NBA Analytics and Information Networks

Marc Chami
York University
Toronto, Canada
whimm337@my.yorku.ca

Indeep Singh
York University
Toronto, Canada
indeep.singh99@gmail.com

Raza Memon
York University
Toronto, Canada
raza9999.ali@gmail.com

ACM Reference Format:
Marc Chami, Indeep Singh, and Raza Memon. 2018. NBA Analytics and Information Networks. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/1122445.1122456

Contents

1 ABSTRACT

The analytical studies in the game of basketball is a massively growing area of observation that has seen major improvements in recent years due to technology and algorithm advancements. Although, all these statistics alone may seem meaningless or unfulfilling due to them simply being standalone numbers. Throughout this paper we will discuss how we can achieve full potential with these statistics by using network and graphing algorithms to visualize each player, how they improve over time, the influence they cause, and the play-styles of any given team. We will complete this task by utilizing the NBA API which grabs data directly from the stats.nba.com website to retrieve any necessary data. Our networks will be based on players within a team, passes made, and events that occured during any game. Where nodes portray players and weighted edges between players will demonstrate passes. Our network analysis results to this experiment concluded how enormous some players' influence may have on a team. The player influence is also very dependent on certain team play styles and further reliant on the progression of players on the team. Throughout this study, we have demonstrated how we determined the progression of players in the NBA league, how we identified play styles, and how players can become more influential in a given team.

Conference'17, July 2017, Washington, DC, USA
2018. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/1122445.1122456

2 INTRODUCTION

From sports betting, coaching, players themselves, to your average fan, sport analytics is a very popular topic in today's society and brings a different crowd to what we call scientific research. Although, the scope of sports analytics has a very broad spectrum and when mentioned can refer to many different topics, we have decided to target a very specific and highly important aspect of sport analytics. One which explicitly entails the positive and negative events that occur in a typical basketball game. More specifically, we will measure in-game performance of each player on their respectful team, using a weighted graph where edges signify events and accumulate a meaningful outcome. Ultimately leading to bigger conclusions based on which our algorithm will draw. Once we reach a substantial amount of analytical data, we will be able to answer questions which revolve around what a specific team's play style is, how their playstyle evolved, and how they will perform against other teams with different or similar play styles. Furthermore, we will be able to create player evaluations and how a player's impact on a team may improve or worsen over time. This can be used in any general case of a basketball game, whether it's WNBA, EuroLeague, NBA, etc. This area of study involving in-game analytics shows no signs of slowing down and only appears to be growing as technology advances, data becomes more relevant and data collection continues to increase. Because of the nature of graphing functionality and the similarity it has to team-based sports, such as basketball, will make this proposal very interesting while creating a useful set of data.

3 MOTIVATION

The motivation for our topic comes in many forms, from personal interest in the sport of basketball, the professional-sport seasons that just finished during these unprecedented times, and especially, realizing the potential and ever-growing industry of sports analytics. This project allowed us to apply and full-fill an idea which was always on the back of our minds. This proposition is especially influenced by the sports betting market and the several claims that refer basketball to a game based on numbers. Throughout any given regular season, outside of physical practice players spend most of their time analyzing other teams and looking at statistics. Several people claim that games are won in the extra work that is put in by the staff and players off the court. Lebron James, arguably one of the best basketball players to ever play the game, is praised for his off court work ethic within the sport because he seems to win most games before they start due to his knowledge about his opponents.

Furthermore, Basketball teams have whole departments dedicated to game statistics and analytics, some of which are not available to the public or have outdated methods. Therefore, our plan was to create and analyze networks that would help us predict outcomes of basketball games, view performance of a single player or team as a whole, all based on in-game statistics. Through the data parsed into our algorithm we will be able to make assumptions on the networks, which will show us how certain teams match up against other playstyles. For the average consumer, this can then be used to help their betting predictions or even personal growth. Both of which are growing businesses. Lastly, the networks will show us how influential certain players are based on metrics that are set by us. This will allow end users to predict MVPs or even their next fantasy draft pick which is becoming more and more popular between sport fans due to the recent pandemic. Such factors listed above have really put an influence on our final decision for this project application and brought to our realization the importance this implementation could have, allowing us to appreciate this opportunity and apply some of our passion to a technical setting.

4 PROBLEM DEFINITION

Our fundamental defining problem exists in the very nature of the statistical data which we are handling. That is, viewing the raw data which basketball associations have to offer, individually, may be confusing and confined to a singular meaning. For example, point stats, rebound stats and give-away stats all may tell their own story of the player alone, but rather, basketball is a team sport and these statistics by themselves do not hold the full picture. We are thriving to take these one-dimensional pieces of data and expand them into visuals which can help users better understand basketball statistics as a whole. This is especially important in today's world because as analytical stats are becoming more involved with monetary value the importance of the enterprise also grows. Expanding the value of these raw statistics will allow us to evaluate the importance of any given player at any given moment, determine play styles of teams, and further understand the progress or weakening of a team/player.

We have decided that the best graph type is one which interacts with other nodes through analytical measurements in hopes to measure the influence one node may have on the rest of the graph as a collective. The main reason why we are able to do this is because we will be dealing with smaller graphs (20 nodes or less). Once we have agreed on this, it was clear that the eigenvector centrality graph would be an ideal requirement. To briefly explain how this graphing technique works with our basketball analytics application, we will be measuring node scores (which are the players) directly in relation to the players' points per game, completed passes and many other positive and negative in-game events. This will be the basis of our upcoming notations and formal definitions.

4.1 Eigenvector centrality

To elaborate more on the eigenvector centrality graph, we will be diving into how each node measurement will be evaluated, without going into mathematical specifics. Essentially, we will construct an event list and attach a positive or negative numerical value to each event which will sway the resulting graph. For example, if Player 1 and Player 2 are both involved in a play where Player 2 nets a three-pointer, both Player 1 and Player 2 will receive a positive numerical value added to their existing value. This same concept will be used for players involved in a negative in-game event (such as turn over, throwing the ball out of bounds, etc). This will essentially boost or diminish a player's overall score measured at the end of the game.

4.2 Playstyles

Playstyles for each team will be evaluated and eventually assigned. This will be completed by observing how a team functions in each of their games played. For example, do they pass the ball equally between all players on the court, do they often pass one superstar player, do they hold possession often, and so on. This function will help assign a team with a specific play style and make random team vs team matchup assumptions more accurate based on previous data.

4.3 Player Importance

With our algorithm set and explained, the use of eigenvector centrality and the identification of a team's play style will help us elaborate on the importance of any given player at any given moment. This will be further determined based on the combination of in-game statistics we digest and how the graph forms around a player based on, most important edges, highest eigenvalues, etc. Using this, we can observe a player's increase/decrease in productivity over multiple seasons or how the player acts in a different team/playstyle, all which can help depict further analytics.

4.4 Potential Issue and Reasoning

Within our fundamental problem roots many other small problems such as the following: Many basketball fans across leagues throughout the world are very passionate about their teams and super star players. This sometimes interferes with the end of season award winners reasoning and creates arguments as to why decisions are made within communities. We want to provide visualization which will entice viewers to research more and observe the meaningful facts compiled together which reason for MVPs, Rookie of the Year, Defensive Player of the Year, etc. We plan to make our resulting network graphs able to demonstrate the most dominant players in any league through essentially showing who makes the most successful plays most often. This will help officiating, fans and shareholders to more accurately depict who is involved in the "most dominant player" conversation.

4.4.1 Constraints.

- The first obvious constraint depicts how analysis will only be based on basketball events happening in-game. That is, each event in the constructed event list will correspond to either a singular player, or multiple players involved in the play. It does not take into account non-statistical data.
- Theoretically, we would like to evaluate forthcoming seasons through previously acquired data. That is, make sure our resulting metrics which we analyze are accurate and for example, will be based on historical statistics. This will help with our next constraint.

- Our resulting data will show an educated assumption about forthcoming seasons, player or game stats but will not be 100% accurate. Although we hope to eventually achieve this goal through previously measured data.
- Every graph will consist of players involved in the matchup and the event nodes such as, success, failure, steal, rebound, inbound, etc. This will be used to demonstrate the resulting outcome of plays and games.

4.4.2 Optimization.

- Optimization will primarily consist of maximizing the amount of data we can run through our algorithm to achieve the most accurate results as possible.
- Creating as many event nodes as possible to help understand which plays result in positive or negative outcomes. That is, maximizing the amount of paths which lead to an event node.

5 RELATED WORK

As sports analysis becomes more prominent and technology advances allowing accurate and more data to be collected studies are becoming plentiful. The hurdles at first were not only the lack of technical advancements but also the infrastructure needed for accurate data collecting was missing. Furthermore, in sports like tennis many critics were against the introduction of technology to determine events and analysing live game data. This could also be seen in a more recent revolution in the game of soccer where goal line technology and Var (Video assistant referee) were introduced. Until this day there are many people skeptical of the use of technology to determine game outcomes, but everyone can agree that video analysis and data is existential in the present to get an upper hand over their opponents.

One of the pioneers in this field are the Toronto Raptors who used machine learning software to predict what actions certain players should have performed in different game scenarios to increase their chances of getting a positive outcome. [1] This was allowed by a new camera-tracking system introduced by SportVU, which notes every movement on the court of each individual and then gets converted into data composed of geometric coordinates. In other words, positioning becomes a measurable quantity rather than a coach's opinion or suggestion. A similar study was used to determine and analyze how teams create and find space for scoring opportunities in soccer.[2] Like the Raptors, advanced machine learning methodologies were used on a season's worth of tracking data. Ghost players were created to imitate individuals and their preferred or most suitable actions. For that they had to prepare their data, role alignment was used to give players roles based on the situation rather than having a set formation and role given to the athletes. In our approach on the other hand, we disregarded positions and solely rely on individual players and the whole team as one unit with no specific roles. Since our goal is not to mimic or find what a player is good at rather to search for a whole team's play style or simply the involvement and influence of a player when attacking the opponents basket.

Furthermore, a study was made on the trajectory and analysis of actions made by soccer players. Their data extraction is based on video analysis, which allows them to create graphs where nodes consist of the players and edges are the temporal correspondence which can be described as how different video images connect with each other. This is needed since during a soccer game several cameras and angles are needed to display the whole game. The difference to our research and problem is that the graphs are made focusing on an individual player rather than how all the players interact with each other on a global basis. A similar approach was taken by Kyoung-Jin Park and Alper Yilmaz. However, in their study they were more focused on the bigger picture and the interaction between different players .[4]This was done by creating social networks, these networks are built in a similar fashion as us. Every event occurring in a game such as a goal or corner is a result of individuals passing the ball to each other. Therefore, in a similar fashion as in the work done before players act as nodes but the connecting edges this time are completed passed between two individuals. Directed graphs are used to demonstrate this, with the directions showing how the ball travelled from Player A to Player B. Due to data being collected from video analysis, they have the ability to extract the distance between two nodes during a play and calculate the closeness. This is not possible in our research, due to the limited amount of data available to us. It would have been a beneficial metric for use since more mathematical calculations could have been made on our data. For example, they noted the closeness between nodes using algorithms to find the geodesic distance, which would allow for more analysis. This particular study was one of our main motivators, since it focused more on the sport as a team game and every player as a piece for the bigger puzzle. They also came across the challenge of deciding what type of graphs to use for their study when choosing directed graphs, this showed us how certain graph characteristics can be used and why they are chosen.

6 METHODOLOGY

We hope to address the problem by tackling each section one step at a time. Beginning with conceptually designing our algorithm which will use the collected data to draw directed eigenvector graphs for each team in a particular game. This algorithm will be heavily reliant on the passes made throughout the game and will be represented through edges while the nodes will exist as players. Moving forward we will then create event nodes which will represent a success or failure play and will only be connected to a player node if and only if a player completes one of the described events. Completing these graphs will essentially give us two very valuable sets of information to observe.

- Who is the most dominant player or most influential player in a team.
- Using further analysis, determine a team's play-style. Do they pass the ball more often, do they let only a couple players isolate, do they hold possession, and so forth.
- Furthermore, comparing multiple graphs can allow us to observe the development of a team or player.

It is important to note that initially the team's play-style will be determined by our personal observations of the graphs, and hopefully, in the long term will be automatically judged by a comparison

machine learning algorithm. Another important principle to mention is that our resulting data will be a compilation of analytical data which is necessary for us to showcase the productivity and effectiveness of a particular player or play style of a team.
Once our algorithm is functioning with test data, we will begin to retrieve data from the official stats.nba website which carries a large amount of advanced stats from an especially wide spectrum of topics. With this came an additional challenge of deciding which data we need to deem necessary and what will be considered unnecessary. Focusing on the necessary data, the consensus will be dependent on the graphs and what we want to be considered a 'successful' or 'failure' play. Initially the amount of plays that will be labelled under one of these categories will be a small list, but eventually we intend to continue to grow the specifications as time moves on. The actual data scraping will be completed using the NBA API which will directly retrieve its data from the official stats.nba website.
Eigenvector centrality is a special type of centrality which we will use to accurately measure the influence of all nodes (players) in our resulting networks. It is the most optimal approach for us to take as it will essentially tell us which player on any given team holds the most pose and effectiveness on the court (leading to potential MVP candidates). We intend for this to work similarly to Google's famous algorithm PageRank, where we measure the number and quality of edges pointing to a node in any of our networks.
For eigenvector centrality we will be understanding how it functions within the graph. Thus we will need to analyze the the existing centrality score equation which can be defined for any vertex v as:

$$x_v = \frac{1}{\lambda} \sum_{t \epsilon M(v)} x_t = \frac{1}{\lambda} \sum_{t \epsilon G} a_{vt} X_t$$

Where λ is a constant, M(v) is the list of neighbouring nodes to vertex v and G is all the nodes in the network. This equation essentially states that we will take the sum of eigenvalues for every neighbouring vertex t linked to vertex v, divided by the given constant λ. And x_t is defined such that the sum of all eigenvalues in the graph is multiplied by a_{vt} which equates to 1 if v is connected to t and 0 otherwise.[5, 6]

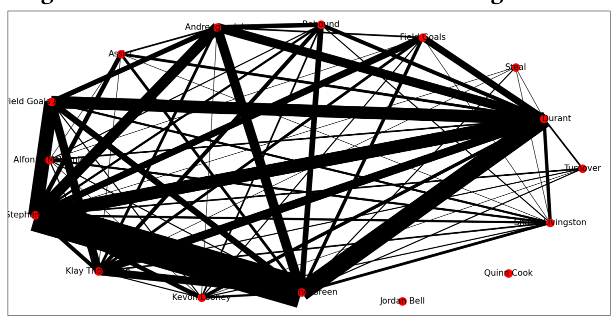
6.1 Limitations Difficulties

Our biggest difficulty was scraping the data when using the NBA API. Initially when trying to retrieve data from a specific game, we found the parameters to be very sensitive with not very much explanation to the error documentation. This took time to research deeper into the API and reveal certain functionalities so we can solve our problems and in turn input the correct parameters. For example, problems emerged when entering the specific date of the game, season, season type, etc. Some difficulties with our approach exist within the eigenvector graphing functionality. Graphing the networks without an eigenvector restriction is much easier as the centrality parameters add another layer of difficulty to our networks. The parameters and visualizations needed to create eigenvector graphs require much more small details in order to retrieve the visualizations we desire.
Additionally, our algorithm is not automated so when finding statistics per game, we will need to find specific dates of the game to retrieve data from the API. Furthermore, if we want to retrieve data based on seasons, i.e. player progression, we will need to pull data from each game that the player played in the season and calculate an average to plot. This will allow us to see progression or any other statistics we need.
A clear limitation is the fact that we are confined to the functionality of the NetworkX API. Although it covers most of our needs, we came across a couple instances where we needed to implement additional functionalities to configure the networks. As we move forward with the project, we will sure come across more limitations and difficulties.

7 EVALUATION

Here is the Golden State Warriors full network for the 2020 season.This is how all of our networks will be structured to best visualize events between players and passes between players. All the data analysis will be originated networks such as these.

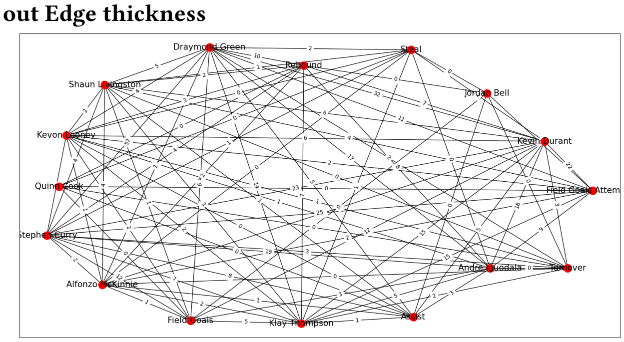
Figure 1: Golden State Warriors 2020 Passing Network



As shown above, the edge between each player portrays the sum of passes to each player (to and from) will thicken as the number gets larger. Same goes with the event nodes, such as rebound, field goals, etc. These are the networks that hold all the data we scrape, analyze, and compare throughout this project.

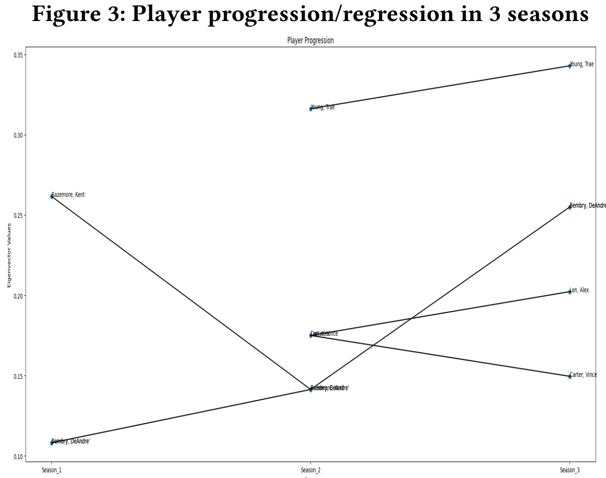
For reference, this is how our networks would look without edge thickness, but rather the weight of edges depicted by number on the actual edges. This is strictly for a visual representation of how our networks would like without edge thickness. The network shown is the same data as the previous network shown with edge thickness.

Figure 2: Golden State Warriors 2020 Passing Network without Edge thickness



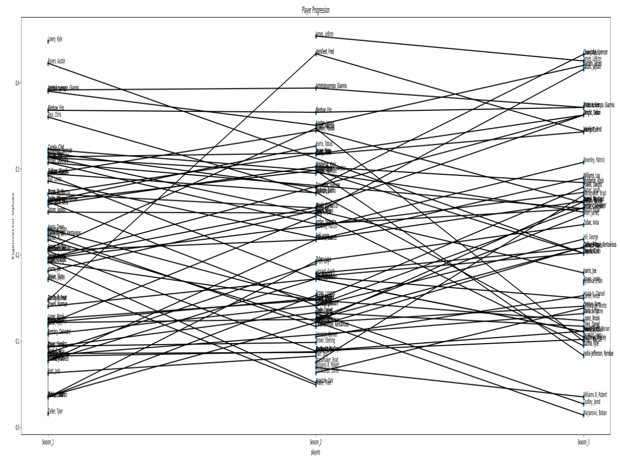
Our data has been retrieved from the official NBA website (stats.nba.com) so we know the results of our algorithm hold true to any specific game statistics. The website provides the information about every event that took place during the game, including points, rebounds, and assists. It also keeps track of the statistics for each player, which would help provide more parameters for them when the algorithm is used to analyze each player individually and create a network graph. For example, to generate a network graph based on the passes from Player A to the rest of his team, the algorithm is required to use the data provided for each player and look at the 'PASS' column to identify how many passes and to whom the ball was passed to. However, the NBA website does not give a downloadable format of the provided data, nonetheless, there are a few available APIs that allow its users to extract any needed information. We decided to go with the API called 'nbaapi', which doesn't only allow the users to scrape the data from the NBA website of specific games, but also to use the result in JSON format and generate the network graphs for it.
Our resulting networks are more or less what we had envisioned from the beginning of our project. Team based networks where nodes represent the players and are connected via passes completed throughout the game. When comparing statistics based on two teams versing each other we ended up being more specific rather than the "success" and "failure" nodes as previously discussed and created nodes like for each type of event. For example, "Steal", "Turnover", "Rebounds", "Field Goals", "Assists", "Field Goals Attempted", etc.

Figure 3: Player progression/regression in 3 seasons



This is an example of how we will draw up player progressions. An experiment we completed involved drawing the progress for each player in the NBA in the past 3 years. This gave us the idea to create the functionality in our algorithm so the user can choose which player they want to draw on the graph and view their progress from a range of previous years.

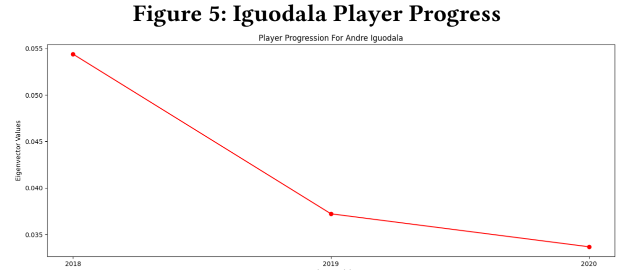
Figure 4: All Player Progression



This is the max amount of players our algorithm limits the user to hold, which is all the players in the NBA that follow restrictions we set (such as, players that played, achieve a certain threshold of eigenvalue, etc.) in the 2017-2018 (season 1), 2018-2019 (season 2), 2019-2020 (season 3). We used this and will continue to use this for observations of player progression. In most cases, we graph less players so we can see clearly the progression or regression of players.

When referring to the graphs we made for analysis, we opted to use bar and plot graphs to illustrate best the eigenvalue influence on which a player has depending on which topic we are discussing.

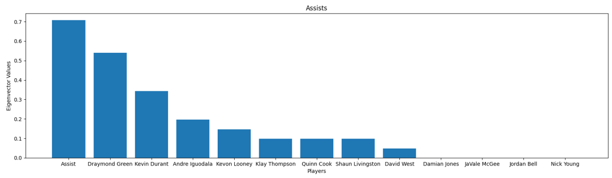
Figure 5: Iguodala Player Progress



This graph analysis shows one of the many ways we can view progress for any given player over multiple seasons. The one above specifically demonstrates the decrease in influence of Andre Iguodala over the past three years.

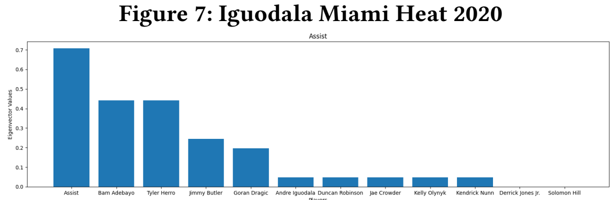
We will use Figure 3 as an example to demonstrate what the data truly demonstrates. More specifically, how Andre Iguodala's individual influence decreased on the respective team from the 2018 season to the 2020 season.

Figure 6: Iguodala Golden State Warriors 2018



In this graph (2018 Golden State Warriors) we can see Andre Iguodala's influence is rather high on this team. He is not the most influential player, behind Kevin Durant and Draymond Green, but he is a close third, especially for his playstyle. In the next graph we can observe how he did not improve when he got traded to the Miami Heat.

Figure 7: Iguodala Miami Heat 2020



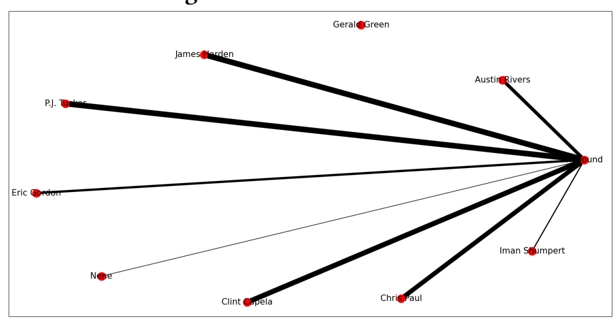
In this graph (2020 Miami Heat Team players), we can see Andre Iguodala's influence decreases substantially. We concluded that this is due to the Miami Heat's play style and the way they rely on players like Iguodala. In Miami, he was more of a "second choice" in the role he played, resulting in a reduced influence and giving him a lower eigenvalue after the trade.

The above graphs tells an important story in the 2018 season when Golden state won the nba championship and placed first in regular season points. In the next season, 2019, when golden state was going for the three peat but lost against our Toronto Raptors after showing a poor performance throughout the playoffs, especially due to injuries. Additionally, they only placed second in the regular season demonstrating their weaker performance to the prior year. We can see through iguodala as he is a well-known playmaker and known as the "glue of the team". When Iguodala was needed the most, his performance declined demonstrating how his team did not perform on the big stage. In the last season, 2020, when he was traded to the Miami Heat, it should be noted that they play a more dominant play-style letting players such as Jimmy Butler and Tyler Herro control the game more. As we can see, Iguodala's influence was much lower in this season which makes us think he does not excel in these types of teams with these types of playstyles. He was not able to apply what he excels in like, shuffling the ball, creating space for players like Curry and Thompson, when he was a Golden State Warrior. By this experiment we have proven that our data is true as it is completely based on data that is directly retrieved from the official NBA statistics website.

7.1 Experiment

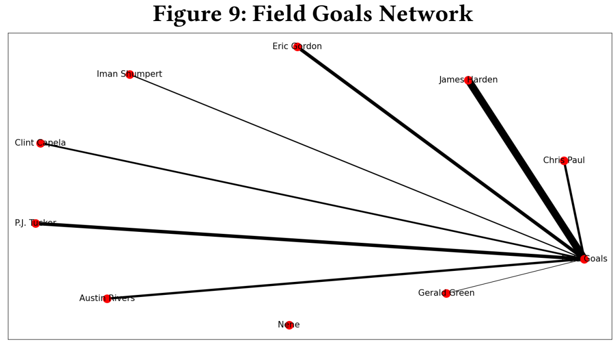
With our premade eigenvector graphs, we will use them to evaluate the play-styles of each team relative to the data provided from the website and conclude the winning team. Naturally this conclusion from the algorithm would need to be evaluated using the data of the games from previous years and comparing the result of the year after the game between the two teams. For example, figure 2 and figure 3 show the playoff game of the 2018-2019 session, between the Golden State and Houston Rockets. These graphs would be further analyzed to predict how the next year's playoff games would look like between the two teams, and who is more likely to win. The prediction of the algorithm would be evaluated from the NBA's website since the data would already be available for session 2019-2020 and show how close or far off the algorithm would be from the actual result.

Figure 8: Rebound Network



For a clearer view here is a network for simply rebound events for the Houston Rockets. We used these when analyzing team play-styles and depicting he has the most rebound collections.

Figure 9: Field Goals Network



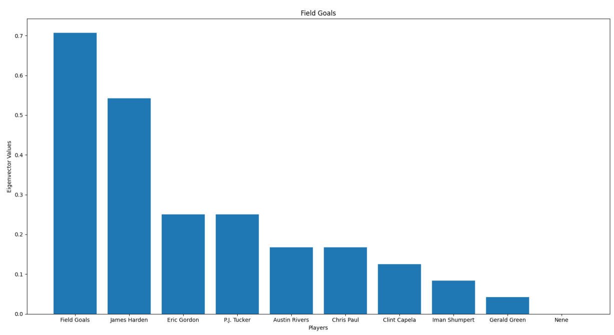
Another cleared view of turnover events. Again we use these for analyzing team play-styles and depicting who has the most field goals. For example in this team (Houston Rockets) we see how it is mostly a single player getting the majority of the points, where in other teams we see points scored by a group of players.

7.2 Data-Set

We aim to make this project public, which would allow the user to enter specific data of a game (such as data and session type) and generate a network graph of that specific game. The datasets are used to specify the specific game so the API finds the exact game and extracts its data from the website.

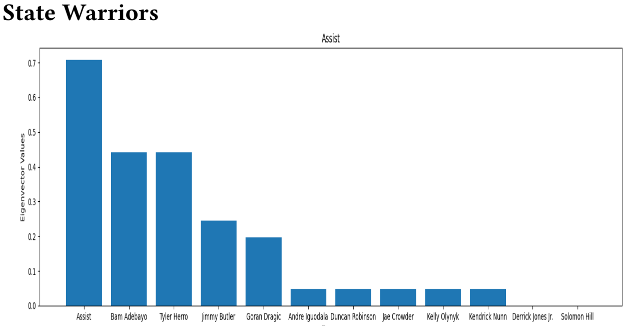
From our initial findings we came across a blank screen because the data sets contained the information about different year's, so the API could not find the exact game that took place. Therefore, it is important for the user to be confident about the dataset that they are entering, because the API is parameter sensitive.

Figure 10: Overall influence for selected players on Houston Rockets



Network Analysis for the Houston Rockets, based on eigenvalues. This one specifically demonstrates how James Harden has a very large influence on how the team performs, compared to any other player. This is also due to the Houston Rocket playstyle which makes James Harden the main focus of their offense.

Figure 11: Overall influence for selected players on Golden State Warriors



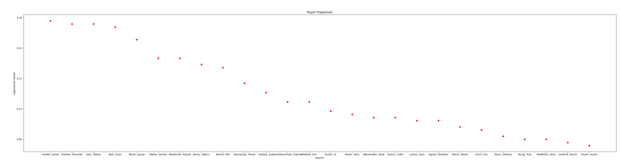
This is overall influence determination for the Golden State Warriors (based on eigenvalue) for each player on the team. The eigenvalue includes all positive and negative influences including passes, goals, steals, etc. (Please ignore the event labels and pay attention to the player demonstrations).

Through figure 7 and figure 8 we can observe how we determine play styles of a team. This is only one example of many but we can see how a team like the Golden State Warriors (figure 8) have a very different play style compared to the Houston Rockets (Figure 7). We can see in figure 8 there multiple players with a high influence. Essentially this portrays how Golden State passes the ball more, splits field points between many players, allows more players to touch the ball, and constantly changes roles between players. This may be the main reason for the triumph success in the league.

Whereas a team like the Houston Rockets (figure 7) have a more solid structure to their game with less variables. Majority of the plays go through one or two players, namely James Harden. This type of play-style seems to be successful although is especially reliant on one player, sometimes two. We observe that this play-style tends to be highly vulnerable to failure if the most important player on the team gets injured or is having an off day.

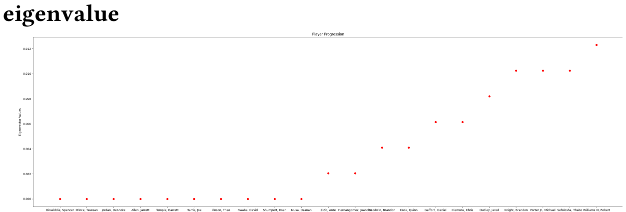
Lastly, we will demonstrate what our work has been leading up to. We are able to plot all players in any league over the span of any given years and average the eigenvalue score to see if improvements are made within.

Figure 12: Top 26 players sorted from highest to lowest eigenvalue



Here we demonstrate all the top players in the NBA based on their influence (eigenvalue score) which is averaged through 5 seasons. We used this to quickly view the top players in the NBA at any given moment so we can compare player to player and view progress of certain players. This helped us determine the league Most Valuable Players as well as the weak links in any team by sorting the graph in ascending order (lowest to highest) rather than descending (highest to lowest). For each season we have observed different MVPs with many different players on the rise and also a constantly changing low end (players with lowest eigenvalue) per season. We observed that players on the low end usually do not stay there for long.

Figure 13: Bottom 20 players sorted from lowest to highest eigenvalue



This graph shows the players that need improvement throughout the league assisting us in determining players that have very low influence in the league. This is just an example showing 20 players with the lowest eigenvalues.

This will be one of the most useful graph analyses as it gives us an overall view of all players and allows us to view how each player is doing in a team while giving us an easy overall view of player influence.

8 CONCLUSION

Our application of information networks based on basketball statistics is a way to apply our passion for all sports, algorithm engineering and the couple relationship between data and networks into a real world scenario. Throughout this project, we have taken raw data which holds a one-dimensional meaning and combined it to other relational and relevant data to create an automated set of networks which represent a basketball player's progression, a whole teams playstyle, define player influence and much more. With the use of an alternated PageRank, eigenvector centrality, weighted edges, we can apply observation and network analysis to retrieve meaningful data in the growing world of sports analytics.

The future works for this experiment are endless. We are hoping to continue this information and network analytic test to see how far we can go with our already set proposal. Some immediate evaluations we hope to attempt are determining how to automate network and graph generation to have an easier method of visualization graphs. Perhaps creating simplistic prompts to input the data we wish to see so we can analyze more graphs at a faster rate, and focus more on how to visualize our graphs in an improved way. Furthermore, we should focus on implementing additional metrics to display a players performance in a more accurate manner. This could include defensive statistics such as interception or blocks. In addition that, using community detection algorithms we could decipher small clusters of nodes that work well together. This idea could be expanded to find out how a player performs with certain line ups in an individual team in comparison to others.

List of Figures

REFERENCES

[1] Zach Lowe. *Lights, Cameras, Revolution.* https://grantland.com/features/the-toronto-raptors-sportvu-cameras-nba-analytical-revolution/.
[2] Hoang M. Le, Peter Carr, Yisong Yue, Patrick Lucey. *Data-Driven Ghosting using Deep Imitation Learning.* http://www.yisongyue.com/publications/ssac2017$_g hosting.pdf$
[3] Jun Miura, Hiroyuki Kubo. *Tracking Players in Highly Complex Scenes in Broadcast Soccer Video Using a Constraint Satisfaction Approach.* https://www.researchgate.net/publication/221369001Tracking$_p layers$ inhighlycomplexscenesinbroadcastsoccer videousingaconstraint satisfactionapproach
[4] Kyoung-Jin Park, Alper Yilmaz. *Social Network Approach to Analysis of Soccer Game.* https://www.researchgate.net/publication/220931287Social$_N etwork$ ApproachtoAnalysisofSoccerGame
[5] *Centrality.* https://en.wikipedia.org/wiki/Centrality
[6] Joseph Rabinoff, Dan Margalit. *Interactive Linear Algebra.* https://textbooks.math.gatech.edu/ila/eigenvectors.html