CSC111 Final Report: The National Basketball Association's Statistically Best Players

Junwei Quan, Vincent Louie

Saturday, April 1, 2023

1 Introduction

The National Basketball Association (NBA) is one of the most popular professional sports leagues in the world. Throughout its history, countless talented players have graced the courts and entertained millions of fans. One of the key aspects of any professional sport is the recognition of exceptional talent through awards. The NBA has several awards that recognize various achievements, such as Most Valuable Player (MVP), Rookie of the Year, and Defensive Player of the Year, among others.

Given the vast number of players and awards, it becomes an interesting challenge to analyze the relationships between players and their achievements throughout their careers. To address this challenge, our project aims to create a visual representation of NBA players and their respective awards, enabling us to examine the correlations and draw insights from the data.

Our main goal is to create an "AwardNetwork" using Python programming that represents the connections between NBA players and the awards they have achieved. This network will be visualized using graphs, allowing us to easily explore the relationships between players and their awards. By examining the AwardNetwork, we hope to answer questions such as:

- Which players have the most awards, and how do their careers compare to others?
- Are there any commonalities among players who have won specific awards?

To achieve our goal, we will use a dataset containing information on the top 80 NBA players and their awards, along with relevant background information such as age, career duration, and Hall of Fame status. We will then implement various Python classes and methods to parse the data, construct the AwardNetwork, and visualize the graph using libraries such as NetworkX, Plotly, and Matplotlib.

Ultimately, our project aims to provide a deeper understanding of the achievements of NBA players and the role that awards play in shaping their careers. We believe that this analysis will not only be of interest to basketball enthusiasts but also contribute to the broader understanding of performance recognition in professional sports.

2 Datasets description

In this study, we utilize two distinct datasets to analyze and compare the performance and achievements of players.

2.1 Official NBA Dataset

The first dataset is sourced from RealGM and we did some addition research to combine the dataset with age, status, year and some factors we have as the attributes of our class, which provides a comprehensive overview of individual player statistics and awards they have. Some of the awards and distinctions covered in the dataset include the Michael Jordan Most Valuable Player (MVP), Bill Russell NBA Finals MVP, All-NBA First Team, Hakeem Olajuwon Defensive Player of the Year and many others. This dataset offers a detailed overview of the players' accomplishments and provides valuable insights into their impact on the sport. The dataset also includes the name of the image icon that is associated to the specific player which is used in the visualization process.

2.2 Custom Awards Points Dataset

The second dataset is a custom creation that focuses on quantifying the value of various awards earned by NBA players throughout their careers.By assigning specific point values to each award, we aim to provide a comprehensive evaluation and comparison of player achievements. The point system takes into account the prestige and significance of each award, with more influential awards receiving higher point values. The dataset also includes the name of the image icon that is associated to the specific player which is used in the visualization process.

3 Computational Overview

Our project uses various types of data to represent the NBA players and their awards. The primary dataset contains information on the top 75 NBA players, including their names, ages, career durations, Hall of Fame status, and the awards they have achieved. We utilize trees and graphs to represent the connections between players and the awards they have won, forming the basis of our AwardNetwork.
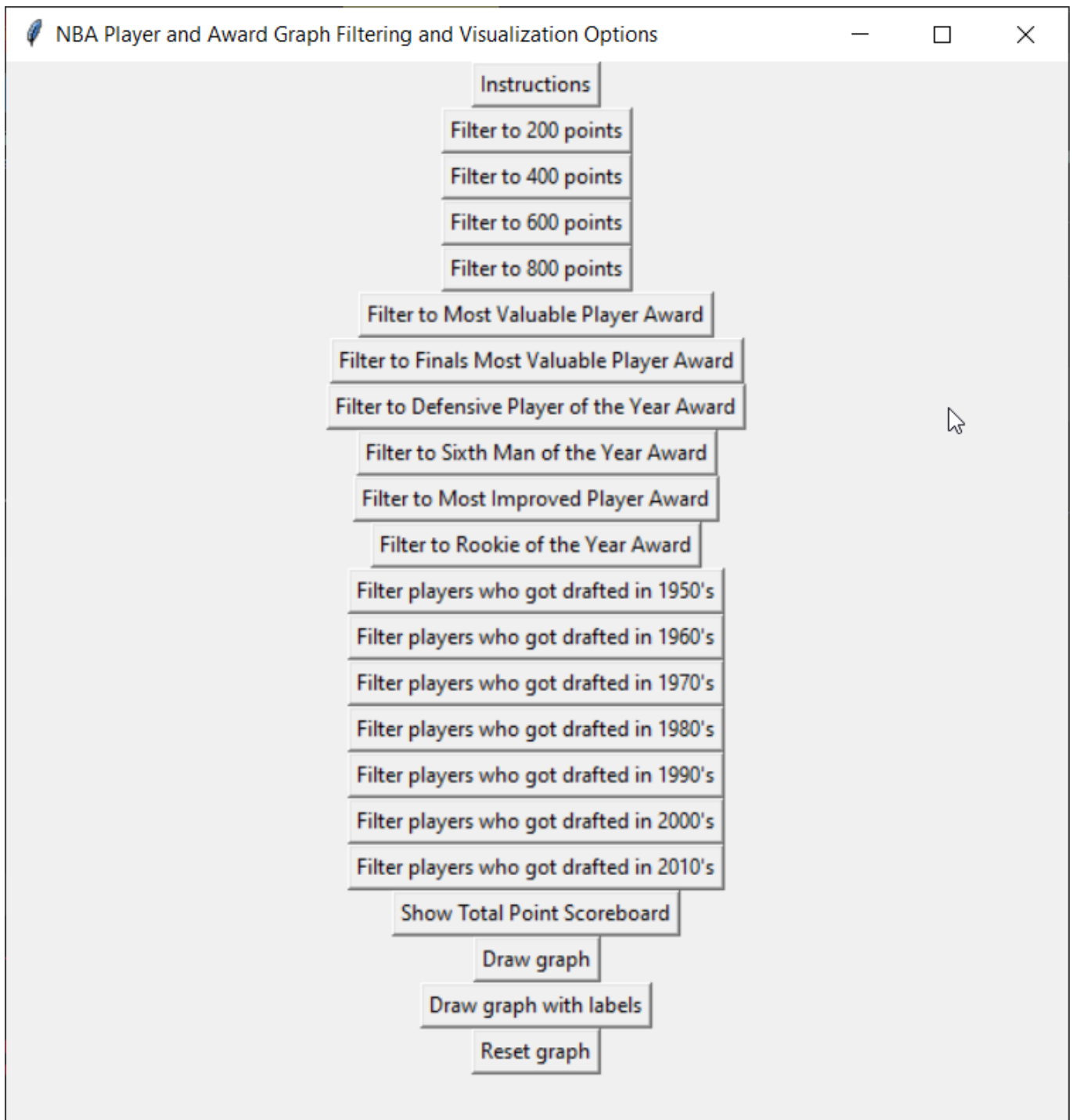
The major computations our program performs include:

1. Data parsing and representation: We read data from CSV files (`player_data.csv` and `award_data.csv`) and store them in lists. We then create custom classes to represent the relevant domain entities, such as `BasketballPlayer`, `Award`, and `Achieved`. These classes help us model the relationships between players and awards, with the `Achieved` class acting as an edge in our graph.

2. Building the AwardNetwork: We construct the AwardNetwork using the parsed data. Our `AwardNetwork` class contains methods to add players (`add_players`), awards (`add_awards`), and their achievements (`add_achieves`). These methods utilize the `NetworkX` library to create nodes and edges, forming the graph.

3. Data aggregation: We compute the total points for each player based on their awards using the `get_total_points` method in the `BasketballPlayer` class. This method calculates the total points for each player by multiplying the number of times they have won an award by the points associated with the award.

4. Visualization: Our program reports the results of the computation by visualizing the AwardNetwork graph using the `NetworkX`, `Plotly`, and `Matplotlib` libraries. The `draw_network` function takes the network and creates a `NetworkX` graph which is used as an input and generates a visual representation of the graph, with nodes representing players and awards, and edges representing the Achieved connections.

5. Library usage: We utilize several libraries to accomplish the tasks mentioned above. The `NetworkX` library plays a central role in constructing the graph and performing graph-related computations. We use functions such as `add_node`, `add_edge`, and `spring_layout` to create nodes, edges, and layout the graph. The `Plotly` and `Matplotlib` libraries are used to create the final visual representation of the graph, with functions such as `imshow`, `axis`, and `show` helping us display images, set axis properties, and present the graph.

Our program effectively models the relationships between NBA players and their awards, providing valuable insights into the careers of these exceptional athletes. The use of trees and graphs, along with the appropriate libraries, allows us to parse, represent, and visualize the data in a meaningful and engaging manner.
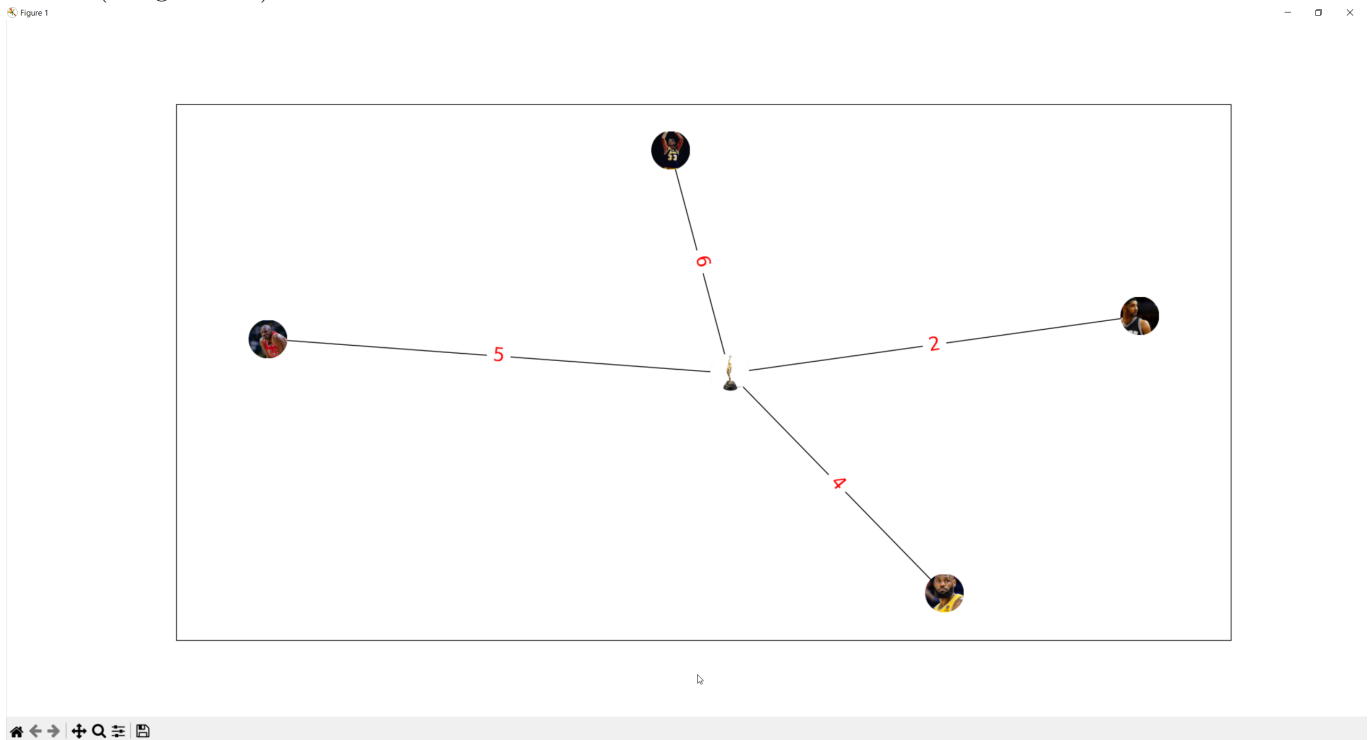
4 Instruction

After downloading the ZIP file which includes everything needed for our program, ensure that everything from the requirements.txt is installed. After this, you will be able to run the main.py file which will give a new window which is our graphical user interface that is used to see the graph visualizations, which will look like this:

Instructions

Filter to 200 points

Filter to 400 points

Filter to 600 points

Filter to 800 points

Filter to Most Valuable Player Award

Filter to Finals Most Valuable Player Award

Filter to Defensive Player of the Year Award

Filter to Sixth Man of the Year Award

Filter to Most Improved Player Award

Filter to Rookie of the Year Award

Filter players who got drafted in 1950's

Filter players who got drafted in 1960's

Filter players who got drafted in 1970's

Filter players who got drafted in 1980's

Filter players who got drafted in 1990's

Filter players who got drafted in 2000's

Filter players who got drafted in 2010's

Show Total Point Scoreboard

Draw graph

Draw graph with labels

Reset graph

In this GUI, you will have many buttons but you should start with clicking the instructions button that will pop up another window to read. After reading this then you should have a good idea of how the GUI and program works. When you draw a graph, there should be another window that appears which displays the plot that you have just drawn, whether it was the original graph or a filtered version. This window can be expanded to get a better view of the graph. The new window with the graph visualization should look something like this (with PyCharm)
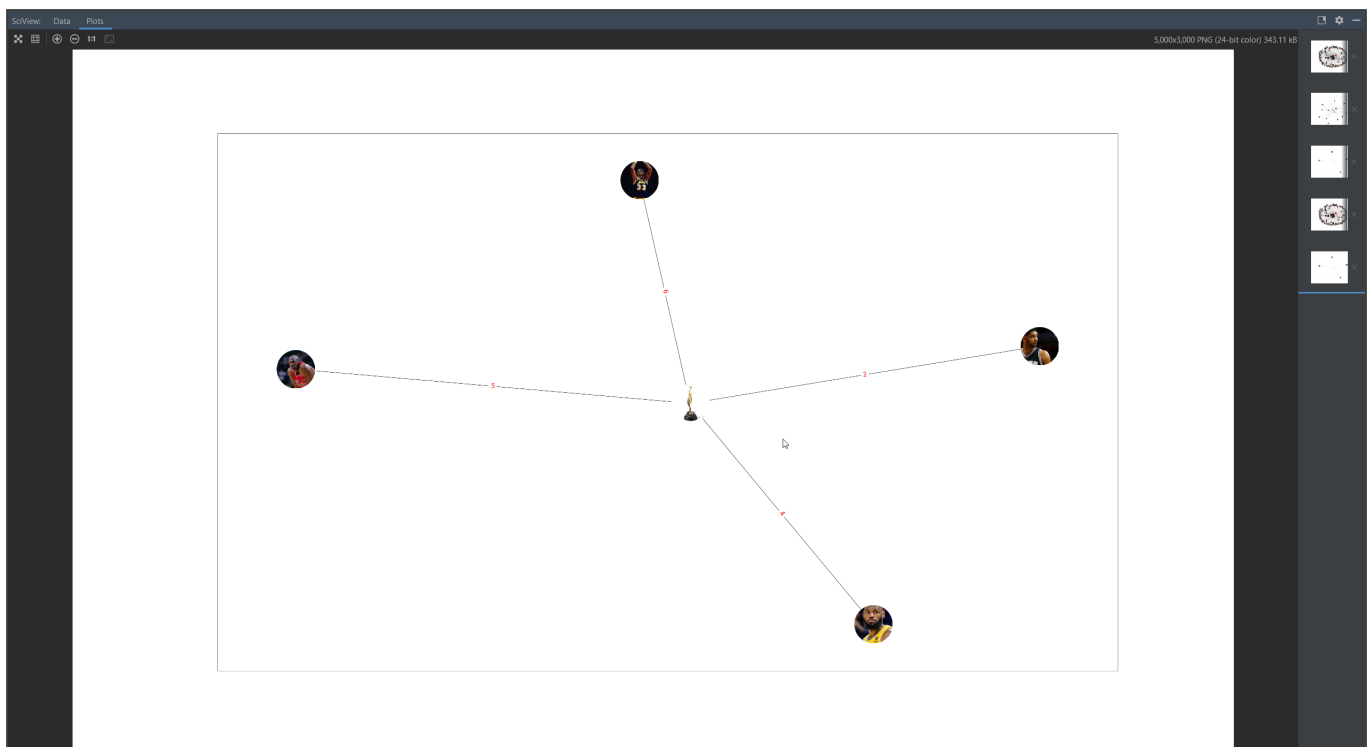
Or this (using console):

You can experiment with the filters, keeping in mind the instructions provided in the GUI instruction message popup.

5 Changes to the Project Plan

Throughout the course of our project, several changes have been made to our initial plan in response to TA feedback, discussions with the instructor during office hours, and our own exploration of ideas. Here, we summarize the main changes we have made to our project since the initial proposal.

5.1 Data Collection and Analysis

In our initial proposal, we focused on analyzing game scores above 30. However, after considering the feedback and thinking about the amount of data required to generate meaningful insights, we feel the case with over 30 game scores is too extreme and can not give a clear insight into players performance. Thus, we decided to focus our analysis on season and history awards. This change allowed us to perform a more comprehensive analysis of player performance and better address our research questions.

5.2 Libraries and Tools

Based on the suggestions in the TA feedback, we have made changes to the libraries and tools we used for our project. We decided not to use Pandas, Pygame, or other libraries for data management, visualization, and interaction with users. Instead, we focused on using Python's built-in data structures and functionalities, as well as NetworkX, Plotly, and Matplotlib for generating visualizations and statistics.

5.3 Bipartite Graphs

We initially planned to create a graph with players and teams as nodes. After learning about stuff in tutorial 8, we realized that our graph was, in fact, a bipartite graph, with players and awards as nodes and achieved as edges. We incorporated this concept into our project, which helped us better understand and analyze the structure of the AwardNetwork.

5.4 Interactive Visualization

Although we considered using Tkinter or other libraries to create an interactive GUI, we ultimately decided to focus on a more static visualization approach using Plotly and Matplotlib. This allowed us to present the data in a clear and concise manner while still providing meaningful insights into player performance and awards record.

5.5 Inline Citations

Following the TA's advice, we have included inline citations in our final submission to provide proper attribution and context for the sources we used in our project.

These changes have strengthened our project by providing a more comprehensive analysis of NBA player performance and their achievements. By addressing the TA feedback and incorporating new ideas, we believe that our project now offers a deeper understanding of the relationships between players and their awards, and the role of these awards in shaping their careers.

6 Discussion

Our primary goal was to create an "AwardNetwork" that represents the connections between NBA players and their awards, allowing us to explore the relationships between players and their achievements throughout their careers. In this section, we discuss the results of our computational exploration, the limitations we encountered, and the potential next steps for further investigation.

6.1 Results and Achievements

The AwardNetwork we created using Python, NetworkX, and other libraries has provided a valuable tool for visualizing and analyzing the relationships between NBA players and their awards. By examining the AwardNetwork, we were able to address our initial research questions:

- We identified players with the most awards and compared their careers to others, revealing key differences in terms of achievements and career trajectories. Specifically, the top two players in all of NBA according to our points is Michael Jordan and Lebron James. Following relatively closely after those two is Kareem Abdul-Jabbar. We can deduce that these are the top three players in NBA history using awards, considering the gap between the top three players and the next player.

- We discovered commonalities among players who have won specific awards, which could provide insights into the skills and attributes required to excel in particular aspects of the game. For example, the top three players, Michael Jordan, Lebron James and Kareem Abdul-Jabbar were all excellent at offense, able to score no matter the defender using their own signature moves. Michael Jordan was able to get a point from mid-range at will, Lebron James is unstoppable once he gets into the paint at full speed and Kareem Abdul-Jabbar was unguardable with his iconic sky-hook. All three players excelled at offence but all three also had impressive defensive skills and the longevity of Kareem and Lebron is extremely impressive. All these skills and attributes of these three players show the immense talent and skill needed to be at the top of the NBA.

Overall, the results of our computational exploration have helped us achieve our project goal and provided a deeper understanding of NBA player achievements.

6.2 Limitations

Despite our project's success, we encountered several limitations during our computational exploration:

- The dataset we used may not be comprehensive enough to capture the entire scope of NBA player achievements, as it only includes the top 75 players and their awards.

- We relied on NetworkX, Plotly, and Matplotlib for visualizations, which, while effective, may not offer the most advanced or interactive visualization options available.

- Our analysis focused on the relationships between players and awards, but we did not dive into other potentially relevant factors, such as player positions, team dynamics, or the impact of coaching staff.

- The dataset was relatively large and because of this, the visualizations were not as good as we wanted them to be with the nodes being very tightly packed together and many edges, making it hard to see which player links to which award. Fortunately, we were able to make filtering for our graph but the visualizations still are not as good as we would want them to be.

6.3 Next Steps

To further enhance our project, several next steps could be explored:

- Expand the dataset to include a more comprehensive list of players and awards, allowing for a more thorough analysis of player achievements across the NBA.

- Implement advanced interactive visualizations using libraries like D3.js or Plotly Dash to provide users with a more engaging and informative experience.

- Incorporate additional variables, such as player positions, team dynamics, and coaching staff, to gain a more holistic understanding of the factors that contribute to player achievements.

- Utilize machine learning algorithms to predict future award winners based on historical data and player performance metrics.

6.4 Conclusion

In conclusion, our project successfully created an AwardNetwork that visualizes and analyzes the relationships between NBA players and their awards. Although we encountered some limitations, our computational exploration has provided valuable insights into player achievements and the role of awards in shaping their careers. By pursuing the next steps outlined above, we can further enhance our understanding of NBA player performance and continue to contribute to the broader study of professional sports achievements.

7 References

[1]   RealGM, "Michael Jordan Awards," RealGM, 2023. [Online].
       Available: `https://basketball.realgm.com/nba/awards/by_player`.
       [Accessed: April 2, 2023].