	#1286
	Moore S.	#7848
Authorities	Kaull J.	#7848
	Ewart S.	#1286
	Fayez A.	#1594

Concerning the tennis rankings, the top players are indeed the most dominant players during the data period.

For instance, for the men rankings, the top PageRank valued players are also the best authorities (which is consistant), and hubs are indeed the next best top players.

Concerning US Fencing rankings, there is a serious discrepancy between strength of fencers on the US circuit relative to their PageRank rankings: there is an average ranking position difference of 5.4 for fencers in both the top 32 US points and top 32 PageRank values).

We can see that 6 fencers are in the top 32 PageRank values, but not in the top 32 of US results. Moreover, we can see that fencers of national rank 1, 2, 3, and 4 in the US, are ranked with PageRank values by 4, 1, 7, and 16 respectively.

An explanation is that US National Fencing rankings are a product of both domestic and International events. Removing international results and only factoring in the highest two domestic results adds thus a significant bias to our ranking predictions. Important to note, however, that only 11 of the 270 analyzed competitors have results that actually affect national rankings. Looking forward, we will attempt to obtain national ranking information sans international points.

However, this also means that certain players are being undervalued (and overvalued) on the US circuit for fencing, relative to their PageRank scores. This has widespread implications, namely for recruiting and national team selection in the United States.

Importantly, there is a lot of randomness inherent in sports competitions. Elimination rounds' results that we had were seeded according to pool rounds, which are randomly assigned. As such, if competitors have a weak pool, they can have a relatively high seeding in the next round, leading to easier opponents, overall. In order to decrease the bias associated to randomness, we need to increase the size of our fencing dataset.

Because of these reasons, the rankings are most reflective of player level in the tennis rankings, less so in the chess ranking (smaller edges per node ratio), and much less so in the fencing rankings.

To get a better picture, we also plotted the cumulative distribution of PageRank scores in Fig. 3.
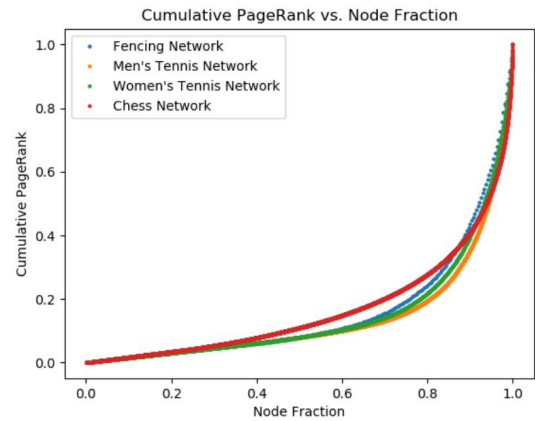

Fig. 3. The distributions of PageRank scores among competitors suggest that fencing is a higher-variance competition than tennis, and chess has a smaller set of elite players than either.
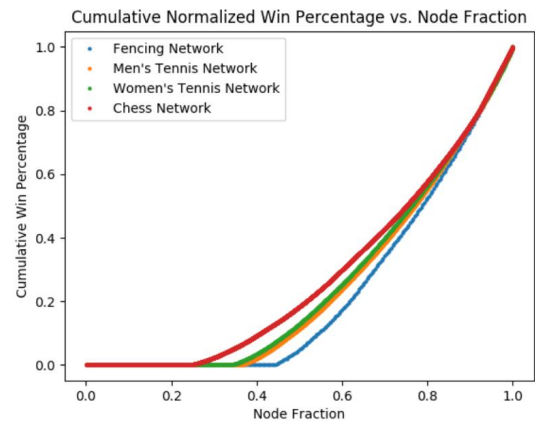

Fig. 4. This figure serves as a counterpoint to the graph of PageRank distribution, plotting an integrated, normalized version of win percentage against node fraction as an alternative metric.

The chess curve deviates from the linear initial trajectory first, suggesting that there is a less distinct division between good and great players than in sports competitions. The other curves appear to sharply increase in slope around the same time, suggesting a rough equivalence in network structure, as stated above.

Interestingly, while chess diverged first, it also stayed at a low value for a larger portion of the nodes, which suggests a significant difference in quality between the great and elite players.

The relative positions of the graphs furthermore agree with the previous observation that chess and fencing have higher

variance in games' outcomes than tennis, with both of them being above the other curves.

The graph of integrated PageRank differs from the graph of integrated win percentage in a few important ways. Although the win percentage scores are normalized to sum to one, win percentage seems to be less descriptive. Especially towards the right-hand side of the graph, it becomes apparent that the difference in PageRank scores is more pronounced than the difference in win percentage scores, suggesting that it might be a less arbitrary, more insightful method for ranking competitors.

Based on the competition networks, tennis performance exhibits less variance than both fencing and chess performance in the selected competitions, while chess exhibits a greater concentration of talent in the hands of a few top competitors. Ultimately, this type of analysis may be more helpful when comparing different leagues of a given sport or different seasons of a given league to identify trends in skill concentration and outcome variance.

D. Temporal Analysis

For this analysis, we used the following networks:
- Men tennis: time-period of one year per network (19 networks from year 2000 to 2018)
- Women tennis: time-period of one year per network (12 networks from 2007 to 2018)
- Chess: time-period of ten months (10 networks)

We computed the alternative characteristic temporal clustering coefficient of the three sequences of networks:

Men tennis	Women tennis	Chess
0.153	0.129	0.0254

As expected, over time, the chess network is much less clustered than the tennis networks, which have similar temporal clustering coefficient.

We can also note that this metric is more realistic than the atemporal alternative clustering coefficients seen above (which consider the networks as static). For instance, the men tennis network had an alternative clustering coefficient of 0.467 and has an alternative characteristic temporal clustering coefficient of 0.153.

We also plotted the variation of some metrics over time like the number of active nodes (not isolated) or the alternative clustering coefficient.

The clustering coefficients stay roughly in the same order of magnitude over time, which is in accordance with the fact that competition rules stay the same.

For the chess network however, the alternative clustering coefficient seems to be not relevant: it contains a too big fraction of isolated nodes and leaves for the first months, leading to a much higher alternative clustering coefficient. This can lead to misinterpretation.

By looking at the plots of edges and active nodes over time of the tennis networks, we can observe two opposite development.
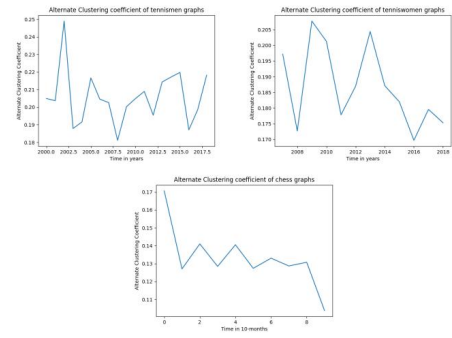

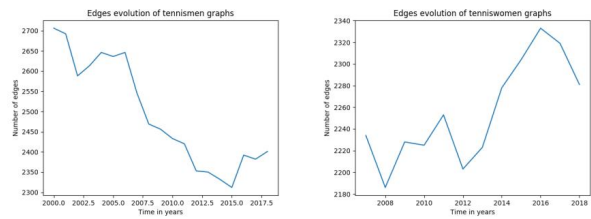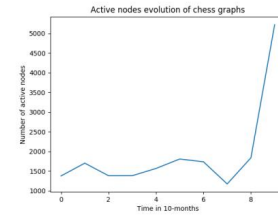Fig. 5.   Alternative clustering coefficient over time.
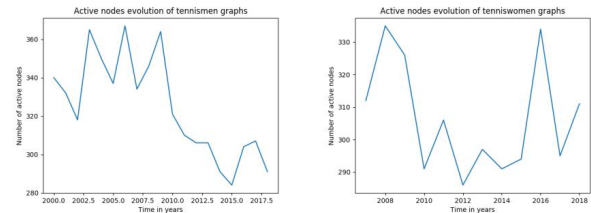

Fig. 6.   Edges over time.

Fig. 7.   Active nodes over time.

The men competition in tennis seems to have become more "elitist" over the last years: with the same number of matches, less nodes are active, meaning that the same tennismen play against each other more often.

It is the opposite for the women competition in tennis, which has gained a lot in tenniswomen diversity over the last years.

This is in accordance to the current state of tennis, where men tennis is dominated by a small pool of players, and where women tennis is becoming more and more popular and less predictable than men competitions.

E. Predicting Outcomes With PageRank

One question of interest related to how players are ranked, is which of two competitors is more likely to win a match between them. We tried using the PageRank scores to develop a model for outcome prediction. The men's tennis data is used

for this purpose, as results can be compared to bookmakers' odds to get a point of comparison.[4]

The data used to compare our results to are the bookmaker's odds for the 2018 US Open, Round of 32 and above (31 matches, a fairly small data set). The bookmaker accurately predicted 23 of 31 of those matches, for an accuracy of 74.2%. The odds also imply a probability, assuming any bet has an expected value of zero. This is an approximation, as an oddsmaker will usually aim for negative expected value. The average predicted probability (i.e. for each match, the implied probability that the winner would win) is 67.6%.

The simplest model is a classifier that predicts the player with the higher PageRank score will win. During the specified US Open data, the player with higher PageRank won 67.7% of the time. This is not as strong as the bookmaker's accuracy but is better than random.

In order to get an estimate of our confidence in that accuracy, we built a logistic regression model where each sample is a match with four features: the favorite's PageRank, the underdog's PageRank, the difference between the two, and the ratio between the two. The labels corresponded to whether the favorite won.

On both a random held-out test set and the US Open data, the logistic regression predictions exactly match the linear classifier. This is expected, as the difference in PageRank is essentially the only input to the model. It does, however, associate a probability with each prediction, which can be used to compare these predictions to the bookmaker's odds.

With only the PageRank information, the logistic model achieves an average predicted probability of 55.5%. While this is better than random, it isn't close to the prediction ability of the bookmakers. One thing to note is that the probabilities for favorites in the logistic model all fall in the range of 65.5% to 66.4%, while we observe favorite probabilities of 53.7% to 97.1% in the bookmaker's data. So our model tends to keep its estimates in a very conservative range; good for big upsets (such as when Millman beat Federer, an outcome with 6.5% probability according to the bookmaker) and bad for most other scenarios. For matches where the winner is almost guaranteed (like that Federer match) we would like to see larger probabilities and we would prefer to see probabilities closer to a coin flip for more uncertain matches. The average of the predictions is good, but the predictions are too narrowly distributed, likely due to the lack of expressiveness available with this single metric.

Ultimately, the bookmaker ends up with better accuracy and average predicted probability. This is likely due to their ability to base predictions on a much larger range of factors. Using PageRank as a catch-all statistic generally produces better predictions than random, but likely won't give you any edge at the betting counter.

VII. CONCLUSION

Individual competitive networks provide an exciting opportunity for exploration and analysis. In our research, we were able to analyze these networks in a variety of different ways –

intra-sport, inter-sport, and over time – in order to answer a variety of different questions – including the efficacy of network ranking systems, the fidelity of competitive networks as a model for competitive fields, and the similarities between different competitive disciplines.

Our results are consistent with our understandings of each sport. This consistency not only validates our process for information retrieval and network modeling, but also provides a rich body of information from which we can draw insights. With our findings, we can infer the structure of a competitive network given the rules of competitive discipline. We can see pertinent differences between each discipline's level distribution and competition process. We can even provide predictive power, albeit not Vegas-tier.

There is still much room for future analysis in this field, especially concerning motif detection (which shows interaction between players) and temporal analysis (e.g. temporal PageRank). Looking forward, as researchers, athletes, and sports-enthusiasts, we are excited and optimistic as to the future of network analysis in the context of sports. Not only does this analysis provide new and interesting insights to an already mature field, but it also provides a whole new paradigm through which to view sport.

We feel this was a very interesting and insightful project, and we all learned a lot from it. Thank you!

REFERENCES

[1] S. Brin, and L. Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Proc. 7th International World Wide Web Conference*, 1998.

[2] F. Korte, and M. Lames. Characterizing different team sports using network analysis. *Current Issues in Sport Science 3*, 2018.

[3] T. U. Grund. Network structure and team performance: The case of English Premier League soccer teams. *Social Networks*, 2012.

[4] P. O.S. Vaz de Melo, V. A.F. Almeida, and A. A.F. Loureiro. Can complex network metrics predict the behavior of NBA teams?, 2008

[5] S. Wernicke. Efficient Detection of Network Motifs., 2006

[6] J. M. Kleinberg. Authoritative Sources in a Hyperlinked Environment, 1999

[7] Y. Shimada, Y. Hirata, T. Ikeguchi, and K. Aihara. Graph Distance for Complex Networks, 2016

[8] M. M. Deza, and E. Deza. Encyclopedia of Distances 2nd Ed., Berlin Heidelberg, 2013)

[9] T. A. Schieber, L. Carpi, A. Díaz-Guilera, P. M. Pardalos, C. Masoller, and M. G. Ravetti. Quantification of Network Structural Dissimilarities, 2017

[10] M. Kaiser. Mean Clustering Coefficients: The role of isolated nodes and leafs on clustering measures for small-world networks, 2008

[11] J. Tang, M. Musolesi, C. Mascolo, and V. Latora. Temporal Distance Metrics for Social Network Analysis, 2009

[4]https://www.oddsportal.com/tennis/usa/atp-us-open/results/