EN.560.653 Introduction to Network Modeling

Final Project Report

Yilin Huang, Honghui Zheng, Haoran Liu

**Abstract**

This project develops a network modeling framework for analyzing NBA player matchups and optimizing team lineup decisions. Using data from the 2020-21 regular season, we construct three types of networks: personal performance networks showing individual player matchups, team-perspective networks illustrating lineup relationships, and optimization networks for lineup selection. The framework employs centrality analysis to evaluate player performance and minimum cost flow analysis to optimize starting lineups. Applied to the Los Angeles Lakers case study, our analysis identifies both defensive vulnerabilities and strengths within the team, while the optimization model suggests optimal defensive lineup combinations against specific opponents. The methodology provides a systematic approach to quantifying player interactions and matchup effectiveness, offering teams a data-driven tool for strategic decision-making. Our results demonstrate the potential of network theory in basketball analytics, providing insights that traditional statistics might overlook. The framework's flexibility allows for application across various performance metrics and strategic objectives.

1. **Introduction**

NBA, known as the National Basketball Association, is one of the most famous sports leagues in North America. It contains 30 teams, which will take a total of 82 games each regular season per team and play-off stage after the regular season for the championship. NBA teams generated $115 billion in the 2022-2023 season. Boston Celtics, the champions of the NBA league in the 2023-2024 season, wins up to a whopping $12.1 million ($804,000 per player). Hence, it’s the priority for every team to make strategies for its lineup while facing different opponents.

This NBA Project will focus on using what we learn in Introduction to Network Modeling and generate tools to help team managers create different tactics for facing different opponents. Managers could create different strategies based on various criteria, such as defense, offense, and more detailed, 3-point defense and offense, and so on. By setting up our various networks, managers could have a deeper understanding of team status in the game before the matchup.

1. **Data Acquisition and Model Preparation**

As this project focuses on different lineups for teams, NBA Player matchup data is used for constructing the basic network for analysis. This project uses python interface NBA.api for data collection. By using matchup data collected from python NBA.api and web crawling from data in website [1], project system network is ready to be set up.

By taking data in regular season 2020-21 as an example, there are a total of 277 players data after data collection process. For every dataset downloaded, it shows all matchup records for that specific defense player versus different opponents in regular season of 2020-2021.

表格

描述已自动生成(continued)

表格

描述已自动生成*Table 1. Dataset Overview for Every NBA Defense Player in the Regular Season of 2020-21*

In every dataset for a defense player, it contains his opponent (Offensive Player) stats, including points (PLAYER\_PTS), total time played (MATCHUP\_MIN), total matchup field goal attempt, and field goal made (MATCHUP\_FTM, MATCHUP\_FTA), and so on. The data cleaning procedure takes place after all players matchup data collected.

Data cleaning is finished to remove the effect of noise in data, such as stats with matchup time less than 120 seconds in the whole season. As stats, like Blocks, Steals, and points made are normalized via 36 minutes time-based data, those stats will be inaccurate as noise if their matchup time is less than 120 seconds. All stats normalization and data cleaning formulas are shown as formula 1. Time-normalized data are shown in Table 1 for columns (UNIT\_MATCHUP, UNIT\_MATCHUP\_PTS, UNIT\_MATCHUP\_TOV.) Meanwhile, the defense index is defined as an example of user-initialized stats for analyzing player defensive ability. Its formula is shown below as formula 2.

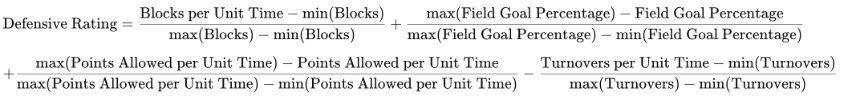
图片包含 文本

已自动生成说明图片包含 文本

已自动生成说明图片包含 文本

已自动生成说明

*Formula 1. Data Normalization formula based on 36 minutes standard*

*Formula 2. Defensive Rating for player*

1. **Network Representation Methodology**

The team lineup tactics tool would be shown as different types of networks in this project. Basically, there are three types of networks that will be constructed in this project.

**3a.** **Personal Performance Network**

The first network is constructed for personal matchup data, showing relations for that player versus different opponents in leagues based on stats criteria such as blocks, scoring percentages, and other different types of data. The network should be displayed as the analyzed player in the middle of the plot, while other opponents stay on the outside loop of the graph connecting to the center analyzed player. Take one of the NBA players, Aaron Gorden, as an example. Its personal defense rating performance network is shown in Figure 1. The red weights for each edge represent the defensive ability of Aaron Gordon versus other players.

图表

已自动生成说明

Figure 1 Aaron Gordon Defensive Rating Personal Performance Network for Regular season 2020-2021

**3b. The Network for Whole League with Team Perspective**

The network for the whole league, with a team perspective, is constructed for team managers to gain a better understanding of the team's lineup relationships with different opponents. The example of the Lakers team lineup for their player's Unit Matchup Block Stats is shown below in Figure 2.



Figure 2 Network Plot for Whole League including Lakers Lineup Players (Yellow Node in the Middle) Block Ability to different players.

To give a more detailed look at Figure 2 on the Lakers line-up, Figure 3 creates the subplot shown in Figure 2. It shows only the Lakers’ blocking stats versus other opponents by removing the irrelevant links in the graph and keeping only the links coming into the players of the Lakers. Color mapping is generated in this graph, showing deeper color if that Lakers player creates more blocks on that opponent than average.

图表, 雷达图

描述已自动生成

Figure 3 Bipartite for blocks against Lakers' players

For the analyzation of blocks made by the Laker’s players against all other players in the NBA league, we removed all the irrelevant links in Figure 2, keeping only the links that reaching from the Lakers’ nodes, thereby creating a bipartite as is shown in Figure 4, that demonstrates the capability of blockings within the Lakers’ players, and the players in the NBA league who are more vulnerable against Lakers’ defense.

图表, 雷达图

描述已自动生成

Figure 4 Bipartite for blocks made from Lakers' players

1. **Network Analysis**

We aim to analyze the designed network through two primary methods: centrality analysis and minimum cost flow analysis. The objective of our analysis is twofold: to evaluate the capabilities and performance of our players and to determine how to optimize the starting lineup based on the opponent's lineup. This analysis is expected to provide valuable insights to guide the team’s decision-making and strategic planning.

**4a. Centrality Analysis**

After several discussions, we concluded that the study of networks themed around NBA players represents a specialized type of network that does not conform to the conventional structures of typical networks. Unlike transportation or river networks, where most of the analytical methods covered in coursework can be directly applied, NBA player networks require careful consideration and substantial time investment in the design and construction of the network itself. Based on our discussions, we believe that centrality analysis is particularly crucial for understanding and analyzing NBA player networks.

We conduct centrality analysis on various networks to evaluate and rank the strengths and weaknesses of players’ different abilities. For instance, in the network focused on blocks, we analyze the centrality of players to identify the most influential blockers and those who are blocked most frequently. Specifically, we calculate the weighted in-degree centrality for the Lakers to determine who gets blocked the most and the weighted out-degree centrality to identify who blocks the Lakers the most. The Laker Player block network is shown in Figure 5.



*Figure 5 Block Against Lacker Player*

The results for centrality analysis of blocking capabilities are as follows:

**Most Blocked Lakers Players:**

1. Kyle Kuzma
2. Dennis Schröder
3. Kentavious Caldwell-Pope
4. Alex Caruso
5. Talen Horton-Tucker

**Top Blockers Against the Lakers:**

1. Draymond Green
2. Giannis Antetokounmpo
3. Kevin Durant
4. Nickeil Alexander-Walker
5. Naji Marshall

**4b. Min Cost Flow Analysis for Optimization in Starting lineup.**

Building on the network representation discussed in previous sections, we implemented a min cost flow analysis to optimize starting lineup selections. This approach provides a systematic way to determine optimal player matchups while considering multiple performance criteria.

The min cost flow algorithm works by modeling the lineup selection problem as a network flow optimization problem. In this framework, we construct a directed graph with a source node connected to all available players on one team (Lakers in our example) and a sink node connected to all players on the opposing team (Magic), where the capacity of each link in the flow network is 1 representing the assignment of the starting lineup. The algorithm then finds the optimal flow through this network that minimizes the total "cost" while satisfying various constraints like having exactly five players in the lineup.

For our specific implementation analyzing Lakers versus Magic matchups, we structured the network with several key components shown in Figure 5. The source node connects to all available Lakers players with zero-cost edges and unit capacity, representing the potential selection of each player. Each Lakers player’s node then connects to each Magic player node, with edge costs representing the defensive liability (measured by the opposing player's scoring capability against that defender). Finally, each Magic player node connects to the sink with zero-cost edges.

图表, 图示

描述已自动生成

Figure 5 Flow Graph of Starting Lineup optimization

The mathematical formulation of our lineup optimization problem incorporates several key constraints to ensure practical and valid solutions is shown below. The objective function minimizes the total defensive liability across all matchups, while the constraints establish the structural requirements of a valid basketball lineup. The formulation requires that exactly five Lakers players be selected, enforced through a source node with total flow equal to five, while flow conservation principles ensure each selected player maintains consistent assignments. Unit capacity constraints on all edges prevent multiple assignments of the same player, and binary constraints on player selection variables guarantee each player is either fully in or out of the lineup. This creates a well-defined framework that captures the essential requirements of basketball lineup construction while seeking optimal defensive matchups against the Magic's starting five.

文本, 信件

描述已自动生成

Where:

*cij*: Scoring capability of Magic player *j* to Lakers player *i*

*xij*: Flow between Laker *i* and Magic player *j*

When applied to our Lakers-Magic example, the algorithm suggested an optimal defensive lineup of Alex Caruso, Kentavious Caldwell-Pope, Kyle Kuzma, LeBron James, and Talen Horton-Tucker. This lineup was selected based on minimizing the total scoring potential of the Magic's starting five.

While we focused on defensive performance in this example, the framework is flexible enough to incorporate different evaluation criteria. By modifying the edge cost calculations, we also optimized for various aspects like the capability of scoring, turnovers, or 3 points metrics. This versatility makes the min cost flow approach a valuable tool for coaches and analysts in developing game-specific strategies.

**5. Conclusion**

Based on the analysis of NBA player matchup data through network modeling, we have developed several valuable tools for NBA team management and player evaluation. The three distinct network representations - personal performance networks, whole league team perspective networks, and optimization flow networks, each provide unique analytical capabilities for understanding player and team dynamics.

The centrality analysis revealed important defensive patterns, identifying both vulnerable players and effective defenders. This information is particularly valuable for defensive strategy planning, as demonstrated in our Lakers case study where we identified both the most blocked Lakers players (led by Kyle Kuzma) and the most effective blockers against the Lakers (with Draymond Green leading).

The min-cost flow analysis framework proved especially powerful, offering a systematic approach to lineup optimization. By successfully modeling the complex relationships between opposing players and expanding over multiple performance metrics, this tool can help coaches make data-driven decisions about lineup selections. The framework's flexibility allows for optimization across different criteria - whether defensive capability, scoring potential, or other performance metrics - making it adaptable to various strategic needs. While our focus was primarily on defensive metrics, the framework is readily adaptable to offensive analysis and other aspects of the game.

This work demonstrates the significant potential of network modeling in sports analytics, providing teams with powerful tools for strategic decision-making that go beyond traditional statistics. As basketball continues to evolve as a data-driven sport, such analytical approaches will become increasingly valuable for maintaining competitive advantages in the NBA.