National University
Of Computer and Emerging Sciences

Business Analytics

Technique

(CS-4062)

Name : Aliasghar

Roll no : 19K-0293

Section : A

Important Questions

Q1) In what context or for whom the performance is being measured?

Q2) What are the objectives or why the performance is being measured?

Q3) Is the measured performance aligning with the objectives?

Q4) How is the performance measured?

Q5) How can we validate our measured performance?

Q6) What's the most important feature/variable involved while measuring performance?

Q7) What will be the method for calculating performance?

Q8) Does the calculation of performance involves single or multiple features/variables?

Measure of performance &

hypothesis and experiments

The performance cannot be directly measured, first of all it should be known that for what audience or stake holders the performance is being measured and what is their interest.

So, hypothesis are in regards to the stake holders that wants to select a team and needs a captain, openers, middle order batsmen and bowlers the hypothesis further defines each category **performance criteria**.

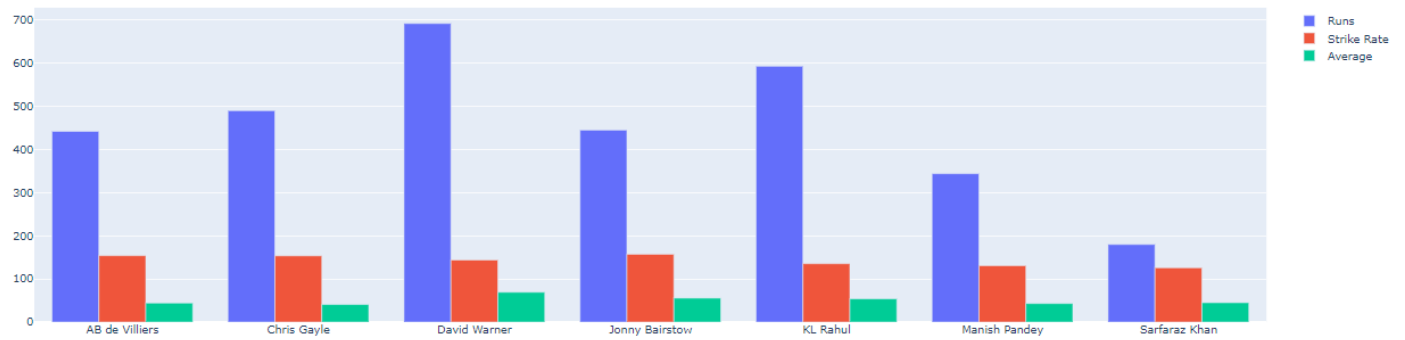
> *NOTE: the data is highly reflecting only the properties related to a batsman,* ***hence****,*
>
> *1) Captains selected are batsmen primarily.*
>
> *2) Assumption for bowlers, i.e., SR is below 120.00.*

Following are the hypothesis made along with the performance metrics:

1) For a captain he may have

- an average above 40.
- at least have a 100 or 50 (to know that he is experienced).
- no. of fours is more compared to sixes (to show that he plays non aggressive and sensible shots).

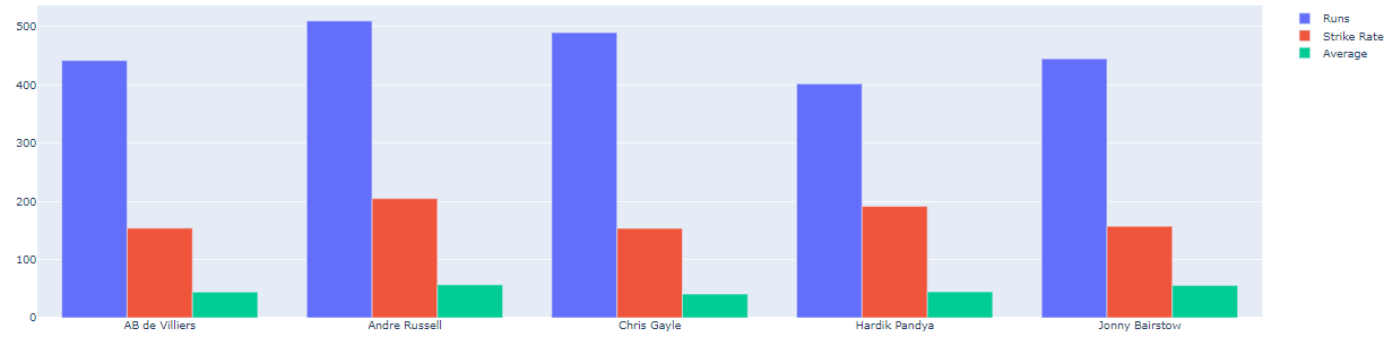
```
No. of potential Captains are:  7
```



2) Openers should be aggressive and make complete use of the powerplay so,

- above 150.00 SR
- above 40.00 Avg

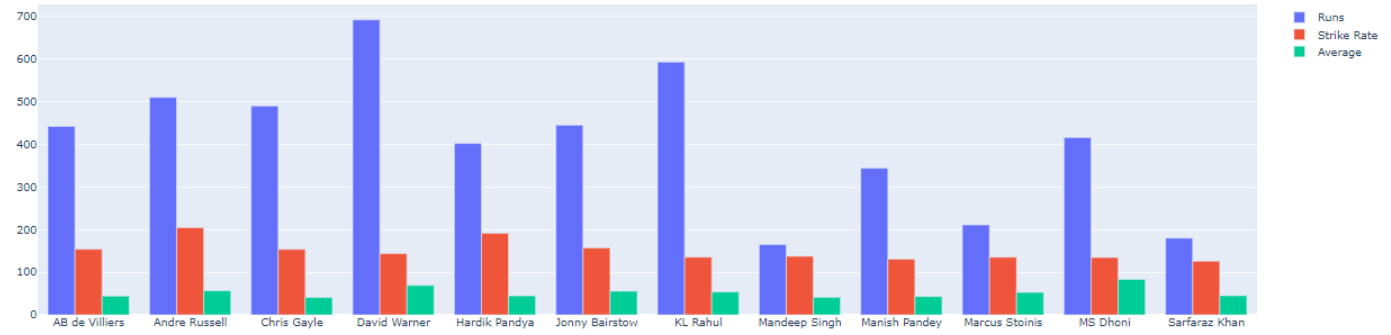
```
No. of potential Openers are:  5
```



3) Middle order batsman main objective is slow and steady batting and chasing the score so,

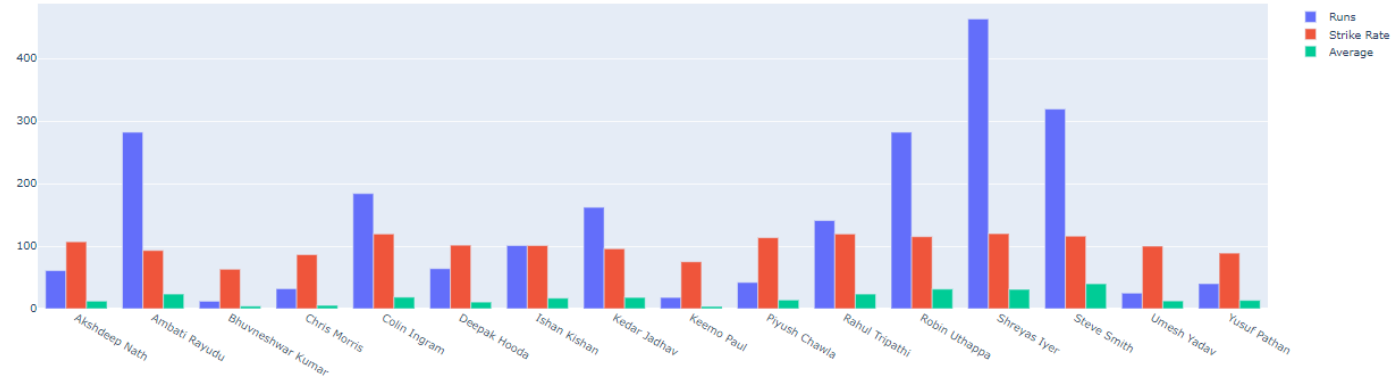
- high averages above 40
- runs above 150

```
No. of potential Middle-Order Batsmen are:  12
```



4) (read Note above) bowlers should have an SR below 120.

```
No. of potential Bowlers are:  16
```



Develop predictive Models

Predictive modelling is done in order to predict the salaries of each Player in accordance with the features/variables pprovided as performance for them.

> ***Note:*** *model used for the following is multi variate linear regression.*

The steps included are as follows:

1) Data pre-processing (cleaning the data)

2) Feature engineering (is the step where predictor variables are decided for the model, and target variable is separated)

- Sno, Team, Player's columns are of no use so we will remove it as we are evaluating individual salaries of players based on performance and textual data is of no use.

- To see which feature/variable has highest correlation with salary.
- Using recursive feature engineering (RFE) we will be selecting the most dependent variables for the model.

3) Training and splitting into 75:25.

4) Evaluating our model performance.

- First, we will select all features for model and evaluate performance.
- Now we will use RFE technique to select most important features, and then evaluate it again to show difference in R2 score.

Knowledge Graphs

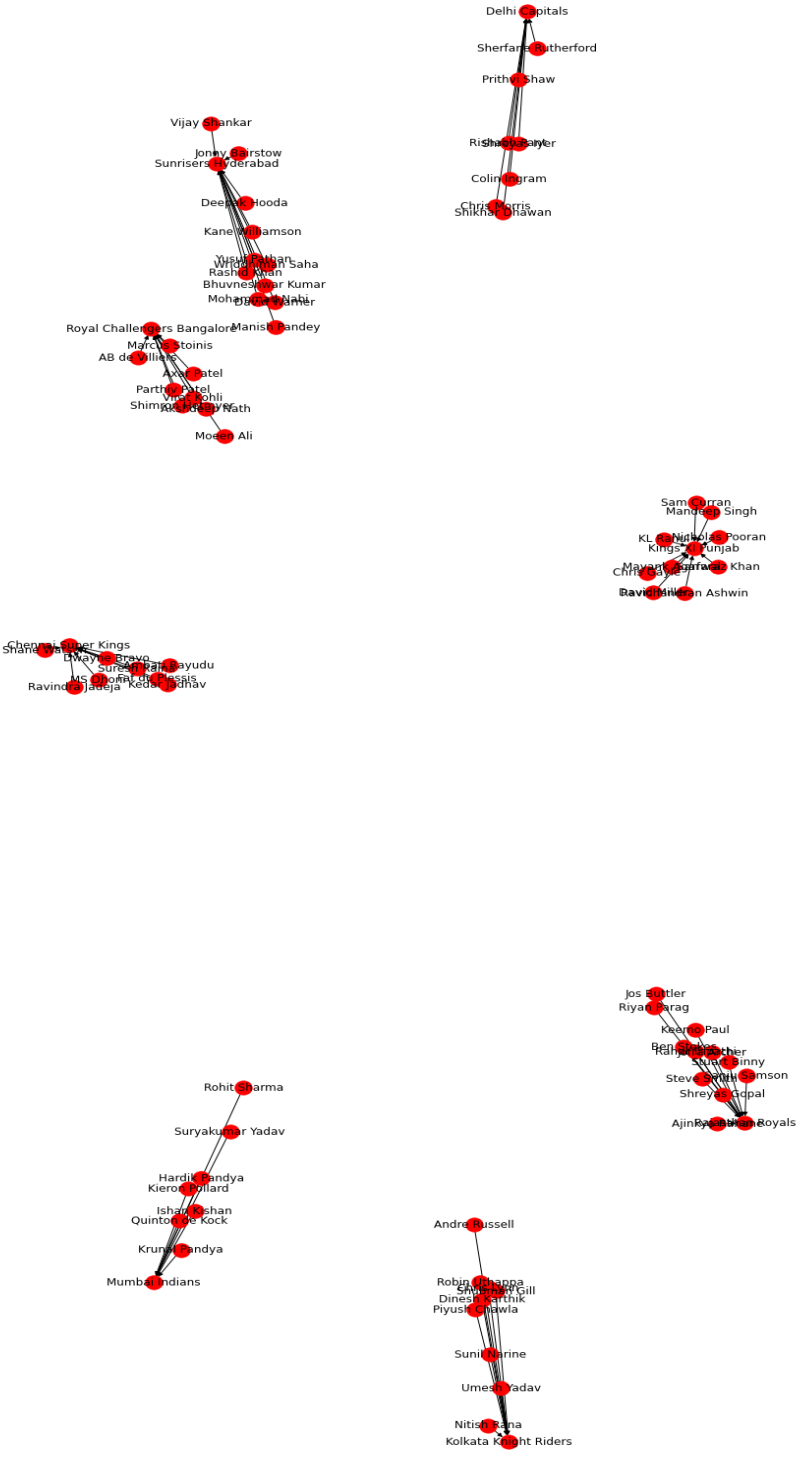
Are entity relationship graphs that show some kind of real-world modelling, it is a powerful way of representing data because Knowledge Graphs can be built automatically and can then be explored to reveal new insights about the domain.

Similarly, here we can show the different hypothesis we created in the form of graphs as all potential players belonging to some category as an example,

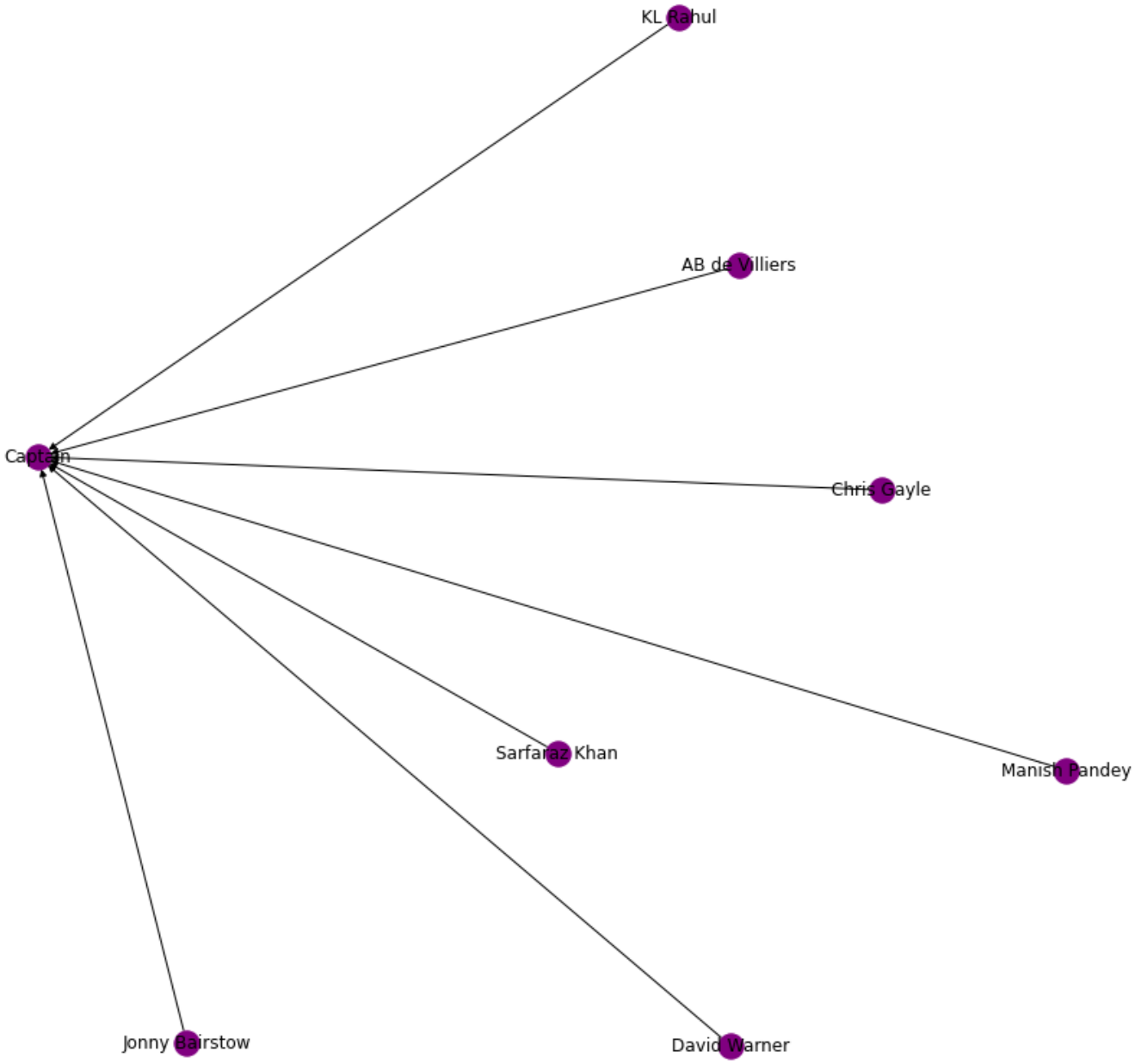
Yousuf Pathan --> is a --> bowler. {Subject} {Predicate} {Object}

Knowledge Graphs are as follows:

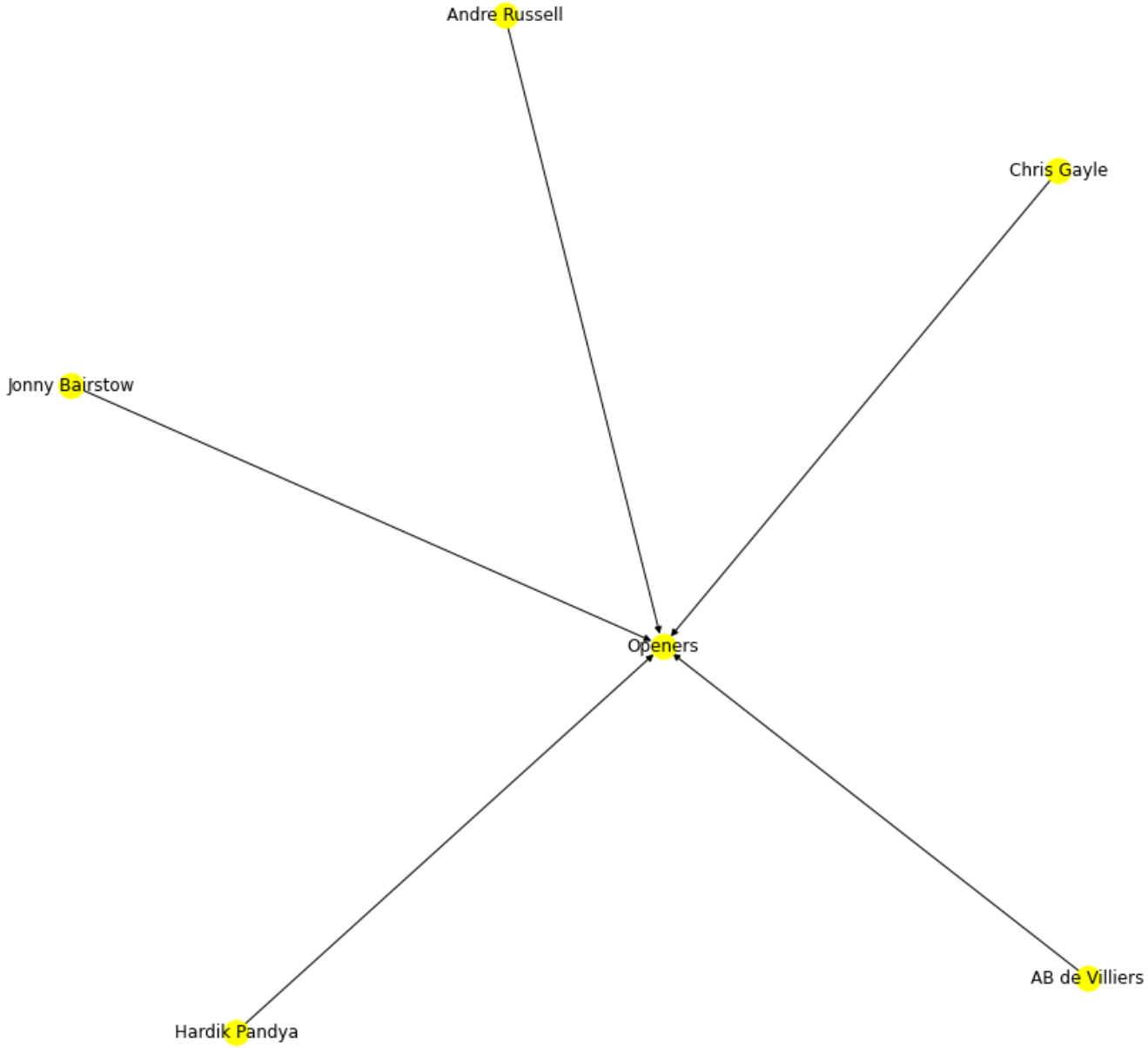
1) Players for each Team



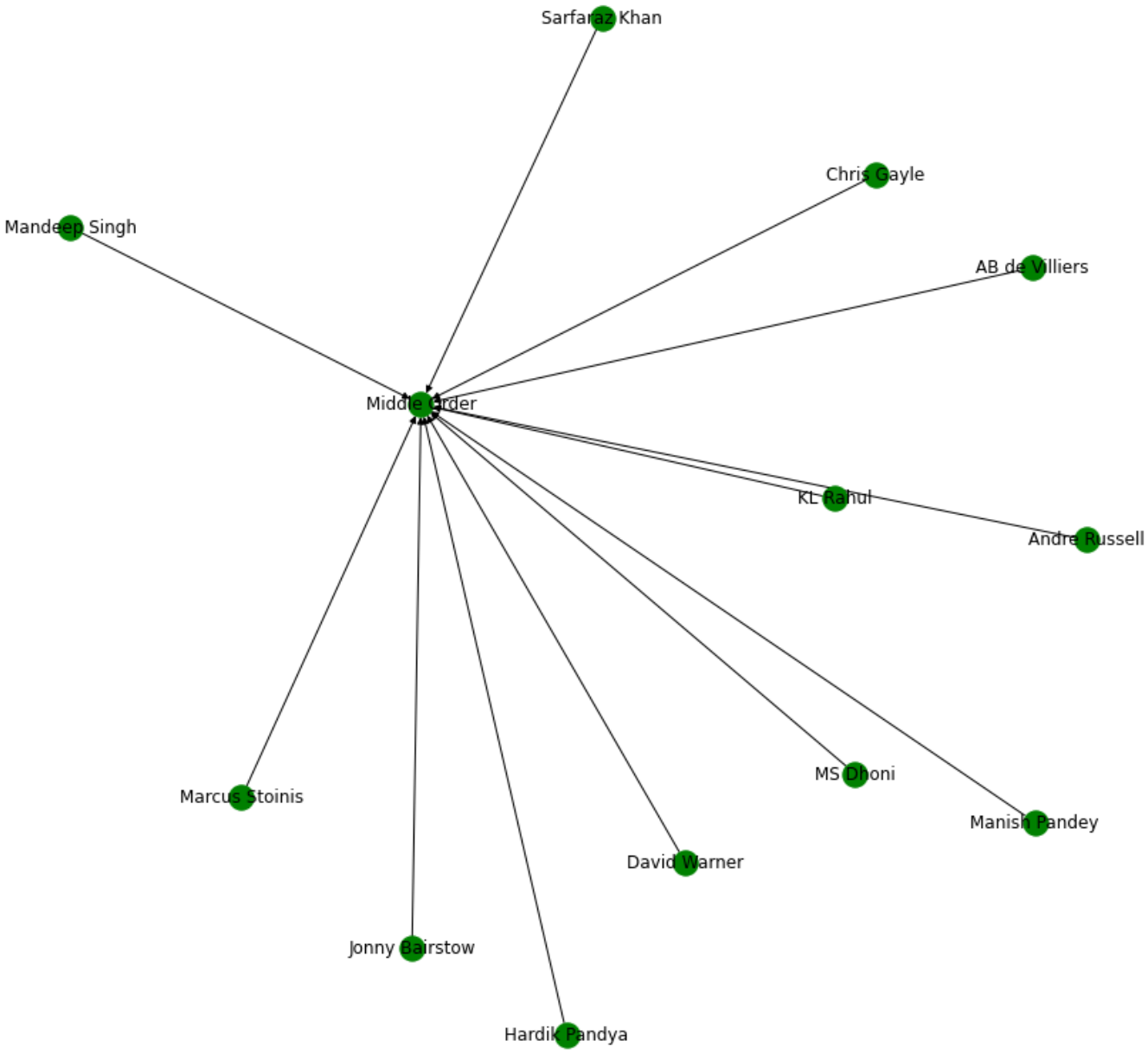
2) Potential players for Captaincy



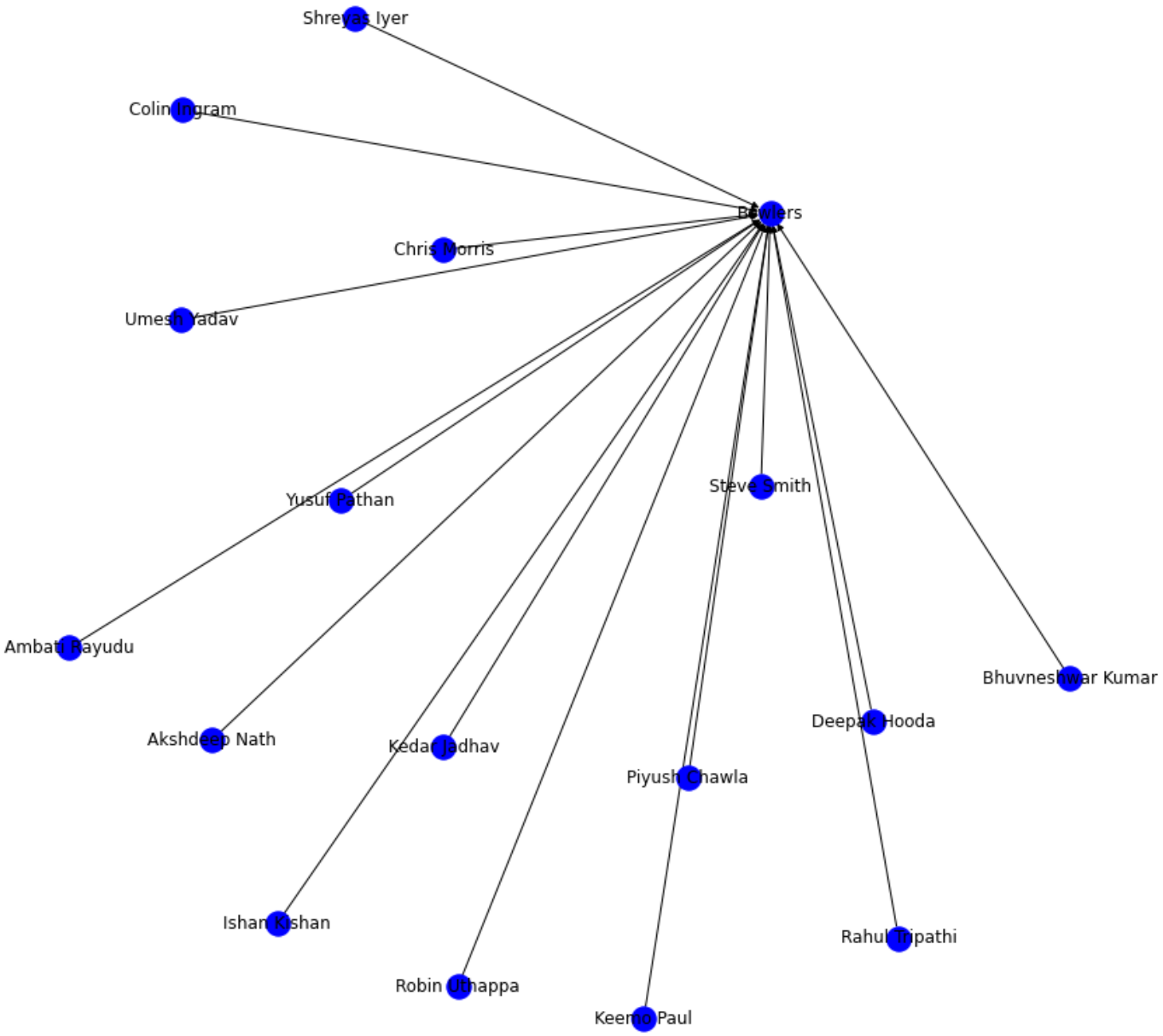
3) Potential players for Opening



4) Potential players for Middle Order



5) Potential players for Bowlers



Thank You