The Evolution of the Optimal NBA Player

Angus Lam
anguslam228@berkeley.edu

May 5, 2021

Table of Contents

I. Introduction

The game of basketball has been around for centuries, initially created as a fun alternative to boring physical education classes. Throwing a ball into a peach basket in the late 1800s had no particular meaning whatsoever, but over the course of its existence, basketball has become one of the most popular sports worldwide. As the game slowly evolved from stationary shooting around the basket to rapid-fire shots from the three-point line, the optimal players started to look and play in many different ways. As an avid fan of the sport, I was intrigued by the vast amount of data that can be collected with today's advanced technology, and wanted to try to take a glimpse at which player archetype is the most important in today's game. Inspired by findings from Muthu Alagappan's 2012 project *From 5 to 13: Redefining the Positions in Basketball*, I'll be studying how today's game has evolved both physically and statistically.

II. Sources and Methods

Sources

I gathered data from multiple sources, including ESPN, Basketball-Reference.com, fivethirtyeight.com, and bball-index.com. Basketball-Reference.com provided basic statistics on each player in the National Basketball Association (NBA), such as Points per Game (PTS), Total Rebounds (TRB), and Assists per Game (AST). The other 3 sources non-linearly combined various basic stats to produce, in short, different numerical 'valuations' of each player. All data used is from the ongoing 2020-21 NBA season, so only around 45 games for each player have been taken into account thus far.

Players are normally categorized into one of five positions: Point Guard (PG), Shooting Guard (SG), Small Forward (SF), Power Forward (PF), and Center (C). Historically speaking, Guards are typically smaller in size and are the designated playmakers and jump-shot takers, while Forwards and Centers are the slow tanks that muscle their way to shoot shots closer to the basket.Therefore, in the stat sheet, we often assume that Guards take more jump shots and mainly focus on offense, while the Forwards and Centers take more shots under the basket and play defense.

Before the implementation of the three point line in 1986, the game was played at a much slower pace, and players were, on average, much larger and slower overall. However, with the rise of the three point shot in the 2010s, the game has radically changed, and the flow of the game has become much different. To tackle this change, I decided to measure each player based on their value for their respective teams, also known as Value over Replacement Player (VORP). This statistic is a derivative of the Box Plus Minus, and essentially measures how much impact a player adds for their team compared to another replacement player. In other words, this statistic gives a general idea of how important a player is for their team.

	Player	VORP	Pos
0	Kareem Abdul-Jabbar	8.7	C
1	Bobby Jones	5.5	PF
2	Bill Walton	5.0	C
3	Julius Erving	4.8	SF
4	Bob Lanier	4.6	C
5	Artis Gilmore	4.5	C
6	Elvin Hayes	4.1	PF
7	Paul Westphal	4.0	PG
8	George Gervin	3.8	SF
9	Billy Knight	3.7	SF

	Player	VORP	Pos
0	Nikola Jokic	7.2	C
1	Stephen Curry	4.6	PG
2	Giannis Antetokounmpo	4.5	PF
3	Luka Doncic	4.5	PG
4	Kawhi Leonard	3.6	SF
5	Zion Williamson	3.5	PF
6	Jimmy Butler	3.5	SF
7	Julius Randle	3.4	PF
8	Damian Lillard	3.4	PG
9	LeBron James	3.4	PG

Figure 2.1.1: Top 10 NBA Players with highest VORP, 1976-77 (L) vs 2020-21 (R)

As shown above, the game in the 20th century was heavily reliant on the large players, with a large proportion of Centers and Power Forwards. However, in today's game, the implementation of the three point line has allowed for much more movement, letting smaller Guards and Forwards thrive. As seen in the 2020-21 top 10 list, only one Center was among the top 10 players with the highest VORP.

Setting Edges and Nodes

Since the data for the entire NBA was too large to plot clearly, I decided to take the top 50 players (based on ESPN's rankings) as source nodes. However, target nodes were not limited to top 50 players, allowing for more connections with different players.

Edges were formed between source and target players if the target VORP scores were within +/- 0.25 of the source VORP scores. Node colors correspond with one of the five positions (PG, SG, SF, PF, C) or cluster (more on this later). Since my comparison is heavily correlated with the implementation of the three point line and its effects, node sizes correlate with either Three Point Attempts per Game (3PA) or Assists per Game (AST).

III. Results

Initial Network Graphs

Network graphs with the same edges and node properties are shown below, with sizes dependent on either 3PA or AST. Labels are of the top 10 VORP players along with a random assortment of target nodes.

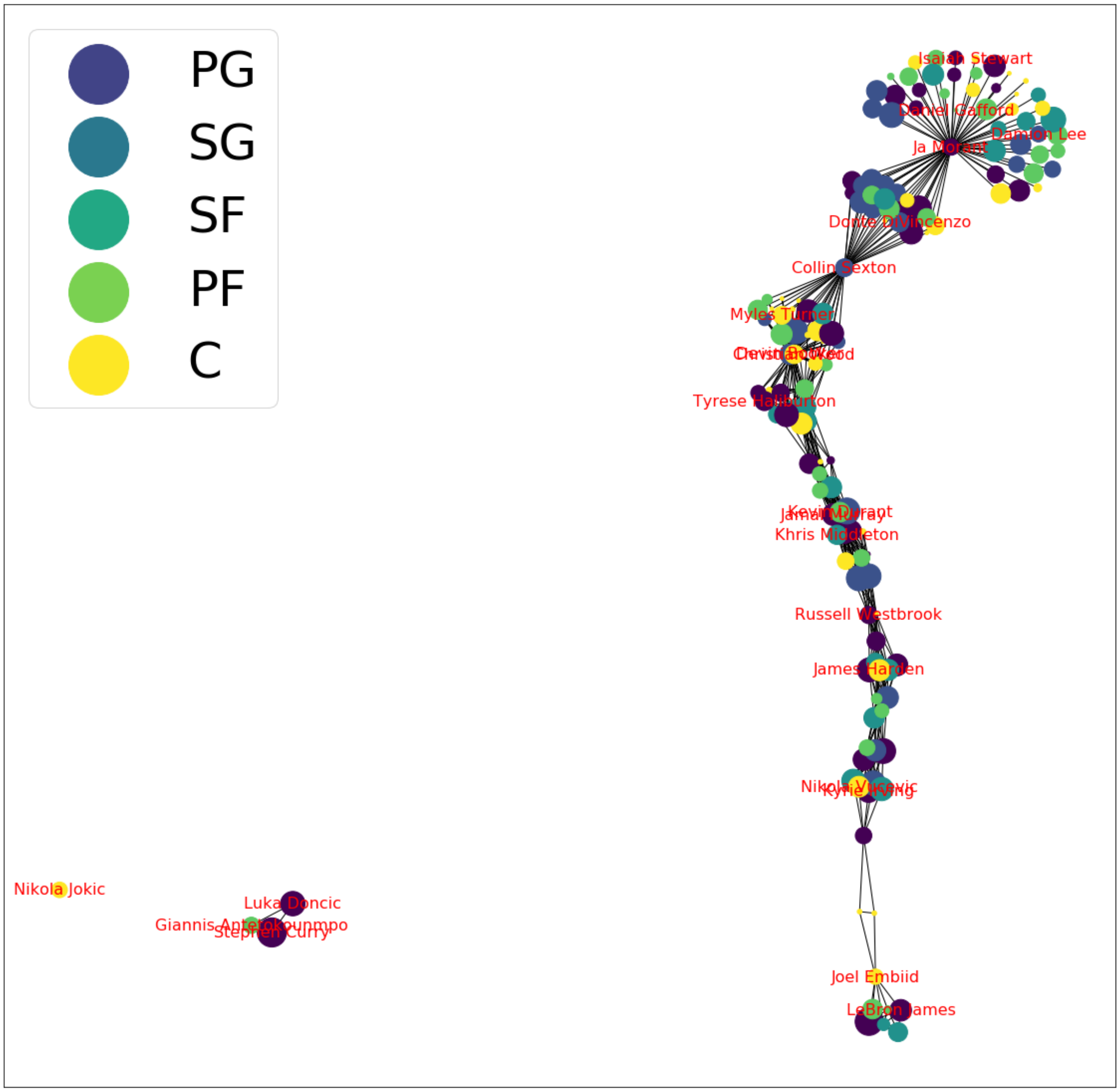

Figure 3.1.1: Network Diagram of Players based on Position, VORP, and 3PA (2020-21)

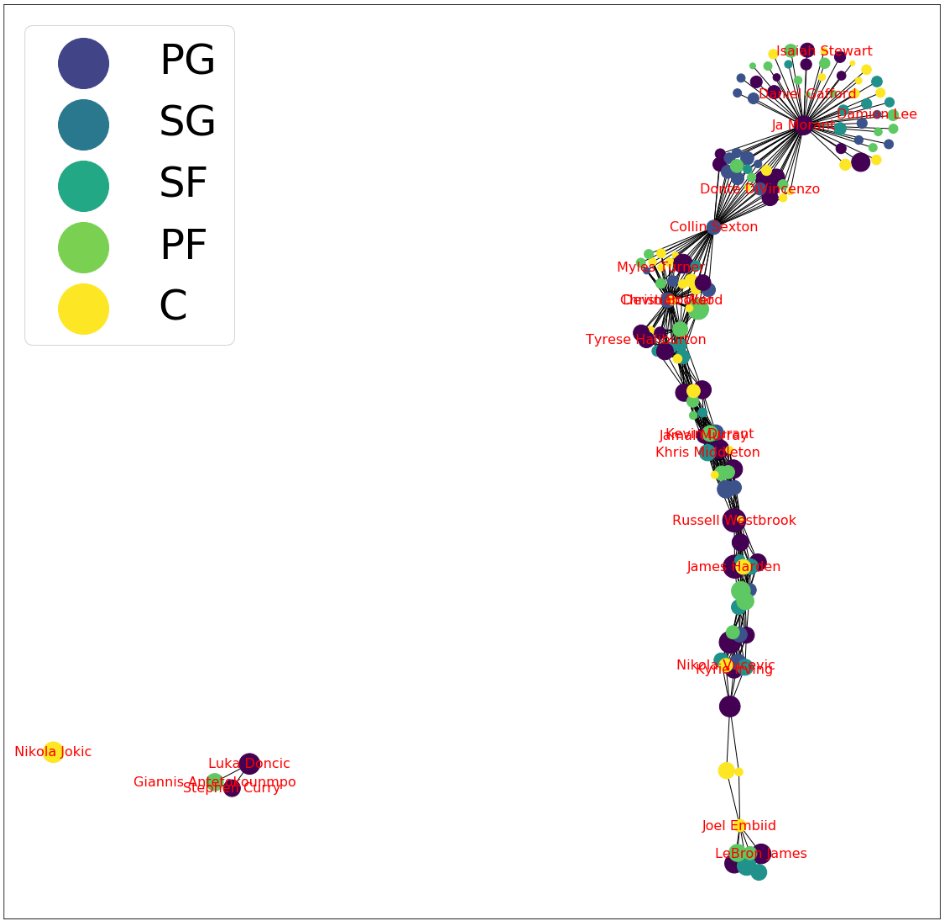

Figure 3.1.2: Network Diagram of Players based on Position, VORP, and AST (2020-21)

There are three groups here that I'd like to focus on:

1. The island of Nikola Jokic
    a. He is the only center in the top 10 highest VORP scores, and leads the pack by a wide margin at 7.2 (Figure 2.1.3). Among all centers, his 3PA of 3.3 is much greater than the 1.34 3PA average for centers. However, more evidently, his APG at 8.7 blows the competition away compared to the average 1.38 APG for Centers. Combining these outstanding statistics with his elite rebounding and scoring, it's no surprise that he stands as the MVP frontrunner for the 202-21 season.

2. The trio of Stephen Curry, Giannis Antetokoumpo, and Luka Doncic
    a. These 3 players are #2-4 from the top 10 VORP list, two of them being guards. Stephen Curry and Luka Doncic take 12.0 and 8.4 threes per game, respectively, much higher than the league average 3.02 3PA. If we take their assists, they also beat the competition, but not by as large a margin as Nikola Jokic: 5.7 and 8.7 respectively, compared to the average 4.2 for PGs. Their value comes from their elite production of prototypical Guard duties of scoring and playmaking.

3. Ja Morant and his surrounding nodes
    a. For the casual basketball fans, many know Ja Morant as the 2019-20 Rookie of the Year and the main star of the Memphis Grizzlies. However, his VORP is not actually very high; in fact, his role player teammate Kyle Anderson has a VORP of 1.8, higher than Ja's 0.6. This could be attributed to his injury earlier in the season, but another big factor could be his lack of a three point shot. He takes an average of 3.7 threes per game, much lower than the elite production of Steph and Luka from above. Not only does he attempt very few, he's also quite inaccurate, sinking an average of 29.5% of such shots, significantly lower than the 35.1% average at the PG position.
    His lack of an efficient three point shot limits his offensive options, and thus his overall value suffers. The nodes surrounding him with similar VORP scores also tend to take limited numbers of threes, such as Isaiah Stewart (C) and Daniel Gafford (PF), implying some correlation between 3PA and players' VORP.

Considering the variety of positions within each cluster in the network graph, I decided to use K-means clustering to essentially 'reclassify' players into five groups and see what mix of players would fall in each group.

Clustering

Since players are still largely measured by the five basic statistics today (PTS, TRB, AST, STL, BLK), I clustered players based on those variables, along with their VORP. The averages for each statistic in each group are shown below:

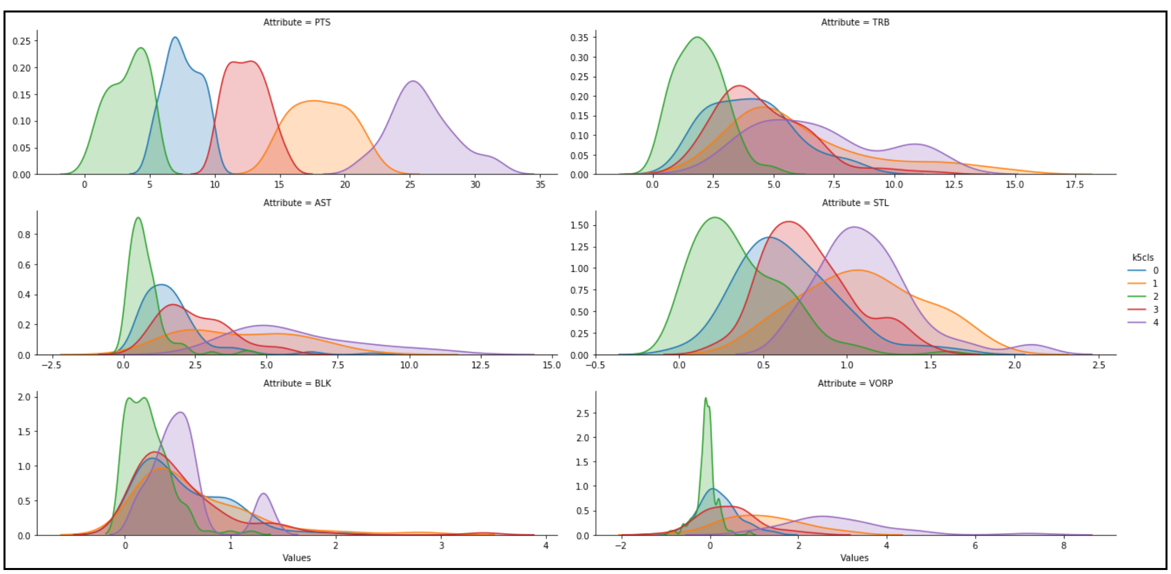

Figure 3.2.1: Distribution Plot of Cluster Variables

k5cls	0	1	2	3	4
PTS	7.541	18.170	3.333	12.467	25.952
TRB	4.106	6.149	1.963	4.529	6.893
AST	1.707	4.074	0.778	2.515	6.003
STL	0.652	1.096	0.362	0.773	1.117
BLK	0.542	0.651	0.226	0.536	0.562
VORP	0.186	1.176	-0.072	0.438	2.889
Count	124.000	47.000	120.000	85.000	29.000

Figure 3.2.2: Clusters with Statistical Averages

k5cls	Pos	
0	C	35
	PF	27
	PG	19
	SF	15
	SG	28

k5cls	Pos	
1	C	8
	PF	8
	PG	14
	SF	6
	SG	11

k5cls	Pos	
2	C	16
	PF	35
	PG	15
	SF	26
	SG	28

k5cls	Pos	
3	C	18
	PF	16
	PG	17
	SF	13
	SG	21

k5cls	Pos	
4	C	4
	PF	4
	PG	10
	SF	5
	SG	6

Fig 3.2.3: Distribution of Player Positions in Each Cluster

By far the most prominent distinction between groups is PPG averages, but different distributions of TRB and AST show the nuances between each cluster. From here, there are two main takeaways:

1. Cluster 4 is composed of the most elite players in the league. Aside from BLK, all other stats are elevated across the board, with the highest average VORP of 2.889. Interestingly enough, a majority of the positions within this cluster are either PG or SG, implying the rising value of smaller guards in today's version of the game.
2. The other 4 clusters have blended skill sets, but specialize in their own areas. For example, Cluster 0 has an unusually high TRB/PTS ratio at 4.1/7.5, or around 55%. Compared to Clusters 1 and 3, which have ratios at or under ~36%, Cluster 0 seems to be more defense-focused, grabbing rebounds instead of scoring points. This observation is supported by the position breakdown within each cluster in 3.2.3 , with the highest concentration of Centers and Power Forwards out of all clusters.

Given these results, I then re-graphed the previous network graph, but instead colored nodes by their clusters.

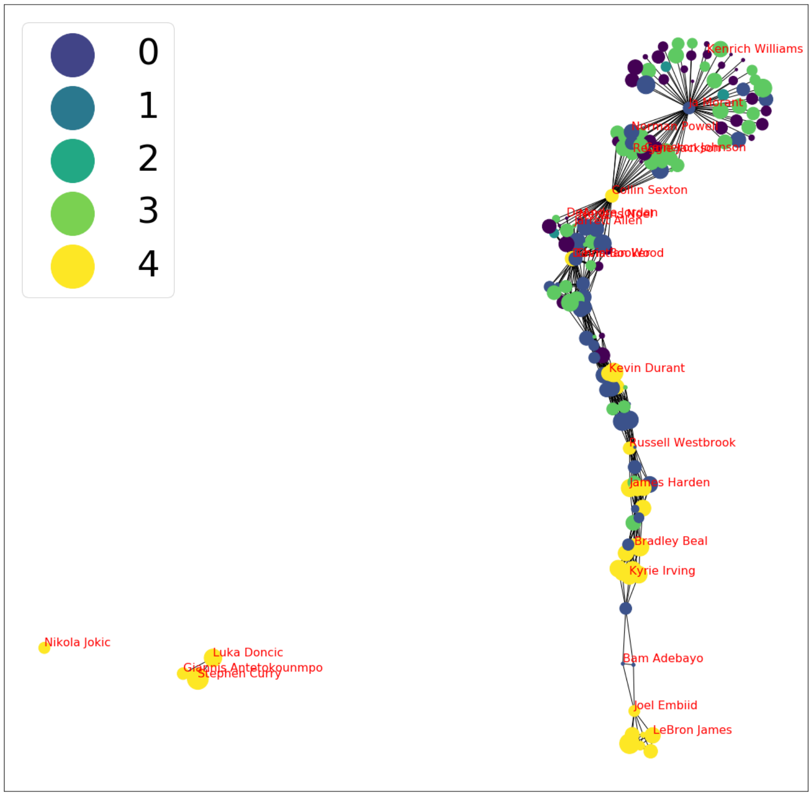

Figure 3.2.4: Network Diagram of Players based on Cluster, VORP, and 3PA (2020-21)

From this graph, we can now clearly see a more distinct separation between the newly classified ‘positions’. These more distinct clusters may signal a better classification method for players based on their statistics rather than their traditional positions.

IV. Conclusion and Future Work

Through my analysis and visualizations, it seems that guards are becoming much more important in basketball than ever before. The game has shifted from the tough, gritty big man battles to a faster, fluid competition. Even with these results, there is still much room for improvement and further research. However, there were many caveats that could not be taken into account.

Perhaps the biggest flaw in this project has been the absence of physical statistics. Size still plays heavily into a player’s statistics regardless of their skillset. A 5’ 9” point guard will very unlikely grab a rebound over a 7’ center, so many physical limitations need to be taken into account. Another caveat has been the nature of the 2020-21 season and its adjustments to the COVID-19 pandemic. Teams are having to adjust to constantly changing game schedules, which could present a heavy mental obstacle that affects their production. More importantly, the availability of players this season are very volatile, with players needing to self-isolate and miss games, constantly shifting responsibilities to different players.

Playing time is also a major factor that would influence such results. Players see different lengths of playing time depending on opponents and injuries, so statistics are heavily influenced as a result. Less playing time means less opportunities to rack up any statistics, and coupled with missing teammates, heavily influences the results shown on the stat sheet. Percentages may be a better metric as it depicts a player's efficiency regardless of playing time.

From here, I hope to take a deeper look into the past and observe how these differences can be better represented. There was no three point line in 1976, so the skills needed would be much different. A major ongoing debate is the dying art of the fundamental two-point shot: why settle for 2 points when you can score 3? I'd like to explore the different playing styles and how those features could be quantified. Teams also have different identities, so each team's needs are unique; one player's VORP on their current team may be much different if they play on teams with different dynamics. There are many other metrics that can be used to connect edges with player nodes, so there's still many directions left to explore.

Basketball is definitely more than a game of statistics, but they can provide a new perspective to the ideal players needed to succeed. Hopefully this project provided casual fans with a surface-level analytical understanding of the evolution of the optimal player in today's game.

V. References and Sources

"2020-21 LEBRON Data - Basketball Index." *Basketball Index*, 3 May 2021, www.bball-index.com/2020-21-lebron-data/. Accessed 4 May 2021.

"2020-2021 NBA Season Leaders - National Basketball Association - ESPN." *ESPN.com*, 2020, www.espn.com/nba/seasonleaders/_/league/nba/page/1. Accessed 4 May 2021.

Alagappan, Muthu. 2012, *From 5 to 13: Redefining the Positions in Basketball*. https://www.wired.com/2012/04/analytics-basketball/. Accessed 4 May 2021.

Basketball Reference. "2020-21 NBA Player Stats: Advanced | Basketball-Reference.com." *Basketball-Reference.com*, 2020, www.basketball-reference.com/leagues/NBA_2021_advanced.html. Accessed 4 May 2021.

Basketball Reference. "1976-77 NBA Player Stats: Advanced | Basketball-Reference.com." *Basketball-Reference.com*, 2021, www.basketball-reference.com/leagues/NBA_1977_advanced.html. Accessed 4 May 2021.

Best, Ryan. "The Best NBA Players, according to RAPTOR." *FiveThirtyEight*, FiveThirtyEight, 12 Jan. 2021, projects.fivethirtyeight.com/nba-player-ratings/. Accessed 4 May 2021.