Network Analysis of NBA Matchups

Abhishek Dhar

Abstract

The purpose of this paper is to create and analyze a network of matchups between NBA players during the 2021-2022 regular season with the goal of gaining insights about the common positional composition of matchups, the best performing players, and the outcomes of playoff games. The results of the analysis, which include a visualization of the network, show that guards and centers often matchup within their own positions while forwards tend to have more flexible matchups, and that the top-ranking players according to the PageRank algorithm are primarily offensive driven - suggesting a bias in the way we determine the matchup winners. We also conclude that while the network is not entirely successful in determining the winning team in a playoff series, expanding the network to span across several seasons could provide a more holistic view of interactions between players. Overall, the research provided some fascinating learnings and opened up new avenues for future work. The code and dataset for this project can be found at: https://github.com/Xenox473/NBA-Network-Analysis

Introduction

Arguing over the greatest player of all time (GOAT) is a frequent tradition across fans of any sport. Whether it's Nadal vs Federer or Messi vs Ronaldo, fans engage in heated debates with one another, presenting statistics that support their claim. The NBA is not exempted from this debate as fans clamor to voice their support for a number of different players (Michael Jordan, Lebron James, Kareem Abdul Jabbar, etc.). The most common statistics used in the argument are often the number of championships, individual accolades, points, or even an "intuitive" approach to quantifying grit. The purpose of this paper is to use network analysis techniques to investigate the merits of determining matchup winners in order to provide some insight into the GOAT debate.

For the uninitiated, a matchup in the context of the NBA is essentially a pair of players from opposing teams that guard one another during a game. Matchups play an important role in the outcome of a game. For example, a team that consists of shorter players might be a really good team because they're able to outscore their opponents by shooting 3 pointers, but they may lose horrifically if they face a much taller team that likes to dunk the ball all the time. As a result, matchups are often exploited - especially during the playoffs when teams repeatedly face one another in a best of 7 series - by forcing opposing players into guarding their desired player.

The idea behind this paper is to try to determine the best player across the league based on the outcomes of the opponents they are matched up with. This is not a novel idea on its own. A similar study by Filippo Radicchi [5] built a matchup network of tennis players in order to determine the best player of the season. However, possibly due to the fact that the NBA is a team sport, the application of this idea to this space is certainly unique and could lead to some promising insights.

The objectives for this project are as follows:

1. Construct a network of matchups.
2. Identify the top ranked players in terms of matchup performances.
3. Group players into communities to determine the most common types of matchups across positions.
4. Test whether the network can provide insights into outcomes of playoff matchups.

Methodology

Data

The matchup data is openly available on the NBA stats page [3]. It was mined using the NBA API client available for Python [4]. It consists of all pairings of opposing players and their recorded stats when guarded by each other. Figure 1 provides examples of the recorded measurements for each matchup. It displays all of a player's stats when guarded by Draymond Green. Since the current regular season was not completed at the time of starting this project, data from the 2021-2022 regular season was mined. In addition to the matchup data, more generic data such as player information and team rosters were also collected from the same source.

Figure 1: Example of the matchup data available on the NBA stats webpage

Network Properties

The network was designed to be representative of all the matchups that occur between players during a regular season. As a result, a node in the network represents an active player and an edge between two nodes represents a matchup between two players. The edges have two properties:

1. Direction: A matchup edge is directed towards the winner of the matchup. In order to win a matchup, a player would have to perform better than their opponent in certain key matchup metrics on the offensive and defensive sides. The direction is towards the winner because that will correctly represent the flow of importance - this will prove important later on when calculating the PageRank value for each node.
2. Weight: The weight of the edge will correspond to the total matchup time of the two players. This allows for the prioritization of primary matchups. For example, winning a 6 minute matchup is more impressive than a 2 minute matchup.

The nodes have certain features as well:

1. Player's name
2. Player's team
3. Player's position

Determining the Matchup Winner

A matchup winner must be determined so that the direction of the edge can be known. To do this, we derived Equation 1 that calculates the matchup score between two players. It does this by summing up the difference between certain stats per second when either player is on the offensive/defensive side. The stats chosen were: Assists, Blocks, Turnovers, Field Goal Mades, 3 Point Field Goals Made, and Shooting Fouls. If the score is positive, player 1 wins the matchup, else player 2 does. The beta parameter exists to credit the defensive player for the relevant stats (Blocks, Turnovers) by being -1 for those specific stats, and 1 otherwise.

$$score = \sum_{Stat \in \{S\}} \left(\frac{Matchup_{1,2}[Stat]}{Matchup_{1,2}[Time_{sec}]} - \frac{Matchup_{2,1}[Stat]}{Matchup_{2,1}[Time_{sec}]}\right) * \beta$$

Equation 1: Calculating the matchup score

Analysis

Once the network is set up, we can analyze it to investigate the aforementioned objectives. The PageRank algorithm is used in order to identify the top ranked players in terms of matchup performances. We chose to use PageRank as opposed to simply counting the in-degree of each node because it's important to also consider the quality of the opponent the player has beaten. For the second objective, we perform community detection on the network using the Louvain method.

To test whether the network can provide insights into outcomes of playoff matchups, the network was converted into a directed bipartite network where every player is connected to a node that represents its team, and edges between team nodes are aggregations of player edges between teams with the directions set towards the team with the most weighted incoming edges. This naive approach was used to predict outcomes of playoff matchups since those game data was not included in the original dataset. Another approach to determining the playoff winners was to calculate the PageRank values of all the playoff teams per conference. The results are interesting and will be described further in the following section.

Results

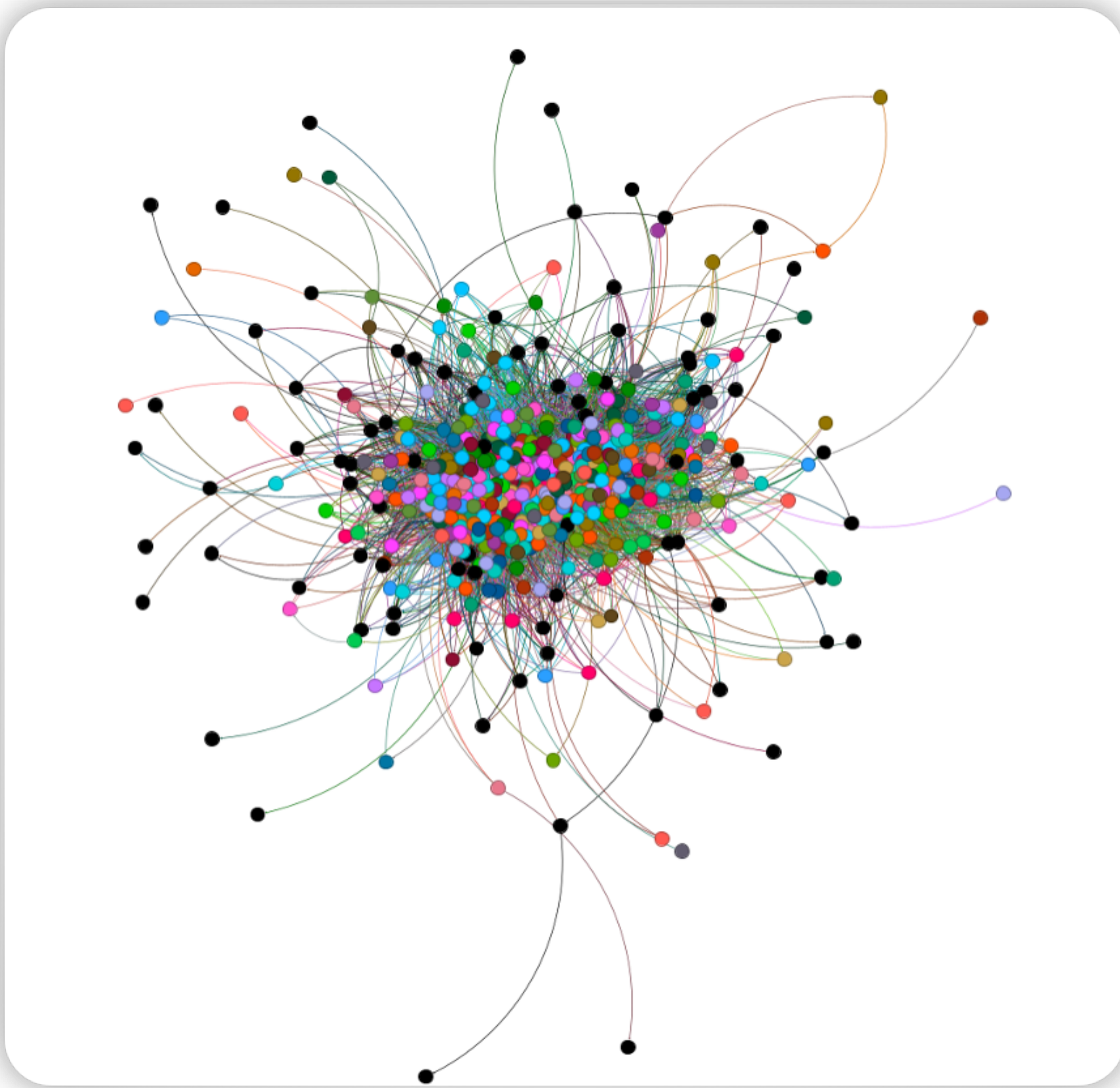

Figure 2: Network Visualization

The Network

A visualization of the network is shown above in Figure 2. It consists of 564 nodes and 24404 edges. The average in-degree was measured to be around 43.36. Figure 3 below shows the various degree distribution measurements for all the nodes. The network is not particularly dense - measured to be 0.077 - or clustered either with an average clustering coefficient of 0.192. Since the clustering coefficient measures how connected a node's neighbors are, this suggests that for the most part, matchups are not transitive. This means that if Player A loses to Player B who then loses to Player C, Player C will not lose to Player A. Moreover, the network was also tested for the Friendship Paradox. The results state that 72% of nodes have an outgoing degree less than the average outgoing degree of their neighbors. This is fairly high, and implies that a player is very likely to lose to a player who's lost more match-ups than they have.

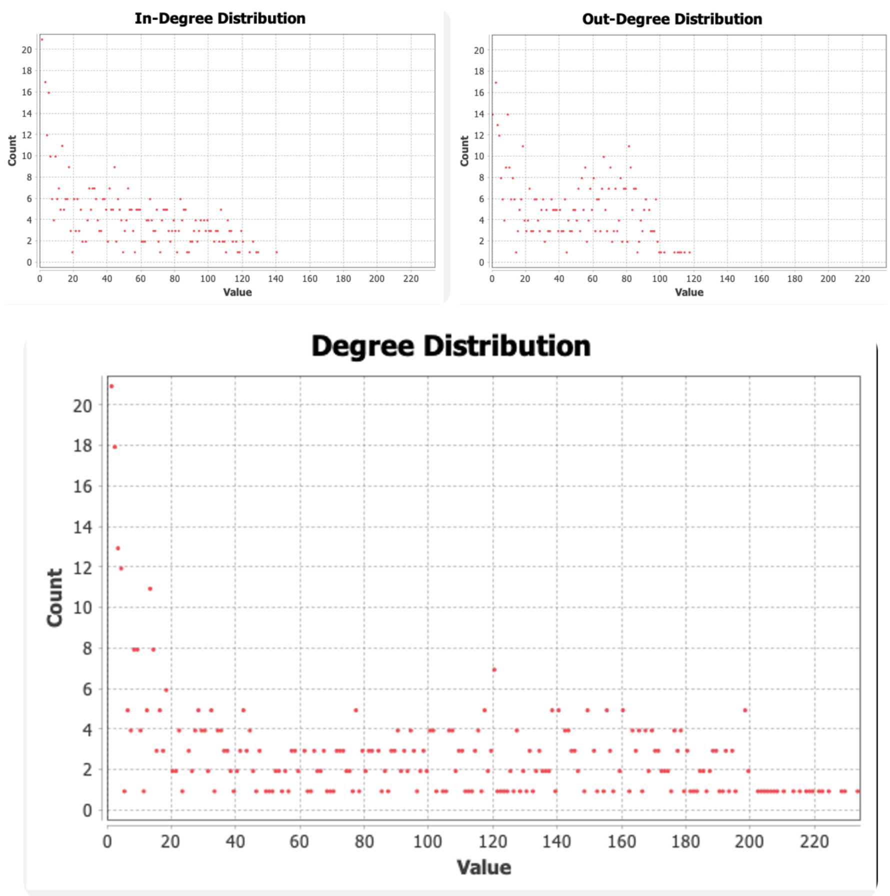

Figure 3: (In/Out) Degree Distribution of the Network

Community Detection

Using the Louvain method, the network was divided into four communities with a Modularity score of 0.282. Figure 4 shows the breakdown of the four communities by the node member's positions. Groups 2 and 3 consist primarily of guards and a few forwards while Group 1 is pretty interesting with an almost equal split of forwards and centers. This can be explained by the fact that some centers and forwards often switch between the two positions depending on the matchups. Overall, the figure shows that forwards are more likely to be exposed to a diverse set of matchups.

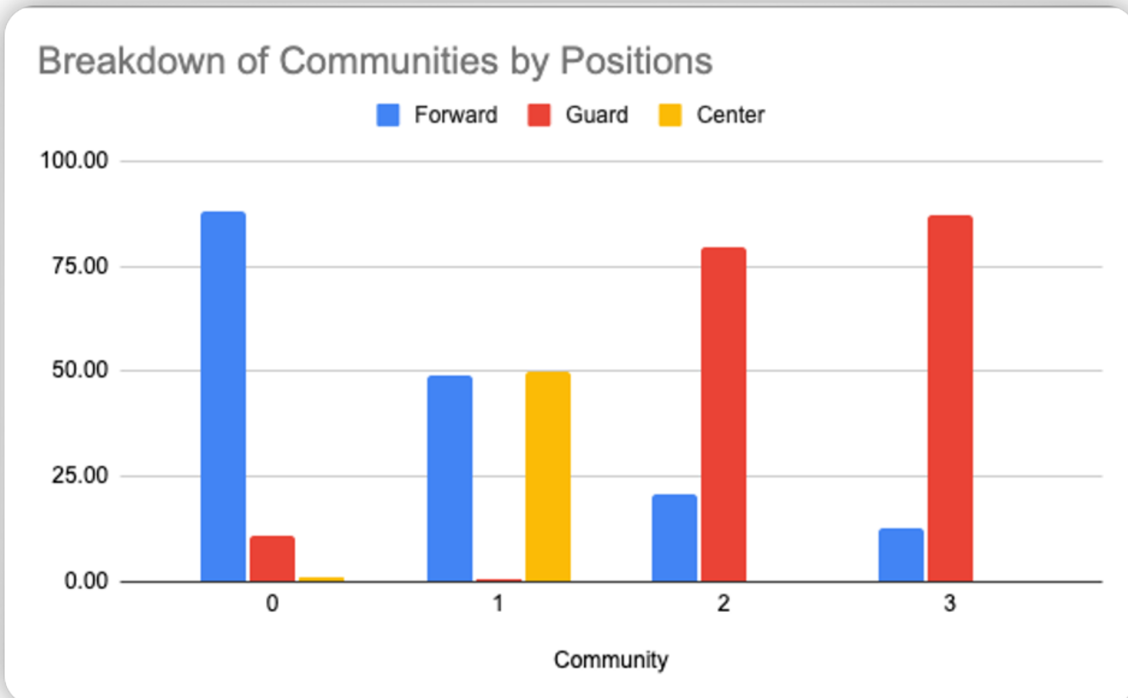

Figure 4: Breakdown of Communities by Positions

Top Ranked Matchup Winners

Table 1 shows the top ranking players in the network according to the PageRank algorithm. The PageRank algorithm is used because we also want to account for the quality of the opponent a player has beaten. Since in-degrees are equivalent to a matchup win, it is fairly simple to implement the algorithm for this network. As shown in the table, the majority of the players are guards. This is probably due to the fact that guards are often the ones in control of the ball for longer periods of time and therefore face a higher number of matchups.

Rank	Name	Position	RAPTOR OFF Ranking	RAPTOR DEF Ranking
1	Jayson Tatum	F	4.0	1.6
2	Dejounte Murray	G	3.1	0
3	DeMar DeRozan	G	3.6	-2.3
4	Donovan Mitchell	G	4.7	-2.7
5	Terry Rozier	G	2.7	0.2
6	Trae Young	G	6.7	-3
7	LaMelo Ball	G	2.7	-0.7
8	Darius Garland	G	4.8	0.3
9	Gary Trent Jr.	G	1.4	-1
10	Russell Westbrook	G	-1.2	2.1

Table 1: Top 10 Players Ranked According to PageRank

In addition to the player names, the table also includes another metric for comparison known as RAPTOR [6]. This metric is developed by FiveThirtyEight and assigns an offensive and defensive score to the player. Comparing the results with this metric, we can see that our method of determining matchup winners seems to be biased towards the offensive player, since most of the players have very high offensive ranking but neutral or low defensive rankings. This could also explain why they are almost entirely guards, since guards tend to be the offensive leaders of the team, while centers and forwards contribute to the defensive side as well.

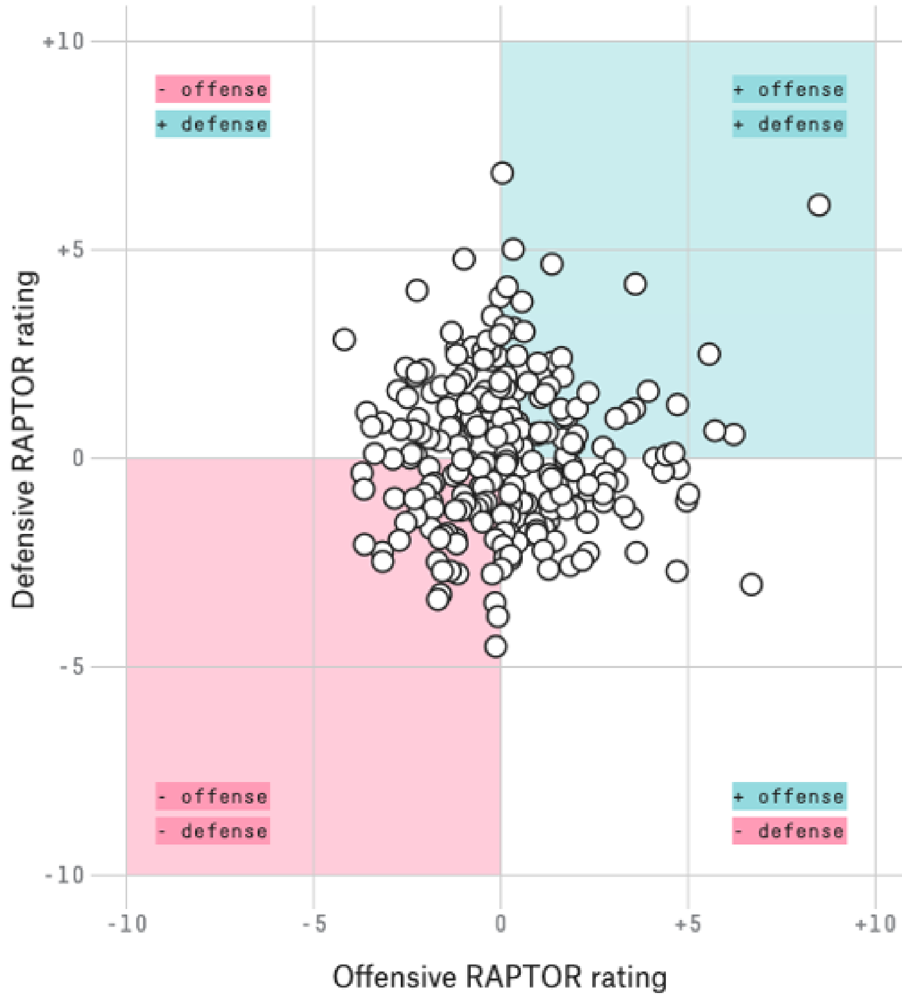

Figure 5: Plot of NBA players according to their RAPTOR scores for the regular season 2021-2022

Playoff insights

Testing whether the network could provide any insights to the outcome of playoff series was done using the bipartite team network. This approach involved identifying the edge and its direction between two teams to determine the series winner. However, only an accuracy of 53% was achieved. On the other hand, splitting the bipartite network into two - one for each conference - and performing PageRank on each subgraph showed some promising results. Figure 6 shows the Eastern and Western playoff team networks with the nodes sized according to their PageRank value. The top 2 teams in both conferences include the previous year's conference champions (PHX, MIL) and the eventual conference champions of that year (GSW, BOS). Further research into this could provide a new method to determine playoff winners.

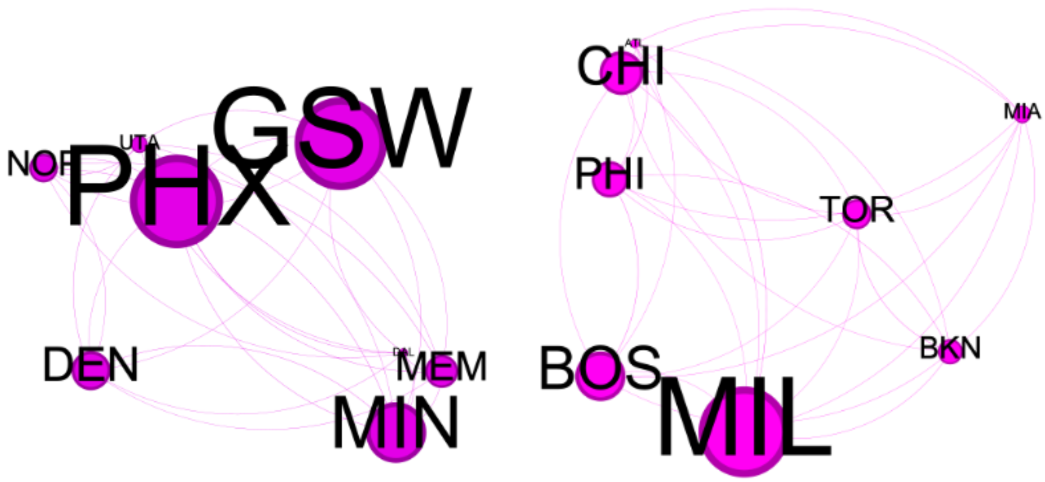

Figure 6: Subgraphs of the Western and Eastern Conference Playoff teams

Per game insights

Another potential use case of this research could be to split the original matchup network into a subgraph of two teams - as shown in Figure 7. Then, through PageRank or another equivalent measure, one could identify the important players or matchups for that particular game.

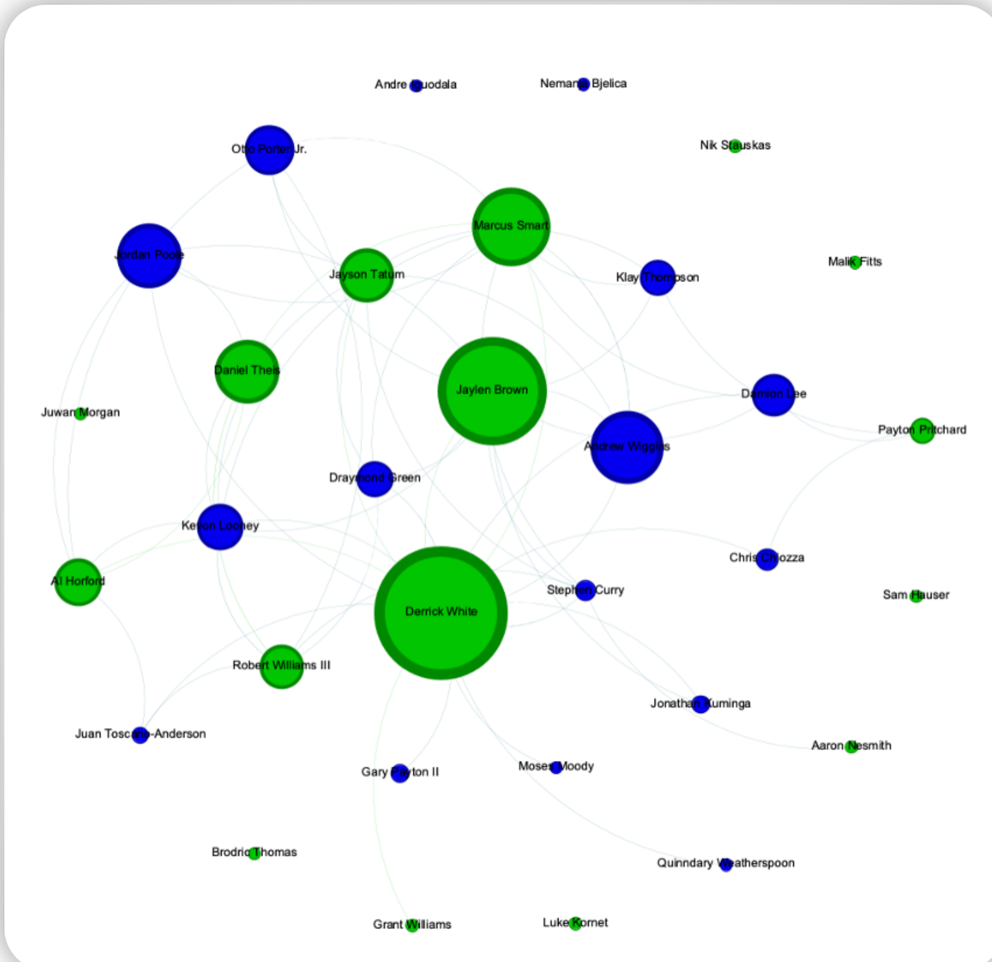

Figure 7: Subgraph of Golden State Warriors and Boston Celtics Players

Discussion

In this paper we set out to build a network of NBA players and the matchups between them in order to see whether we could gain any insights. One of the insights was through the community detection that showed that guards and centers are usually matched up with players in their same position while forwards have no restrictions on the positions they're matched up with.

Through the identification of the top players, we discovered that the algorithm we used to determine a matchup's winner was biased towards offensive stats. Adding more types of stats to the equation such as rebounds and steals could help balance the bias.

We also showed that our network was not entirely successful in determining the winning team in a playoff series. Usually, two players would play against each other at most 4 times during the regular season. This isn't a lot of data, and could make our algorithm susceptible to anomalous performances. Expanding the network to span across several seasons could provide a more holistic view of interactions between players. Finally, another way to increase the accuracy would be to have per game predictions use dynamic subgraphs that change according to the lineup for that particular game in order to account for injuries, absences, etc.

References

1. Basketball player's value evaluation by a networks-based variant … https://arxiv.org/pdf/2012.15734.pdf. Accessed May 5, 2023.
2. Basketball lineup performance prediction using network analysis | IEEE … https://ieeexplore.ieee.org/abstract/document/9073145/. Accessed May 5, 2023.
3. Players. Official NBA Stats | Stats | NBA.com. https://www.nba.com/stats. Accessed May 5, 2023.
4. Swar. Swar/NBA_API: An API client package to access the apis for nba.com. GitHub. https://github.com/swar/nba_api. Accessed May 5, 2023.
5. Radicchi F. Who is the best player ever? A complex network analysis of the history of Professional Tennis. PLOS ONE. https://doi.org/10.1371/journal.pone.0017249. Accessed May 5, 2023.
6. Ryanabest. The best NBA players, according to Raptor. FiveThirtyEight. https://projects.fivethirtyeight.com/nba-player-ratings/. Published May 5, 2023. Accessed May 5, 2023.

Acknowledgements

Thanks to Professor YY for recommending papers to reference.