
Abstract

This research proposes a cricket analysis using a collection of datasets. In order to get meaningful information, we evaluate qualities such as best batsman, best bowler, best batter to bat with (partnerships), and a comparison of various IPL teams and their players. Nodes represent the players, and their runs/wickets are represented by edges connecting nodes. In addition, the paper offers "Game Theory in Cricket," which explains which strategy or response is optimal when encountering a specific ball and for the bowler which ball is best against a particular batsman to maximize payoffs. We used social network analysis and concepts like the PageRank algorithm, Degree distribution, Game theory, Etc. We created a variety of graphs to visualize better and gain meaningful insights.

Keywords: Cricket analysis, Social network analysis, Graphs , Gephi, Game Theory, Pagerank.

1 Introduction

Consider the cricket game, which generates vast amounts of data every year. Each team's technical analyst analyses the data to work on specific elements as a team or individually. One method is to use social network analysis.

The Indian Premier League is the most well-known cricket league in the world. We acquire many datasets from Kaggle and cricsheet that include information about various areas of cricket, such as a ball to ball data, match win data, one IPL season data, commentary data, and many more. We attempted to analyze and visualize IPL data to provide answers to some cricketing questions and get significant insights.

2 Dataset

We used three datasets in our Analysis.

1. Cricsheet Data

The ball by ball data of each and every match played in the history of IPL from cricsheet.org.

2. Kaggle IPL Dataset

This dataset gives us quite a lot of useful information, such as: The batting average in different formats, Number of times the batsman scored 100 or more than it, Highest scores of the batsman, Number of matches

played.

3. IPL 2019 Commentary Data The dataset consists of a summary of all IPL 2019 matches.

About Variables

The variables in the match-level data are self-explanatory. Variables in ball-level data

- Over-No - Represents the ball number in a specific over
- Over-Score - Runs scored. The variable also consists of extras and wickets taken. For example, 1b means 1 bye. w means wicket.
- Short-comm - A brief commentary. Consists of batsman and bowler names
- Commentary - Full commentary text
- Bold-Comm - Consists of bold words highlighted in the commentary

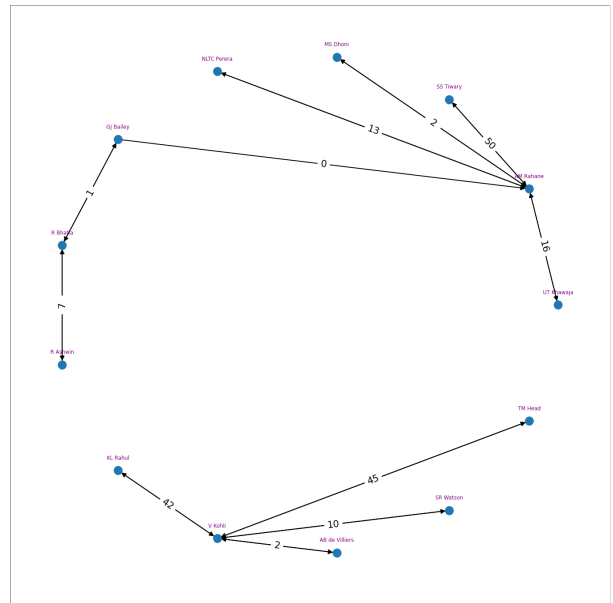
3 Most valuable batsman of particular season/series or match

Creating nodes and edges, where nodes represent players and edges indicate unique game behaviours, is one of the more exciting ways to look at the IPL data.

The batsmen and bowlers are nodes in the batsmen-bowler network depicted in Figure, and the directed weighted edges between them represent the runs scored off a bowler by the batsman. Orange, red and blue are used to colour the borders of the edges with higher weights.

The red edges represent batsmen who have scored more than 25 runs, the orange edges suggest batsmen who have scored more than 15 runs, and the blue edges indicate players who have scored fewer than 15 runs.

Virat kohli, for example, has 30 runs off A Zampa, which is represented by an edge from A Zampa's node to Virat kohli with a weight of 30. Same as for Orange and Blue edges. A Rahane has 19 runs off CJ Jordan, which is represented as a weighted orange edge from CJ Jordan's edge to A Rahane.



Pagerank

[('V Kohli', 0.13883323546725787),
('AM Rahane', 0.1000159268301116),
('SS Tiwary', 0.08717002549588658),
('KL Rahul', 0.06628412478992703),
('SR Watson', 0.06004648817405449),
('MS Dhoni', 0.045602306675803866),
('UT Khawaja', 0.04482163903363839),
('NLTC Perera', 0.04098370549926644),
('TM Head', 0.03916928571780246),
('R Bhatia', 0.038498932748557854),
('R Ashwin', 0.0375850636490481),
('GJ Bailey', 0.03196991624433954),
('AB de Villiers', 0.03055345679070464),
('STR Binny', 0.02980823661045011),
('Parvez Rasool', 0.02980823661045011),
('CJ Jordan', 0.02980823661045011),
('VR Aaron', 0.02980823661045011),
('YS Chahal', 0.02980823661045011),
('AB Dinda', 0.02980823661045011),
('RP Singh', 0.02980823661045011),
('A Zampa', 0.02980823661045011)]

For Ipl Match (year) Virat Kohli was indeed the player of the match as also ranked at the top by Pagerank. Closely follows by AM Rahane.

4 Bowler part

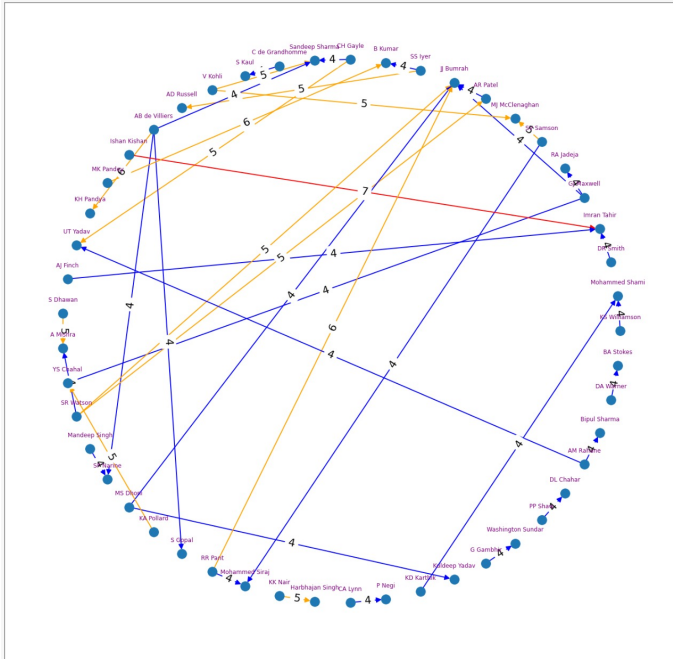


Figure.1

- Nodes → Player (batsman and bowler)
- Edges → wickets taken by player
- weighted graph
- Red edge = 7 or above wickets
- Orange edge = 5 or 6 wickets
- Blue edge = 4 wickets
- K Rabada has a red edge
- If we know 3-4 batsman of opponent team. we can format our team in such a way that for each batsman, we have a dominating bowler.

Here, we have taken the data of 251 matches and tried to plot the graph between the bowler and batsman. This graph will help us to understand that which bowler is dominating more batsmen we have kept 4 different parameters if a bowler has taken the wicket of same batsman less than 4 the edge between the bowler and batsman should not be visible. If a bowler has take the wicket of the batsman 4 times than the color of the edge will be blue. Similarly yellow color for 5 and 6 time and Red color for more than 6 time.

With this data we can plan for a particular batsman. Like what type bowlers are the weak point of batsman. Which bowler will be more efficient against the batsman of particular team. If we have the data of other team's batsman in playing 11 we can arrange the bowling order in such a that we have a multiple plans for batsman to get his wicket.

5 IPL partnership analysis of 2019 season of IPL

IPL is becoming more and more competitive year by year. If a team wants to win a match the team have to relay on both of their bowlers and batmen. As we are observing year by year the batting score of 1st inning is exponentially increasing. So, as a cricket enthusiast we have to keep look at batting performance of each team year by year. Cricket teams are paying too much amount for a data analyst to analyse and visualise the data of the batting performance and finding the weak spot of particular batsmen so that a batsmen can improve. Also, to pick a perfect batsman for a perfect place is one of the hardest things to decide. We have formatted graphs of batting partnership of different players in team to reduce the load of coach and captain to choose the beast batsmen for the team for both of the roles

(1) Supporting

(2) Run scorer

We have formatted a circular graph containing the information about batsmen in the nodes. These all graphs are directed graphs arrow to a particular batsman represent the major contribution in the partnership. In that we are determining the major run scorer in the partnership. The out arrow represents the minor contribution in the partnership formation. But as a team management should be more focus on minor contributor too. Because these are the batsmen who provides the freedom to the main run scorer to play their natural game (whether it is aggressive or defensive). With that minor contribution visualization, we can see the contribution of lower order batsmen too. Also, by this data we can see the overall characteristic (in terms of who is scoring more run) of the team as well.

We have built 4 graphs for 4 teams

- (1) CSK
- (2) MI
- (3) SRH
- (4) DC

By looking at the sorted indegree, out degree and total degree of the player we can clearly distinguish between both types of the batsmen. Through we also can pick best 2-3 batsmen which can give more contribution to team for IPL mega auctions.

Nodes → Player (either batsman or bowler)

Edges → Runs scored by both batsman (Partnership)

Directed graph

Here Direction (arrow) represent that contribution of that batsman is more.

Chennai Super Kings(CSK)



degree

Total Degree	In Degree	Out Degree
[('AT Rayudu', 9), ('MS Dhoni', 8), ('SR Watson', 7), ('SK Raina', 7), ('KM Jadhav', 7), ('RA Jadeja', 7), ('DJ Bravo', 7), ('M Vijay', 6), ('F du Plessis', 5), ('MJ Santner', 5), ('SN Thakur', 4), ('DL Chahar', 3), ('MM Sharma', 1), ('SW Billings', 1), ('DR Shorey', 1), ('Harbhajan Singh', 1), ('Imran Tahir', 1)]	[('KM Jadhav', 5), ('SR Watson', 4), ('AT Rayudu', 4), ('MS Dhoni', 4), ('F du Plessis', 4), ('SK Raina', 3), ('RA Jadeja', 3), ('DJ Bravo', 3), ('MJ Santner', 3), ('M Vijay', 3), ('SN Thakur', 2), ('DL Chahar', 1), ('Harbhajan Singh', 1), ('MM Sharma', 0), ('SW Billings', 0), ('DR Shorey', 0), ('Imran Tahir', 0)]	[('AT Rayudu', 5), ('SK Raina', 4), ('RA Jadeja', 4), ('MS Dhoni', 4), ('DJ Bravo', 4), ('SR Watson', 3), ('M Vijay', 3), ('KM Jadhav', 2), ('DL Chahar', 2), ('SN Thakur', 2), ('MJ Santner', 2), ('MM Sharma', 1), ('F du Plessis', 1), ('SW Billings', 1), ('DR Shorey', 1), ('Imran Tahir', 1), ('Harbhajan Singh', 0)]

analysis

- Most In degree - KM Jadhav
- Most Out degree - Ambati Rayudu
- Most Total Degree - Ambati Rayudu

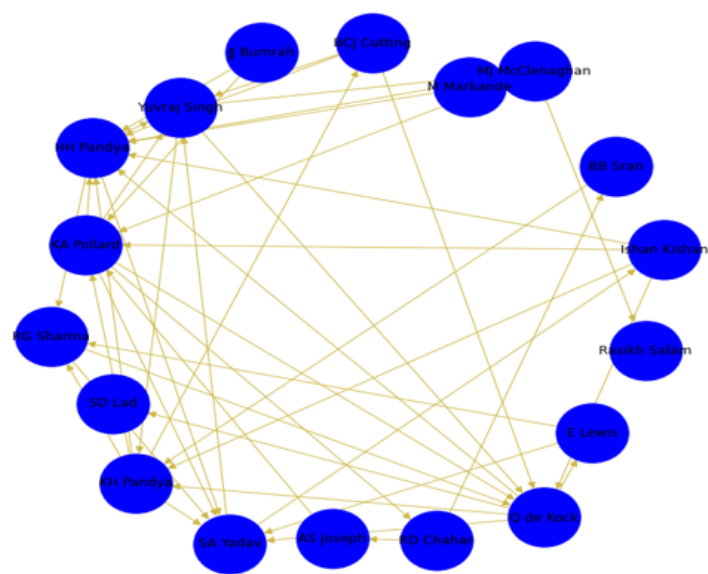
Here as we can see in the image of total degree more than 9 batsmen having total edge number greater than 6. This is showing that in csk there is no dominance of a particular batsman in the team. All the batsman of the team is contributing in the total score of teams. The total and out degree of AT Rayudu is highest in the team. So, as a coach's perspective we can say that it is better to place AT Rayudu above in batting order to get maximum support and other

batsmen can play along with AT Rayudu superior support.

- Most In degree - Hardik Pandya

Mumbai Indians(MI)

Graph



- Most Out degree - Q de kock

- Most Total Degree - Hardik Pandya

Here the total degree and in degree of HH Pandya is the highest in the team. These mean that HH Pandya dominated the team in terms of batting partnership. With the high count of in degree HH Pandya was the best run scorer as well in Mumbai Indians team. So, as a conclusion we can say that to not retain HH Pandya in team in 2022 season might be the reason behind 8 consecutive loss of Mumbai Indians team. If we look at total degree of team only 6 batsmen are contributing more in the partnership. So, we can say that MI batting order is more dependent on 5-6 batsmen.

Degree

Total Degree	In Degree	Out Degree
[('HH Pandya', 11), ('Q de Kock', 10), ('KA Pollard', 10), ('KH Pandya', 9), ('SA Yadav', 8), ('Yuvraj Singh', 7), ('RG Sharma', 5), ('MJ McClenaghan', 5), ('Ishan Kishan', 5), ('BCJ Cutting', 4), ('RD Chahar', 3), ('E Lewis', 3), ('M Markande', 2), ('JJ Bumrah', 2), ('AS Joseph', 2), ('BB Sran', 2), ('Rasikh Salam', 1), ('SD Lad', 1)]	[('HH Pandya', 8), ('SA Yadav', 6), ('Q de Kock', 5), ('KA Pollard', 5), ('Yuvraj Singh', 4), ('KH Pandya', 4), ('RG Sharma', 3), ('MJ McClenaghan', 2), ('BCJ Cutting', 1), ('Rasikh Salam', 1), ('Ishan Kishan', 1), ('RD Chahar', 1), ('AS Joseph', 1), ('SD Lad', 1), ('E Lewis', 1), ('BB Sran', 1), ('M Markande', 0), ('JJ Bumrah', 0)]	[('Q de Kock', 5), ('KA Pollard', 5), ('KH Pandya', 5), ('Ishan Kishan', 4), ('Yuvraj Singh', 3), ('HH Pandya', 3), ('BCJ Cutting', 3), ('MJ McClenaghan', 3), ('RG Sharma', 2), ('SA Yadav', 2), ('M Markande', 2), ('JJ Bumrah', 2), ('E Lewis', 2), ('AS Joseph', 1), ('BB Sran', 1), ('Rasikh Salam', 0), ('SD Lad', 0)]

SunRisers Hyderabad (SRH)

Graph

Analysis



The degree distribution of SRH team is kind of similar to CSK more than 7 batsmen have more than 7 total degrees in the team. Surprisingly, Spin bowler Rashid Khan has highest out degree and second highest total degree in the team. This shows the capability of batsman to support well to other batsmen and if need this player can finish well with other tailender batsmen as well. This data is supporting current batting performance of rashid Khan in Gujarat Titans team. We can also say that is SRH to not retain Rashid Khan is the biggest loss to the team.

Degree

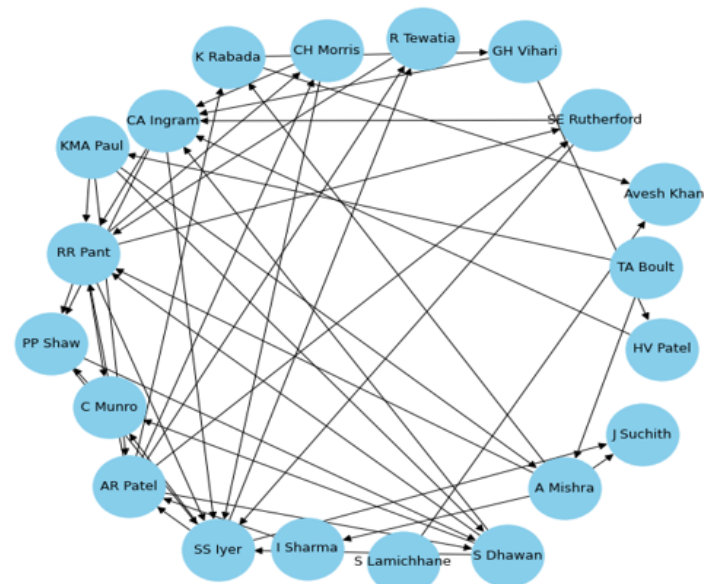
Total Degree	In Degree	Out Degree
(('V Shankar', 10), ('Rashid Khan', 10), ('DA Warner', 9), ('MK Pandey', 9), ('DJ Hooda', 9), ('YK Pathan', 8), ('KS Williamson', 8), ('Mohammad Nabi', 8), ('B Kumar', 7), ('JM Bairstow', 5), ('Abhishek Sharma', 5), ('Shakib Al Hasan', 4), ('S Kaul', 3), ('Sandeep Sharma', 3), ('WP Saha', 3), ('MJ Guptill', 2), ('RK Bhui', 1), ('KK Ahmed', 1), ('Basil Thampi', 1))	(('DA Warner', 6), ('MK Pandey', 6), ('V Shankar', 5), ('DJ Hooda', 5), ('Mohammad Nabi', 5), ('Rashid Khan', 4), ('B Kumar', 4), ('JM Bairstow', 3), ('YK Pathan', 3), ('KS Williamson', 3), ('Shakib Al Hasan', 3), ('Sandeep Sharma', 2), ('Abhishek Sharma', 1), ('WP Saha', 1), ('MJ Guptill', 1), ('Basil Thampi', 1), ('S Kaul', 0), ('RK Bhui', 0), ('KK Ahmed', 0))	(('Rashid Khan', 6), ('V Shankar', 5), ('YK Pathan', 5), ('KS Williamson', 5), ('DJ Hooda', 4), ('Abhishek Sharma', 4), ('DA Warner', 3), ('MK Pandey', 3), ('Mohammad Nabi', 3), ('B Kumar', 3), ('S Kaul', 3), ('JM Bairstow', 2), ('WP Saha', 2), ('Sandeep Sharma', 1), ('RK Bhui', 1), ('KK Ahmed', 1), ('Shakib Al Hasan', 1), ('MJ Guptill', 1), ('Basil Thampi', 0))

Analysis

- Most In degree - DA Warner
- Most Out degree - Rashid Khan
- Most Total Degree - Vijay Sankar

Delhi Capitals (DC)

Graph



Degree

Total Degree	In Degree	Out Degree
[('RR Pant', 11), (('SS Iyer', 10), (('AR Patel', 9), (('CA Ingram', 8), (('S Dhawan', 7), (('A Mishra', 6), (('PP Shaw', 5), (('KMA Paul', 5), (('CH Morris', 4), (('K Rabada', 4), (('C Munro', 4), (('SE Rutherford', 4), (('R Tewatia', 3), (('GH Vihari', 3), (('HV Patel', 2), (('Avesh Khan', 2), (('I Sharma', 2), (('J Suchith', 2), (('TA Boult', 2), (('S Lamichhane', 1)]	[('SS Iyer', 6), (('RR Pant', 6), (('CA Ingram', 5), (('PP Shaw', 3), (('S Dhawan', 3), (('AR Patel', 3), (('C Munro', 3), (('R Tewatia', 2), (('CH Morris', 2), (('K Rabada', 2), (('Avesh Khan', 2), (('A Mishra', 2), (('SE Rutherford', 2), (('J Suchith', 2), (('KMA Paul', 1), (('GH Vihari', 1), (('HV Patel', 1), (('I Sharma', 1), (('S Lamichhane', 0), (('TA Boult', 0)]	[('AR Patel', 6), (('RR Pant', 5), (('S Dhawan', 4), (('SS Iyer', 4), (('KMA Paul', 4), (('A Mishra', 4), (('CA Ingram', 3), (('PP Shaw', 2), (('GH Vihari', 2), (('CH Morris', 2), (('K Rabada', 2), (('SE Rutherford', 2), (('TA Boult', 2), (('R Tewatia', 1), (('HV Patel', 1), (('S Lamichhane', 1), (('I Sharma', 1), (('C Munro', 1), (('Avesh Khan', 0), (('J Suchith', 0)]

Analysis

- Most In degree - SS Iyer
- Most Out degree - AR Patel
- Most Total Degree - RR Pant

As we can see in the total degree distribution of Delhi team, we may see the similarity between MI and DC. Because in DC RR Pant is dominating the team with highest total degree, second highest out degree and in degree as well. Also, like MI team DC's total degree is limited to around 5-6 batsmen. So, we can say that the team is more dependent upon 5-6 player to score runs.

6 Game Theory in Cricket

The concept helps determine decision-making in cricket. For batters, always make some strategy to face balls, such as which shot to play against which ball to

get higher payoffs. Bowlers make strategies such as if some batter has strength in playing certain shots (such as the pull shot and Rohit Sharma), then which ball is better to bowl against them to get the most payoffs.

This article takes an IPL 2019 dataset to investigate game theory in cricket. We separated information as per getting payoffs for batsman and Bowler while playing shots and balling a ball, respectively.

For the batsman, his strategy is to play a

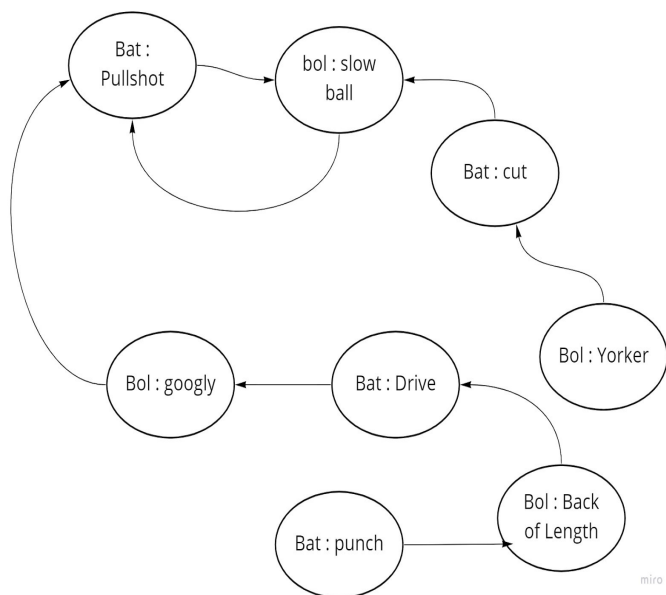
- Pull Shot
- Cut Shot
- Drive Shot
- Punch shot

And for the Bowler, his strategy is to ball a

- Back of a length ball
- Googly ball
- Yorker ball
- Slower ball

After filtering the data, we get a Figure 1 data table that gives both payoffs.

From the Payoff tables we draw a **best response graph** which is shown in Figure



best response graph

We address some of the question with the help of the best response graph which are listed below.

Q- What is the best response for the batsman when bowler ball a Back of length?

→ **Drive Shot**

Q- What is the best response for the batsman when bowler ball a Googly?

→ **Pull Shot**

Q- What is the best response for the batsman when bowler ball a Yorker?

→ **Cut Shot**

Q- What is the best response for the batsman when bowler ball a Slower ball?

→ **Pull Shot**

Q- What is the best response for the Bowler when Batsman has Strength in playing Pull Shot?

→ **Slower ball**

Q- What is the best response for the Bowler when Batsman has Strength in playing Cut Shot?

→ **Slower ball**

Q- What is the best response for the batsman when Batsman has Strength in playing Drive Shot?

→ **Googly or Slower ball**

Q- What is the best response for the batsman when Batsman has Strength in playing Punch Shot?

→ **Back of a Length ball**

Amazing Insights

As we can see, out of the four best responses, 3 has a slower ball. Therefore, we can conclude that the bowler should deliver a slower ball to maximize his pay-off regardless of a batsman's power.

Nash Equilibrium

Property	Pull Shot	Cut Shot	Drive	Punch
Back of a Length	(0.09,1.31)	(0.01,1)	(0,1.67)	(0.04,0.81)
Googly	(0.08, 1)	(0.15 ,0.73)	(0.11,0.72)	(0,0.63)
Yorker	(0,0)	(0,2.3)	(0.04, 1.48)	(0,0)
Slower Ball	(0.12, 1.22)	(0.19, 0.81)	(0.11,0.94)	(0,1.07)

Table 1:
Payoffs of Batsman and Bowler

The best response graph shows that the game has one Nash equilibrium between a Slower ball and a Pull shot. If the bowler bowls a slower ball, the batsman should play the Pull shot to maximize his payoffs, and if the batsman has strength in playing a Pull shot, then the bowler should ball a Slower ball to maximize his payoffs.

7 Related Work

We can use social network analysis to expand our analysis to **team formation**. In the IPL, like in every cricket match, team selection is always a subject of concern. We may construct a team in various ways, one of which is to form a team based on the opponents' team. If the other side has one bowler or batsman, choose a batter who has played a better inning against that bowler and a bowler who has already dismissed an opponent's batters several times. In other words, choose a player who has more power than the opposing player. Another option is to choose team members based on their individual and team performance, where team performance refers to a partnership or a team player. In our first study, we created a graph that indicates how certain players are stronger than others, an analysis that we can apply to team formation and other cricket analyses.

8 conclusion

This study focuses on numerous cricket analyses employing social network analysis and its features, such as the PageRank algorithm, Game theory, Degree distribution, Etc. When we analyze a dataset, we discover certain essential players who may not appear to be contributing much yet have a significant influence on a team. Sometimes we obtain fantastic insights, such as the reason for MI's continuous eight losses in 2022 is because Hardik Pandya, the team's primary player in prior seasons, was not taken. We can acquire the dominating ball of a particular season and some of the finest strategies for the player to play using Game theory.

9 Languages and Libraries Used

- Python (Google Colab) - Data processing, Visualization and and analysis
- Excel - Data filtering and for Game theory
- NetworkX - Plotting Graphs and finding degree

References

- [1] Akbar Hussain1, Yan Qiang. (n.d.). Region-wise ranking for one-Day international . Retrieved from <https://pdfs.semanticscholar.org/6784/5b>
- [2] Dr. R. Ramesh. Game theory-nash equilibrium and its applications, from <http://www.journalstd.com/gallery/5-oct2019special.pdf>

[3] Network Analysis Python for IPL 2019. Analytics Vidhya. (2020, June 14). Retrieved from <https://www.analyticsvidhya.com/blog/2020/02/network-analysis-ipl-data/>