CNT 5805 Final Project
Project Team 2 - NBA Trade Impact Analysis

Project Member Names:
Bailey LaRea
Nadine Rose
Emmanual Nsiye

1. **You and your teammate(s) are to report on your real-world dataset and the network analysis you ran on it. (maximum of one page)**
    a. **Summarize your project. Include such items as your motivation for selecting your topic. What was the source of your dataset (provide a link if it was obtained online)?**

    This study uses statistical analysis methods and network visualization to evaluate whether a small market team has the chance to win the Finals. To conduct this analysis, salary amounts, all-star player transactions, media market rankings, and power rankings are assessed across the seasons during 2017 - 2020.

    The motivation for this study is to see what happens to small market teams when they make big signings. This topic is not normally discussed on ESPN, NBA app, or CBS sports, so we are interested in studying the impact that trades and signings have on small market teams vs large market teams. We want to know if there will be a way for a small market team to go far in the playoffs/win a championship.

    The source of the dataset for creating the network was obtained online on the NBA official website through their published player transactions. (https://www.nba.com/stats/transactions/).

    b. **Explain your purpose (e.g., inform, persuade, educate, entertain, predict, etc.) for analyzing this network. Use at least one of the terms from the preceding sentence for your purpose.**

    The purpose of analyzing this network is to inform those on the impact All-Star signings have on small versus large market teams.

    c. **List and explain the research questions of your project.**

    Does being a small market team affect end-of-year power rankings with respect to the power rankings at the beginning of the season? We will measure this by going to ESPN to find the power rankings pre and post-season for every year 2017-2020.

    From our analysis, what are the chances of a small market team winning the NBA championship?

How does a trade or a signing impact a team record (Win vs Loss)? We will measure this by looking at the win vs loss for the season prior to the trade and the end of the season with the trade.

How does an all-star signing or trade impact a team midseason? To answer this question, we will reference the changes in NBA vegas betting odds after an all-star signing and/or trade. Found on sportsoddshistory.com

Do specific trades impact the media market rankings? (from preseason to next season) Using this website https://www.sportsmediawatch.com/nba-market-size-nfl-mlb-nhl-nielsen-ratings/ in addition to reddit. We found a website for every year of media market rankings and they are all in terms of media market size by viewership of the team.

How does a team's market size influence the trades or signings they make? We will find the price of the contract on . We will also find the salary caps and cash of every team for every year on https://www.basketballinsiders.com/chicago-bulls-team-salary/chicago-bulls-salary-archive-2017-18/.

2. **Now open your file using Gephi or any other network analysis software you used.**
    a. **How many nodes and edges are in your graph, is it directed or undirected, weighted or unweighted?**

For the project, there are four different networks each representing different seasons of trades in the NBA. In years where all teams were recorded participating in any trade or player transactions on NBA.com, there will be 31 nodes. 30 nodes representing all the teams in the league and one node representing the Free Agency market. In years where not all teams were recorded involved in any transactions on NBA.com, our networks will have less than the standard 31 nodes.

Edges will vary on the number of trades per season:

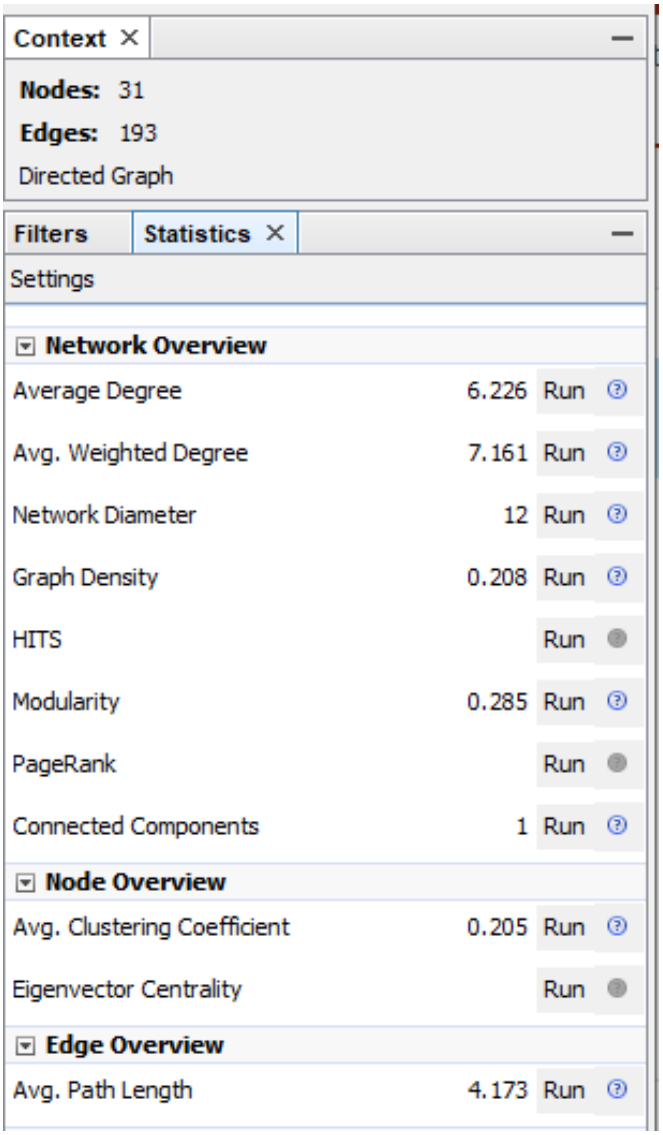
- The season of 2017 - 2018: 193 edges.
- The season of 2018 - 2019: 218 edges.
- The season of 2019 - 2020: 257 edges.
- The season of 2020 - 2021: 244 edges.

The graph is directed to show the player movements from one node to another. Weights in our network will be defined as the following:

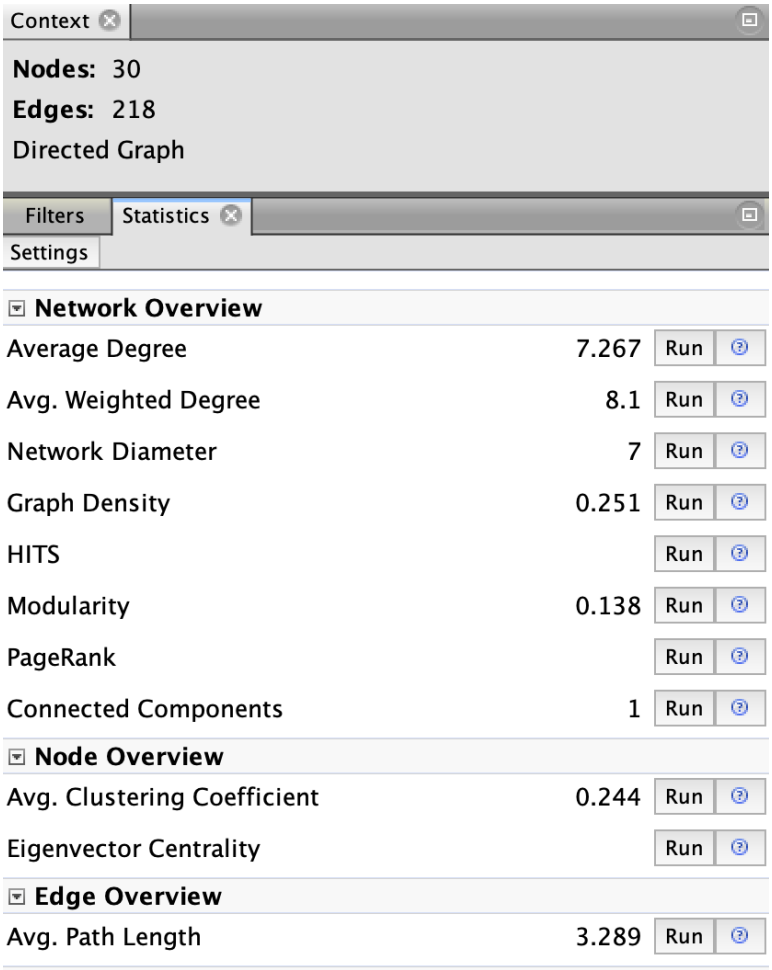
An All-Star player that shows an improvement on the betting odds from pre-trade to post trade will be defined with a weighted edge of 3.0.

An All-Star player that shows no improvement or decline on the betting odds from a pre-trade to post-trade will be defined with a weighted edge of 2.0.
Remaining players that are not All-Stars will be defined as a weight of 1.0.

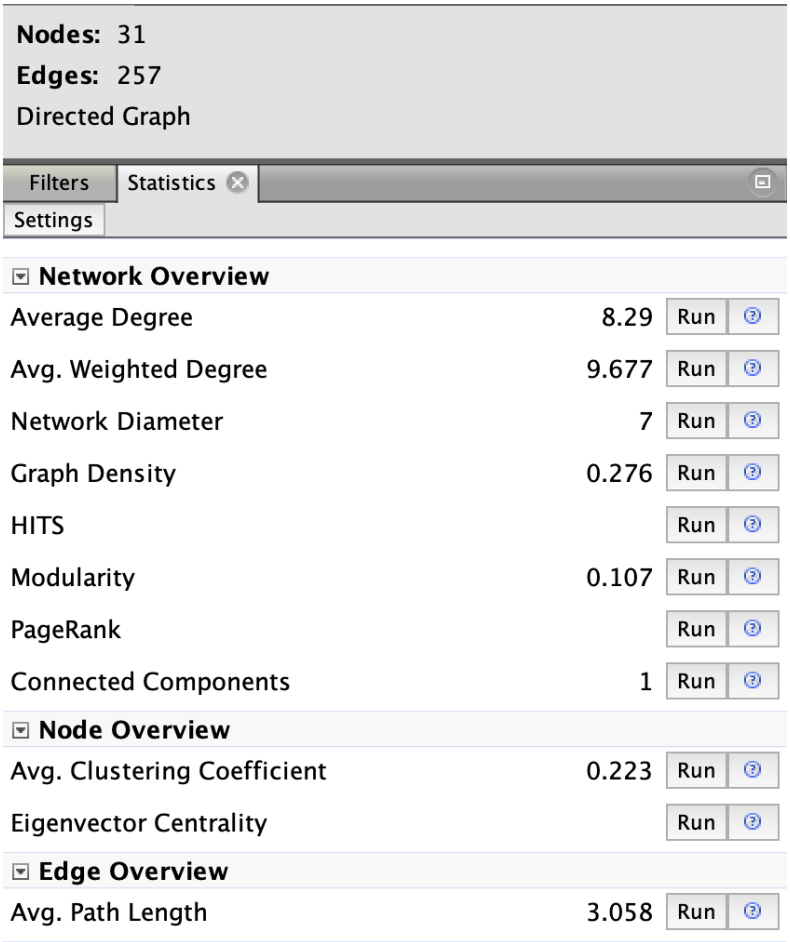
b. **Run all the Statistics, Network Overview in Gephi and show a screenshot of it. 2017-2018:**



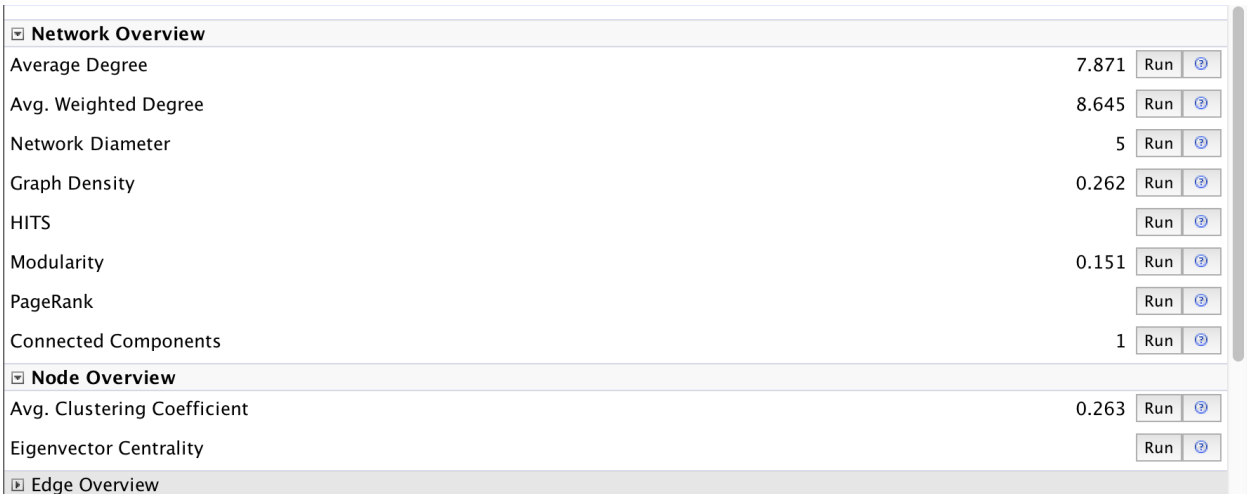
2018-2019


Context
Nodes: 30
Edges: 218
Directed Graph
Filters
Statistics
Settings
Network Overview
Average Degree 7.267 Run
Avg. Weighted Degree 8.1 Run
Network Diameter 7 Run
Graph Density 0.251 Run
HITS Run
Modularity 0.138 Run
PageRank Run
Connected Components 1 Run
Node Overview
Avg. Clustering Coefficient 0.244 Run
Eigenvector Centrality Run
Edge Overview
Avg. Path Length 3.289 Run

2019-2020

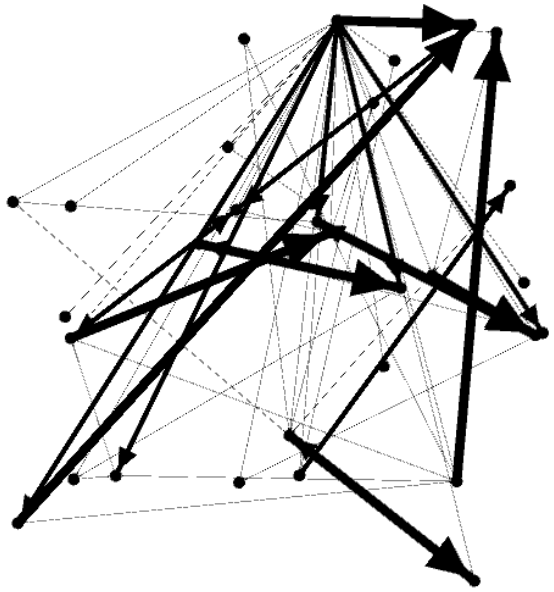


2020-2021



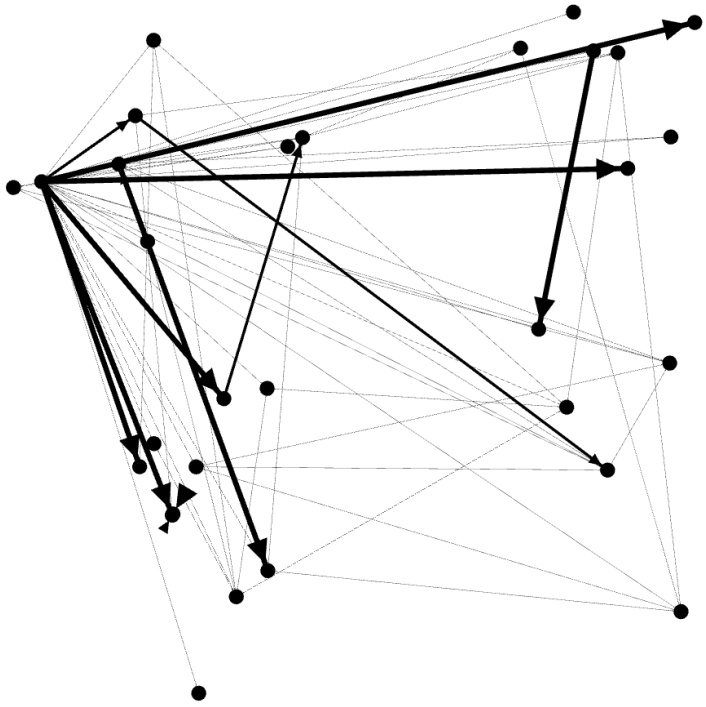
c. **What can you ascertain from the initial graph you see? Include a diagram of the initial graph in your report.**

2017-2018: 31 nodes and 193 edges, one thing that jumps out is that the higher weighted edges are easily recognizable. Overall, in the graph we notice that there is a possible existence of a node in the center where most of the edges are sourced from. However, there is nothing we can ascertain for our research into small/large market team trade involvement based on this.

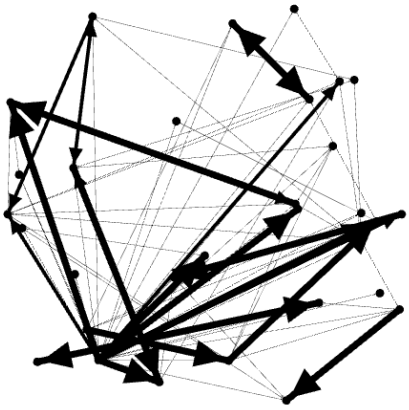


2018-2019: After importing the dataset to Gephi, from the context section one can identify that there are 30 nodes and 218 edges. From how weight was defined in the network, one can see that there are all-stars that improved the betting odds after a trade took place. However, at the current status of the graph it is difficult to differentiate the all-star trades that improved betting odds after a trade compared to the all-star trades that

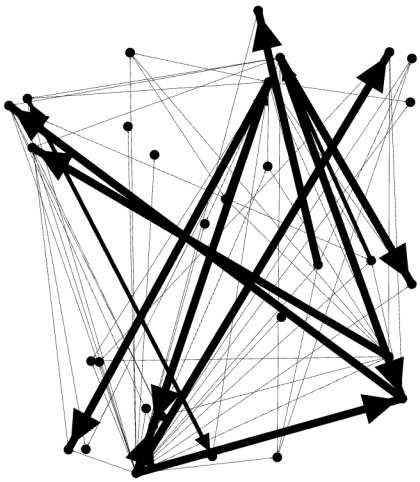
showed no improvement on the betting odds.



2019-2020: There are 31 nodes 257 edges. This is a directed graph with what appears to have several all star trades, judging by the heavier weight on several of the edges. This graph does not reveal any obvious communities. There certainly appears to be some hubs towards the lower portions of the graph. However, it is difficult to make this out with so many of the edges overlapping and no formatting applied just yet.



2020-2021: There are 31 nodes present in this network. It is a directed graph with 244 edges and 14 weighted edges for the All-Stars. The heavier weights are for the All-Stars that positively impacted the odds for their new teams. You can't discern any hubs or communities from this network.



3. **Now run three to five layout algorithms.**
   The following layouts for each network were ran:

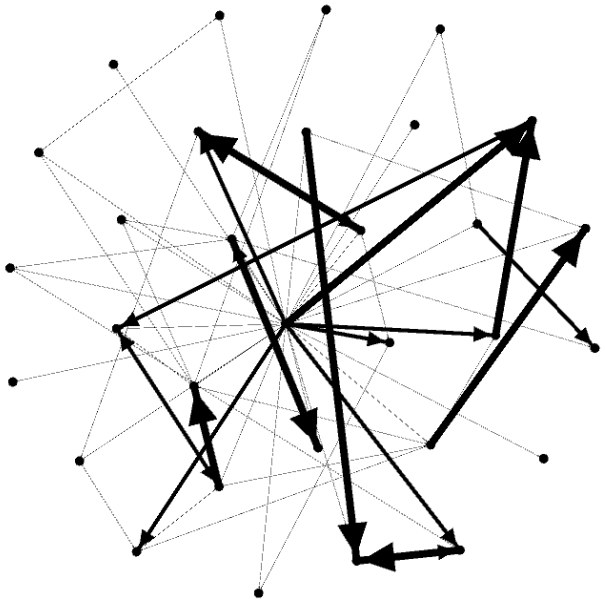
   Fruchterman Reingold (Screenshot was captured after letting layout run for 10 minutes) - The edges are spread apart in a way that makes the edges overlap less. As a result, this network forms a network in a circular shape where the original network forms a square shape. The nodes can be found equidistantly dispersed forming a radius 360 degrees

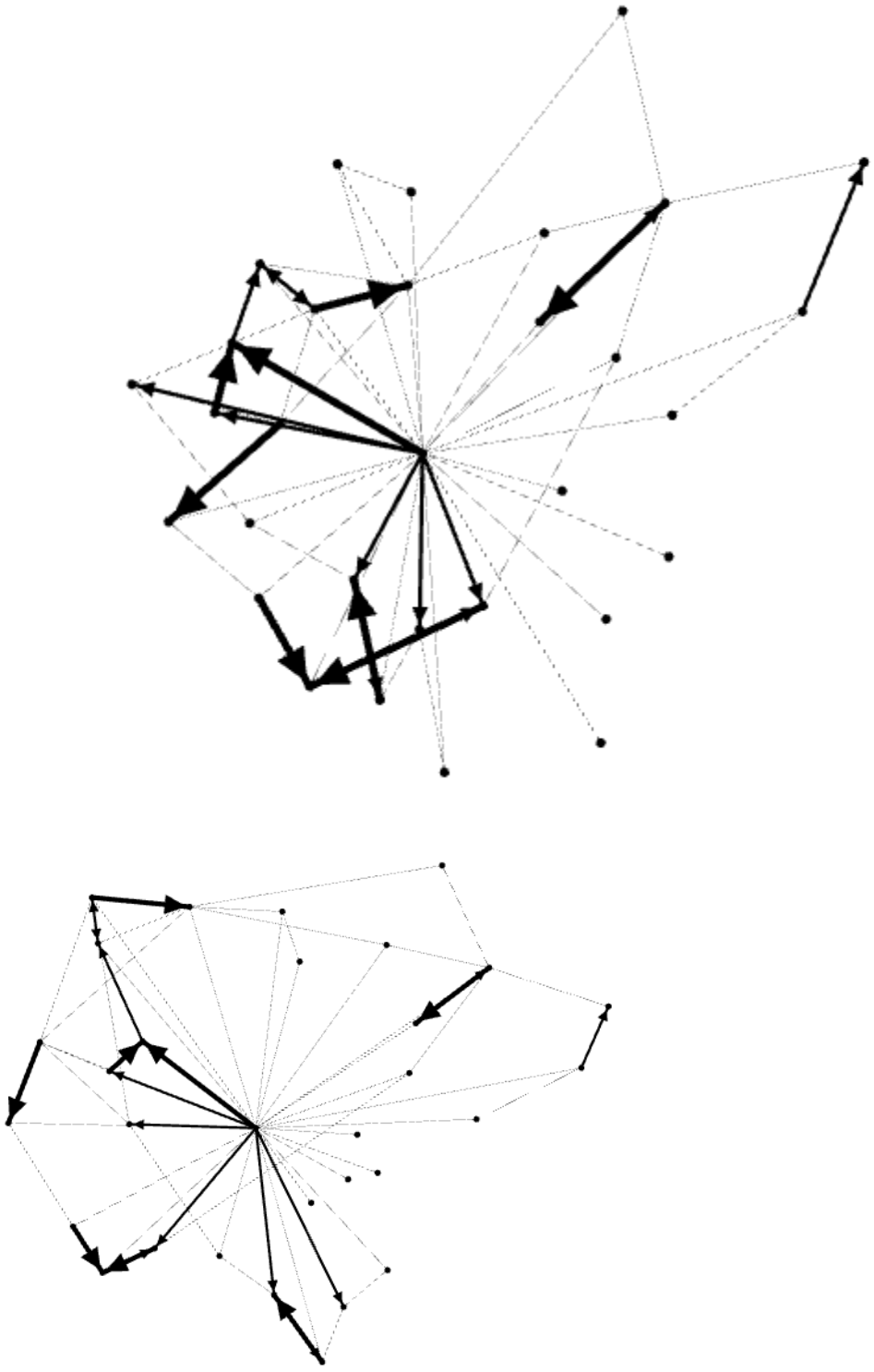
around the center of the graph.. The nodes towards the center seem to be a part of communities and have many links which in turn means those nodes in the center will have higher degrees. With a network of this smaller size with less than or equal to 31 nodes total, this layout could be helpful.

Yifan Hu - When comparing the Yifan Hu layout to the original default layout, the Yifan Hu layout spreads the nodes out away from the center. However, they are not spread equidistantly as Fruchterman Reingold. The original layout defaulted from Gephi keeps the nodes contained in a square-like shape. The new disbursement of the nodes can allow the clusters and communities to be more visible for analysis. With this layout, it is easier to identify nodes with smaller degrees as many appear to be located on the outer portions of the graph.

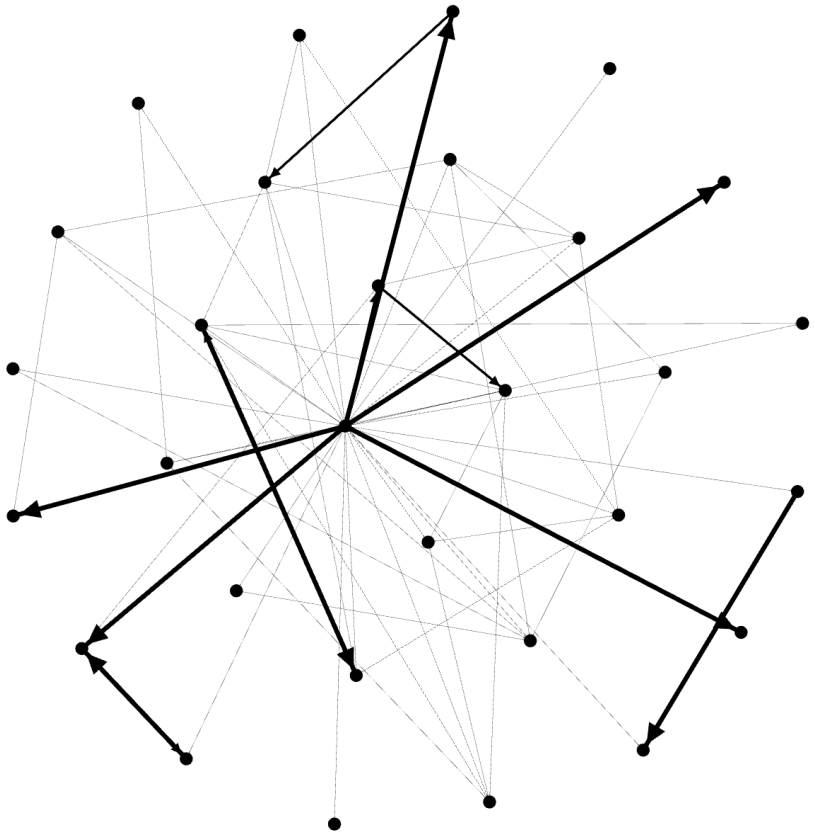
Force Atlas - (This layout was zoomed in by using the Expansion layout tool. The screenshot was captured after letting it run for 10 minutes). This layout is closest in comparison to the Yifan Hu layout. The related nodes appear to be located closer together, while nodes that are unrelated are located farther apart. Since the network is small, the closeness of the nodes makes it hard to view the entire graph without using the Expansion Tool. It is difficult to visualize any communities or clustering since the closeness of the nodes almost gives this network the dreaded “hairball” effect that often goes with this layout.

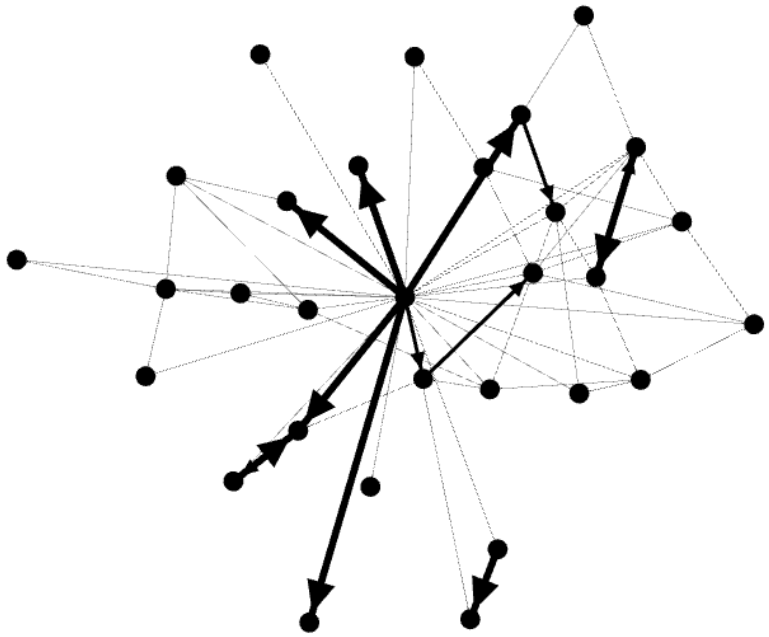
2017-2018:

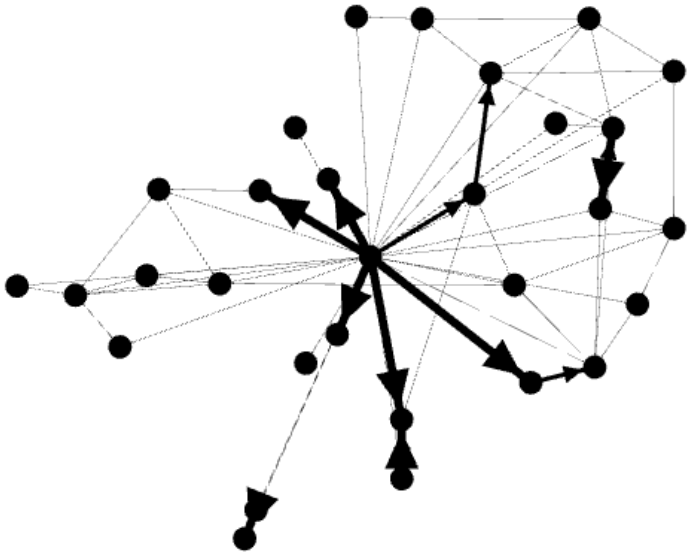




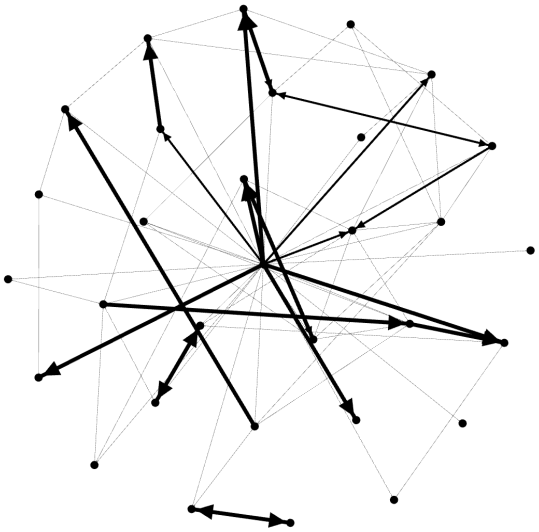
2018-2019:

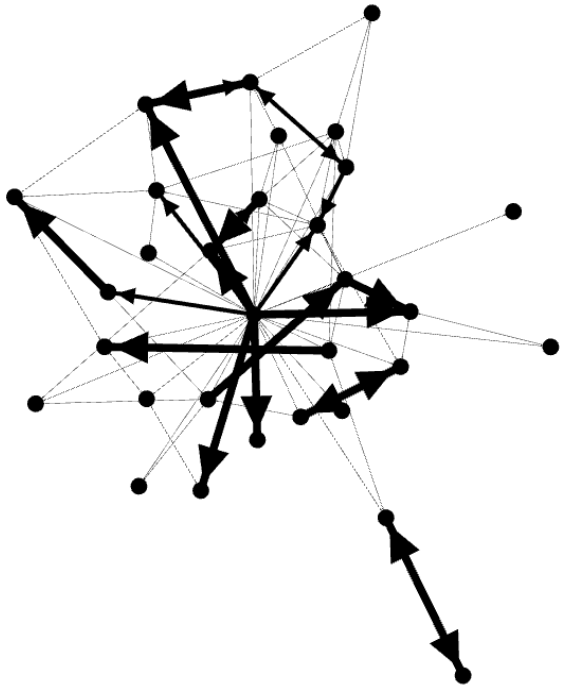


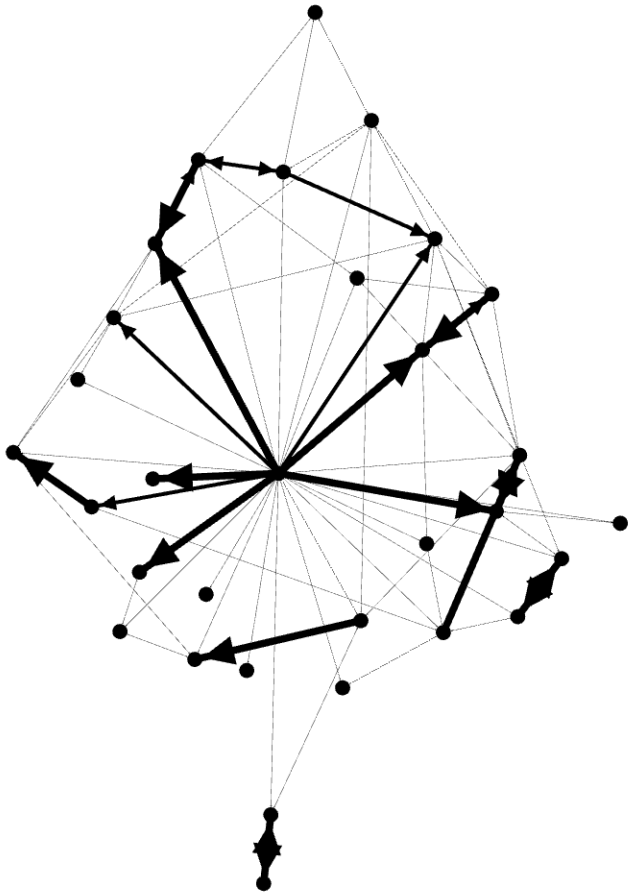




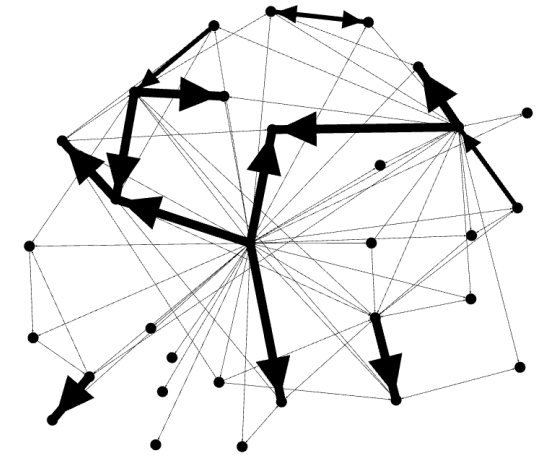
2019-2020:

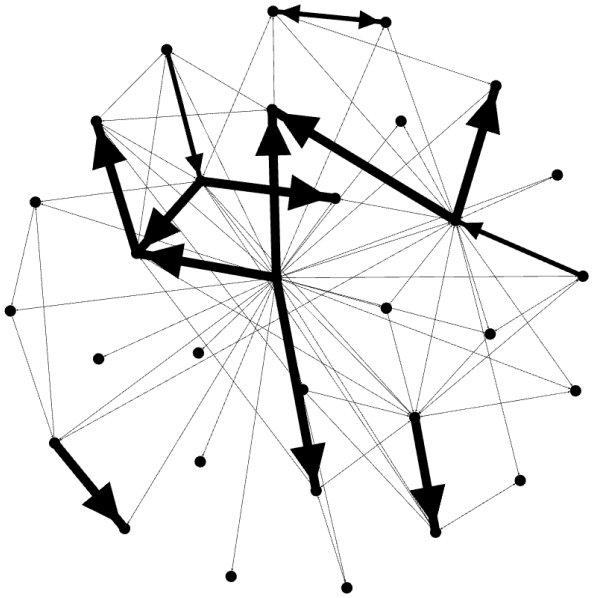


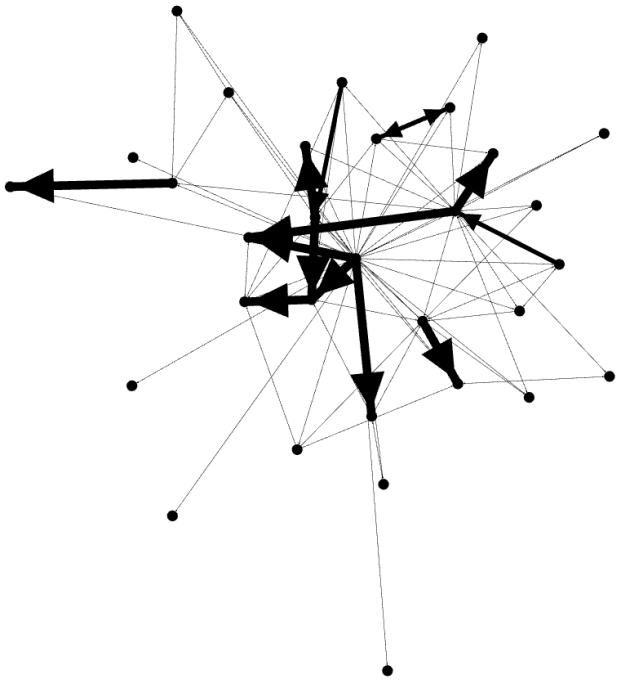




2020-2021:





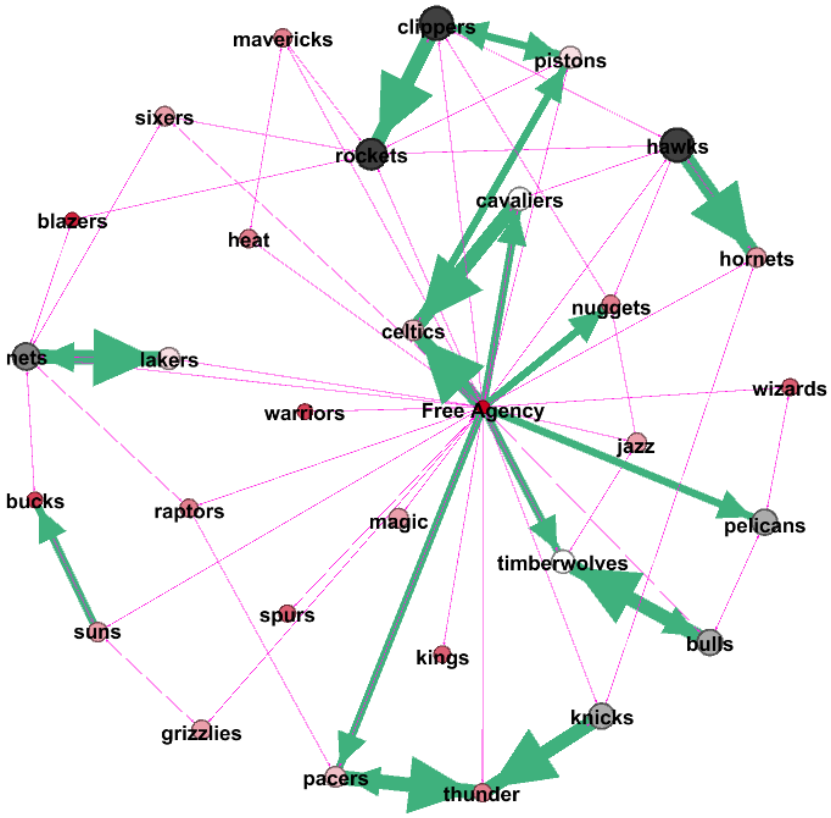


a. **At this point, which layout seems the most useful and why?**

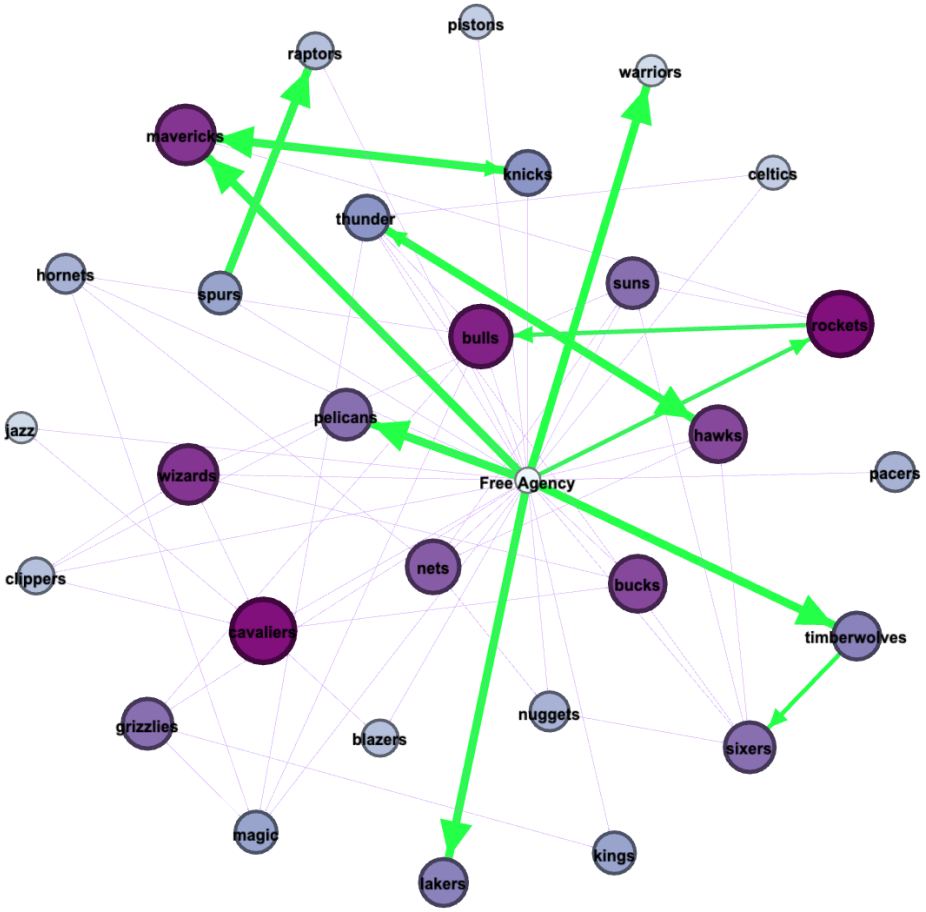
After running the different layouts on all 4 networks representing different years of trades, Fruchterman Reingold seems to be the most useful for further analysis. Visually, this layout places the hubs towards the center of the spherical layout. Additionally, with the hubs in the center of the graph, it is easier to see the different directions of the trades. The circular shape of the graph makes it easier to ascertain the degree of each node placed on the circumference of the graph.

4. **Now add some emphasis to your diagram by sizing, node coloring, and naming the nodes.**
   a. **Show a screenshot of this action**.

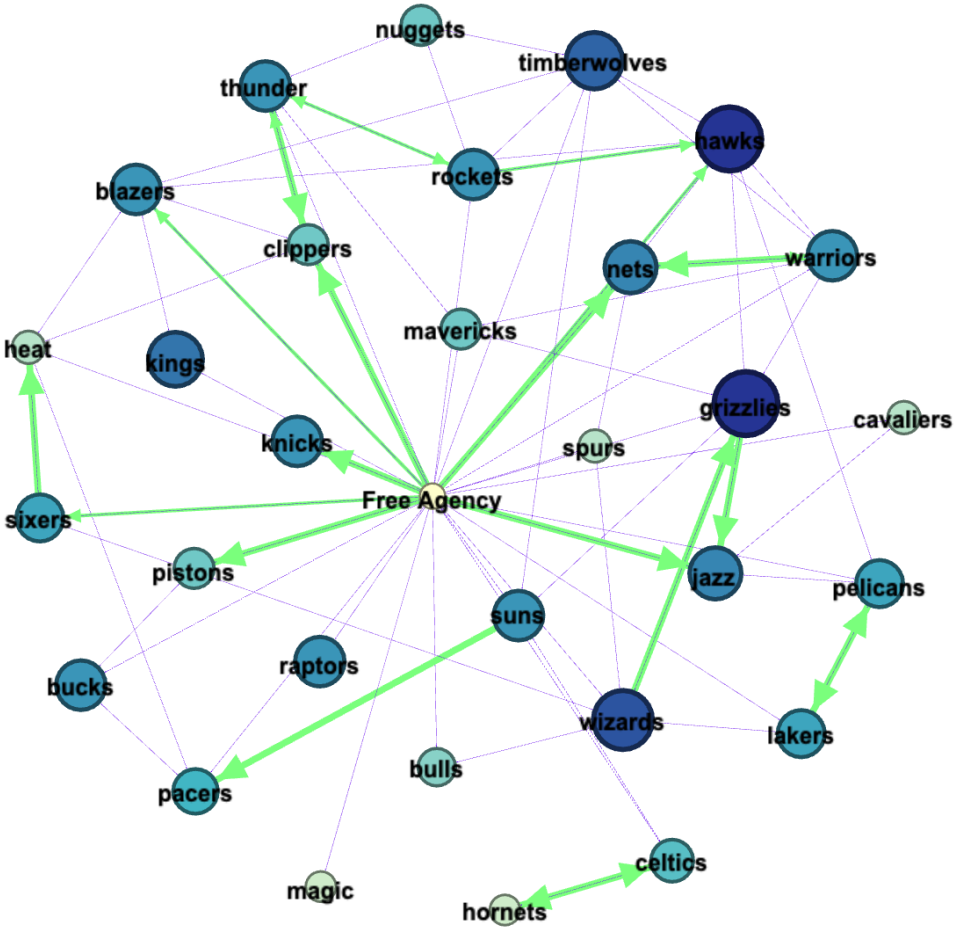
2017-2018:


clippers
mavericks
pistons
sixers
rockets
hawks
cavaliers
blazers
heat
hornets
nuggets
celtics
nets
lakers
wizards
warriors
Free Agency
jazz
bucks
raptors
magic
pelicans
timberwolves
spurs
suns
bulls
kings
knicks
grizzlies
pacers
thunder

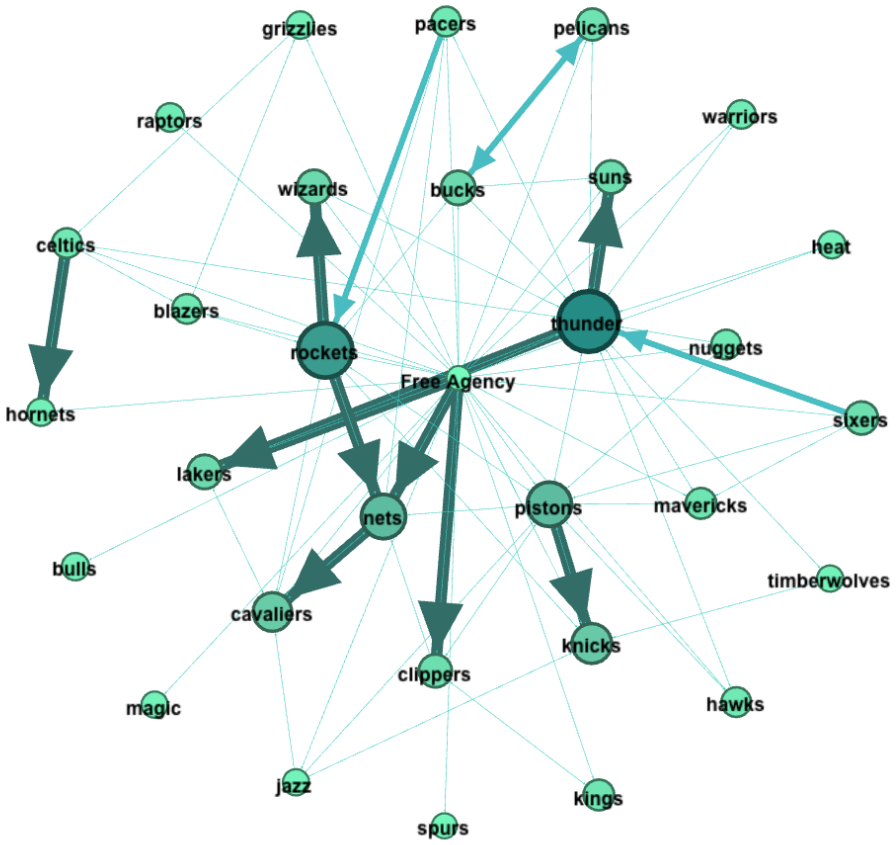
2018-2019:


pistons
raptors
warriors
mavericks
knicks
celtics
thunder
hornets
spurs
suns
rockets
bulls
pelicans
jazz
hawks
wizards
Free Agency
pacers
clippers
nets
bucks
timberwolves
cavaliers
grizzlies
nuggets
blazers
sixers
magic
kings
lakers

2019-2020:


nuggets
timberwolves
thunder
hawks
rockets
blazers
clippers
nets
warriors
mavericks
heat
kings
grizzlies
cavaliers
knicks
spurs
Free Agency
sixers
pistons
jazz
pelicans
suns
raptors
bucks
wizards
lakers
bulls
pacers
celtics
magic
hornets

2020-2021:



b.

What have you learned from your new diagram?

As stated in the preliminary report, a small market team is identified as a team in the bottom 15 of the media market rankings. A large market team is in the top 15. For all years we can observe that the free agency node has the smallest in-degree. The size of the nodes is not dependent on whether a team is considered in a large or a small market. In a similar way the all-stars trades edges are evenly distributed amongst small and large market teams.

To format each weighted, directional network, node size was adjusted according to the In-Degree rankings, edge colors are partitioned by all star versus non all star trades, and node labels are on display. Node color is determined by the In-Degree rankings.

5. **Now run some statistics about the network such as average degree, density, betweenness, Eigenvector centrality, clustering, etc. You decide which ones are most relevant.**

    a. Explain what statistical results you found. Show graphs, plots, and numbers generated by your software and explain what impact it had on your analysis.

    From the statistics run, shown in the plots below and the overview shown in question 2, all networks are relatively not dense with a density statistic of ~0.2 for each network. The cluster coefficient is closer to 0 than 1 so this also confirms the lack of density within this network. This would mean that more trades could have been formed but teams opted to remain close to their local communities.

    Plots and reports generated from the statistics by year:

2017-2018:

The network is broken down into 6 communities.

Modularity Report

Parameters:

Randomize: On
Use edge weights: On
Resolution: 1.0

Results:

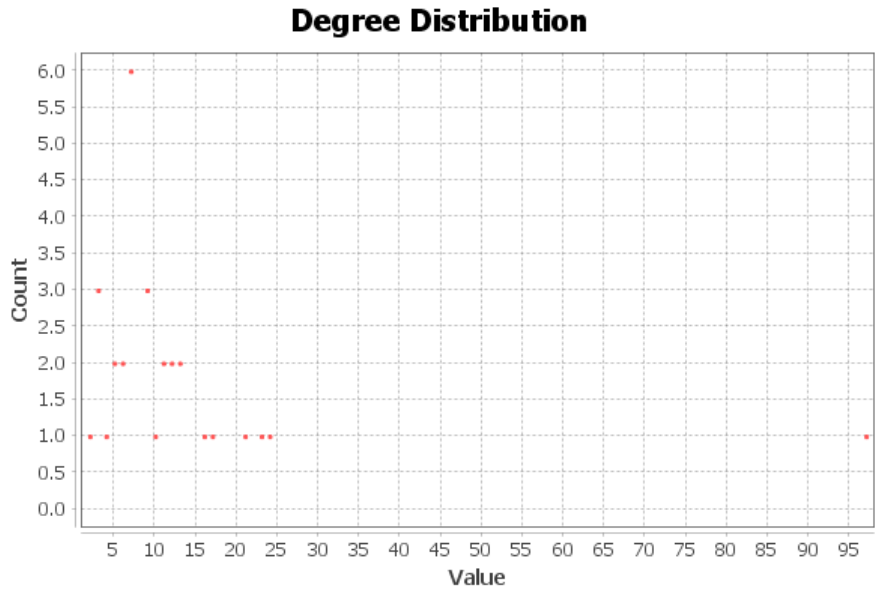
Modularity: 0.285
Modularity with resolution: 0.285
Number of Communities: 6

It is a scale free network due to the shape of the graph. This makes sense given the hub that is the free agency node.

Degree Report

Results:

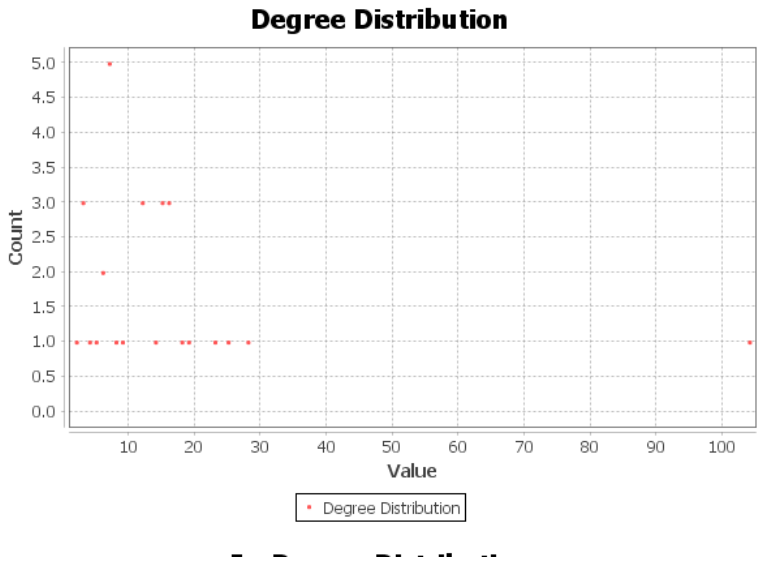
Average Degree: 6.226



Weighted Degree Report

Results:

Average Weighted Degree: 7.161



We notice that when we consider the weight, the average weight degree goes up, which might be an indication that the nodes which have higher weighted edges, have more edges on average.

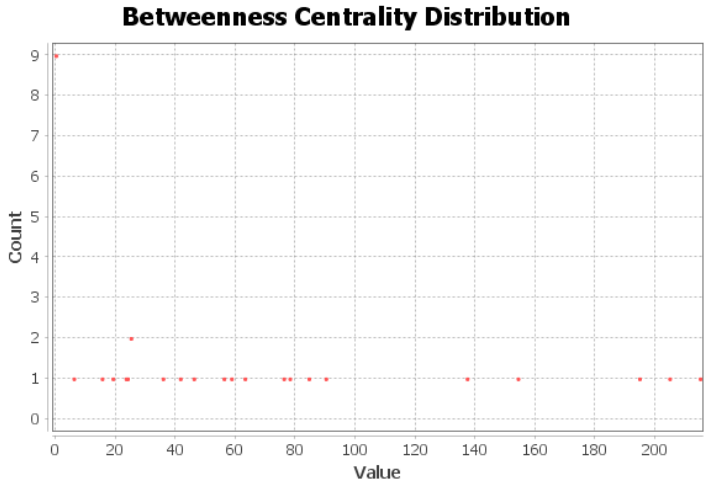
Graph Distance Report

Parameters:

Network Interpretation: directed

Results:

Diameter: 12
Radius: 0
Average Path length: 4.172675521821632



The average path length being this short implies that trades are frequent between teams without distinction of market size.

2018-2019:

There are four communities.

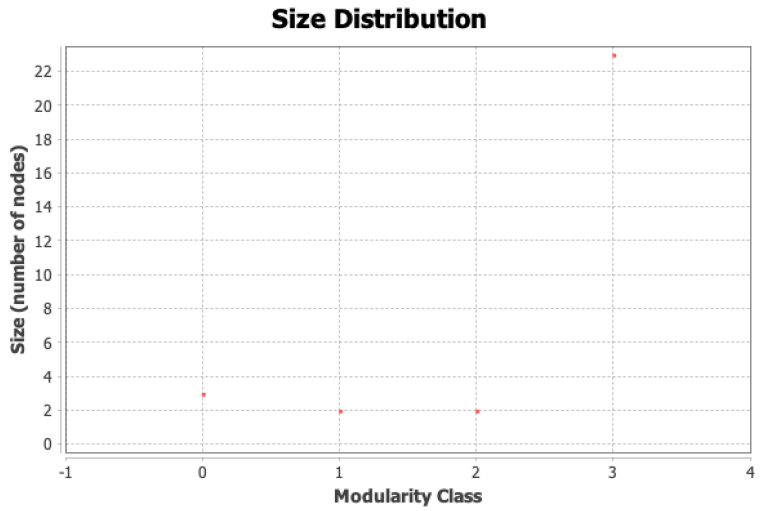
Modularity Report

Parameters:

Randomize: On
Use edge weights: On
Resolution: 1.0

Results:

Modularity: 0.090
Modularity with resolution: 0.090
Number of Communities: 4

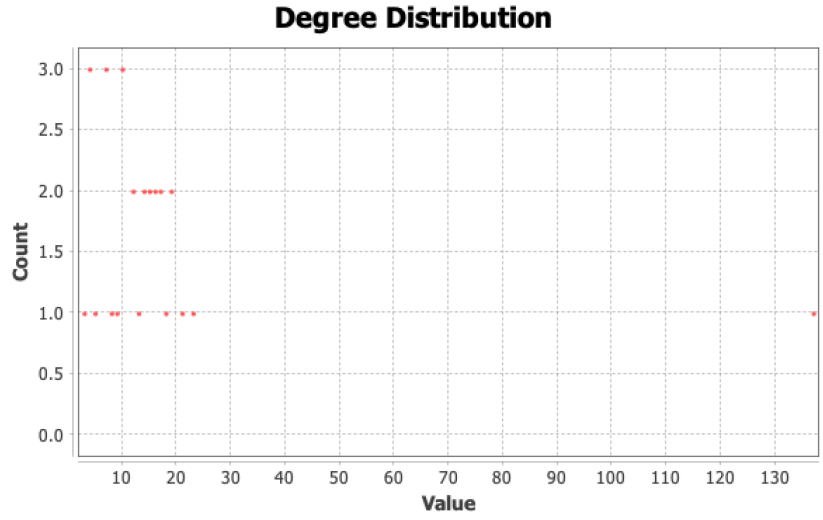


This is also a scale free network.

Weighted Degree Report

Results:

Average Weighted Degree: 8.100



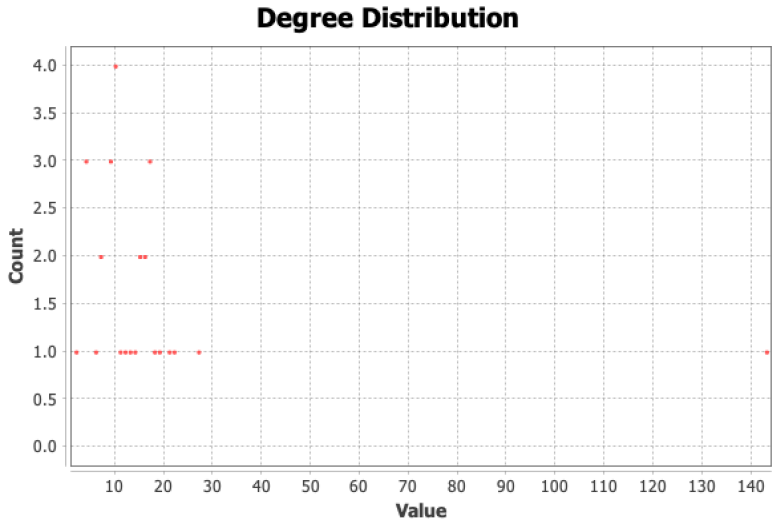
2019-2020:

This is also a scale free network due to the curve shape of the graph.

Degree Report

Results:

Average Degree: 8.290



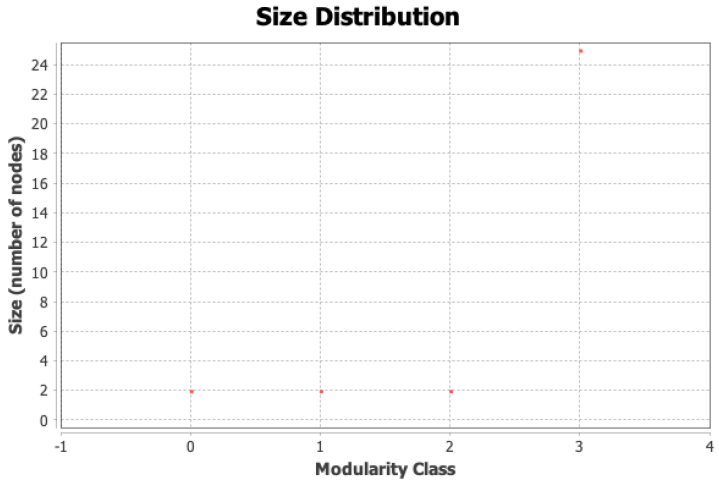
Modularity Report

Parameters:

Randomize: On
Use edge weights: On
Resolution: 1.0

Results:

Modularity: 0.107
Modularity with resolution: 0.107
Number of Communities: 4

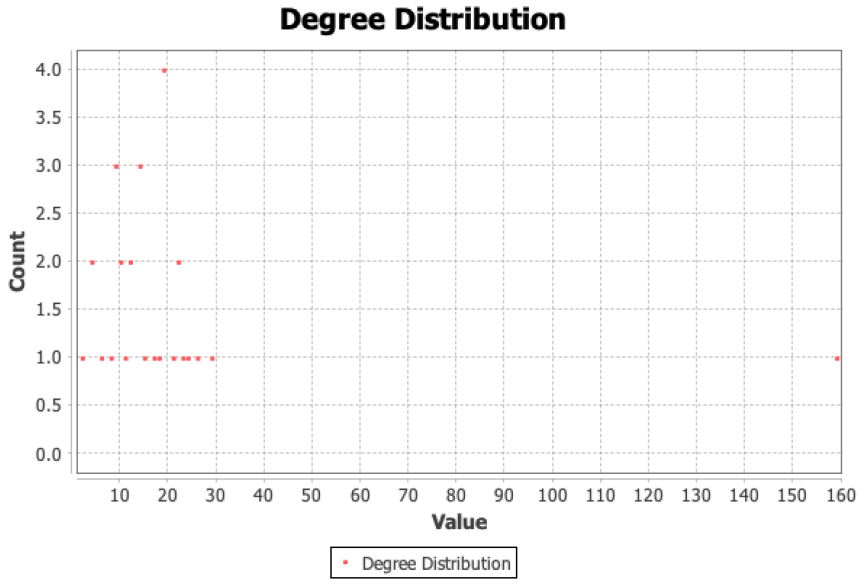


There are four communities.

Weighted Degree Report

Results:

Average Weighted Degree: 9.677

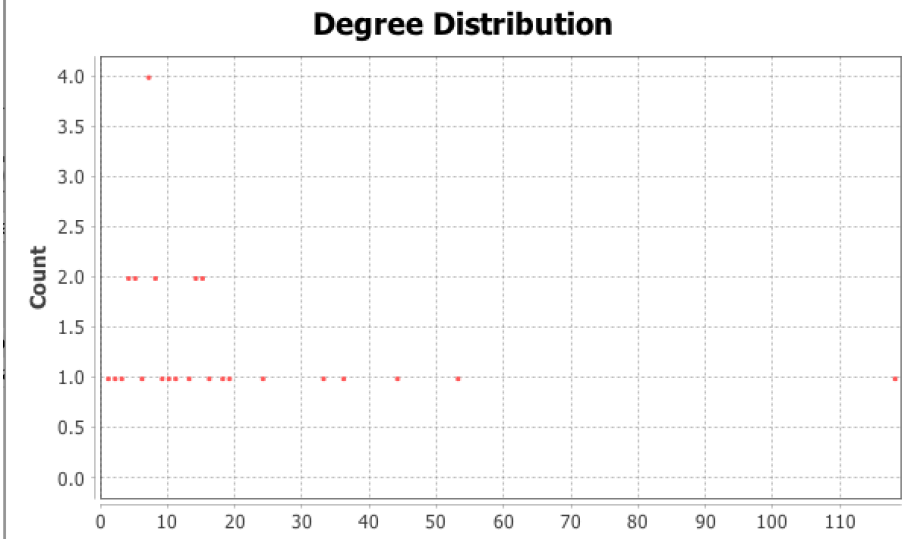


2020-2021:

This is a scale free network.

Results:

Average Weighted Degree: 8.645



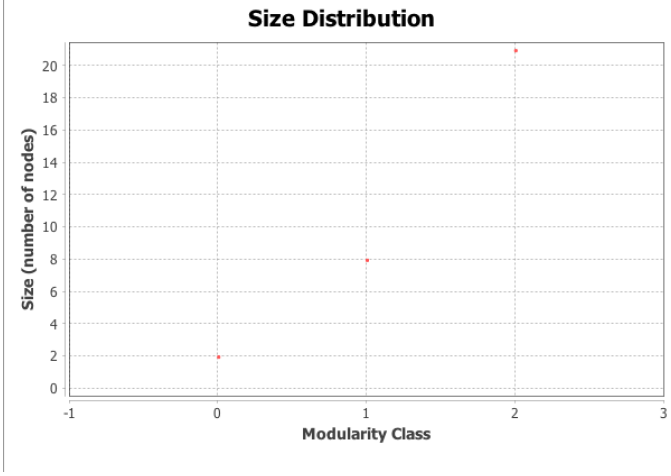
There are 3 communities.

Parameters:

Randomize: On
Use edge weights: On
Resolution: 1.0

Results:

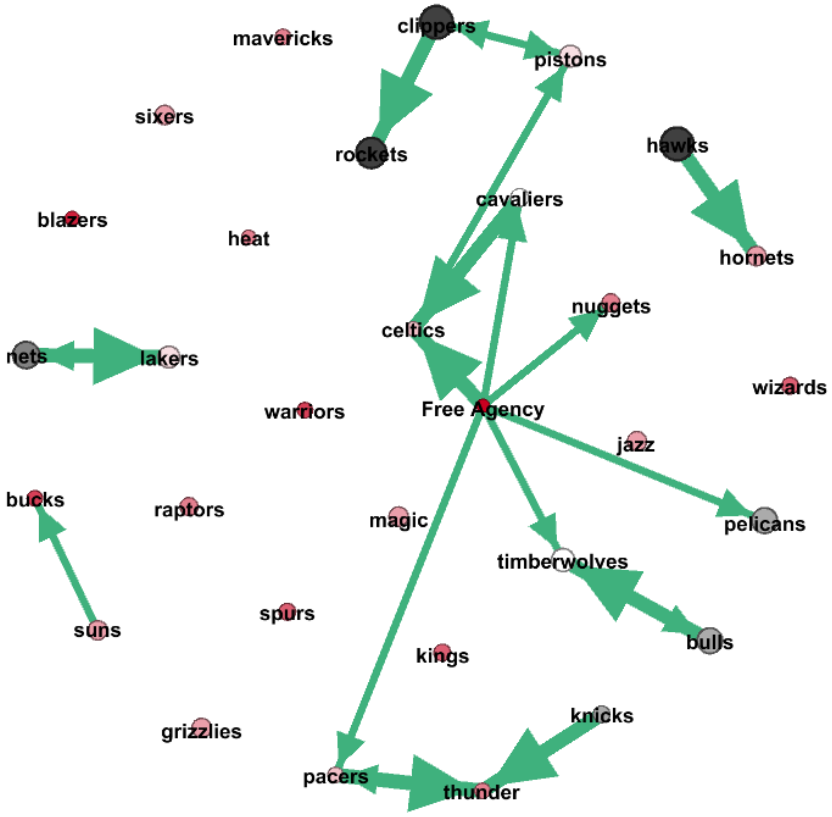
Modularity: 0.151
Modularity with resolution: 0.151
Number of Communities: 3

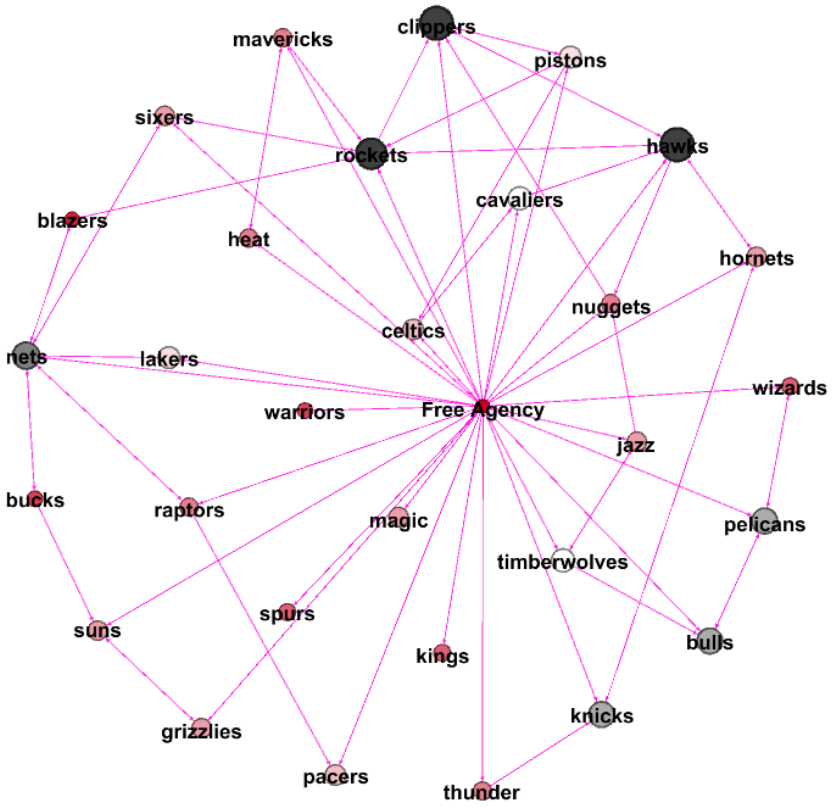


6. **Now run some filters on the network. You decide which ones are most relevant.**
    a. **Explain what results you found. Show graphs, plots, and numbers generated by your software and explain what impact it had on your analysis.**

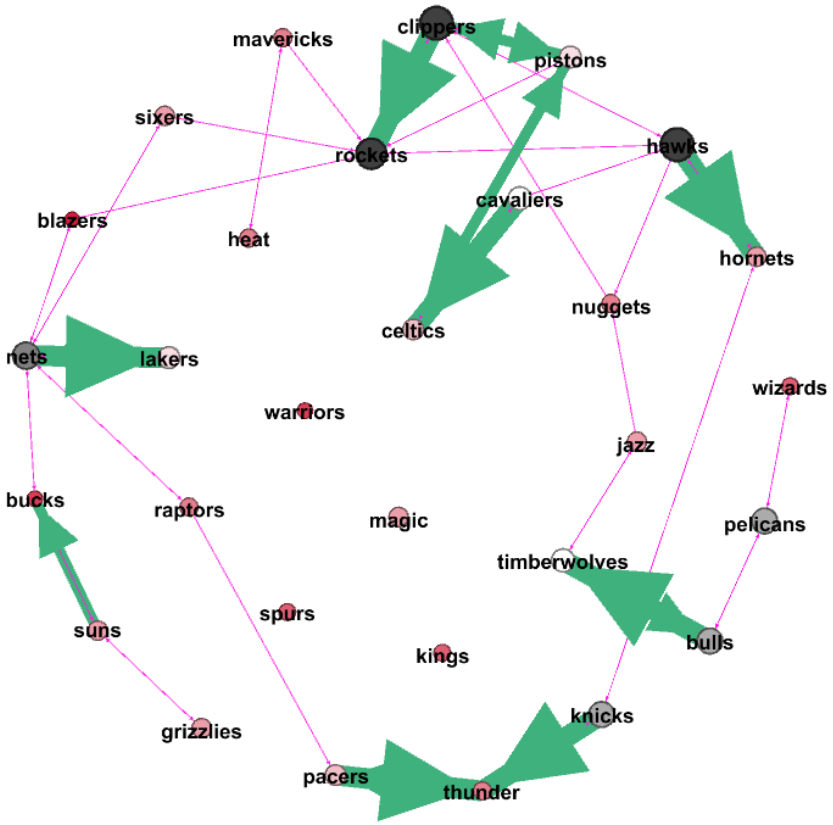
2017-2018:

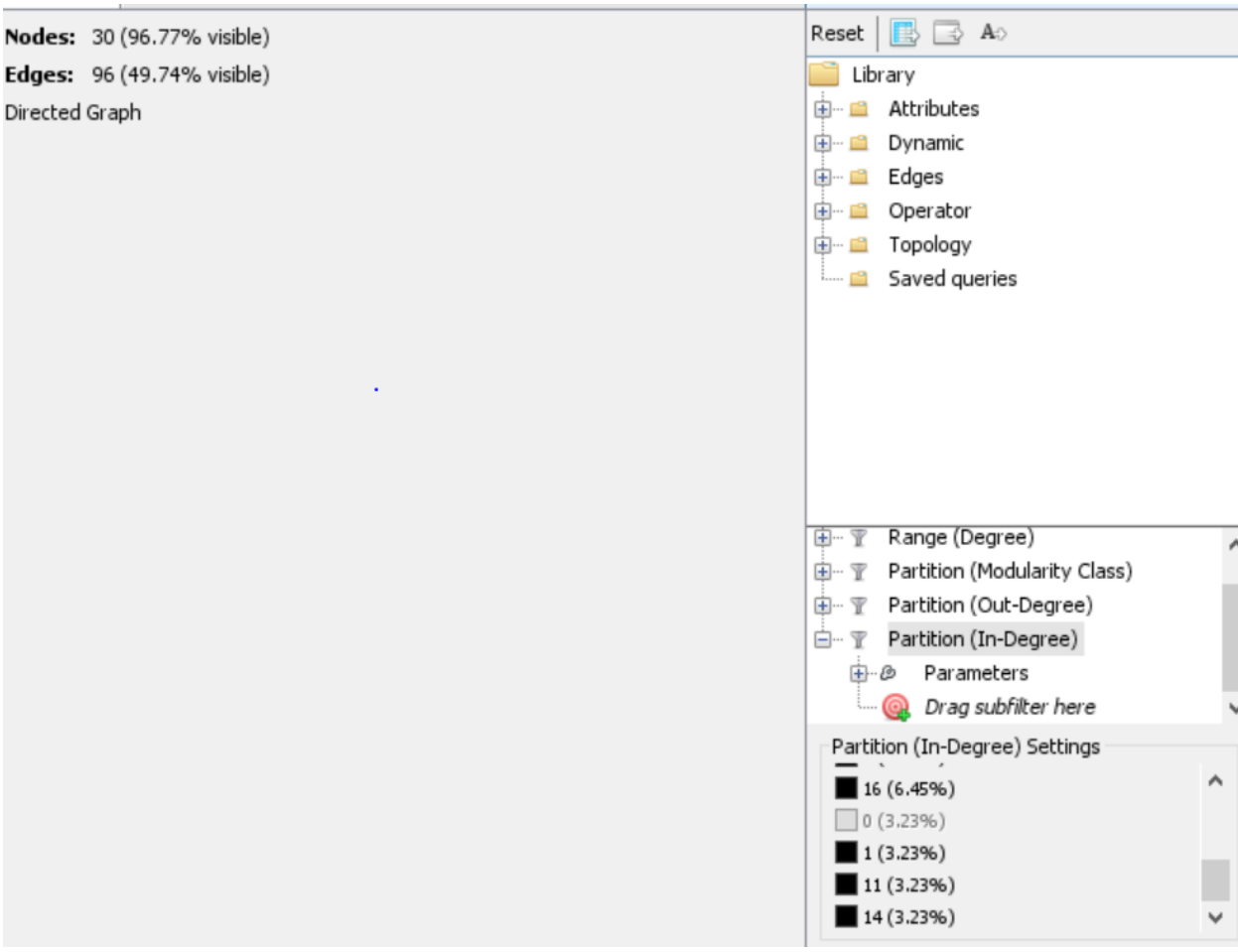
Partition for all stars: Display the edges which involve an all star trade (Green) or not. We notice a lot of movement in small market teams for all stars trades.



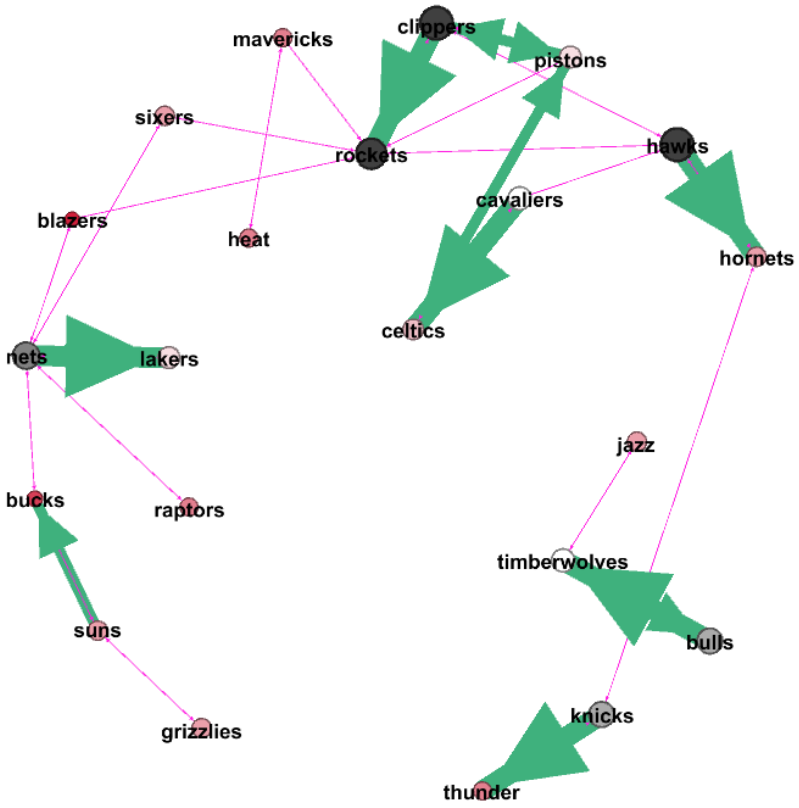


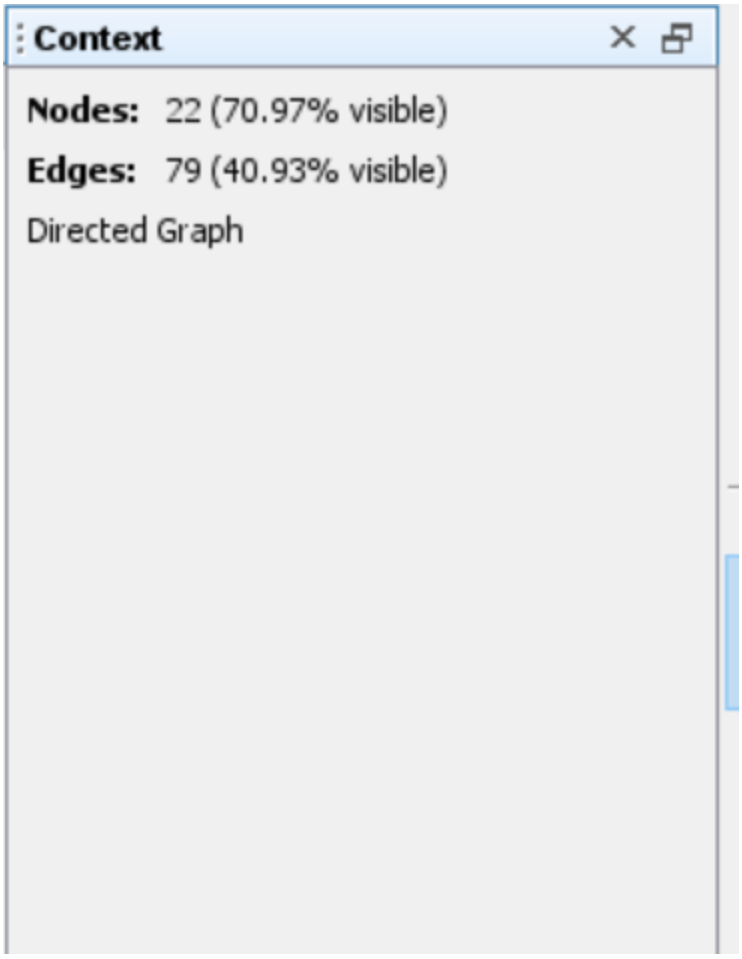
Partition in-degree: One node (Free Agency) combines for over 50% of the edges. Here we show the rest of the nodes and there is no distinction between small and large markets.


clippers
mavericks
pistons
sixers
rockets
hawks
cavaliers
blazers
heat
hornets
nuggets
celtics
nets
lakers
wizards
warriors
jazz
bucks
raptors
magic
pelicans
timberwolves
spurs
suns
bulls
kings
knicks
grizzlies
pacers
thunder

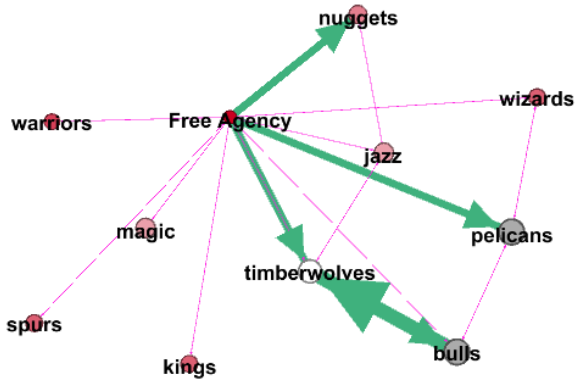


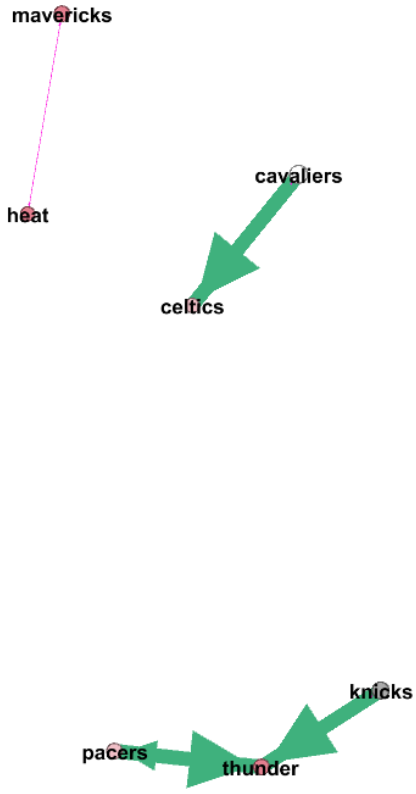
Partition out degree: We see that 22 nodes combine only to 79 edges out of 193. Of those nodes only 9 nodes are large market teams.


clippers
mavericks
pistons
sixers
rockets
hawks
cavaliers
blazers
heat
hornets
celtics
nets
lakers
jazz
bucks
raptors
timberwolves
suns
bulls
knicks
grizzlies
thunder



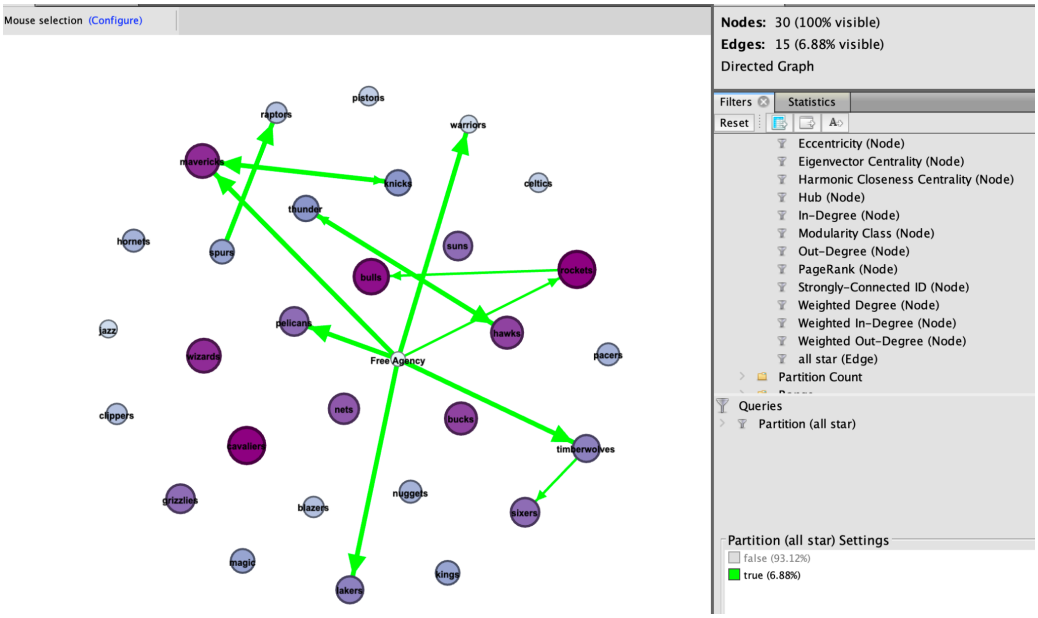
Partition modularity: Here displayed are 4 of our 6 communities, with the largest in the first picture and the 3 smallest in the second picture. 7 (out of 15) of our small market teams find themselves in the largest community.


nuggets
wizards
warriors
Free Agency
jazz
magic
pelicans
timberwolves
spurs
bulls
kings

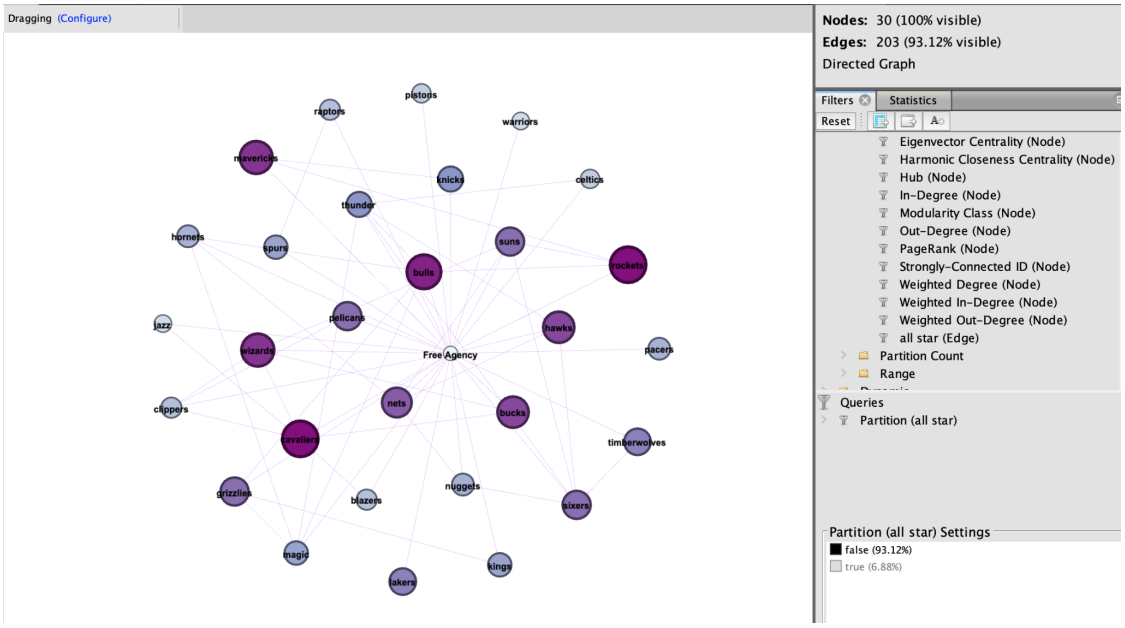


2018-2019:

Partition by All-Star Attribute - This filter is only displaying the edges that involve All-Star trades. From the context section it states that there are only 15 out of the original 218 edges in the network that represent the all-star trades. Only about 7% of players this season are all-stars.

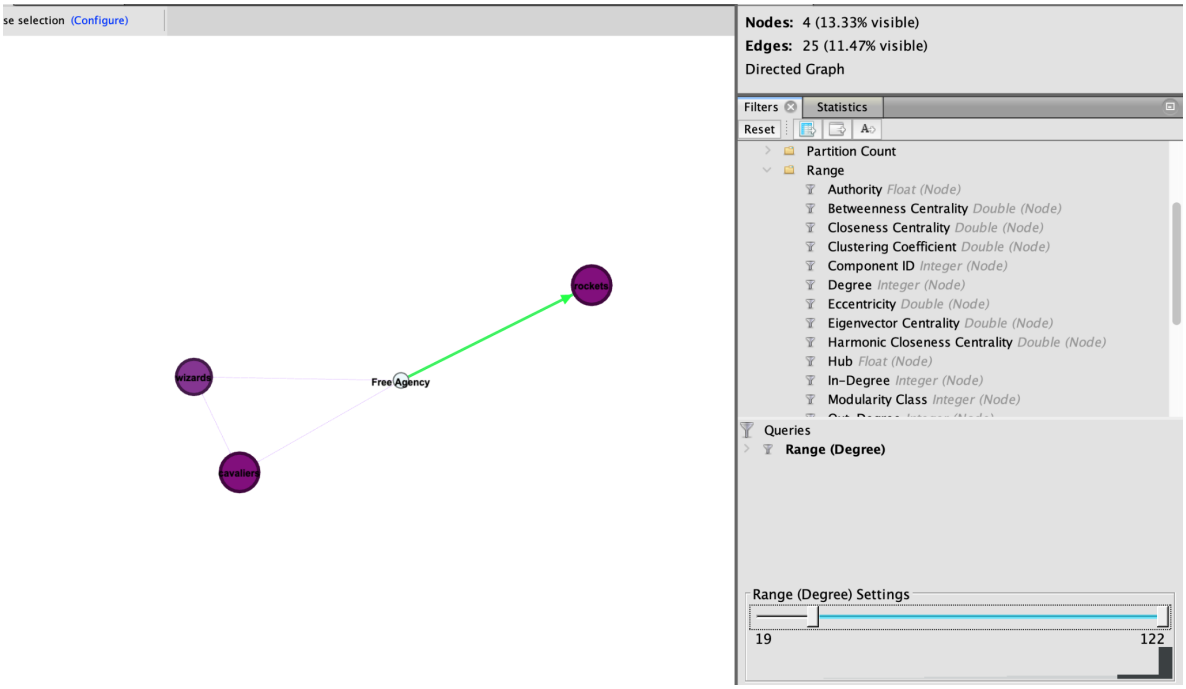


This filter is showing the edges that are not involving all-star trades. From the context section all nodes remain visible but only 93.12% edges are showing. This means 93.12% of the network consists of non all-star player transactions.

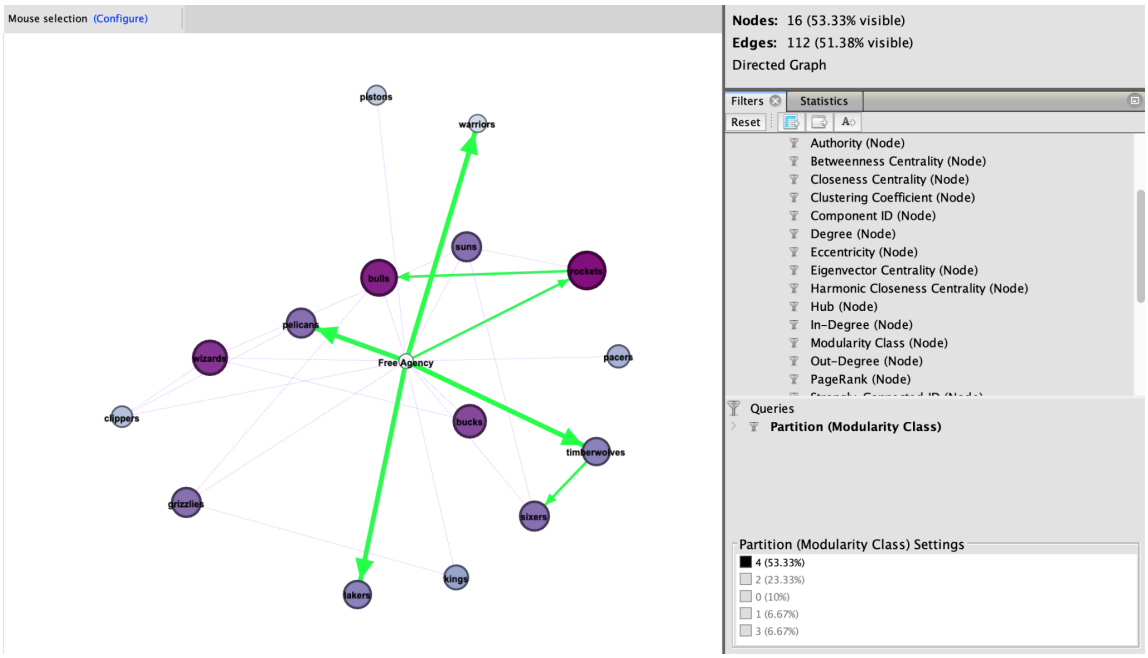


Range Degree Filter - This filter is showing nodes with degrees in the range of 19 - 122. The Free Agency Node has the highest degree. In addition, it has the highest out-degree but the lowest in-degree. The reason for this is that many players got claimed off free agency, but our study did not include waiving players. The next three nodes that have a

larger degree are the Wizards, Cavaliers, and the Rockets.

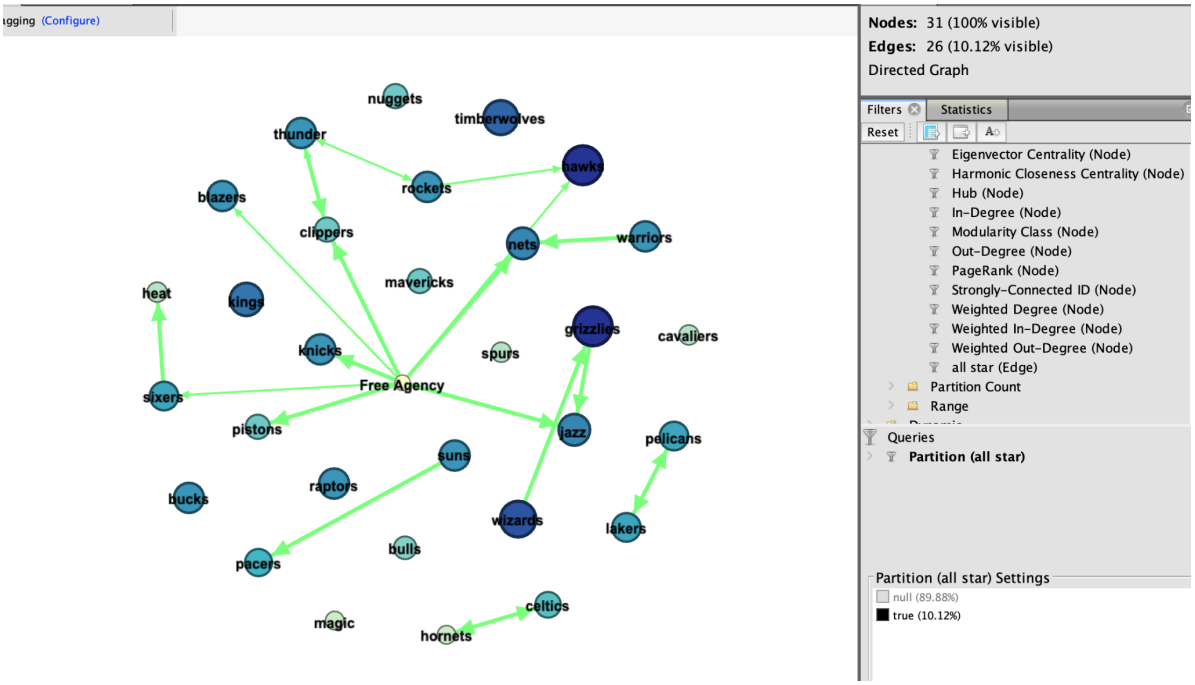


Partition Modularity Class: This filter shows the modularity class 4. After running the modularity statistic, it is revealed that there are five communities. This filter only shows one of the five communities. It is also the biggest community that makes up over 50% of the network. There seems to be an even split between small and large market teams.

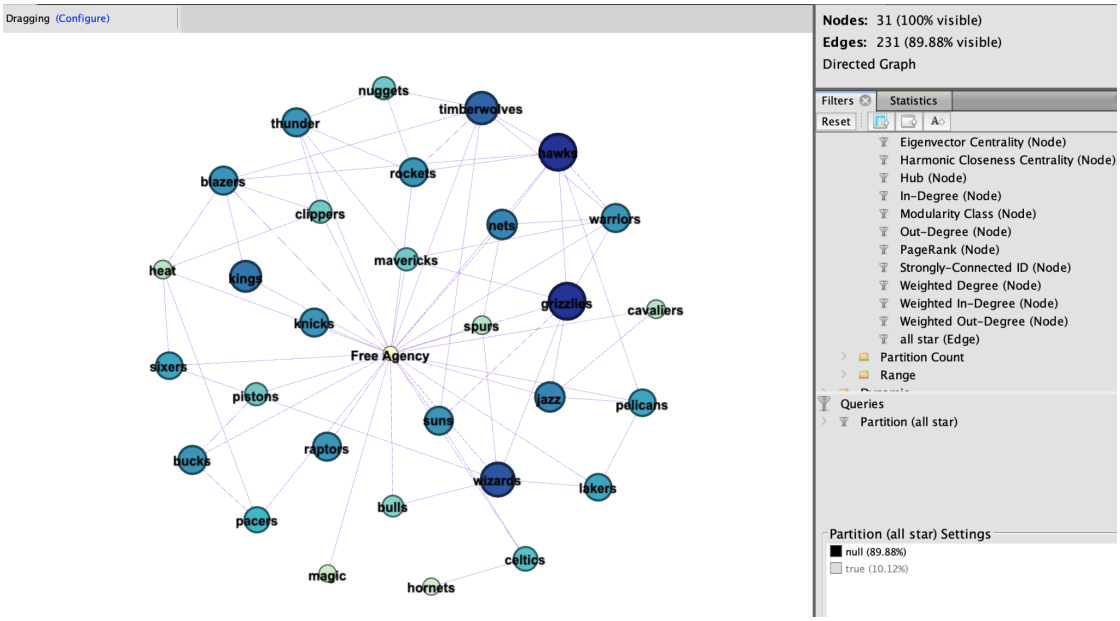


2019-2020:

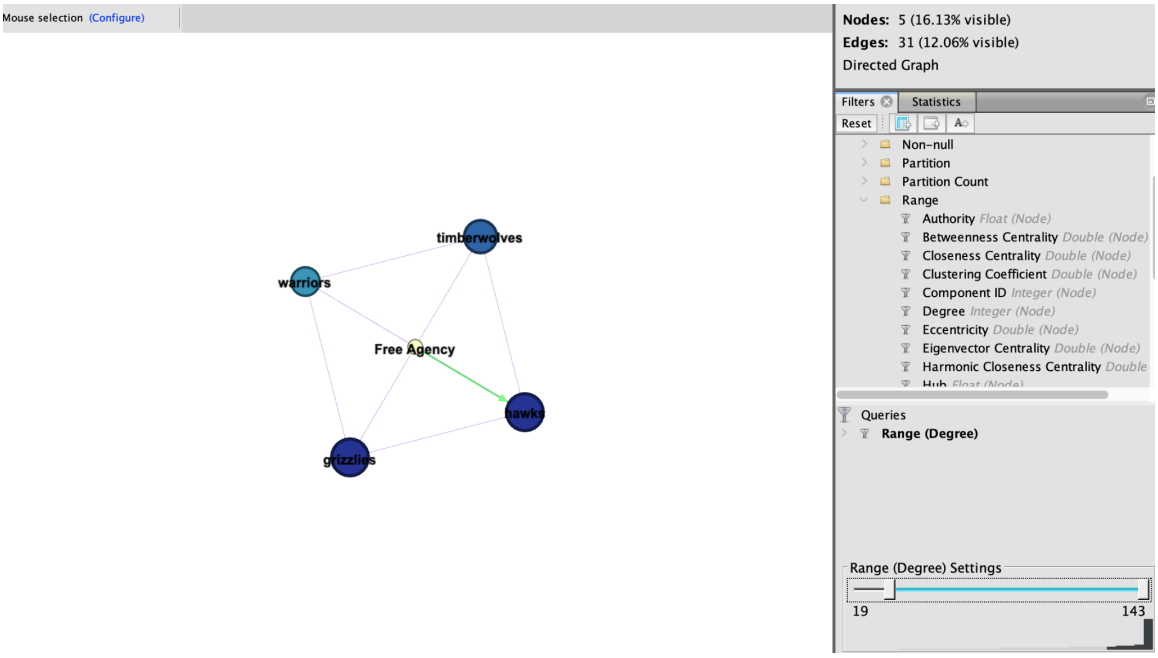
Partition All-Star Edge Attribute - This filter will partition the graph to only show edges that are related to all-star trades. From what is filtered, it is revealed that there are 10.12% all-stars within this network.



This image below shows all edges that are related to non all-star players. For this network 89.88% of trades are non all-star trades. All nodes remain after using this filter because the filter will only apply to the edges.

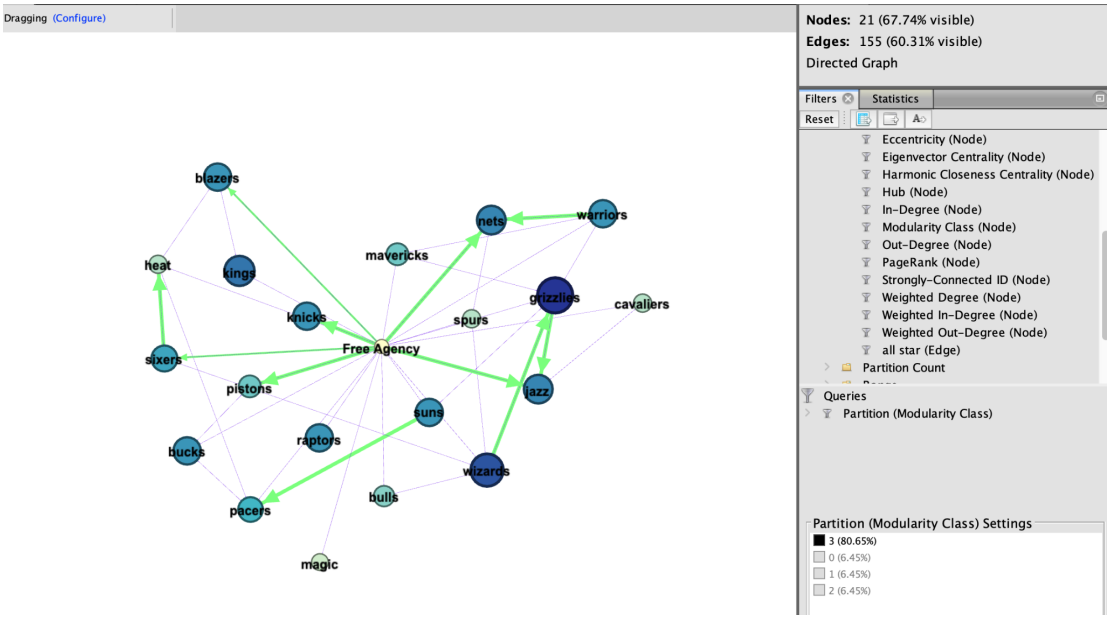


Range Degree Filter - This filter is only displaying nodes with degrees in the range of 19 to 143. This means that these nodes deal with the most trades in the network overall. Free Agency will have the smallest in-degree, however it has the highest out-degree of 143. This occurs because most of the players in this study were claimed from Free Agency.

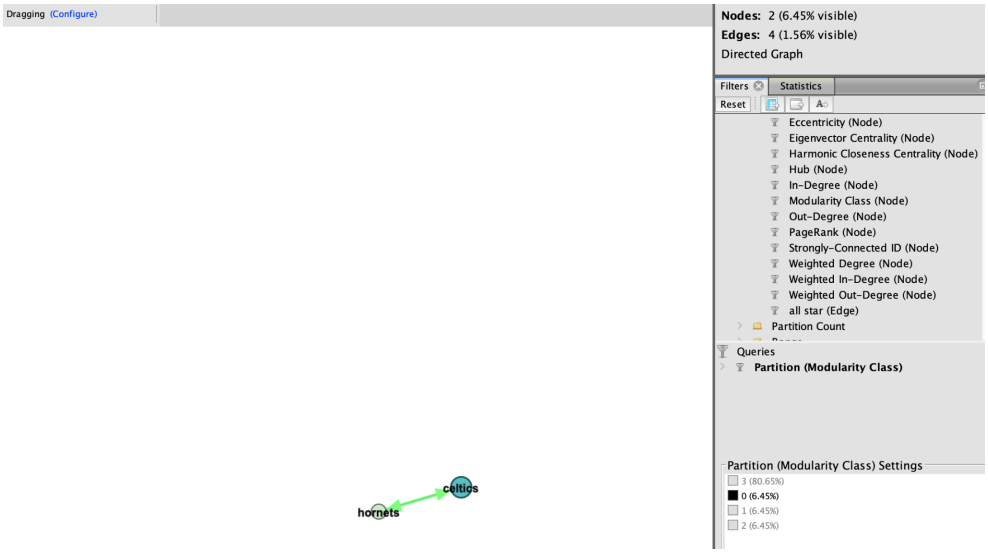


Partition Modularity Filter - After running the modularity statistic, there are 6 identifiable communities. The free agency node forms a hub in the network as it has the most edges.

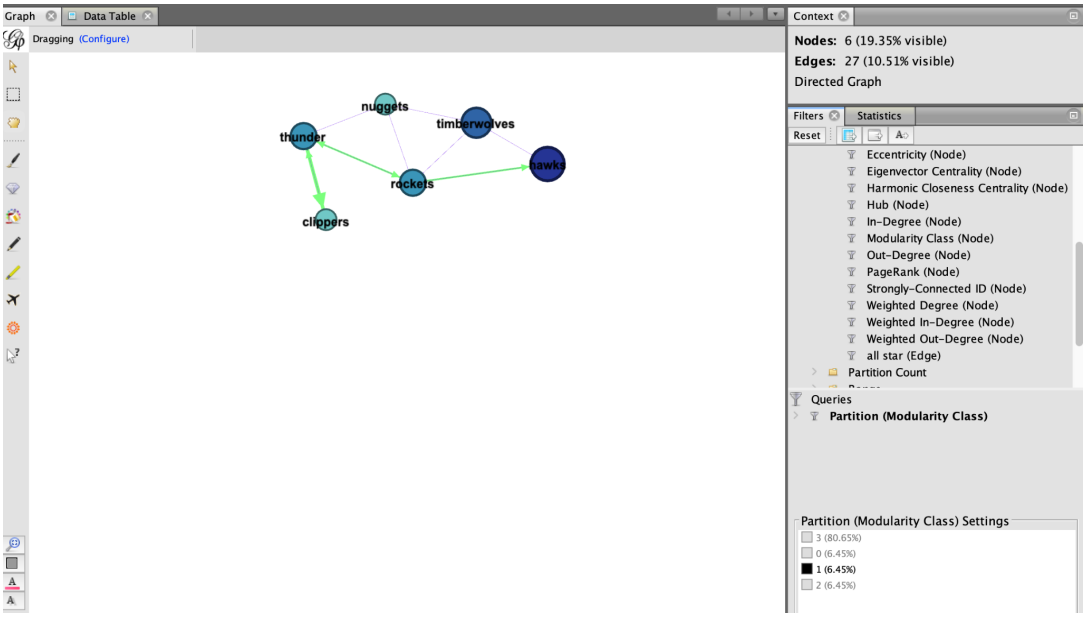
This image below shows the modularity class 3 that includes free agency. This is the largest community in the network.



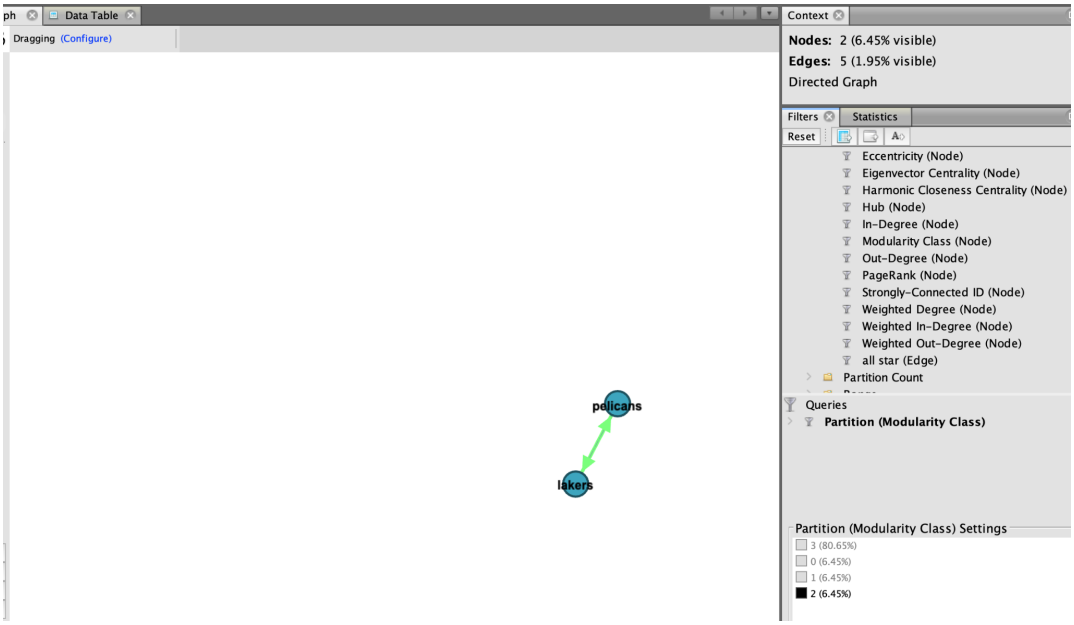
The next image shows modularity class 0 which only includes the Hornets and the Celtics. From our formatting, there are only all star trades occurring between these two nodes.



The next image shows modularity class 2. There are several all star trades occurring within this community. However, the Nuggets and the Timberwolves do not have any all-star trades occurring within this community.

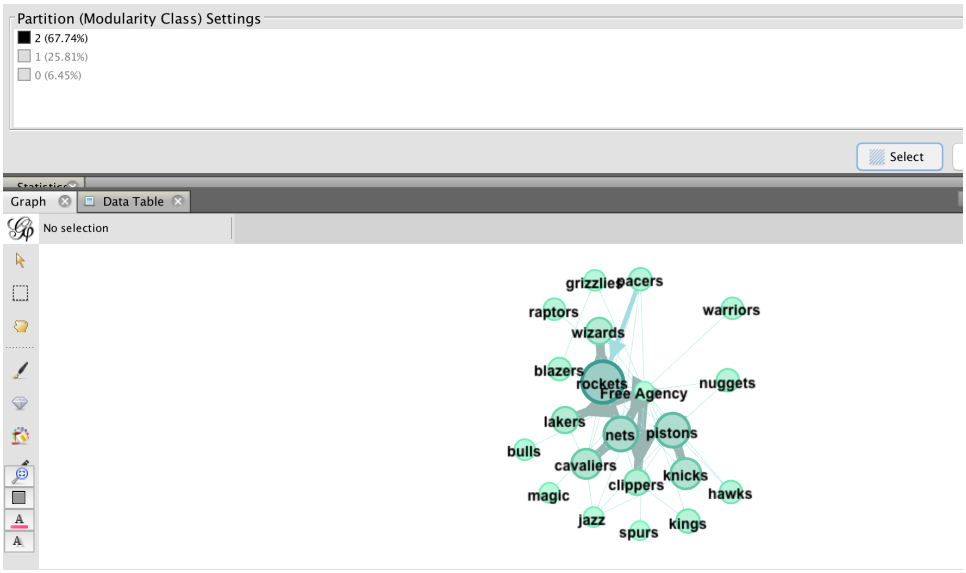


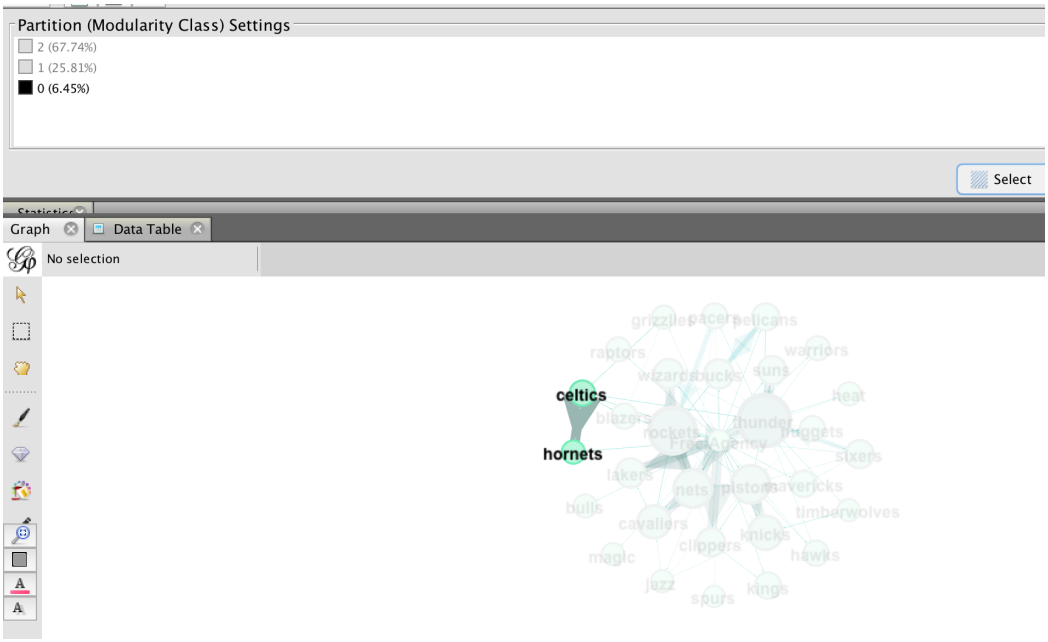
The last partitioned modularity class is class 2. Again, this community appears to involve only all star trades.

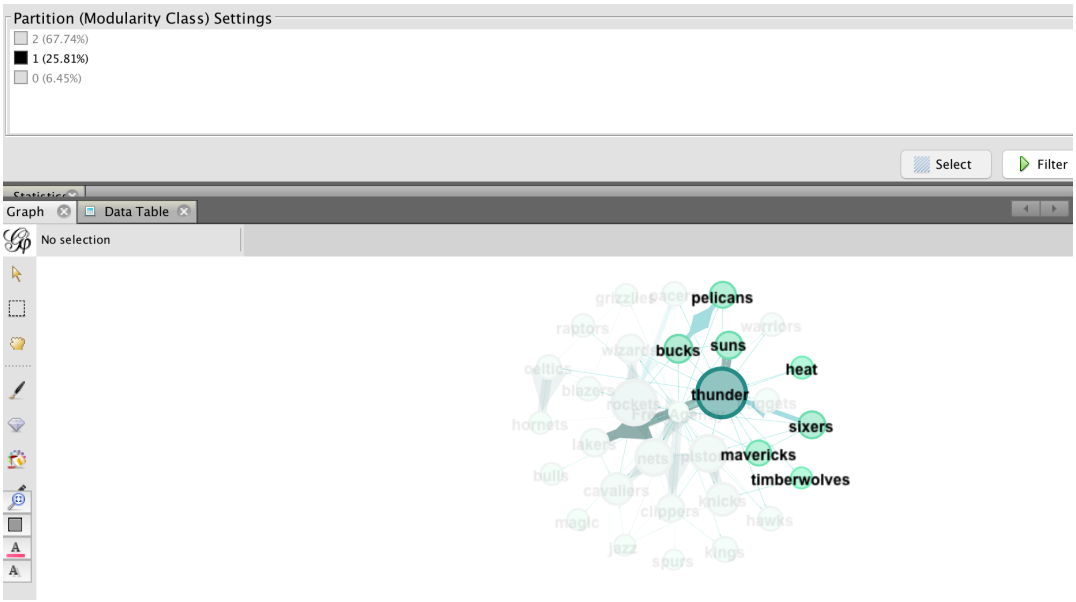


2020-2021:

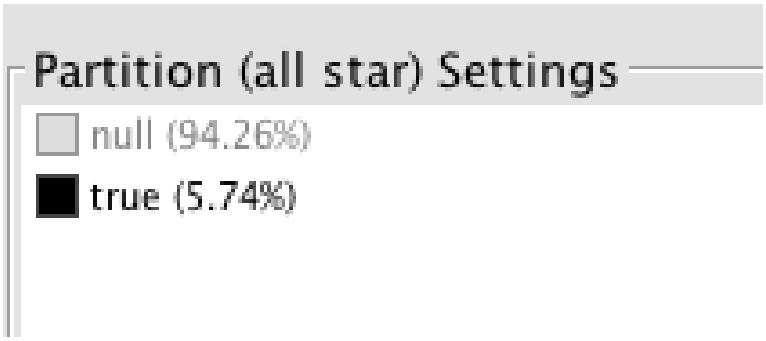
Partition Modularity: There are 3 communities present in this network. They are split with both small and large market teams present in every community. The largest community makes up about 70% of the network.


Partition (Modularity Class) Settings
2 (67.74%)
1 (25.81%)
0 (6.45%)
Select
Graph
Data Table
No selection
grizzlies
pacers
raptors
warriors
wizards
blazers
rockets
nuggets
Free Agency
lakers
nets
pistons
bulls
cavaliers
knicks
clippers
magic
hawks
jazz
kings
spurs

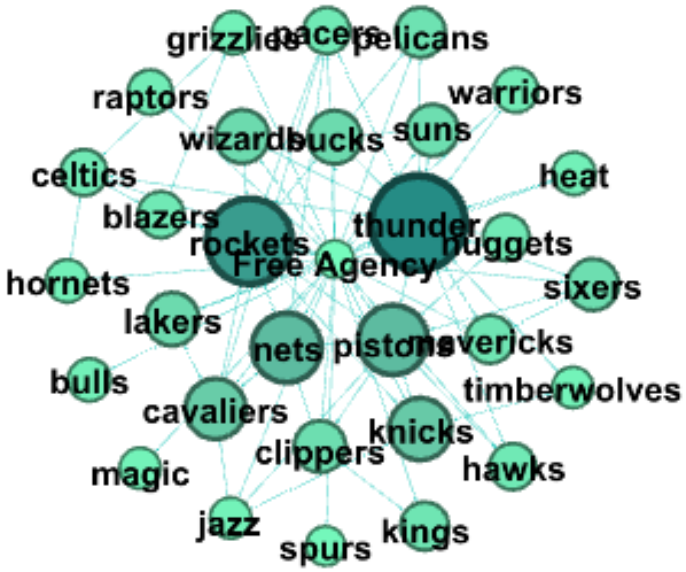

Partition (Modularity Class) Settings
2 (67.74%)
1 (25.81%)
0 (6.45%)
Select
Graph
Data Table
No selection
celtics
hornets

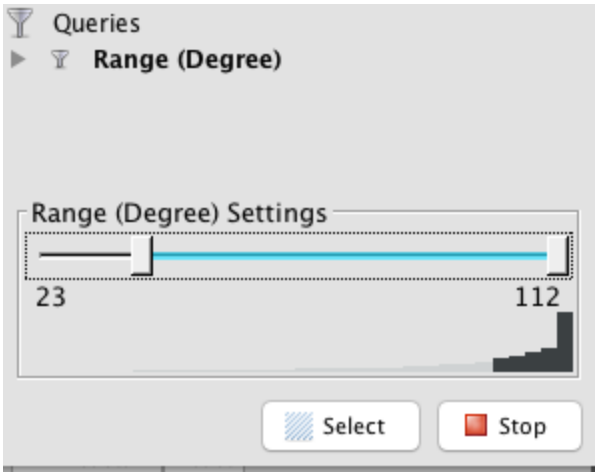


All-Star Partition: 95% of the players traded in this season were not considered all stars. The 5% of all stars is lower than the previous years.

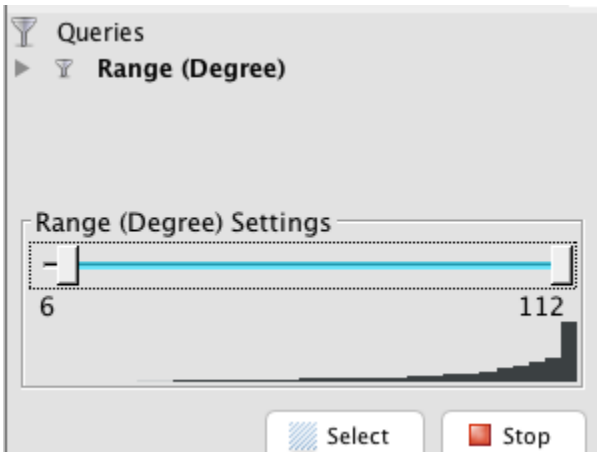


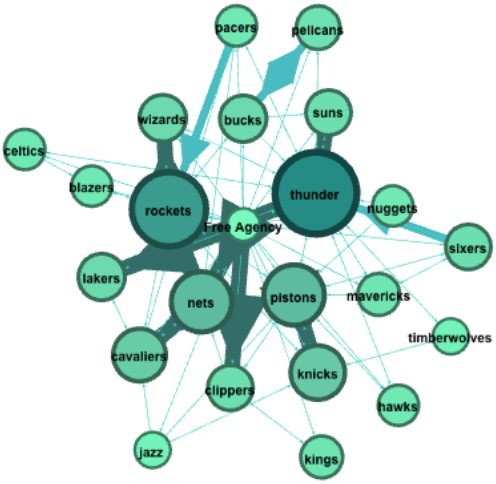
Range Degree: We can see that the free agency node has the highest degree as it has the most edges coming off of it. This occurred because most players in this study were claimed off free agency. The majority of nodes have a degree less than 6. Only four team nodes have a range over 20.

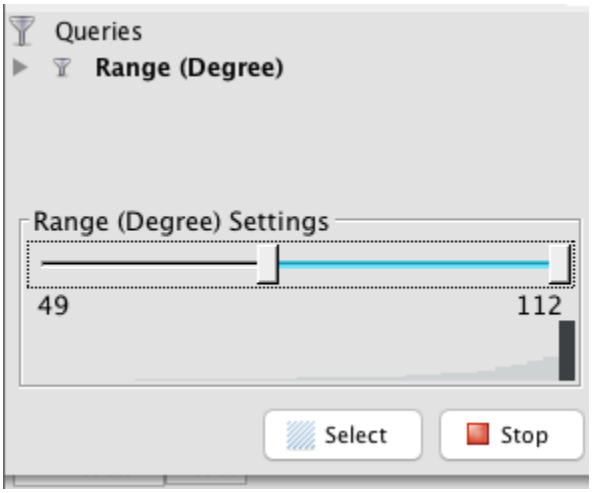

grizzlies
pacers
pelicans
raptors
warriors
wizards
bucks
suns
celtics
heat
blazers
thunder
rockets
nuggets
Free Agency
hornets
sixers
lakers
nets
pistons
mavericks
bulls
timberwolves
cavaliers
knicks
clippers
magic
hawks
jazz
kings
spurs


Queries
Range (Degree)
Range (Degree) Settings
23
112
Select
Stop


rockets
thunder
Free Agency
nets
pistons


Queries
Range (Degree)
Range (Degree) Settings
6
112
Select
Stop





7. **Address what network features you found such as communities, giant clusters, homophily, robustness, spreading phenomena, etc.**
    a. **Explain any features you found and what relevance they had in your study**

    We found that free agency has most of the edges and the highest out-degree given that it's the largest pool you can select a player from in the league, so it's bound to be connected to most everything. The teams that had a large in-degree were not specific to any conference, division, or market size. All stars make up between 5-10% of trades and free agency signings on average. It is not unexpected given how many players are in the league. The communities found within these networks are close to evenly split between small and large market teams, meaning there is no specific type of team that trades with another. We cannot conclude homophily as we do not know any characteristics about the team or players such as positions the team wants and the positions the players play. Also we cannot know the salary caps of these teams to see who they would be willing to pay for. There were some giant clusters that were composed of 60-80% of nodes.

8. **Results and contemplation.**

a. **What concrete results did you achieve in doing your project? Address this in terms of your research questions.**

Through this analysis it was identified that there is no statistical advantage of being a large market team when it comes to winning the Finals. Through a t-test we came to the conclusion that the only advantage a large market team has over a small market team is that after they pick up or trade for an all-star their odds jump more than the small market, but in the end the odds are relatively the same.

Additionally, small market teams spend about the same on all stars as large market teams do. Bar charts show the average amount in salary and max cap contracts are within a few hundred thousand of each other in most cases. A t-test could not be performed as normality could not be established with this small of a sample size. One trend to notice is that the majority of teams that spend an increasing amount on their max contract over the four years typically will have a decreasing average salary. One team that defies this is the Pacers. They spent the same amount on their max contract both years, but in the second year their average salary skyrocketed, this could be caused by adding more all stars to the roster and spending closer to that cap on more people.

There were multiple instances of an increase in rank in the media market from between all seasons. However, there was only one change in rank that took a small market team to be large and vice versa. This was between the Timberwolves and the Pistons. The Timberwolves finished 15 while the Pistons finished 16. All other increases or decreases in ranking did not cause a change in the large or small market pool of teams.

Regarding wins and losses the Pacific Division seems to be the strongest and they are made up of all large market teams except one who finished below .500. Other than that there does not appear to be a trend regarding small or large market teams performing better throughout the years.

Power rankings do not appear to have a correlation to market size. There is an equal chance the large market or the small market will move up or down. However a large market team landed in the number one spot starting the season all four years.

b. **Address your contemplation about the project. Include, for example, other data that could have made your study more relevant, the difficulty in dividing the workload between team members, data preparation, what was unexpected, etc.**

In order to enhance the study, more data could be collected on player transactions so that more insightful statistical analysis could be performed with a larger sample. Additional data that could be collected was player injuries and how that impacts a team's performance.

The main obstacle that was faced was data cleaning and data preparation. After looking for sources that recorded player transactions, we struggled to scrape the data so we had to start entering the data manually. This would involve typing 1000s of rows of data manually per year that we analyzed. After some time, our team discovered a way that would allow us to scrape the player transactions using Python with the requests and json libraries. Additionally, several types of player transactions had to be filtered out of our output. For example, if a player signed a multi-year contract it would not count towards a trade. We were only interested in capturing player movements from one team to another, as well as players that came out of free agency onto a new team.

A secondary obstacle was finding a reliable dataset for media markets. There is not a lot of public data regarding media market rankings from year to year.

The third obstacle that was encountered was time constraints. There was not enough time to learn the proper time prediction models to forecast the team's result from the data that was collected for this study.

Dividing the workload was not difficult since each person had at least one year they were responsible for. However, as a result, it created a heavier workload to answer each report question in terms of the 4 different networks in our study.

References

2019-20 NBA Trade Tracker. (n.d.). NBA.Com. https://www.nba.com/2019-20-trade-tracker

AceOdds. (n.d.). *Odds Converter - Decimal, Fraction, American & Find Probability*. https://www.aceodds.com/bet-calculator/odds-converter.html

"Big Five" Pro Sports Teams Ranked by Media Market (2019–2020 rankings). (2019, November 27). Reddit. https://www.reddit.com/r/MLS/comments/e2imlo/big_five_pro_sports_teams_ranked_by_media_market/

NBA Media Market Rankings (2017). (2017, July 3). Reddit. https://www.reddit.com/r/nba/comments/6l03z5/nba_media_market_rankings_2017/

NBA Media Market Size Rankings 2018. (2018, March 9). Reddit. https://www.reddit.com/r/nba/comments/837czn/nba_media_market_size_rankings_2018/

Pincus, E. (2021, August 19). *Chicago Bulls Salary Archive – 2017/18*. Basketball Insiders | NBA Rumors And Basketball News. https://www.basketballinsiders.com/chicago-bulls-team-salary/chicago-bulls-salary-archive-2017-18/

SOHistory, LLC. (2016, November 24). *Archived NBA Futures Odds |SportsOddsHistory.com*. SportsOddsHistory.Com | Live Odds and Archived Futures Lines of the Super Bowl, World Series & More. https://www.sportsoddshistory.com/nba-odds/

Sports Media Watch. (2021, August 27). *Sports TV market size chart (NBA, NFL, MLB, NHL)*. https://www.sportsmediawatch.com/nba-market-size-nfl-mlb-nhl-nielsen-ratings/