Final Report

Andrew Fu, Aidan Orr, Adam Rogers

Due 5/13/2022

Introduction

The National Basketball Association is one of the most viewed sports leagues around the world and is quickly growing in popularity. At that level of play, teams look to squeeze out every possible advantage. With the rise of basketball analytics departments across the NBA, strategies and game-play have evolved to become even more efficient. Furthermore, NBA players have increased in skill on average, with players having more diverse and well-rounded skill-sets. Many NBA pundits frequently bemoan that NBA games have devolved to one-man stage shows against a porous defense – a world-class athlete attempting a stunning, highlight reel move from mid-court to score the basket on his own, while his four teammates step to the side for the sole, limited purpose of drawing the defense away from the paint. Is offensive teamwork and passing the ball on the wane and, if so, what are the implications? We want to use data science to visualize any potential changes in games, particularly on the offensive side.

One big change we have noticed is the way players "assist" each other. We have noticed that players from all positions seem to be contributing more assists as the league is becoming less position-oriented on offense. The traditional roles each position is thought to fill have become less rigid. There are five positions in the NBA. Traditionally, The point guard, who we have already mentioned, is generally the smallest player on the court and is thought to contribute the most assists by facilitating good shots for other players. The shooting guard is also smaller and is generally understood to be the best shooter on the floor. The two forwards (small forward and power forward) are mid-sized and thought to have a relatively balanced skill set. Finally, the center is traditionally the largest player on the floor and is known to score lots of points close to the basket. (*What are the 5 Positions in Basketball?*) However, it is clear this dynamic is changing. The introduction of the three point line has seen modern basketball players of all sizes (particularly point guards) becoming much more adept scorers, disrupting the traditional offenses dominated by point guards who primarily assisted on plays as opposed to score. (*The Modernization of NBA Offenses and Why Small Ball Is Here to Stay*) For example, this past year's MVP (Most Valuable Player), Nikola Jokic, averaged nearly 8 assists per game. (*Nikola Jokić Stats*). Jokic is a center, the position generally filled by the tallest player on the floor and is not known for contributing a lot of assists. Proving this point, Shaquille O'Neil, one of the most famous centers of all time who was active during the late-90s and early-2000s, only averaged 2.5 assists per game over the course of his career. (*Shaquille O'Neil Stats*)

Broadly, these changes in the league have lead us to ask the following question: Has there been a shift in NBA offensive strategies over time? In order to tackle this broader question, we posed a more specific question we could tackle using data analysis: What are the assist relationships between the different positions? To answer this question, we tracked how NBA assist patterns between the 5 positions have shifted over time.

Data and Methods

We captured the data from the *hoopR* package. Specifically, the package includes both 1) the functions load_nba_pbp() which gives us a data table with every play from every game of an indicated season and 2) espn_nba_player_box() which gives us a data table with the position of each player of an indicated game.

When calling for the play-by-play data in a season, we are given every play from every game from that year. Since we are only interested in assists, we had to filter the data for the word "assist" and perform string extraction using regular expressions to transform the data into a form with a passer column, a shooter column, and a column indicating the game ID code.

Player positions can change between games, but not within a game. Therefore, we needed to grab each player's position for each game in a given season. Since the espn_nba_player_box() only takes game IDs as arguments, we looped that function for every game in a given season. We did this by initially creating a table for an arbitrary game in 2018 (filtered out later) which has two columns for each player, one for passer and one for shooter, and the game ID. At the start of the loop, a new table was made for the first game in the season, combined with the 2018 table, and given the same object name. Then, when the second game was called, that table of data from the second game was combined with the previously established table with the 2018 game and the first game. This process was repeated for every game in a specified season. Finally, the 2018 game was filtered out, giving us a data table with every player and their position for every game in a certain season.

We then joined this position table with the player assist table once to assign the passer position and joined another time to assign the shooter position. Finally, we grouped the data by passer and shooter position and counted the number of interactions with certain positions and NAs removed. Specifically, a few players listed as general "guards" or "forwards" were filtered out since we are looking at the five main NBA positions (center, point guard, shooting guard, power forward, and small forward). Finally, since this chunk takes about an hour to run per season, we wrote CSVs for each year's data which will be read for later visualizations.

This table was used to create the igraph with weights influencing a node's hub and authority score. A hub is a node which has many outgoing links. An authority is a node with many incoming links. For our data, hubs represent players who assisted on many plays, while authorities were players who scored many points off of assists. Hub and authority scores are calculated using an algorithm. First each node is assigned a hub score and authority score of 1. Then the hub score is updated as the sum of authority scores of the nodes it points to. Similarly, the authority score is updated as the sum of hub scores of the nodes that point to it. The hub score is then normalized by dividing each hub score by the square root of the sum of the squares of all hub scores. The authority score is similarly normalized by dividing each authority score by the square root of the sum of the squares of all authority scores. This process is repeated as necessary (See the *HITS algorithm*). The final network was plotted twice with node size depicted by that node's hub score in one visualization and node size depicted by authority score in the other.

That same table was also transformed into a matrix with correct column and row names with the matrix() function. Using the levelplot() function in the *lattice* package, we plotted a heatmap with a scale and correct row and column orientation displayed. The scales were kept the same among assists and field goals between different years since this data did not significantly change per season.

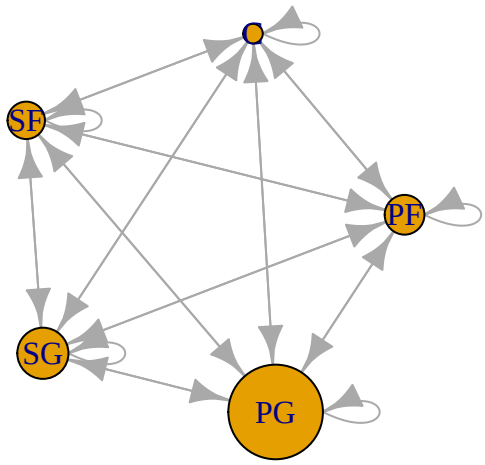
Results

Networks

In the 2003 hub network, we can see point guards do most of the assisting, while the other positions have comparably small contributions to assisting. The authority network shows that power forwards make most field goals after being passed to, while shooting guards and small forwards have the next largest effect.

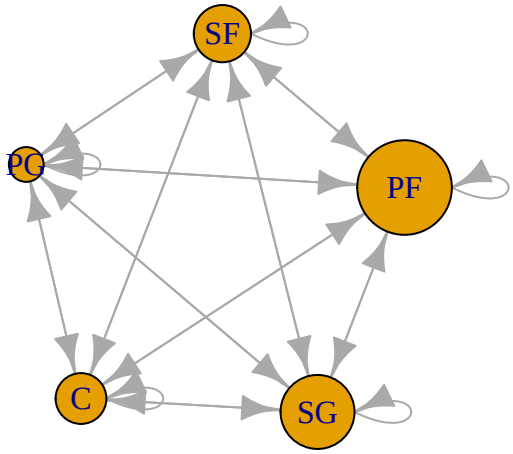
```
# Read CSV
assist_pos_2003 <- read_csv("assist_pos_2003.csv", show_col_types = F)

# Store elements of table
assist_el_2003 <- assist_pos_2003 %>%
 select(passer_pos, shooter_pos) %>%
 as.matrix()

# Create igraph from elements
assist_el_igraph_2003 <-
  graph_from_edgelist(el = assist_el_2003, directed = TRUE)

# Store edge weights from count
assist_el_igraph_2003 <- set_edge_attr(graph = assist_el_igraph_2003,
                   name = "count",
                   value = assist_pos_2003$count)

# Calculate hub score for each position
hs_2003 <- hub_score(assist_el_igraph_2003,
                weights = E(assist_el_igraph_2003)$count)$vector
# Calculate authority score for each position
as_2003 <- authority_score(assist_el_igraph_2003,
                     weights = E(assist_el_igraph_2003)$count)$vector
# Plot hub score for each position
plot(assist_el_igraph_2003, vertex.size=hs_2003*50,
     main="2003 Season Hubs")
```

2003 Season Hubs



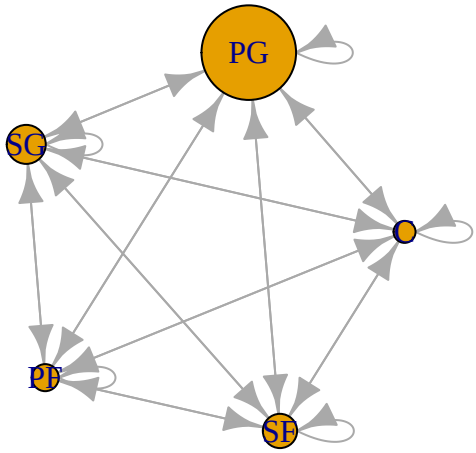

```
# Plot authority score for each position
plot(assist_el_igraph_2003, vertex.size=as_2003*50,
     main="2003 Season Authorities")
```

2003 Season Authorities

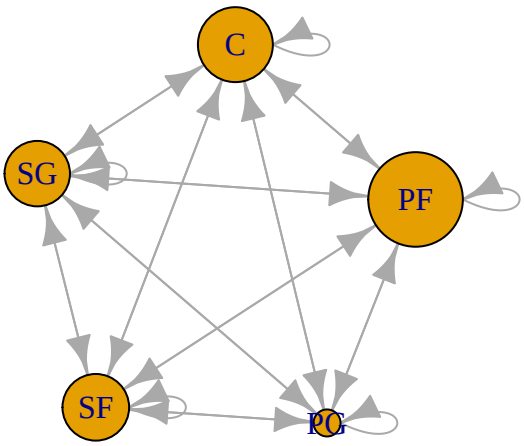


Similar to 2003, the 2011 hub network shows that point guards are the primary contributor to assists while the other positions have little effect. The authority network from 2011 shows again that power forwards make most field goals after being passed to, while shooting guards, small forwards, and centers have the next largest effect. The only difference between 2003 and 2011 is that point guards have a more polarizing effect on assists and centers contribute more to making shots after being assisted.

2011 Season Hubs

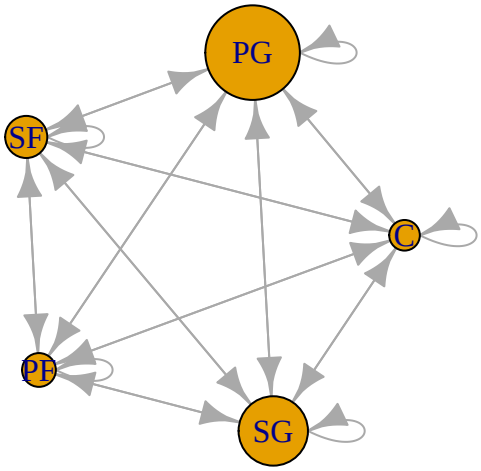


2011 Season Authorities

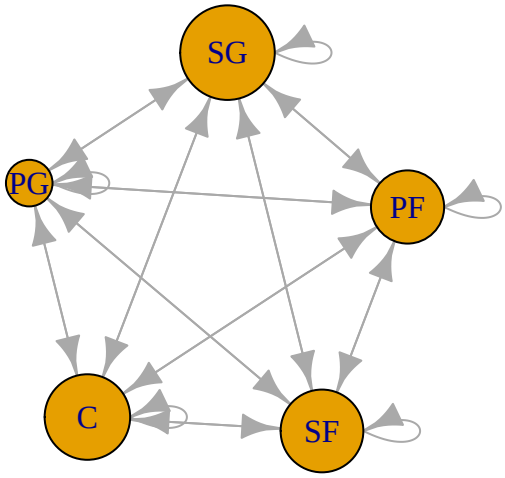


Finally, the 2022 hub network shows that point guards are still the primary position contributing assists, and shooting guards contribute more to assists that season. The authority network shows that shooting guards, small forwards, power forwards, and centers all have a similar amount of field goals after being assisted.

2022 Season Hubs



2022 Season Authorities

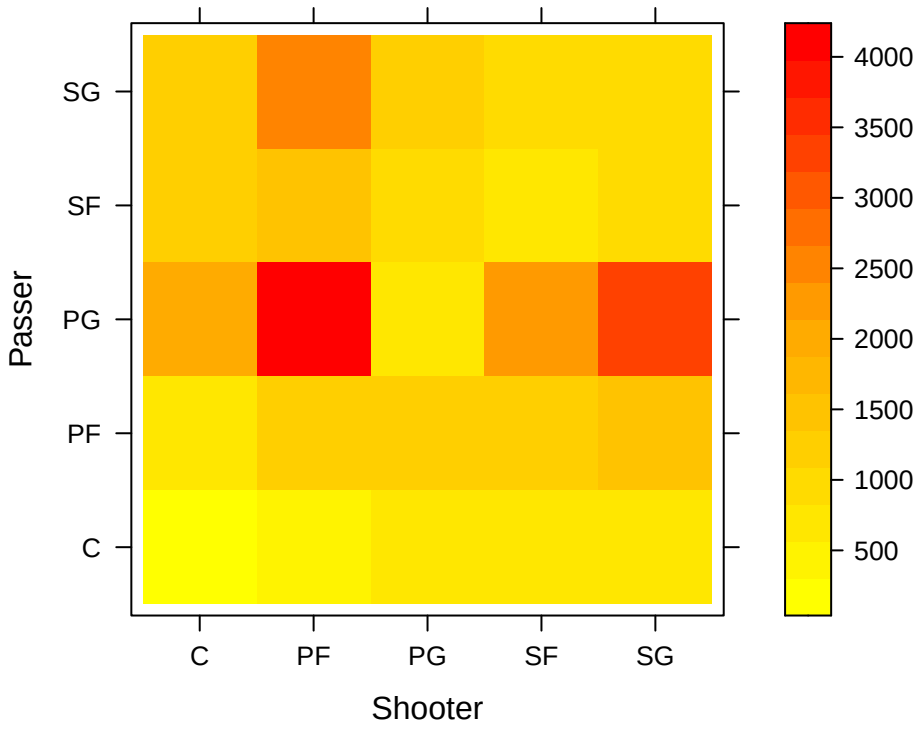


We found that hubs and authority plots were not the best way of visualizing our data since the data we are looking for is quantitative. The qualitative network does not tell the entire story of the data. While we can easily compare nodes to each other in size, there is no indicator of the scale of the hub and authority scores. Additionally, while we can see which positions passed or shot the most, the relationships between two positions is not clear. In our process, we tried setting edge size or color to the number of links between nodes, but the 50 directed edges between our 5 nodes made the network unintelligible.

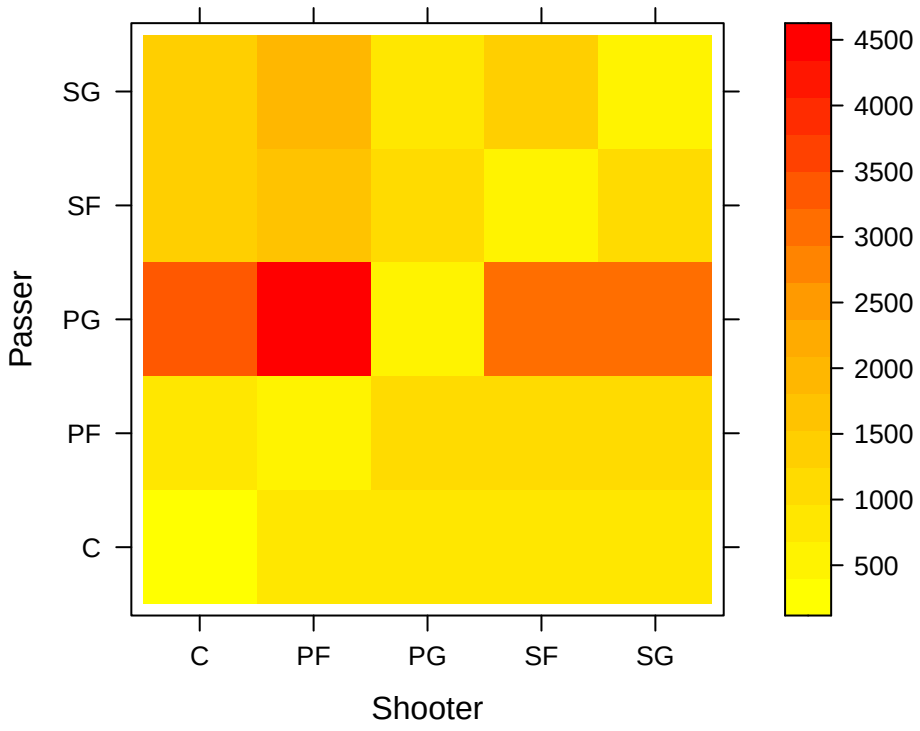
Heatmaps

The heatmap visualization was much better for simultaneously exploring differences in both passer and shooter weights among all positions. Importantly, one can clearly see the assist relationships between two certain positions. For example, point guards in 2003 and 2011 are the primary players contributing assists. However, in the 2022 season, shooting guards and small forwards play a much more impactful role in assisting. Additionally, the diagonals tell us how the league has changed in recent years. In 2003 and 2011, there were few interactions between two players of the same position on the court. In 2022, however, there is a strong relationship from one shooting guard to another.

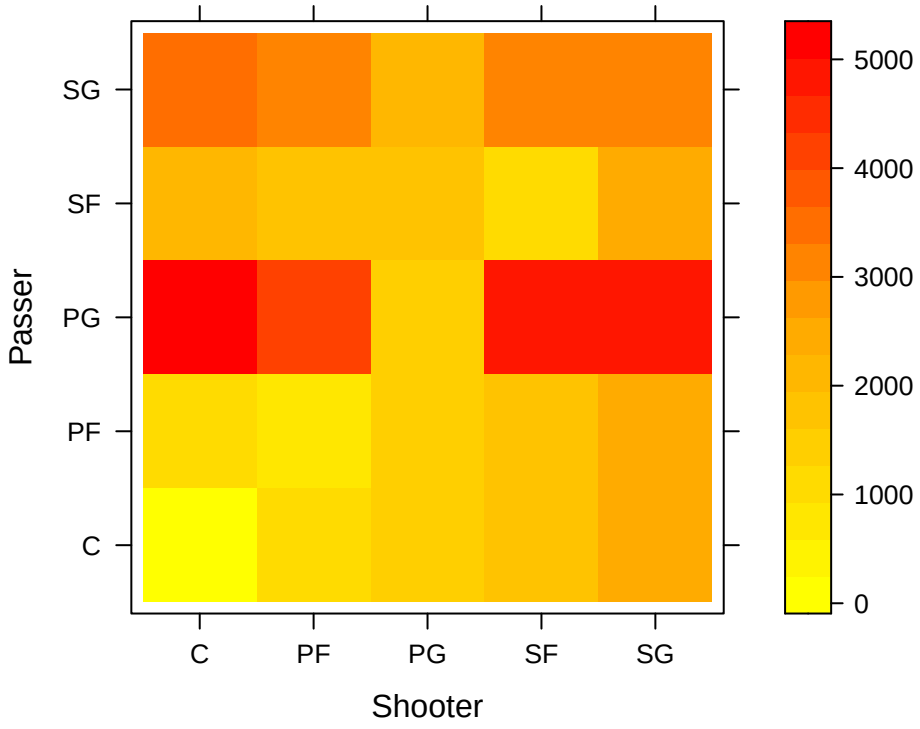
```
# Transform tables into matrices with correct row and column names
mat_2003 <- matrix(assist_pos_2003$count, nrow = 5,
                   dimnames = list(c("C", "PF", "PG","SF","SG"),
                                   c("C", "PF", "PG","SF","SG")))
mat_2011 <- matrix(assist_pos_2011$count, nrow = 5,
                   dimnames = list(c("C", "PF", "PG","SF","SG"),
                                   c("C", "PF", "PG","SF","SG")))
mat_2022 <- matrix(assist_pos_2022$count, nrow = 5,
                   dimnames = list(c("C", "PF", "PG","SF","SG"),
                                   c("C", "PF", "PG","SF","SG")))

# Set color palette
col.1 <- colorRampPalette(c('yellow', 'orange', 'red'))

# Draw heatmap with correct titles and axis labels for 2003
levelplot(mat_2003, col.regions = col.1,
          main = "2003 Season Heatmap", xlab = "Shooter", ylab = "Passer")
```




```
# Draw heatmap with correct titles and axis labels for 2011
levelplot(mat_2011, col.regions = col.1,
          main = "2011 Season Heatmap", xlab = "Shooter", ylab = "Passer")
```




```
# Draw heatmap with correct titles and axis labels for 2022
levelplot(mat_2022, col.regions = col.1,
          main = "2022 Season Heatmap", xlab = "Shooter", ylab = "Passer")
```


2022 Season Heatmap
Passer
SG
SF
PG
PF
C
C
PF
PG
SF
SG
Shooter
5000
4000
3000
2000
1000
0

Conclusion

Admittedly, one big limitation to our work is that we only tracked three years of assist data (2003, 2011, and 2022). The earliest assist data we could find for a complete season was 2003, so we could only analyze trends over the course of the past ~20 years. With more computational power we could analyze every single year of NBA data, allowing us to trace exactly when the key shifts in assist patterns occurred.

Nevertheless, our data analysis concluded that positions are becoming much more dynamic and flexible. Decades ago, each of the five positions called for very specific body types and skill sets – guards were shorter, better ball handlers, and more adept at perimeter shooting. Centers, in contrast, lacked the dribbling and outside shooting skills, so they planted themselves under the rim for rebounds and dunks.

One important caveat to note is that although a basketball team has 5 players on the floor and there are 5 positions, the team does not need to have one of each position on the floor at all times. As the 2022 heatmap indicates, t is much more common to see two shooting guards or two point guards on the court at the same time, a strategy often called "small ball" that is intended to increase an offense's flexibility and speed.

As the game evolved, however, those skill set requirements became more fluid. For example, the archetypal point guard in earlier years was a short, quick, highly technical player who can shoot long range. This season, 6' 11" Nikola Jokić often filled the role of the traditional point guard in one game and power forward or center in the next.

References

Outside of the packages learned in class, we used the *hoopR* and *lattice* packages.

Outside sources for basketball info:

Biderman, D. (2009, April 1). The NBA's Most Misleading Number. Wall Street Journal. https://www.wsj.com/articles/SB123855027541776617

Cheung, M. (2019, October 30). POINT CENTERS, A NEW BREED IN THE NBA. Sports Analytics Group Berkeley. https://giphy.com/embed/404

HITS algorithm. (2021). In Wikipedia. https://en.wikipedia.org/w/index.php?title=HITS_algorithm&oldid=1008455356

Nikola Jokić Stats. (n.d.). Basketball-Reference.Com. Retrieved May 12, 2022, from https://www.basketball-reference.com/players/j/jokicni01.html

Shaquille O'Neal Stats. (n.d.). Basketball-Reference.Com. Retrieved May 12, 2022, from https://www.basketball-reference.com/players/o/onealsh01.html

Fenichel, A, (2021, March 25). The Modernization of NBA Offenses and Why Small Ball Is Here to Stay | The Analyst. Retrieved May 12, 2022, from https://theanalyst.com/na/2021/03/the-modernization-of-nba-offenses-and-why-small-ball-is-here-to-stay/

What are the 5 Positions in Basketball? (2021, March 12). Https://Sportsintherough.Com/. https://sportsintherough.com/blog/what-are-the-5-positions-in-basketball/