
COMPLEX NETWORK ANALYSIS IN CRICKET

RAG PATEL (201701008) | PURVIL MEHTA (201701073) | BHARGEY MEHTA (201701074)

1 Abstract

This report describes the different *Complex Network Analysis (CNA)* to understand the performance of each player within the team. We construct 3 networks to analyse the player-player relationship within the team. We study the community structure provided by Newman, community roles, different centrality indices such as Page-Rank (Performance Index based on partnership), Closeness, Betweenness, Clustering-Coefficient. We further analyse the performance, relative importance and the effect of removing the player from team, based on the different centrality score.

2 Introduction

The main goal of this network analysis is to understand the interactions of individual team members within the team. Cricket team is the composed of 11 players interacting with each other with same goal of winning a game. Our goal is to analyse how these players co-ordinate with each other in a match.

In cricket, series of batsmen play in partnership when they face a bowler. All the batsmen contribute to the overall score in a different manner depending upon their order. Better the opening partnership, not only do they score more runs, they also put more pressure on the opponent. The middle order partnerships are either defensive, protecting their wickets for the next day in test match, or aggressive, focusing on scoring as many runs as possible. We also show with our analysis that the batsman with higher batting average may not be always the best player in a team.

We analyse this batting partnership between players within the top 5 teams in the history of cricket like *India, Australia, England, South Africa*. We analysed the data of International Test matches between 2008 to 2020. We construct an *unweighted-undirected Batting Partnership Network (BPN)* by adding an edge between two players if they batted together in at least once match in this duration.

3 Ranking using PageRank

The traditional measure to determine the quality of a batsmen and bowlers is to rank them according to their career batting and career bowling averages. The career batting average for a batsman is defined as

$$C_{Ba} = \frac{R}{I}$$

where R is the number of runs scored by him and I is the number of innings where he batted. Hence a higher C_{Ba} implies a better batsman. The career bowling average of a bowler is defined as

$$C_{Bo} = \frac{R}{D}$$

where R is the number of runs conceded by him and D is the number of wickets. Thus a lower C_{Bo} implies a better bowler.

3.1 Page Rank

The performance score of individual player is defined as

$$p_i = (1 - q) \sum_j p_j \frac{\omega_{ij}}{s_j^{out}} + \frac{q}{N} + \frac{(1 - q)}{N} \sum_j p_j \delta(s_j^{out})$$

where ω_{ij} is the weight of the link from i to j . p_i is the page rank score assigned to player i . s_j^{out} is the out-strength ($s_i^{out} = \sum_j \omega_{ij}$). q is the free popularity that is awarded to each player and N is the total number of players in our dataset.

For *PIB* and *QIB* network, ω_{ij} is defined as $\omega_{ij} = s_j^{in} - s_i^{in}$ such that $s_j^{in} > s_i^{in}$. For *BPN* network, ω_{ij} is defined as the weight of the edge from i to j .

3.2 Performance Index for Batsman - PIB

A batsman that always plays against weak bowlers should be ranked lower (in spite of a higher C_{Ba}) as compared to a batsman that always plays against strong bowlers (in spite of a lower C_{Ba}). Hence a new measure of interaction between bowlers and batsman has been proposed. The *PIB* on a batsman against a particular bowler is defined as the

$$PIB = \frac{A_{Ba}}{C_{Bo}}$$

A graph with directed edges (from bowler to batsmen) with weights equal to *PIB* is constructed. Then a projection on the batsmen is taken where an edge exists between the 2 batsmen when they batted against at least one common bowler. The weight of the edge is equal to the difference of their in-strength in the original graph and the direction is from the low in-strength to high in-strength batsmen.

Here A_{Ba} is the career batting-average of the batsman against that bowler and C_{Bo} is the career average of that bowler. So if A_{Ba} is high then it obviously implies a better batsman but a lower C_{Bo} means a better bowler and hence the *PIB* increases to reflect the fact that the quality of the runs scored against this bowler is more.

The performance rank of any batsman should not be only depend upon the batting average but also on the type of bowler against whom the runs are scored. For example, The career batting average of *AN Cook* is 42.83 and *HM Amla* is 45.52 as per our data. Still, the page rank of *AN Cook* is higher than the *HM Amla*. Similarly *V Kohli* has scored total 7240 runs in which approximately 3200 runs are scored against the strong teams such as *Australia* and *England*. Whereas *Australia's SPD Smith* has scored 7150 runs in which approx 4200 runs are scored against the strong teams like *India* and *England*. Because of this, *SPD Smith* get better rank as compared to *V Kohli*.

3.3 Quality Index for Bowlers - QIB

Using a similar argument, the QIB for bowlers is defined as

$$QIB = D \frac{C_{Ba}}{C_{Bo}}$$

where C_{Bo} is the career bowling average of the bowler and D is the number of times the batsman is dismissed. When the bowler balls against a better batsman, the quality of the dismissals increases. The term in the denominator captures the inherent quality of the bowler since a lower C_{Bo} indicates a better bowler.

A projection similar to the case of the network constructed using *PIB* is also constructed here as well.

The career bowling average of *R Ashwin* is 24.42 and *HM-RKB Herath* is 26.46 which shows that the *R Ashwin* is the better bowler as compared to *HMRKB Herath*. But the batting average against *Herath* is 24.25 and *Ashwin* is 22.64. Since *Herath* has dismissed the batsmen with good batting average, the page rank of *Herath* is higher than the *Ashwin*.

3.4 Performance Index on the BPN

We constructed weighed-directed networks for batting partnership between batsmen in all teams where there is an directed edge between two players i.e A to B and B to A if they batted together in at least one match. The weight on the edge is equal to the fraction of runs individual player has scored in the total partnership formed by those two players. Let n_A is the total runs made by player A in partnership with player B and n is the total runs made by them then there will an directed edge between B to A with weight as

$$I_{BPN} = \frac{n_A}{n}$$

In case of *BPN*, middle order players such as *MS Dhoni* has made partnership with top order and lower order players, whereas the player with better batting average like *VVS Lakshman* has not played with different order of players. Thus, the page rank of *MS Dhoni* is higher than the *VVS Lakshman*. Refer Table 1

4 Centrality

Betweenness centrality is a measure which indicates the extent to which a node lies on a path to other nodes which in our context would mean which players are crucial to the structure of the partnerships in a team. Also, team with a 2 or 3 players with high betweenness centrality and others with low is more vulnerable to the loss of these players' wickets.

Ideally, the centralities should be uniformly divided so that it doesn't impact the structure of the network too much. We found that in the case of South Africa, top 2 players are significantly more central than even the 3rd ranked batsman. In the case of *India* and *Australia*, the centralities are more uniform if not perfectly uniform.

Closeness Centrality is a measure of how easy it is to reach a given node in the network. In cricket, closeness centrality would mean how closely a player is connected to other players. If a player has a higher closeness centrality than the player is allowed to bat at any position.

In our analysis, we found that the player *MG Johnson*, batted at position no. 8, is the 3rd closest player in the *Australian*

Type	Page Rank	C_{Ba}	Player
PIB	0.16419	42.83	AN Cook
	0.0606	45.52	HM Amla
	0.03863	42.06	LRPL Taylor
	0.0231	46.26	KS Williamson
	0.02077	51.08	AB de Villiers
QIB	0.18766	24.27	JM Anderson
	0.08336	26.64	SCJ Broad
	0.04235	21.76	DW Steyn
	0.02799	26.46	HMRKB Herath
	0.02666	24.22	R Ashwin
BPN India	0.04193	49.93	V Kohli
	0.03853	45.63	CA Pujara
	0.03607	34.6	MS Dhoni
	0.03192	41.22	VVS Laxman
	0.03151	46.48	SR Tendulkar
BPN England	0.04383	16.18	SCJ Broad
	0.03501	44.2	JE Root
	0.03233	37.46	IR Bell
	0.03225	42.83	AN Cook
	0.03012	36.3	BA Stokes

Table 1: Ranking based on Page Rank

Country	Betweenness	C_{Ba}	k_i	Player
India	0.08253	05.63	36	I Sharma
	0.07134	49.93	38	V Kohli
	0.07039	24.37	30	R Ashwin
	0.07014	34.59	34	MS Dhoni
	0.05722	45.62	35	CA Pujara
Australia	0.09047	18.41	39	MG Johnson
	0.07599	55.42	40	SPD Smith
	0.05426	45.14	37	MJ Clarke
	0.04458	08.48	36	NM Lyon
	0.04332	40.57	32	MEK Hussey

Table 2: Ranking based on Betweenness Centrality

team. This is simply because *Johnson* has scored 5-fifties and 1-hundred. In contrast to this, the player with highest closeness centrality for team India is a prominent batsman i.e. *V Kohli*. He has scored 7240 including 43-centuries and 7-double-centuries during which he had ample opportunities to form partnerships with various players.

Country	Closeness	C_{Ba}	k_i	Player
India	0.62868	49.93	38	V Kohli
	0.61256	05.63	36	I Sharma
	0.6048	45.62	35	CA Pujara
	0.59724	34.59	34	MS Dhoni
	0.5688	24.37	30	R Ashwin
Australia	0.5287	55.42	40	SPD Smith
	0.5287	45.14	37	MJ Clarke
	0.5287	18.41	39	MG Johnson
	0.51525	29.49	35	BJ Haddin
	0.51092	08.48	36	NM Lyon

Table 3: Ranking based on Closeness Centrality

We found out that the most central player does not necessarily have the highest batting average. For example, *I Sharma* from India has played from 2007 to 2019 (which coincidentally is almost the entire duration of the dataset that we used) with various player. Since it is unweighted network, there will be an edge from one player to another player if they batted even once in any match and so it fails to capture the importance of the batsman in their team. Similar argument also holds for Australian bowler *NM Lyon*. Also, *V Sehwag* from India has 46.64 batting average with low-betweenness centrality. Also, the player with high degree is not necessarily has high betweenness. For example *Harabhajan Singh* has degree 24 with low betweenness centrality of 0.0099.

5 Community Structure and roles

We analysed community structure within the each team using *Modularity Maximization* proposed by *Newman*. According to our data, we found 4 communities in the Australian team and England team and 2 communities in the Indian Team. We also calculated the *Modularity Score (Q)* of each team. We found the z-score of each player and classified the players into hubs and non-hubs. We found that the z-score is relatively high suggesting that the structure of the network is statistically significant. We expected 3 communities within the each team due to the fact that there are top, middle and lower order batsmen playing in a match. Top order batsmen form a community since they exclusively play with top order and middle order batsmen (since lower order by definition means that the players above have been bowled out). Similarly the lower order batsmen exclusively play with the middle order and the lower order batsmen. Hence we expected 3 communities with a middle order community connected with 2 others which are not connected by themselves or weakly connected.

For example, *V Kohli* and *CA Pujara* played in the same era and batted at number 3 and 4 respectively and belong to the same community according to our data. In contrast, *MS Dhoni* and *V Kohli* also have played together for many years, but they both belong to different communities. This is because they often batted at number 3 and number 7 respectively and hence don't necessarily form partnership with the same set of people.

We further analysed the z-score and participation score of each node and classified each player as hub or non-hub. We calculated the within-community degree score, called the z-score, using

$$z_i = \frac{k_i - k_{s_i}}{\sigma_{k_{s_i}}}$$

where k_i is the total degrees of player i within the community. k_{s_i} and $\sigma_{k_{s_i}}$ is the average within community degree and standard deviation of the within community degree to which i belong. The participation coefficient is given by

$$P_i = 1 - \sum_{s=1}^M \left(\frac{k_{s_i}}{k_i} \right)^2$$

where M is the total community count. Here P_i captures the connection of a player with different communities. Hence a high P_i indicates a node that "participates" in various communities and not only it's own and vice versa.

Based on this analysis, z-score denotes how player is connected to within community. A high z-score indicates that a

node is well connected to other nodes in it's community and vice versa. If z-score > 1.75 then the player is classified as hub (since he serves as a way for other players to reach his community) otherwise he is classified as non-hub. We further classified players into sub categories (structural Role) of hub and non-hub depending upon the participation score. If a player is classified as hub and having high participation coefficient would imply that the player is versatile to play at any position and has made many partnership with different players. These players are called as connector or kinless to the team or NOT having a tendency to play with a specific position in the order.

Country	Size	Mean C_{Ba}	Prominent Player
India	31	26.23	V Kohli, KL Rahul, HH Pandya
	Q = 0.2946	27	20.44 SR Tendulkar, V Sehwag, VVS Laxman
Australia	37	24.3	SPD Smith, AC Voges, UT Khwaja
	Q = 0.3173	36	21.23 AC Gilchrist, ML Hayden, RT Pointing
		19	27.71 JM Fields, SJ Coyte
		4	24.56 TA Copeland, RJ Harris
England	37	23.01	JE Root, CR Woakes, JM Bairstow
	Q = 0.3519	32	19.81 KP Pietersen, A Flintoff, AN Cook
		18	17.36 HC Knight, AE Jones
		6	22.45 CJ Jordan, LE Plunkett

Table 4: Community Structure in BPN network

- Hubs are classified in three different roles.
 - **Provincial Hubs:** (Most partnerships belong within community) $P_i \leq 0.30$
 - **Connector Hubs:** (Forms partnership with several but not all communities) $0.3 < P_i \leq 0.75$
 - **Kinless Hubs:** (Forms partnership Uniformly with all communities) $P_i > 0.75$
- Non-hubs are classified in four different categories. These players have less within community connections compared to hubs.
 - **Ultra Peripheral non-hubs:** (Forms close to 0 partnerships outside) $P_i \leq 0.05$
 - **Peripheral non-hubs:** (Doesn't form many partnerships outside) $0.05 < P_i \leq 0.62$
 - **Connector non-hubs:** (Forms partnership with few but not all communities) $0.62 < P_i \leq 0.8$
 - **Kinless non-hubs:** (Forms partnership with many communities) $P_i > 0.8$

With our analysis, we do not have any player in kinless hub category while we have *Virat Kohli* and *VVS Laxman* which appears in the connector hub category. Indian test specialist *CA Pujara* has batting average of 45.625 but it appears in peripheral non-hub category, not in connector hub.

India's *MS Dhoni*, *RG Sharma* and *AK Rahane*, England's *AJ Strauss*, *KP Pietersen* and *AN Cook* and Australia's *SR Watson*, *UT Khawaja* and *RT Pointing* have very good batting average but they did not belong to connector hub category since they mostly played at a fixed batting position. They all belong to peripheral non-hub category. We also observed that specialist openers and lower-order bowler appears in ultra-peripheral non-hub category, for example Australian openers such *AN Gilchrist* and *M Hayden* belong to the same category.

6 Assortativity

We analyse the degree-degree correlation (Assortativity Coefficient) of BPN network which measures the likeliness of the network to connect same or different degree node.

$$A = \frac{1}{\sigma_q^2} \sum_{jk} jk(e_{jk} - q_j q_k)$$

where $q_k = \sum_j e_{jk}$ and $\sigma_q^2 = q_k \sum_k k^2 - (\sum_k k q_k)^2$. Here e_{jk} is the probability that a randomly chosen edge has nodes with degree j and k at either end. If ($A > 0$) then the network is said to be assortative and if $A < 0$ then the network is said to be disassortative.

Country	India	Australia	England	S Africa
Assortativity	-0.12769	-0.03183	-0.17353	-0.2698

Table 5: Assortativity countrywise

For the team India, we find that $A = -0.128$ suggesting that the network is **weakly disassortative**. Similarly for team Australia, $A = -0.03$. It is also expected because partnerships between any one player to various other players are only possible when other players get dismissed early. It is very unlikely that all the players get dismissed very early or two players never get dismissed in most of the matches. Thus partnership record of one player is likely to be different from a randomly chosen other player. Thus, we get weakly disassortative network.

7 Clustering Coefficient

We analysed the clustering coefficient for BPN network. C_i is the clustering coefficient defined as the fraction of edges in the neighbourhood of node i to the maximum possible edges in the neighbourhood of i . The overall clustering coefficient is given by

$$C = \frac{\sum_{i=1, k_i > 1}^N C_i}{N}$$

Clustering coefficient measures the density of inter connected nodes. The density of the graph is defined as the fraction of total edges to the total possible edges. If clustering coefficient is more than the density suggests that the graph is more interconnecting forming small local clusters in the network. For the BPN of India, we find that the density is 0.17 and the C is 0.63 indicating that the team has some sort of clusters of top, middle and lower order players. Similar argument hold for Australian team with density = 0.12 and $C = 0.68$.

8 Conclusion

We analysed the data of international test matches across different countries from 2008 to 2020 on three graph structures, *PIB*-network, *QIB*-network and *BPN*. Our analysis not only provides

Country	India	Australia	England	S Africa
Coefficient	0.63381	0.68152	0.6993	0.71022

Table 6: Clustering Coefficient countrywise

the information about the team performance but also highlight individual player's performance in the team.

The batting average or bowling average is not the only measurement of player's popularity or rank. Thus we introduce a new ranking scheme based on the *Page-Rank* algorithm which also takes into account the bowlers' strength against which a batsman has scored and batsman's strength whose wicket a bowler has taken. We found that the closeness centrality denotes the most crucial players in the team whereas betweenness centrality denotes the flexibility of a player in the team. In conclusion, career averages are not the true reflection of a player's rank.

The existence of communities in the BPN network suggest that there are patterns in the way batsmen form a partnership. We also observed that the somewhat contemporary players and players who play at the same position belong to the same community. We observed that *AN Cook*, considered to be one of the best test player with more than 10000 test runs till date is not in connector-hub category due to fact that he majorly used to open in a match.

This helps us in understanding which player performs better when partnering with some other player, which in turn produces a balanced team. Thus, our *complex network analysis(CNA)* provide us a quantitative way to quantify the overall performance of a batsman or bowler in a tournament.

In a tournament, we can apply CNA to determine the contenders for the *Man-of-the-Match* by considering network of that particular match and *Man-of-the-Series* award by considering whole network of the tournament. Also, in a national team selection against tour of some other country, we can form network of our team and that other team in order decide the squad as well as *Playing-11* for particular matches.

Also, we analysed that the players who has most closeness centrality score or the high page rank can be considered as crucial player and removal of these players may decrease the percentage of winning a match.[1] [2]

References

- [1] Satyam Mukherjee. "Complex Network Analysis in Cricket: Community Structure, Player's Role and Performance Index". In: *Advances in Complex Systems* 16.08 (2013). DOI: 10.1142/S0219525913500318.
- [2] Satyam Mukherjee. "Quantifying individual performance in Cricket — A network analysis of batsmen and bowlers". In: *Physica A: Statistical Mechanics and its Applications* 393 (2014). DOI: <https://doi.org/10.1016/j.physa.2013.09.027>.