EMORY
GOIZUETA BUSINESS SCHOOL

NBA Social Network Effects: The NBA Player and Team Performance Analysis

ISOM 673:
Dr. Demetrius Lewis

Written by:
Russel (Sicheng) Shen
Conner (Haotian) Jin
Daniel Byun
Feifan Gu

ISOM 673

Research Report

Network Analysis of The National Basketball Association: Player and Team Success Prediction

Student:

Conner (Haotian) Jin

Feifan Gu

Russel Shen

Xiao Chen

Under supervision of

Dr. Demetrius Lewis

Emory University

Goizueta Business School

Conner Jin, Feifan Gu, Russel Shen, Xiao Chen

Network Analysis of The National Basketball Association: Player and Team Success Prediction

---

CONTENTS

Chapter 1: Introduction to the National Basketball Association

NBA, the National Basketball Association, is a men's professional basket league in North America. The NBA is composed of 30 teams: 29 in the United States and 1 in Canada. The current league organization divides thirty teams into two conferences of three divisions with five teams each. To reflect the population distribution of the United States and Canada as a whole, most teams are in the Eastern half of the country; thirteen teams in the Eastern Time Zone, nine in the Central Time Zone, three in the Mountain Time Zone, and five in the Pacific Time Zone. In fact, the NBA is considered one of the four major professional sports leagues in the United States and Canada and is widely considered to the premier men's professional basketball league in the world. Furthermore, NBA players are the world's best-paid athletes by average annual salary per player.

The NBA's regular season runs from October to April, with each team playing 82 games (41 home and away) and the playoffs extend into June. A team faces opponents in its own division four times a year (16 games). Each team plays six of the team from the other two divisions in its conference four times (24 games), and the remaining four teams three times (12 games). Finally, each team plays all the teams in the other conference twice (30 games). Potentially, this asymmetrical structure could improve a team's win rate over another through the strength of schedule.

The NBA playoffs begin in April once the regular season is over and we are left with the top eight teams in each conference. In every round, the best of seven formats is instilled in each matchup. Similarly, the final playoff round is also the best of seven series with the final two winners, one from each conference. This is what we call the NBA Finals, and this is held annually in June.

This concludes the brief summary of what the NBA is and how these games are scheduled and organized. However, to transition to how the NBA relates to social networks, we would have to discuss the newly founded motivation for data analytic in the basketball association. To elaborate, the NBA is undergoing massive revolution and data analytic is a new way for NBA teams to develop strategies and to become championship-contending teams. To name one of the champions (no pun intended) for data and statistical analysis in the association, Daryl Morey, the General Manager of the Houston Rockets has a background in statistics. He has been integrating extensive data analytic into the game and the Rockets have been participating in every playoff the last six years. From observing the success of the Rockets, many teams started to imitate what the Rockets are doing, which could explain why the game has shifted to more of a three-point shooting game (with James Harden, Houston Rockets, an all-time leader in unassisted three-point makes).

To refer to the first sentence of this report, the goal of our research, we want to investigate the relationship between the mobility, or the centrality measure in players' co-working network, and the player's career path; as well as the relationship between the aggregated team-level player mobility and the teams' winning rate. To do this, we raise several hypotheses in the two sessions below, run regression to analyze the relationship and finally provide a suggestion of what might be the best strategy to secure a higher win rate.

EXPLORATORY DATA ANALYSIS

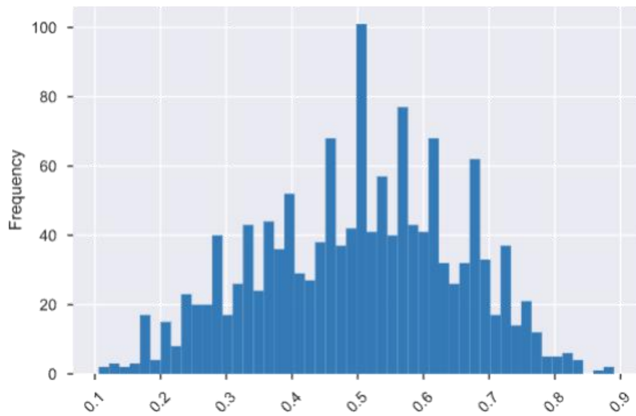
1. Original datasets

All the datasets used in this report is from Data.World[1]. The first dataset is "team_performance_perseason_to2017.csv". This data contains 1417 records of NBA teams' wins and losses and their wining rate from 1946 to 2017. The second data set, "player_1985to2018.csv" contains 14163 records of NBA players' unique id, salary and team of each season from 1985 to 2018. The third one is "player_perseason_1978to2016.csv". It contains 17729 records of NBA players' information of all kinds for every year including 'PER', which is a significant variable to measure a player's ability from 1978 to 2016. The fourth one is "playoff.csv". It contains 525 records of playoff results from 1985 to 2019.

Below we briefly explore several important outcome variables and predictor variables that we used and studied.

The winning percentage of regular seasons is one of the most natural proxies for a team's success and one of the most crucial outcome variables in the data set. The distribution of the winning percentages among teams over the years is approximately normal with more flatness. In addition, from the histogram below (FIG. 1), the distribution is nearly symmetrical at x=0.5. This is reasonable because the winning percentages for the teams are zero-sum. That is, for each game, the only results are a loss or a win. Therefore, it should be approximately normal against the symmetric line at 0.5.

[1] "DataWorld Home Page." *Data.world*, data.world/.



Quantile statistics

Minimum	0.106
5-th percentile	0.244
Q1	0.39
Median	0.512
Q3	0.61
95-th percentile	0.7322
Maximum	0.89
Range	0.784
Interquartile range	0.22

FIG. 1. Distribution and stats of regular-season winning percentage. On the left, the count of winning percentage shows a normal distribution look; On the right, the median is 0.512 and minimum and maximum are symmetrical about the median.

Data on won playoff rounds might also be informative because playoff games are elimination rounds. Hence, each advancing match up during the playoffs ranks higher in the hierarchy in games. Therefore, it is reasonable to replace these characters of rounds to factor numbers: 1, 2, 3 and 4.

Value	Count	Frequency (%)
Eastern Conf First Round	140	26.7%
Western Conf First Round	140	26.7%
Western Conf Semifinals	70	13.3%
Eastern Conf Semifinals	70	13.3%
Western Conf Finals	35	6.7%
Eastern Conf Finals	35	6.7%
Finals	35	6.7%

FIG. 2. Frequency of playoff winning rounds. Count of teams going to semifinals shrinks to half comparing to that of the first round and shrinks again when counting Conf Finals. Finals have the same count number as Conf Finals since Finals teams consist of 1 from Western Conf and 1 from Eastern Conf.

PER, which stands for Player Efficiency Rating, is an indicator of the ability of players. It is an official measure for player performance; such as, taking shot accuracy, blocks, steals, turnover rate, and minutes played, etc. into account. The PER measure is originally a transformation from the efficiency measure created by statistician Martin Manley[2]. The formula for this efficiency measure is listed below:

$$Efficiency = \frac{PTS+REB+AST+STL+BLK-MissedFG-MissedFT-TO}{GP}$$

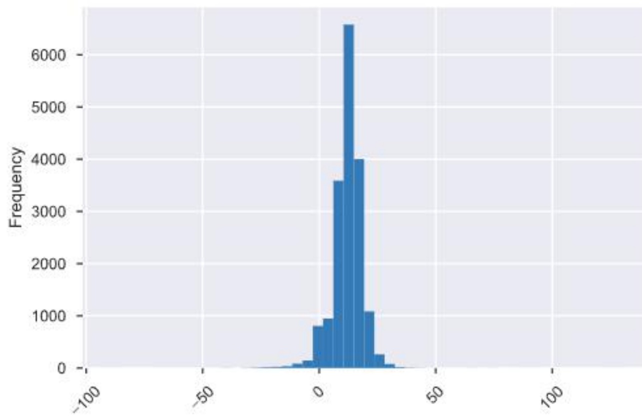
T

he PER measure slightly adjusts the efficiency measure by scaling into the same minutes played for all the players. From the variable description, a league-average PER is always 15.00, which permits comparison of player performance across seasons[3]. However, from our exploratory analysis of this variable, the median of the variable is only 12.6, with an approximately normal distribution. The explanation for this might be because there are huge outliers from both tails in the dataset. The largest right tail outlier is 129.1 and the smallest left tail outlier is -90.6. This is because PER is calculated with 'minutes played' in the denominator. Hence, large outliers for PER is a result of dividing a player's efficiency by a small value of the number of minutes he played. For example, Liggin's PER –129.1– is the highest single-season mark in NBA history, which is earned because he was only on-field for one minute[4].

[2] Manley, M. "Basketball heaven 1990 edition." *Bantan Doubleday Deli Publishing Group, New York* (1989).

[3] "2019-20 Hollinger NBA Player Statistics - All Players." *ESPN*, ESPN Internet Ventures, 2019, insider.espn.com/nba/hollinger/statistics.

[4] Fischer-Baum, Reuben. "The Greatest Season In NBA History*, Brought To You By DeAndre Liggins." *Deadspin*, Deadspin, 28 Sept. 2016, deadspin.com/the-greatest-season-in-nba-history-brought-to-you-by-1564448303.

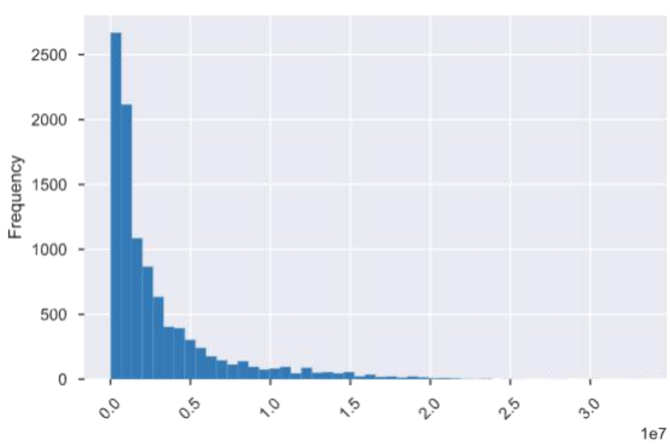


Quantile statistics

Minimum	-90.6
5-th percentile	0
Q1	9.4
Median	12.6
Q3	15.7
95-th percentile	20.9
Maximum	129.1
Range	219.7
Interquartile range	6.3

FIG. 3. Distribution and stats of PER. the number is around 12 and symmetrical to x=12, and there could be outliers as large as 129.1 or as small as -90.6. PS: official regulation of average PER of each season is 15, suggesting that our original dataset could be incomplete.

The last important indicator for an athlete success is his salary. The distribution of the players' salary is approximately an exponential distribution. This shows that most of the NBA athletes got basic salary. Only few of them, who are probably known as NBA stars enjoyed a high salary. This distribution is reminiscent of the distribution of the U.S. household income.



Quantile statistics

Minimum	5000
5-th percentile	150000
Q1	638679
Median	1480700
Q3	3672646
95-th percentile	11556316.25
Maximum	33140000
Range	33135000
Interquartile range	3033967

FIG. 4. Distribution and stats of NBA player's salary. On the left, the histogram shows an exponential distribution look. On the right, the specific stats show that the salary ranges from 5,000 to 33,140,000 within our original time range.

2. Cleaned and transformed datasets

To efficiently use all the datasets, we considered the overlap of each dataset's year range: 1985 to 2018, with necessary supplementary data. After data cleaning and variable transformation, we combined the four datasets listed above to several large datasets for network analysis. The network constructed by players is formed under the following rationales:

- Players build a collaborative relationship through playing with each other in a team.
- The relationship is not likely to decay after one player in a pair transferred to another team.

An edge list for the player network is created, which contains 104716 records. This edge list is the base, in which we construct different kinds of social networks between players or teams later. Below shows the first few lines of the edge list:

player_id.x	season_end	team	player_id.y
abdulka01	1985	LAL	coopemi01
abdulka01	1985	LAL	johnsma02
abdulka01	1985	LAL	kupchmi01
abdulka01	1985	LAL	lestero01
abdulka01	1985	LAL	mcadobo01
abdulka01	1985	LAL	mcgeemi01
abdulka01	1985	LAL	rambiku01
abdulka01	1985	LAL	scottby01
abdulka01	1985	LAL	sprigla01
abdulka01	1985	LAL	wilkeja01

FIG. 5. Head of ‘edge_list’. Each row records a player in a team in a year and his teammate.

Chapter 2 Player Network and Mobility

ABSTRACT

NBA players choose to play in a team that they desire to join. In this report, NBA players and their mobility are investigated through network analysis. We found that for high-level players, better-performed players are more likely to have higher mobility, although the transferring practice does not necessarily improve the players' performance. In contrast, for low-level players, poorly performing players are more likely to be transferred, but they are more likely to improve after the transfer.

INTRODUCTION

NBA basketball professionals achieve personal success through refining personal skills and playing in a team that ranked higher in the association. NBA player transfer market, as well as free agents provide the opportunity for an NBA athlete who performs well but does not plays in a successful team originally to move towards success in their career. After joining another team, a player might be able to collaborate with higher-quality teammates and thus further refining his personal skills[5]. In this way, a player could form a virtuous cycle that leads to self-achievement.

On the other hand, a player could be transferred passively. This could happen when a player's contract with a team ended but the player did not perform well in the previous season. Therefore, the player would have to leave and transfer to another team to continue his career. In this sense, the player might still have the chance to improve his skills onward. However, there will be no immediate synergy among the traded player and his new high-ranking teammates. This is

[5] Siddiqi, DJ. "Lakers Coach Discusses Brandon Ingram's Growth." *Los Angeles Lakers*, 247sports, 7 Aug. 2018, 247sports.com/nba/los-angeles-lakers/Article/Los-Angeles-Lakers-coach-discusses-Brandon-Ingrams-growth-120470628/.

because new players need some time to get assimilated with their new teams and play style. In contrast, given enough time, a player can easily improve over time with his new team.

The third type of transfer happens when a player performed excellently in the previous season. Managers or coaches from another team might consider paying a premium to have the highly skilled player on their team, in hopes of improving the team's winning rate for the next season. However, since a player needs time to familiar with different team's competing style, transferring highly skilled player to another team might not necessarily improve the player's performance. Additionally, it is also possible that a player who got a lucky break had an excellent season that was attributable to outside forces and not his own. Thus, after transferring teams, the player might return to his lower normal performance.

In this paper, we investigated which factor contributes to mobility, as well as how mobility influences a player's career. We raised the following hypothesis to test:

- **Low-level players who have worse performance are more likely to be transferred.**
- **High-level players who have better performance are more likely to be transferred.**
- **Low-level players transfer teams to improve skills.**
- **High-level players who transferred would have worse performance.**
- **Player achieve better salary through transfers.**

RESEARCH DESIGN AND MEASUREMENT

We use NBA players’ co-working experience network as strategic research setting to test our theoretical claims. Regression is the primary tool in this study. Regression methods provides robust and statistical relationship indication between variables. We could analyze how the dependent variable influences the outcome at different levels through deploying regression methods. Panel linear model and panel autoregression model are also considered in our study since panel models provide reasonable ways to eliminate the influence on players from the team they belong to. Autoregression model could also implement reasonable control for time-related trends.

1. Dependent Variables I: Centrality Measures in Player Network

The first dependent variables we chose are centrality measure in the player networks. Since the player network is derived through constructed connections when players collaborate in a team, the centrality of a player measure how many players the focal player has collaborated with. A higher centrality score means that the player has been in the same team with more people. Therefore, since if a player has transferred into other teams, he will have the opportunity to collaborate with more other players. Thus, the centrality measure is a good proxy for the mobility of a player. Through using centrality measures as the dependent variable, we will be able to analyze what contribute as a factor that influences a player’s mobility. In this study, degree centrality is a primary choice of network centrality measure because count of relationships made by forming teammates is a natural indicator of player mobility.

2. Independent Variables I: Player Performance Statistics and Player Salary

Player performance statistics provides information about how well a player compete in NBA matches. The performance statistics we picked include PER, TS, ORB, DRB, AST, STL, BLK, TOV, and USG. These metrics depicts a player's overall skill level in each season. According to our thesis, we would expect high-level players (definition in the next session below) with higher performance statistics and low-level players with lower performance statistics would more likely to be transferred. A player's salary could also play an effect in the mobility of an athlete. Since transferring usually end up with a higher salary for the player, a team would also consider the player's original salary before making transferring decisions.

3. Player Level Splitting Criterion: High-level and Low-level

Since PER measure should have a league-level average of 15.00, we used PER of 15 as a splitting criterion for distinguish high-level and low-level players. We would expect players with a PER over 15 to have better skills or better collaboration with teammates than players with a PER under 15.

4. Dependent Variables II: Player Ability Improvement and Player Salary

The second set of dependent variables we chose are player ability (increase in PER measure) and the salary that a player received. These two measures label the success of an individual player. Implement regression on those two measures could help derive insights about what metrics could help improve an NBA player's personal skills and salary. This could be very interesting and insightful in terms of helping an athlete develop personal strategy for his career.

5. Independent Variable II: Player Mobility

According to our theory, we would expect a player that wish to improve and realize self-achievement to have a higher tendency to move from team to team. In addition, player transferred passively due to underperformance would also have the tendency to pay more effort and improve performance in the next season. Therefore, using player mobility, which is measured by centrality scores in the player network, as an independent variable in the regression could be helpful in deriving insights on whether or not mobility help players improves.

6. Control Variable: Minutes Played, Age, Team and Year

As stated in the Exploratory Data Analysis session, the PER measure and other performance measure could be influenced by how much time a player was on-field during the season. Hence, the variable minutes played are include as a controlling variable. The age of a player is a proxy of a player’s physical condition. If a player is relatively older than other players in the NBA league, he might not be able to improve skill purely due to human physical constraint. Therefore, age should also be taken into account. We also included team and year as a controlling factor in the regression. The choice is made based on the rationale that a specific team or a specific year might show as a special year for NBA such that overall league has an increase in mobility.

RESULTS AND DISCUSSION

1. Low-level players who have worse performance are more likely to be transferred.

To explore the first hypothesis, the first analysis investigates that for players with a PER measure lower than 15, whether there is a negative relationship between the players performance measure and centrality measures of a player. First, we only included player with PER lower than 15 because we only want to analyze the effect of performance on mobility for lower level players. We regressed the degree centrality of players on the player’s normalized salary and different performance measure, with controlling variables of team and year.

```
Call:
lm(formula = degree ~ normalize + PER.x + `TS%` + `ORB%` + `DRB%` +
    `AST%` + `STL%` + `BLK%` + `TOV%` + `USG%` + factor(Tm) +
    factor(year), data = mobility_stats)

Residuals:
      Min        1Q    Median        3Q       Max
-0.041444 -0.011078 -0.002025  0.009002  0.086662

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      6.107e-02  2.109e-03  28.955  < 2e-16 ***
normalize       -1.989e-02  1.367e-03 -14.549  < 2e-16 ***
PER.x           -2.479e-04  4.070e-05  -6.091 1.16e-09 ***
`TS%`           -2.981e-03  1.818e-03  -1.639 0.101155
`ORB%`          -9.932e-05  4.400e-05  -2.258 0.023997 *
`DRB%`          -3.391e-06  3.191e-05  -0.106 0.915375
`AST%`           2.242e-05  2.396e-05   0.935 0.349566
`STL%`           3.440e-04  1.577e-04   2.181 0.029191 *
`BLK%`          -1.907e-04  9.075e-05  -2.101 0.035649 *
`TOV%`           1.357e-05  2.366e-05   0.574 0.566261
`USG%`          -5.706e-05  3.327e-05  -1.715 0.086389 .
factor(Tm)BOS   -4.136e-03  1.123e-03  -3.684 0.000231 ***
factor(Tm)BRK   -1.508e-02  3.579e-03  -4.215 2.52e-05 ***
factor(Tm)CHA    4.319e-04  1.431e-03   0.302 0.762887
factor(Tm)CHH   -1.233e-02  2.088e-03  -5.905 3.63e-09 ***

  factor(year)2010 -9.324e-03  1.700e-03  -5.483 4.27e-08 ***
  factor(year)2011 -4.505e-03  1.672e-03  -2.694 0.007075 **
  factor(year)2012 -1.336e-02  1.732e-03  -7.715 1.32e-14 ***
  factor(year)2013 -3.544e-03  1.693e-03  -2.094 0.036304 *
  factor(year)2014 -1.361e-02  1.746e-03  -7.799 6.86e-15 ***
  factor(year)2015 -9.923e-03  1.689e-03  -5.876 4.32e-09 ***
  ---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

  Residual standard error: 0.01583 on 10379 degrees of freedom
  Multiple R-squared:   0.15,     Adjusted R-squared:  0.1436
  F-statistic: 23.48 on 78 and 10379 DF,  p-value: < 2.2e-16
```

TABLE 2.1 Linear Regression. The outcome variable is the mobility of low-level players and the predictor variable are normalized salary, PER, TS%, ORB%, DRB%, AST%, STL%, BLK%, TOV% and USG%, including control for time fixed effects and team fixed effects.

As we can see from the regression result, normalized salary, PER, ORB% and BLK% have a negative and significant coefficient; whereas STL% has a positive and significant coefficient. The remaining performance measure does not show clear significance. This shows that **for player with lower-level skills, those players with less salary and worse performance are more likely to be transferred.** This could due to the reason that as a team, coaches and managers want to only keep the players that help the team win as an entity. Therefore, they have the tendency to **transfer underperformed lower level players to other team as a strategy of**

increasing overall team strength. The fact that more detailed performance measures are less significant might be because teams does not care much about lower-level player's detailed statistics itself since they only want to infer the potential for that player to become better. In this way, a team could get rid of lower-level players that might not have tendency to improve when the players stay in the same team.

2. **High-level players who have better performance are more likely to be transferred.**

Compare to lower-level players, the transferring market for high-level players might be a different situation. Our theoretical hypothesis yields that for high-level players, the better performance, the more mobility the player would have. Therefore, for this time, we filtered the dataset by PER higher than 15 to only consider the effect of performance for high-level players. Similarly, we regressed the degree centrality of players on the player's normalized salary and different performance measure, with controlling variables of team and year.

```
Call:
lm(formula = degree ~ normalize + PER.x + `TS%` + `ORB%` + `DRB%` +
    `AST%` + `STL%` + `BLK%` + `TOV%` + `USG%` + factor(Tm) +
    factor(year), data = mobility_stats)

Residuals:
      Min        1Q    Median        3Q       Max
-0.042009 -0.009087 -0.001514  0.007272  0.080298

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.892e-02  2.953e-03  19.950  < 2e-16 ***
normalize       -1.460e-02  1.086e-03 -13.439  < 2e-16 ***
PER.x            3.950e-04  7.128e-05   5.542 3.16e-08 ***
`TS%`           -1.684e-02  3.309e-03  -5.090 3.74e-07 ***
`ORB%`          -1.853e-04  5.634e-05  -3.289 0.001013 **
`DRB%`           3.839e-05  4.089e-05   0.939 0.347916
`AST%`          -1.572e-04  2.631e-05  -5.976 2.47e-09 ***
`STL%`          -1.755e-04  2.131e-04  -0.824 0.410146
`BLK%`          -9.513e-04  1.339e-04  -7.105 1.40e-12 ***
`TOV%`           1.410e-04  4.955e-05   2.846 0.004446 **
`USG%`          -1.876e-04  4.878e-05  -3.846 0.000122 ***
factor(Tm)BOS    2.353e-03  1.414e-03   1.664 0.096111 .
factor(Tm)BRK   -1.795e-02  5.107e-03  -3.514 0.000446 ***
factor(Tm)CHA    7.381e-03  2.102e-03   3.512 0.000449 ***
```

```
factor(year)2012 -9.104e-03  1.824e-03  -4.991 6.25e-07 ***
factor(year)2013 -1.229e-02  1.820e-03  -6.750 1.67e-11 ***
factor(year)2014 -1.404e-02  1.851e-03  -7.584 4.06e-14 ***
factor(year)2015 -1.177e-02  1.765e-03  -6.665 2.98e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01314 on 4322 degrees of freedom
Multiple R-squared:  0.2092,    Adjusted R-squared:  0.1949
F-statistic: 14.66 on 78 and 4322 DF,  p-value: < 2.2e-16
```

TABLE 2.2 Linear Regression. The outcome variable is the mobility of high-level players and the predictor variable are normalized salary, PER, TS%, ORB%, DRB%, AST%, STL%, BLK%, TOV% and USG%, including control for time fixed effects and team fixed effects.

The regression results show that normalized income, PER, TS%, ORB%, AST%, BLK%, TOV% and USG% are also statistically significant in predicting mobility of high-level players. Among those predictor variables, normalized income, TS%, ORG%, AST%, BLK%, and USG% have negative and significant coefficient; whereas the remaining variables have positive and significant coefficients. This indicates that **for high-level players, those who have a better overall performance and lower income are more likely to be traded.** This is also reasonable for sports teams. While considering hiring a talent from another team, coaches and managers would want to select the best high-level players with the lowest cost. Since normally transferring for a high-level player requires the accepting team to pay a premium for the player's skills, team always want to **select players that perform the best with the lowest previous salary. In this way, they could achieve increase in team strength by introducing better players with minimal cost.** The coefficient of TOV% is much more plausible as teams may trade a player with too many turnovers even though his PER is high. This could be because once the high-level player have astonishing performance, his mistakes might be overlooked or compensated by his contribution towards winning.

3. Low-level players transfer teams to improve skills.

In addition, we also want to analyze that how mobility of players could help a player improve throughout his career. First, similarly, we want to analyze how mobility could help a low-level player to improve their skills. We only included player with PER lower than 15. Then we regressed the increase in PER measure compared to the previous year on the degree centrality, minutes played during the previous season, the age of the player, with a team fixed controlling variable.

```
lm(formula = per_increase ~ deg + mp_lag1 + Age + factor(Tm),
    data = p1)

Residuals:
    Min      1Q  Median      3Q     Max
-74.735  -1.749   0.435   2.294  59.925

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.268e+00  5.345e-01   2.372 0.017699 *
deg            2.160e-03  6.091e-04   3.546 0.000393 ***
mp_lag1       -5.030e-04  7.747e-05  -6.493 8.98e-11 ***
Age           -6.807e-02  1.608e-02  -4.235 2.32e-05 ***
factor(Tm)BOS -6.404e-01  4.620e-01  -1.386 0.165737
factor(Tm)BRK -9.149e-02  1.094e+00  -0.084 0.933343
factor(Tm)CHA  1.312e-01  5.746e-01   0.228 0.819390
factor(Tm)CHH -6.507e-01  6.216e-01  -1.047 0.295232

factor(Tm)SEA -7.156e-01  5.295e-01  -1.351 0.176617
factor(Tm)TOR -3.647e-01  4.847e-01  -0.752 0.451847
factor(Tm)UTA -3.830e-01  4.755e-01  -0.805 0.420585
factor(Tm)VAN -6.105e-01  7.391e-01  -0.826 0.408824
factor(Tm)WAS -2.197e-01  4.961e-01  -0.443 0.657944
factor(Tm)WSB -7.799e-01  7.114e-01  -1.096 0.273012
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.215 on 7396 degrees of freedom
Multiple R-squared:  0.01398,   Adjusted R-squared:  0.008645
F-statistic: 2.621 on 40 and 7396 DF,  p-value: 1.183e-07
```

TABLE 2.3 Linear Regression. The outcome variable is the increase in performance measure of low-level players compared to a previous year. The predictor variables are the mobility (degree centrality) of the player, the minutes played during the previous season, and the age of a player including control for team fixed effects.

The regression results show that all the predictor variables have statistical significance. In particular, the mobility, which is the degree centrality in the player network have positive and significant coefficients, whereas minutes played in the previous season and age have negative and significant coefficients. **This shows us the fact that the more transfer that a low-level player involved in, the better chance that player would have to improve his personal skill and performance. In addition, human health condition does contribute as a factor of a player's ability to improve.** The result confirms our thesis and provide insightful guideline for lower-level players to improve themselves. **Lower-level player would be able to gain more experience at another team, and thus transferring to other teams could help a player who does not have great skills to have better performance. In addition, since basketball, like all sports, younger player would have a better chance to improve his skill levels.** At a team level, professional basketball teams could introduce other team's low-level players if they believe that player could have a potential and synergy that fits the teams competing strategy.

4. **High-level players who transferred would have worse performance.**

On the other side, high-level players, who already become the best talents in basketball, might not experience the same situation since they do not have much to improve in terms of basketball skills. Their performance relies more on collaboration with teammates that have similar playing styles and synergies as a team when combing strengths. Therefore, this time we filter all the players by the criterion of PER bigger than 15. Then we regressed the increase in PER measure compared to the previous year on the degree centrality, minutes played during the previous season, the age of the player, with a team fixed controlling variable.

```
Call:
lm(formula = per_increase ~ deg + mp_lag1 + Age + factor(Tm),
    data = p1)

Residuals:
    Min      1Q  Median      3Q     Max
-31.301  -1.931  -0.115   1.560  67.227

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    9.647e+00  6.142e-01  15.707  < 2e-16 ***
deg           -3.244e-03  6.475e-04  -5.010  5.7e-07 ***
mp_lag1       -1.694e-03  7.502e-05 -22.583  < 2e-16 ***
Age           -1.357e-01  1.859e-02  -7.301  3.5e-13 ***
factor(Tm)BOS -6.993e-01  4.907e-01  -1.425   0.1542
factor(Tm)BRK -1.431e+00  1.276e+00  -1.122   0.2620
factor(Tm)CHA -8.786e-01  7.808e-01  -1.125   0.2606

 ---
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Residual standard error: 4.065 on 3626 degrees of freedom
 Multiple R-squared:  0.1803,    Adjusted R-squared:  0.1713
 F-statistic: 19.95 on 40 and 3626 DF,  p-value: < 2.2e-16
```

TABLE 2.4 Linear Regression. The outcome variable is the increase in performance measure of high-level players compared to a previous year. The predictor variables are the mobility (degree centrality) of the player, the minutes played during the previous season, and the age of a player including control for team fixed effects.

The regression results show that all the predictor variables have statistical significance. In particular, the mobility, which is the degree centrality in the player network have minutes played in the previous season and age have negative and significant coefficients. This indicates that **for high-level players, transferring to another team might not necessary leads to better performance, given that minutes played during the last season and age of the players have already been considered.** This finding is actually counter-intuitive and insightful. Teams that hire players with high skills in the hope of improving overall team performance. **However, due to the fact that when a player transfer into another time, he needs time to get familiar with his teammates and the new competing strategy of the new team. High-level player**

transferred to another team actually perform worse than those stayed in the same team. Therefore, beyond the performance measure, a team should also consider how much a player fit the team's collaboration style before they consider a transfer. If they only focus on the performance measure, the player they buy might actually perform worse after the transfer.

5. **Player achieve better salary through transfers.**

After we have examined how mobility influences a player's performance from the perspective of a team manager, we want to further analyze what kind of factors would affect a player's salary from the perspective of a player. This time, we regressed the salary on the PER measure during the previous season, minutes played during the previous season, the age of the player and the degree centrality, with a team fixed controlling variable.

```
Call:
lm(formula = salary ~ deg + per_lag1 + mp_lag1 + Age + factor(Tm),
    data = p1)

Residuals:
      Min        1Q    Median        3Q       Max
-12995040  -2031564   -579834   1151755  24818047

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   -5.371e+06  2.842e+05 -18.899  < 2e-16 ***
deg            9.964e+03  3.056e+02  32.602  < 2e-16 ***
per_lag1       1.657e+05  6.879e+03  24.082  < 2e-16 ***
mp_lag1        7.217e+02  4.056e+01  17.792  < 2e-16 ***
Age            1.030e+05  8.230e+03  12.510  < 2e-16 ***
factor(Tm)BOS  5.787e+05  2.331e+05   2.483 0.013052 *
factor(Tm)BRK  4.086e+06  5.721e+05   7.143 9.67e-13 ***
factor(Tm)CHA  1.277e+06  3.102e+05   4.116 3.88e-05 ***
```

TABLE 2.5 Linear Regression. The outcome variable is the salary. The predictor variables are the mobility (degree centrality) of the player, the PER measure during the previous season, the minutes played during the previous season, and the age of a player including control for team fixed effects.

The results show that all the predictor variables have statistical significance and all the predictors have positive coefficients. **This firstly demonstrates that the higher the mobility, which is the**

degree centrality, the more salary the player will receive. Secondly, the PER and minutes played of the player during the previous season also positively influence the salary. Finally, the age of the player could also influence the salary. Players who are older generally have higher salaries. Part of the finding is kind of surprising to us. We expected players stay in the same team for a longer time will have higher salaries as NBA stipulates the maximum salary a player can receive is positively related to the number of years he stays in one team. **However, the other explanation also works here. Many players change their teams as a result of higher salaries offered by other teams, especially for players who are at a high level. Beyond this counter-intuitive result, PER and minutes played during the previous season have a positive influence on the salary as teams prefer to offer higher salary to high-level players and high-level players always play longer time. Besides, it's plausible that older players will have higher salaries as the contract size for rookies is limited.** From the perspective of a player, to get a higher salary, he should try to get higher PER and play more time each game. When other teams offer a contract, it's a good choice to take it as the salary will often be higher than that offered by the current team.

Chapter 3 Player Network and Team Success

ABSTRACT

The NBA is one of the most followed sports championships. In this report, NBA athletes and teams are analyzed through investigating networks between athletes. There are two types of player transfers: one type that transfers because the player is high quality and move to join a better team, the other type that transfers because the player is low quality and disruptive. In this analysis, we found that in general, team with less player transfers is more likely to have a higher winning rate, although minimal, targeted transfers help a team to win more.

INTRODUCTION

NBA basketball players transfer market is huge. In 2018-19 season, there's more than 400 player transactions in the association[6]. Since the minimum salary for NBA athletes is currently $582,180, each one of the transfer decisions that a team makes is a relatively influential decision in financial terms[7]. Therefore, it is important for a team to transfer players wisely that lead to future success.

NBA athletes transaction can be briefly classifies into two distinct categories: one, which we called "proactive transfer", is defined as a team make a decision to introduce a highly-skilled player, in the hope of improving team performance; the other type, which we called "reactive transfer", is defined as when a team sells players that does not perform well in the previous season.

---

[6] "List of 2018–19 NBA Season Transactions." *Wikipedia*, Wikimedia Foundation, 27 Nov. 2019, en.wikipedia.org/wiki/List_of_2018%E2%80%9319_NBA_season_transactions.

[7] "NBA Minimum Salary." *Basketball Insiders*, www.basketballinsiders.com/nba-salaries/nba-minimum-salary/.

Both of those transferring practice are team-based decision with the goal of achieving higher winning rate in the next season. One interesting question raised from this practice is that whether hiring a player from another team would help a team win more or not. If yes, then what type of players that is transferred could be a good choice for a team.

To investigate this topic, we raised the following hypothesis:

- **Teams that have more members transferred from other teams tend to have lower winning rate.**
- **Few, targeted transferring practice helps increase winning rate; while massive transferring decreases winning rate.**

RESEARCH DESIGN AND MEASUREMENT

We use NBA team network as a strategic research setting to test our theoretical claims. We construct this team network by creating relationship between teams with player transfer history. More specifically, a team A would have a relationship towards team B if team A has a player that was part of team B previously. By implementing this network configuration, we want to capture the relationship of "knowledge". That is, if team A has a player that was in team B before, team A would have a source of knowledge of understanding the competing strategy of team B. This knowledge may or may not leads to the actual win for team A in the match between A and B, but one would expect this type of knowledge could play an effect. In addition, the centrality measure of the team network also infers the amount of transfer practice that the team has involved. In this sense, we could also analyze whether transferring practice would help a team increase its overall winning rate or not.

Similarly, Regression is the primary tool in this study since it provides robust and statistical relationship indication between variables. Panel autoregression models are also used to eliminate the team-based effect over the time.

1. Dependent Variable: Winning Rate of a Team

The most important dependent variable we choose for the team level network analysis is the winning rate of a team. Achieve higher winning rate is the primary goal for an NBA team. In this analysis, we use winning rate of a team as an outcome variable to derive insights about what factors might be crucial for a team to rank higher and win more in the NBA matches. More specifically, following our assumption, whether involving more in player transfer could help a tea, build better collaboration.

2. Independent Variable: Out Degree of the Team Network, the Average, Minimum, Maximum Player Strength on a Team, and the Total Salary of the Team

The first independent variable we choose for the regression analysis, as stated in the hypothesis, is the out degree in the team network. We picked out degree from all the centrality measures because out degree measures how many players that a team have is from other teams. This measure would be able to capture the level of transferring practices that a team participated in. In addition, the level of skills that players in a team have. This is reasonable since the stronger players on a team are, the more chance that a team would have to win. In this study, we used PER to measure the strength, or how skilled is each player. And we used the average, minimum, and maximum player strength on a team to aggregate the team level powerfulness measure. Another factor that should be considered is that how much money a team have. A team with more money would be able to have better equipment and coaches to support their players. Hence, the total amount of salary that a team gives its players is a good proxy for how much money a team was able to raise.

3. Control Variable: Team and Year with Panel Linear Model.

Since a specific team might have higher winning rate in a specific year purely because their strategy happens to be the best for that year. We used the panel linear model to control for the year factor and team factor.

RESULTS AND DISCUSSION

1. **Teams that have more members transferred from other teams tend to have lower winning rate.**

To explore the first hypothesis, the first analysis investigates whether a team could achieve higher winning rate through transferring practices. We regressed the winning rate of a team on the outdegree centrality, the average, minimum, and maximum strength of players on that team, and the total salary that the team gives its members as a proxy of how much money the team We also used a panel linear model to have a team fixed and a year fixed controlling variable.

```
> summary(reg1)
Twoways effects Within Model

Call:
plm(formula = wr ~ outdegree + avg_strength + min_strength +
    max_strength + total_salary, data = team, effect = "twoways",
    model = "within", index = "Team")

Unbalanced Panel: n = 26, T = 3-26, N = 574

Residuals:
      Min.    1st Qu.     Median    3rd Qu.       Max.
-0.2469162 -0.0576876  0.0025014  0.0545589  0.2488889

Coefficients:
                Estimate  Std. Error t-value  Pr(>|t|)
outdegree    -1.5414e-04  6.3670e-05 -2.4209  0.015827 *
avg_strength  9.7233e-02  6.2202e-03 15.6319 < 2.2e-16 ***
min_strength -8.1958e-03  2.8965e-03 -2.8296  0.004842 **
max_strength  1.4652e-02  1.5422e-03  9.5012 < 2.2e-16 ***
total_salary  1.4538e-09  4.5403e-10  3.2021  0.001448 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    11.333
Residual Sum of Squares: 4.2005
R-Squared:      0.62936
Adj. R-Squared: 0.59
F-statistic: 175.915 on 5 and 518 DF, p-value: < 2.22e-16
```

TABLE 3.1 Linear Regression. The outcome variable is the winning rate of a team during each season. The predictor variables are the amount of transfer (out degree centrality) that a team makes, the average, minimum and max strength of players on the team, total salary a team offers its players, including control for team fixed effects and year fixed effects through panel regression.

From the regression result, we could see that the out degree and the minimum strength of players on the team have negative and significant coefficients; whereas average player strength, maximum player strength, and the total salary offered by the team have positive and significant coefficients. This basically illustrates that **teams that introduce player from other teams more tend to have lower winning rate overall. In addition, team with better players on average or have a better most skilled player on the team can lead to a higher winning rate. In contrast, a worse lowest-level player might help the team improve winning rate. Also, a team that**

have more money tend to wine more. This could provide a lot of insights towards building a winning team in the future. First of all, **transferring player from other team was not an optimal strategy for building a winning team.** Players from other teams have to take effort towards many trivial works other than training to transfer from a team to the other. For example, a player should finish paper works for housing, food, and contracts, etc. In addition, they also need to take time to get familiar with new teammates' playing styles. Therefore, having transferred player not necessary help the team improve in the team performance. Secondly, **a high-level player, or the best player on the team could have leading effect and help improve the overall winning rate**. Having a highly skilled best player on the team could build up a role model for other players to learn from. This could effectively improve the team's overall collaboration. Thirdly, **having a worse weakest player on the team could also help improve overall winning rate.** The lowest-skilled players on the teams normally do not have the chance to participate in matches frequently. Instead, the skill level of the lowest-skilled player could be an indicator of how much effort the team is paying to training their own players instead of buying players from another team. Last but not the least, **raising more fund help improve the team's overall winning rate.** This could be explained by the fact that with more money, the team could hire more coaches or design personal training plans for their players and the whole team. Therefore, the team will be stronger if they could raise more fund for the team.

1. **Few, targeted transferring practice helps increase winning rate; while massive transferring decreases winning rate.**

To further test the relationship between the team-level mobility and winning percentage, we used a 'loess' smother to plot and investigate the non-linear relationship between team-level mobility of high-performance players and the winning percentage. Loess stands for "locally estimated scatterplot smoothing" – it fits a locally weighted regression line over the underlying scatterplot, so it provides a tool to observe nonlinear relationship.

The results are shown below:

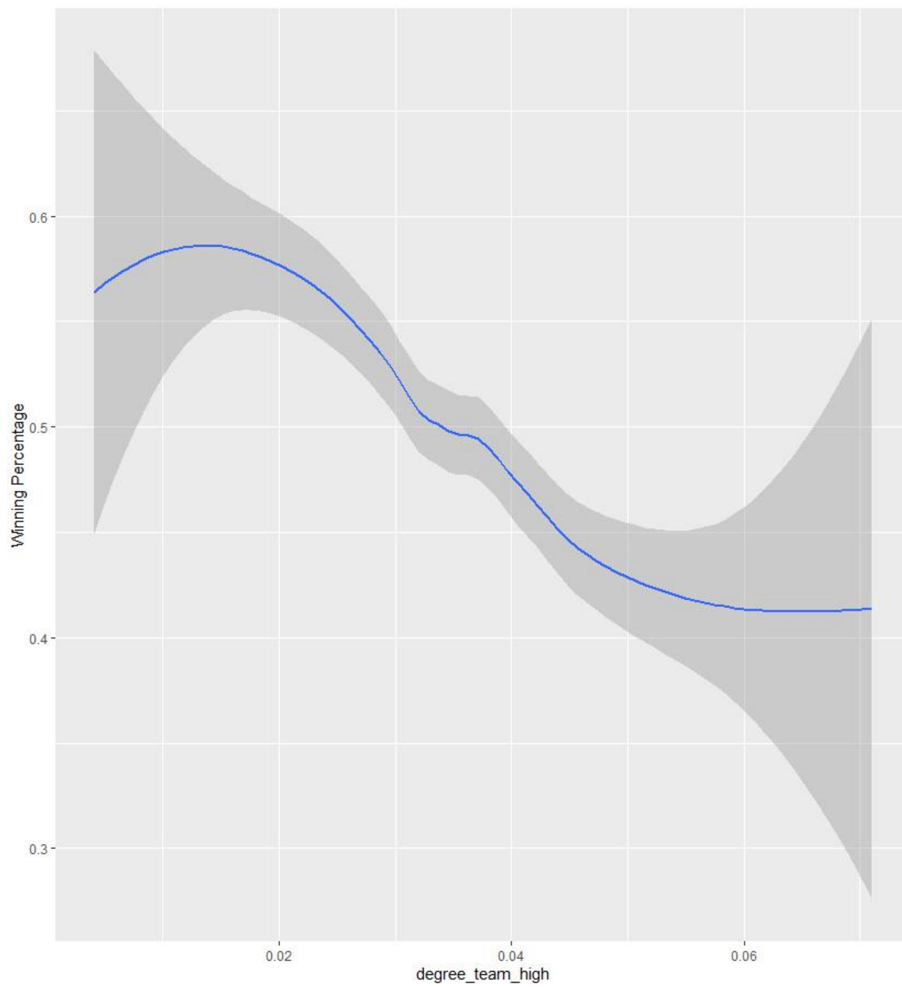


Figure 3.1 Loess Smoother. On the x-axis is the team-level mobility for players and on the y-axis is the winning percentage.

From the plotted graph, we can see the similar trend demonstrated by the previous regression result that the team-level mobility of high-performance player is negatively related to the winning percentage. **However, it the level of degree is very low, which means that the team only involves in minimal amount of player transfer, the winning rate increases as**

transferring transaction count increases. This reflects the finding we have in the previous section. Team which generally **train young, less skilled player by themselves to have better player in the future tend to win more. However, if the team need certain type of talent urgently, they would find the talent through the player market. This type of targeted, detailly studied transfer shows a team's ability to analyze their strength and weakness, which in turn leads to a higher winning rate.**

Conclusions

In this study, we analyzed NBA player network at both player level and team level.

During the player level network analysis, relationships between players are defined as co-working experiences in the same team. Regression on the mobility, which is measured by centrality of players in the network, as well as the increase in each individual player's skill level are performed. We found that low-level players who have worse performance are more likely to be transferred, while High-level players who have better performance are more likely to be transferred. This is because NBA teams introduce better high-level players, as well as transfer out worse low-level players, in order to achieve higher winning rate in the next season. We also found that low-level players improve skills through transferring to different team and learn from a broader range of players; while high-level players who transferred tend to have worse performance. In addition, player could also achieve better salary through transfers. This could be explained by the fact that low-level players have more things to learn from other players since they are less experienced; while high-level players need time to fit in a new team, and thus transferring to a new team compromises performance.

During the team level network analysis, relationships between teams are defined as the situation when a team have a player that was part of another team previously. We found that teams that have more members transferred from other teams tend to have lower winning rate; while few, targeted transferring practice helps increase winning rate; while massive transferring decreases winning rate. This could be a guideline when a team makes transferring decisions. For an NBA team, train players from beginning rather than buying players from other teams would be much more beneficial in term of improving winning rate. However, sometimes targeted transfer which address specific weakness of the team could also improve the team performance as a whole.

Further work could be done include more detail analysis of how player network contributes as a factor throughout a player's career. More specifically, transferring into different team could be based on different reason for each individual player and have different players. We could split transfer activity into different categories to perform deeper analysis.

Exhibits

1. Code for Data Cleaning

```r
library(data.table)
library(dplyr)
library(igraph)
player <-fread('player1985to2018.csv')
team <- fread('teamperformanceperseason -2017.csv')
player_season <- fread('playerperseason1978-2016.csv')
abbrev <- fread('team_abbrev.csv',header = FALSE)

team <- team[team$Team != '',]
team$Team[team$Team == 'Blackhawks'] <- 'Hawks'
team$Team[team$Team == 'Supersonics'] <- 'SuperSonics'
team$Team[team$Team == 'Trail Blazers'] <- 'Blazers'
team$Team[team$Team == 'Bobcats'] <-  'Hornets'
team$Team[team$Team == 'Braves'] <-  'Clippers'
team$Team[team$Team == 'Bullets'] <-  'Wizards'
team$Team[team$Team == 'Zephyrs'] <-  'Wizards'
team$Team[team$Team == 'Bobcats'] <-  'Hornets'
team$Team[team$Team == 'Royals'] <-  'Kings'
player$team[player$team == 'New Jersey Nets'] <- 'Brooklyn Nets'
player$team[player$team == 'Washington Bullets'] <- 'Washington Wizards'
player$team[player$team == 'Charlotte Bobcats'] <- 'Charlotte Hornets'
player$team[player$team == 'New Orleans Hornets'] <- 'New Orleans Pelicans'
player$team[player$team == 'New Orleans/Oklahoma City Hornets'] <- 'New Orleans Pelicans'
player$team[player$team == 'Kansas City Kings'] <- 'Sacramento Kings'

player <- inner_join(player,abbrev,by = c('team' = 'V2'))
player_season <- player_season[,c(-19,-24)]
pl_season <- player_season[,c(1,2,3,30)]
players <- data.table(inner_join(player,pl_season,by=c('player_id' = 'Player ID','season_end' = 'Year','V1'='Tm')))

# player dictionary
players <- players[,c(2,5,8)]


# edge list
players_for_edge_list <- players
edge_list <- data.table(left_join(players_for_edge_list,players_for_edge_list,by = c('V1','season_end')))
edge_list <- edge_list[player_id.x != player_id.y,]
edge_list[,season_lag1 := shift(season_end,1,type = 'lag'),by=.(player_id.y)]
edge_list <- data.table(left_join(edge_list,players,by=c('player_id.y'='player_id','season_lag1'='season_end')))
edge_list[,season_lag2 := shift(season_end,2,type = 'lag'),by=.(player_id.y)]
edge_list <- data.table(left_join(edge_list,players,by=c('player_id.y'='player_id','season_lag2'='season_end')))
names(edge_list) <- c('player_id.x','season_end','team','player_id.y','lag1','lag1_team','lag2','lag2_team')
edge_list <- data.table(data.frame(edge_list)[,c(1,2,3,4,6,8)][!is.na(edge_list$lag1_team)&!is.na(edge_list$lag2_team),])

player_season <- data.frame(player_season)
player_season <- player_season[,c(1,2,4,5,6,7,17,30)]
salary <- data.frame(player)[,c(2,3,5)]
player_table <- left_join(player_season,salary,by=c("Player.ID"="player_id","Year"="season_end"))
```

2. Code for Exploratory Data Analysis

```python
import pandas as pd
import pandas_profiling
import numpy as np


# importing the data
df = pd.read_csv('player1985to2018.csv')


profile = pandas_profiling.ProfileReport(df)
profile.to_file("player1985to2018.html")
```

3. Code for Player Network

```r
#-----------------------------------
edge_player<-as.matrix(unique(edge_list,by=c('player_id.x','player_id.y','team')))
head(edge_player)
edge_player<-as.data.table(edge_player)
edge_player$season_end<-as.numeric(edge_player$season_end)
player_mobility<-as.data.frame(matrix(ncol = 6))
colnames(player_mobility)<-c('degree','between','close','eigen','player_id','year')

for (y in unique(edge_player$season_end)){
  year_dt<-edge_player[season_end<=y&season_end>=y-1,c(1,4)]
  g_sub<-graph_from_edgelist(as.matrix(year_dt),directed = FALSE)
  dt1<-data.frame(degree=degree(g_sub,normalized = T),between=betweenness(g_sub,normalized = T),close=closeness(g_sub,normalized = T),
                  eigen=eigen_centrality(g_sub)$vector)
  dt1$player_id<-rownames(dt1)
  dt1$year<-y
  player_mobility<-rbind(player_mobility,dt1)
  print(y)
}

player_mobility<-na.omit(player_mobility)
#divide into high/low performance
player_mobility<-left_join(player_mobility,player_stat_1985_2015[,c(1,3,4)],by=c('player_id'='Player ID','year'='Year'))
player_mobility<-as.data.table(player_mobility)
player_mobility[PER>=15,performance:='high']
player_mobility[PER<15,performance:='low']

#team list
team_list<-unique(player[,c(2,5,8)])


player_mobility_team<-left_join(player_mobility,team_list,by=c('player_id'='player_id','year'='season_end'))
#some player who is in the last year network has retired, so delete
player_mobility_team<-as.data.table(player_mobility_team)
team_mobility_high<-player_mobility_team[performance=='high',.(mean(degree),mean(between),mean(close),mean(eigen)),by=.(year,V1)]
colnames(team_mobility_high)<-c('year','team','degree_team_high','between_team_high','close_team_high','eigen_team_high')
team_mobility_low<-player_mobility_team[performance=='low',.(mean(degree),mean(between),mean(close),mean(eigen)),by=.(year,V1)]
colnames(team_mobility_low)<-c('year','team','degree_team_low','between_team_low','close_team_low','eigen_team_low')
team_mobility<-inner_join(team_mobility_high,team_mobility_low,by=c('year','team'))
#colnames(team_mobility)<-c('year','team','degree_team','between_team','close_team','eigen_team')
team_mobility<-na.omit(team_mobility)
team_mobility$degree_team<-team_mobility$degree_team_high+team_mobility$degree_team_low

reg_team_mobility<-left_join(team,team_mobility,by=c('season_end'='year','V1'='team'))
reg_team_mobility<-na.omit(reg_team_mobility)
```

```r
reg_team_mobility<-na.omit(reg_team_mobility)

reg1 <- lm(`Winning Percentage` ~ degree_team_high+ degree_team_low  +  factor(V1)+factor(season_end),data = reg_team_mobility)
summary(reg1)

library(ggplot2)
ggplot(reg_team_mobility,aes(x=degree_team_high,y=`Winning Percentage`))+geom_smooth(method = 'loess',se=T)

####what kind of player has the largest mobility, divide by low/high performance
player_stat_1985_2015<-player_season[Year>=1985&Year<=2015,c('Year','Tm','Player ID','PER','TS%','ORB%','DRB%','AST%','STL%','BLK%','TOV%','USG%')]
mobility_stats<-left_join(player_mobility,player_stat_1985_2015,by=c('year'='Year','player_id'='Player ID'))

write.csv(player_stat_1985_2015, file = "player_stats.csv")

#salary of players
player_salary<-as.data.table(player[,c(2,3,5,8)])
player_salary[,min:=min(salary),by=.(season_end)]
player_salary[,max:=max(salary),by=.(season_end)]
player_salary[,normalize:=(salary-min)/(max-min)]

mobility_stats<-left_join(mobility_stats,player_salary,by=c('player_id','year'='season_end'))
mobility_stats<-na.omit(mobility_stats)
mobility_stats<-mobility_stats[mobility_stats$performance=='low',]


reg3<-lm(degree~normalize+`PER.x`+`TS%`+`ORB%`+`DRB%`+`AST%`+`STL%`+`BLK%`+`TOV%`+`USG%`+factor(Tm)+factor(year),data=mobility_stats)
summary(reg3)
#plm
library(plm)
mobility_stats<-as.data.table(mobility_stats)
mobility_stats<-unique(mobility_stats,by=c('player_id','year'))
colnames(mobility_stats)[11:18]<-c('TS','ORB','DRB','AST','STL','BLK','TOV','USG')
mobility_stats$per_square<-(mobility_stats$PER.x-15)^2
summary(plm(degree~normalize+per_square+TS+ORB+DRB+AST+STL+BLK+TOV+USG,data = mobility_stats, effect = "twoways", model = "within", index = c('year')))

write.csv(mobility_stats, file = "mobility_stats_player.csv")
#salary and mobility
mobility_salary<-left_join(player_mobility,player[,c(2,3,5,7)],by=c('year'='season_end','player_id'))
mobility_salary<-na.omit(mobility_salary)
mobility_salary<-as.data.table(mobility_salary)
mobility_salary[,total_salary:=sum(salary),by=.(year)]
mobility_salary$salary_norm<-mobility_salary$salary/mobility_salary$total_salary

reg2<-lm(salary_norm~degree+factor(year)+factor(team),data=mobility_salary)
summary(reg2)
```

4. Code for Team Network

```r
team_network<-data.frame(outdegree=NA,indegree=NA,closeness=NA,betweeness = NA,eigen_centrality = NA,year=NA,team=NA)[numeric

for(k in 1985:2016){
  edge_temp <- edge_list[season_end == k & (team!=lag1_team | team!=lag2_team) ]
  edge_temp <- data.frame(edge_temp)
  temp1 <- edge_temp[,c(3,5)]
  names(temp1) <- c('focal','neighbor')
  temp2 <- edge_temp[,c(3,6)]
  names(temp2) <- c('focal','neighbor')
  edge_temp <- rbind(temp1,temp2)
  edge_temp <- edge_temp[edge_temp$focal != edge_temp$neighbor,]
  #edge_temp <- edge_temp[!duplicated(edge_temp),]

  g <- graph.data.frame(edge_temp,directed = TRUE)
  data <- cbind(V(g),degree(g,mode='out'),degree(g,mode='in'),closeness(g),betweenness(g),eigen_centrality(g)$vector,rep(k,le
  data <- data.frame(data)[,2:7]
  data$team <- rownames(data)
  names(data) <- c('outdegree','indegree','closeness','betweeness','eigen_centrality','year','team')
  team_network <- rbind(team_network,data)
}

library(MASS)

team[,c(1,2,4)]
team$Year <- unlist(lapply(team$Year, function(x) gsub('\\*','',x)))
abbrev$V3 <- unlist(lapply(abbrev$V2, function(x) tail(strsplit(x,split=" ")[[1]],1)))
team <- left_join(team,abbrev,by=c('Team'='V3'))[,c(1:4,22:23)]
team <- team[!is.na(team$V2),]
t_dict <- player[,c(4,5)]
t_dict <- t_dict[!duplicated(t_dict),]
team <- left_join(team,t_dict,by=c('Year'='season'))
team <- team[!is.na(team$season_end),]
team <- left_join(team,team_network,by=c('season_end'='year','V1'='team'))
team <- team[!is.na(team$outdegree),]
team$diff <- team$outdegree - team$indegree
reg1 <- lm(`Winning Percentage` ~ outdegree,data = team)
summary(reg1)


temp <- edge_list[,.(coop_year = length(duplicated(.SD))[1]),by=.(season_end,team),.SDcols = c("player_id.x","player_id.y")]
temp <- temp[order(season_end),][order(team)]
y <- temp$season_end
temp <- temp[,.(coop_cum = cumsum(coop_year)),by=.(team)]
temp[,lag2 := shift(coop_cum,2,type='lag'),by=.(team)]
temp[,diff_cum := coop_cum-lag2]
temp$season_end <- y

team <- inner_join(team,temp,by=c('season_end','V1'='team'))

reg1 <- lm(`Winning Percentage` ~ outdegree + coop_cum + win_lag1,data = team)
summary(reg1)
```

```r
team <- data.table(team)
team <- team[order(season_end),][order(Team)]
team[,win_lag1 := shift(`Winning Percentage`,n=1,type='lag'),by=.(Team)]
dup <- team[,.N,by=.(Team,season_end)][team[,.N,by=.(Team,season_end)]$N > 1,]
team <- team[Team!='Hornets' & Team !='Knicks']

ns <- names(team)
names(team) <- c(ns[1:3],'wr',ns[5:17])
library(MASS)
reg1 <- plm(wr ~ outdegree + coop_cum  + win_lag1 ,data = team,effect = 'twoways',model='within',index='Team')
summary(reg1)
team <- na.omit(team)
team2 <- data.frame(team)[,c('outdegree','coop_cum','wr','win_lag1','Team','season_end')]
reg2 <- panelAR(wr ~ outdegree + coop_cum + win_lag1 , data = team2, panelVar='Team',timeVar='season_end',autoCorr = 'psar1',p

player_season <- player_season[player_season$MP > 500]
player_strength <- data.table(player_season[,c("Tm","Player ID","PER","Year")])
player_strength <- player_strength[,.(avg_strength = mean(`PER`,na.rm=TRUE)),by=.(Tm,Year)]
player_strength_top <- data.table(player_season[,c("Tm","Player ID","PER","Year")])
player_strength_top <- player_strength_top[,.(max_strength = max(`PER`,na.rm=TRUE)),by=.(Tm,Year)]
player_strength_min <- data.table(player_season[,c("Tm","Player ID","PER","Year")])
player_strength_min <- player_strength_min[,.(min_strength = min(`PER`,na.rm=TRUE)),by=.(Tm,Year)]

#player_strength_top5 <-data.table(player_season[,c("Tm","Player ID","PER","Year")])
#player_strength_top5 <- setDT(player_strength_top5)[player_strength_top5[,.I[head(seq_len(`PER`), 5)], .(Tm,Year)]$V1]



team<- inner_join(team,player_strength,by=c('V1'='Tm','season_end'='Year'))
team<- inner_join(team,player_strength_top,by=c('V1'='Tm','season_end'='Year'))
team<- inner_join(team,player_strength_min,by=c('V1'='Tm','season_end'='Year'))
reg1 <- lm(wr ~ outdegree + coop_cum + avg_strength+max_strength +min_strength,data = team)
summary(reg1)
reg1 <- plm(wr ~ outdegree + coop_cum + avg_strength+max_strength +min_strength,data = team,effect = 'twoways',model='within',
summary(reg1)

team_rich <- data.table(player)[,.(total_salary = sum(salary)),by=.(V1,season_end)]
team<- inner_join(team,team_rich,by=c('V1','season_end'))
reg1 <- lm(wr ~ outdegree + coop_cum + avg_strength+max_strength +min_strength+total_salary,data = team)
summary(reg1)
reg1 <- plm(wr ~ outdegree + avg_strength+max_strength +min_strength+total_salary,data = team,effect = 'twoways',model='within
summary(reg1)

team <- team[order(team$season_end),]
team <- data.table(team)
team[,avglag1 := shift(avg_strength,1,type='lag'),by=.(Team)]
team[,minlag1 := shift(min_strength,1,type='lag'),by=.(Team)]
team[,maxlag1 := shift(max_strength,1,type='lag'),by=.(Team)]
team <- na.omit(team)

reg1 <- plm(wr ~ outdegree + avglag1+maxlag1 +minlag1 +total_salary,data = team,effect = 'twoways',model='within',index='Team'
summary(reg1)
```