

Network Analysis of Batting Partnerships in Cricket

Nisarg Patel — 201501134 and Aayush Agrawal — 201501139
*Dhirubhai Ambani Institute of Information & Communication Technology,
Gandhinagar, Gujarat 382007, India
IT-454, Introduction to Complex Networks*

This paper applies the network methods for analyzing the interaction between the members of a team. Here we take into account the interaction of batting partnerships in International Cricket Matches. We took help of the paper[1] and implemented it to get the different results mentioned in it. We calculate various network parameters for this network as studied[2] and find out the behaviour of the network. We generate the batting partnership network for each country and apply centrality measures to deduce inferences like the players that participate the most in partnerships, not necessarily are the most important or central players and the most central players may not have the highest batting average. We try to differentiate eras of players by community detection method and find out the roles played by each of the player in the team.

I. INTRODUCTION

Cricket is an outdoor sports game played between two teams in which each team has 11 players. Cricket is played on a mostly circular ground having a pitch of 20 metres and with a wicket at each end each comprising 2 bails balanced on 3 stumps. Each match comprises on two innings. The objective of the game is simple, a team should score more number of runs than its opposition while the opposite teams restricting the score and dismissing the batsmen. There are 3 prominent forms of cricket - 'Test', 'One Day International' and 'T20 International'[3]. In this paper, we take the network for Test cricket because it is considered as the highest standard of cricket. Each test match is played for 4 or 5 days. It is the longest format of the game, played for more than 140 years. Each match comprises of 4 innings, two innings per team. In each innings batsmen try to score as many runs as possible within their quota of 10 wickets and after the end of 4 innings, the team with maximum number of runs wins the match.

In this paper we consider the Test matches played by the top six historical teams- India, South Africa, England, New Zealand, Australia, West Indies. All renowned cricket team have different approach to their game-play, batting line-up and bowling order. Teams like Australia and England depends on fast bowlers while subcontinental teams like India or West Indies depend on slow bowlers. In this paper we apply tools to understand the importance of each player within a team and the methods of link formation between the batsmen.

Thus, even though cricket is a team sport, we analyze an individual player's effect on the game and team dynamics based on its interaction with other players of the team. We apply network analysis on such set of interactions which lead to formation of a connected network, primarily batting partnership interactions in a team.

II. NETWORK CONSTRUCTION

We analyze the data of batting partnership (available from cricinfo website[4]) in Test cricket between 1877 and June 2018. Two batsmen are connected if they formed a batting partnership in at least one match. An undirected and unweighted batting partnership network is generated for each country. Also to note that partnership networks are always restricted to countries. Two batsmen of different countries do not bat together.

III. ANALYSIS OF NETWORK

In this section we analyze the batting partnership as a complex network of interaction of two batsmen. We calculate various measures for network analysis for the top countries. The results are shown as table in Fig.(1). We observe that the average degree of all the teams are similar except South Africa. Degree is one of the centrality measures of the network. It reflects how many partnerships a batsmen is involved in. So if a batsmen has high betweenness centrality it would mean that he/she is vital to hold partnerships. However, if degree centrality of such a batsmen is less, the number of such partnerships would be less. This would result in the team being uneven since number of important partnerships are less. We can deduce that a team with high betweenness to average degree ratio could have too dependence on some player. On average, India's ratio for the entire dataset comes to be 0.0006, however during 1995 - 2004, it comes out to be 0.0022 which may highlight a high dependence on some players rather than a whole team near 2000s.

Country	Number of Nodes	Number of Edges	Average Degree	Clustering Coefficient	Assortative Coefficient	Average Shortest Path Length
India	285	1882	13.21	0.62	-0.16	3.48
South Africa	332	1741	10.49	0.62	-0.09	4.28
England	680	4557	13.4	0.58	-0.07	4.68
New Zealand	272	1901	13.98	0.6	-0.05	3.99
Australia	449	3046	13.57	0.61	-0.09	5.4
West Indies	311	2045	13.15	0.6	-0.12	3.81

FIG. 1: Networks measures for different teams.

The degree distributions of top teams are shown in Fig.(2). From this, we can see that the degree distribution neither follows a normal curve, nor a power law. We can also see that the degree distribution of South Africa decays faster as compared to other countries. That is because, South Africa faced was banned from international cricket from 1971 to 1991. Thus players were not able to form partnerships with new players more frequently than compared to other team.

The clustering coefficient(C_i) of a node i is defined as ratio of number of links shared by its neighboring nodes to the maximum number of possible links among them. The average clustering coefficient(C) of a network is the mean of clustering coefficients of all the nodes. The average clustering coefficient captures the global density of interconnected nodes in a network. We can see from 1 that all of the countries have nearly same value of clustering coefficient(≈ 0.60). This indicates that the networks are highly clustered. We generated an Erdos-Renly random graph[5] ($G(n, p)$) with average number of nodes($n = 388$) and average number of edges($m = 2529$). Thus the probability of an edge between two vertices:

$$p = \frac{2m}{n(n-1)} \quad (1)$$

Thus we get from Eq.(1) $p = 0.034$. Thus generating the random graph $G(n, p)$, we see that the correlation coefficient($C = 0.03$) is very lower than the original network.

The assortativity coefficient(A) measures the tendency of a network to connect vertices with the same or different degrees. If $A > 0$ ($A < 0$) the network is said to be assortative (disassortative). It has been observed that social networks are assortative and technological and biological networks are disassortative. We find that all the countries have $A \approx -0.10$ indicating the weak disassortativity of the batting partnership networks.

We evaluate the average shortest path length L between a given node and all other nodes of the network. We observe that the average shortest path length of these networks are of the same order of the random graph generated by the ER model ($L = 2.62$). Thus the networks display small-world properties.

From Fig(3), it is visible that 'A Ward' of England has the highest betweenness centrality however, it has lowest of averages and degree among the others who have high betweenness. On the other hand Sir DG Bradman who has the highest average does not have the highest betweenness. Thus, it is not necessary that a player's centrality and batting average be proportional. Concluding this we can state that involving in important partnerships does not convert to higher batting averages and so runs in a partnership cannot be equally weighted edges.

Community Structures and Roles:

Communities are groups of nodes are more densely connected to each other than the remaining of the network. Thus communities in a network are the densely connected

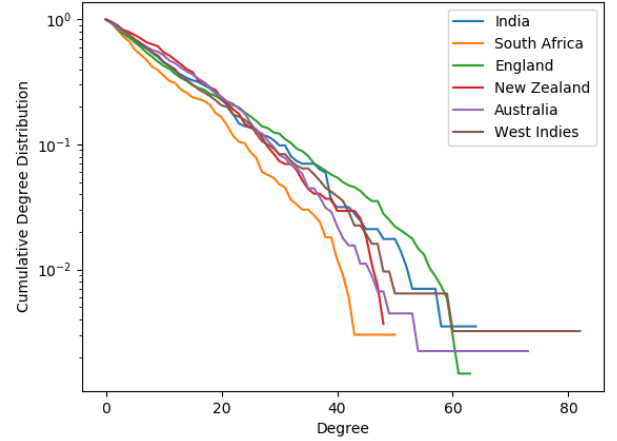


FIG. 2: Degree Distributions for different teams.

Country	Betweenness Centrality	Degree	Batting Avg	Player
India	0.31	48	31.05	N Kapil Dev
	0.3	33	35.5	P Roy
	0.22	65	53.78	SR Tendulkar
	0.11	51	51.12	SM Gavaskar
	0.1	52	17.77	A Kumble
South Africa	0.42	32	32.85	JP Duminy
	0.3	51	40.77	HW Taylor
	0.24	38	48.88	B Mitchell
	0.14	25	43.66	EAB Rowan
	0.12	41	29.78	AW Nourse
England	0.24	15	37.46	A Ward
	0.16	51	32.75	APE Knott
	0.11	60	42.58	GA Gooch
	0.11	52	22.53	JE Emburey
	0.09	61	58.45	WR Hammond
New Zealand	0.31	44	32.22	BE Congdon
	0.20	40	27.16	Sir RJ Hadlee
	0.19	34	37.82	JG Wright
	0.17	46	40.1	B Sutcliffe
	0.16	46	26.28	AC Parore
Australia	0.41	35	48.41	RN Harvey
	0.33	37	46.81	RB Simpson
	0.3	35	99.94	DG Bradman
	0.19	38	31.83	GM Wood
	0.17	42	22.65	WAS Oldfield
West Indies	0.23	43	60.83	GA Headley
	0.18	83	51.37	S Chanderpaul
	0.18	37	6.97	LR Gibbs
	0.17	30	56.68	CL Walcott
	0.16	38	42.29	DL Haynes

FIG. 3: Betweenness centrality sorted player statistics.

within the group and sparsely connect between different groups.

We perform the community detection based on the Girvan-Newman algorithm. The Girvan-Newman algorithm detects communities by progressively removing edges from the original network. The connected components of the remaining network are the communities. Instead of trying to construct a measure that tells us which edges are the most central to communities, the Girvan-

Newman algorithm focuses on edges that are most likely "between" communities. But it also runs slowly, taking time $O(m^2n)$ on a network of 'n' vertices and 'm' edges, making it impractical for networks of more than a few thousand nodes

We found 4 communities in the giant component (GC) of the BPN for various teams. Table in Fig.(4) shows the mean batting averages of each communities along with the 95% confidence interval of each community for each team. We find that most of the 95% confidence intervals of each communities for a country overlap meaning that there is no statistically significant difference in the mean bating average between communities. We get the most important players of each community by calculating the within-community degree z-score[6] given by the following equation:

$$z_i = \frac{k_i - k_{s_i}}{\sigma_{k_{s_i}}} \quad (2)$$

where k_i is the number of edges of node i to other nodes in its community s_i , k_{s_i} is the average of k_i over all of the nodes in s_i and $\sigma_{k_{s_i}}$ is the standard deviation of k_i in s_i . Next we distinguish nodes based on their connections to nodes belonging to different communities. As defined in Ref. [6], the participation coefficient P_i of a node i is given as

$$P_i = 1 - \sum_{s=1}^M \left(\frac{k_{i_s}}{\sigma_{k_i}} \right)^2 \quad (3)$$

where k_{i_s} is the number of links of node i to nodes in community s and k_i is the degree of the node i . And then, based on the z -score we identify the hubs and non-hub nodes. If $z \geq 2.5$ the nodes are classified as hubs while non-hubs are identified with $z \leq 2.5$. The hubs and non-hubs are further classified based on the participation coefficient P . Non-hubs are divided into four roles :

- (R1) "ultra peripheral nodes" - Nodes with all their edges within their community ($P \leq 0.05$)
- (R2) "peripheral nodes" - Nodes with most of their links within their community ($0.05 < P \leq 0.62$)
- (R3) "non-hub connector nodes" - Nodes with many links to other community ($0.62 < P \leq 0.80$)
- (R4) "non-hub kinless nodes" - Nodes with links homogeneously distributed among all communities ($P > 0.80$)

Hubs are divided into three roles :

- (R5) "provincial hubs" - Hub nodes with majority of edges within their community ($P \leq 0.30$)
- (R6) "connector hubs" - Hub nodes with many connections to most of other communities ($0.30 < P \leq 0.75$)

- (R7) "kinless hubs" - Hub nodes with links homogeneously distributed among all communities ($P > 0.75$)

We apply the role-classification approach to the communities detected from Girvan-Newmans algorithm. For each player in the BPNs of different teams we calculate the within-community degree z_i and P_i . We computed the categories for Indian players. We found that there are no players who fall in the category of 'R3', 'R4' and 'R7'. There are only 6 players in hubs category. Out of which only "Sachin Tendulkar" is the connector hub of India. From England 5 players are from the connector hub and the legendary batsman Don Bradman of Australia is not a connector hub. Thus most of the connector hubs are the batsman playing at various positions and having a long career. Thus the roles played by a batsman in the network gives us an idea about the batting position of the player.

IV. CONCLUSIONS

To summarize, we studied the network properties of batting partnership network of different teams in the history of test cricket(1877-2018). We found that the batting partnership network displays small world properties and are weakly disassortative. We also found out that the player with high average is not necessarily the player with high betweenness or degree. We found out the communities within the network using Girvan-Newman for each country and mathematically derived the important players in each community of the team. We also found out the roles played by various node within a network and its role in the game.

Country	Size	Mean	95% CI	Prominent players
India	97	21.49	[18.40, 24.57]	['R Dravid', 'SR Tendulkar', 'VVS Laxman']
	84	20.17	[17.22, 23.10]	['SM Gavaskar', 'N Kapil Dev', 'DB Vengsarkar']
	69	18.67	[15.42, 21.92]	['PR Umrigar', 'VL Manjrekar', 'GS Ramchand']
	35	18.9	[13.97, 23.83]	['S Wazir Ali', 'L Amarnath', 'VM Merchant']
South Africa	98	22.2	[19.35, 25.05]	['JH Kallis', 'SM Pollock', 'MV Boucher']
	118	19.72	[17.06, 22.38]	['JHB Waite', 'AD Nourse', 'DJ McGlew']
	96	13.94	[11.64, 16.25]	['AW Nourse', 'HW Taylor', 'JH Sinclair']
	20	6.2	[3.01, 9.40]	['AB Tancred', 'FW Smith', 'WH Milton']
England	245	20.36	[18.83, 21.89]	['DI Gower', 'AJ Stewart', 'GA Gooch']
	147	21.81	[19.28, 24.33]	['WR Hammond', 'FE Woolley', 'EH Hendren']
	136	21.32	[18.93, 23.71]	['MC Cowdrey', 'TW Graveney', 'TG Evans']
	152	19.69	[17.41, 21.97]	['J Briggs', 'TW Hayward', 'JT Tyldesley']
New Zealand	116	21.79	[19.46, 24.12]	['DL Vettori', 'BB McCullum', 'SP Fleming']
	65	21	[17.77, 24.22]	['BE Congdon', 'Sir RJ Hadlee', 'MG Burgess']
	66	16.41	[13.31, 19.50]	['B Sutcliffe', 'JR Reid', 'AR MacGibbon']
	25	21.01	[14.53, 27.49]	['ML Page', 'CS Dempster', 'GL Weir']
Australia	128	23.82	[21.05, 26.60]	['SR Waugh', 'RT Ponting', 'MJ Clarke']
	107	22.26	[19.59, 24.93]	['KD Walters', 'RW Marsh', 'AR Border']
	162	20.07	[17.55, 22.59]	['WAS Oldfield', 'JM Blackham', 'SE Gregory']
	52	23.31	[19.65, 26.97]	['RN Harvey', 'RR Lindwall', 'KR Miller']
West Indies	64	23.45	[19.50, 27.39]	['CH Lloyd', 'AI Kallicharran', 'HA Gomes']
	126	19.52	[17.18, 21.86]	['S Chanderpaul', 'BC Lara', 'MN Samuels']
	73	23.13	[18.87, 27.39]	['FMM Worrell', 'GS Sobers', 'ED Weekes']
	48	20.68	[15.59, 25.76]	['GA Headley', 'LN Constantine', 'GC Grant']

FIG. 4: Community roles and prominent players

- [1] S. Mukherjee *Complex Network Analysis in Cricket : Community structure, player's role and performance index*, 2012
- [2] M.E.J. Newman, *Networks: An Introduction*, Oxford University Press, 2012
- [3] <http://www.icc-cricket.com>

- [4] <http://www.espnccricinfo.com>
- [5] P. Erdos and A. Renyi. On the evolution of random graphs. Math. Inst. Hung. Acad.Sci. 5, 7 (1960)
- [6] R. Guimer'a, L. Amaral J. Stat. Mech.: Theor. Exp. P02001 1-13 (2005)

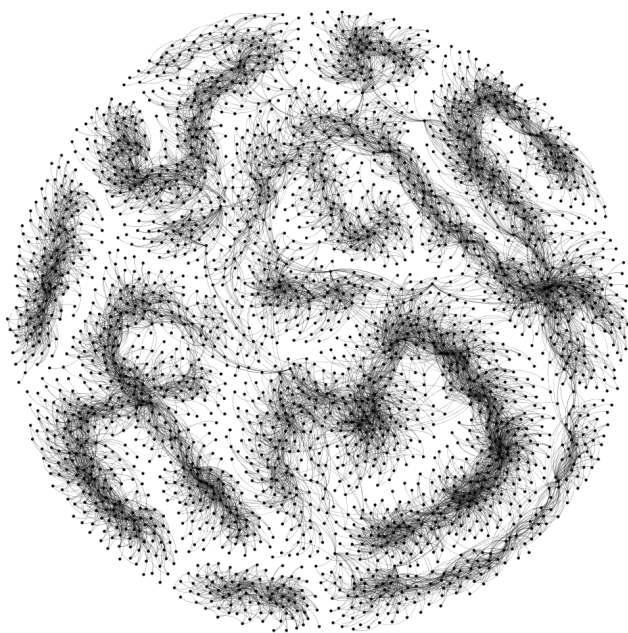


FIG. 5: All players' batting partnership network. The dense black clouds represent each team. You can find some of these dense clouds connected. This is due to the fact that some cricketers have played for different countries thus connecting them.

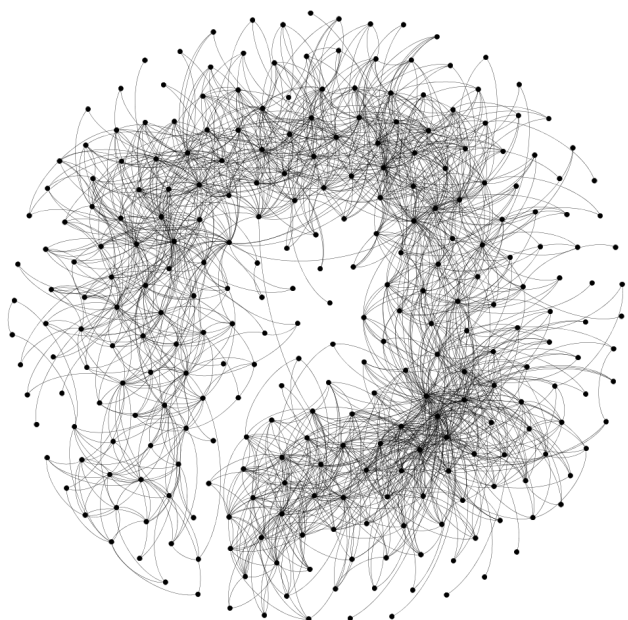


FIG. 6: The batting partnership network for all Indian players