Complex Network Approach to Cricket Matches – Expert Opinion and Individual Ranking

Satyam Mukherjee[1,2]
[1]Kellogg School of Management, 2145 Sheridan Rd, Evanston, IL 60208
[2]Northwestern Institute on Complex Systems, 600 Foster St, Evanston, IL 60208

Abstract

We demonstrate in this paper the use of tools of complex network theory to quantify the performance of players in Cricket matches. Using partnership data made available by Cricinfo during the Indian Premier League 2014, we generate batting partnership network (BPN) for each team, in which nodes correspond to batsmen and links represent runs scored in partnerships between batsmen. The resulting network display a graphical summary of the pattern of run-scoring by each team, which helps us in identifying potential weakness in a batting order. We use different centrality scores to quantify the performance, relative importance and effect of removing a player from the team. The individual ranking is then compared with Man of the Match, based on expert opinions. Our analysis can be extended to study the style of play of different teams – some teams display distributed performance among the players, while other teams display centralized performance, where team performance is centralized around a particular player.

Key Words: Network Analysis, Cricket, Individual performance, Man of Match

1. Introduction

In recent years there has been an increase in study of activities involving team sports. Time series analysis has been applied to football (1,2), baseball (3,4), basketball (5,6,7) and soccer (8,9). Again, a model-free approach was developed to extract the outcome of a soccer match (10). The study of complex networks has attracted a lot of research interests in the recent years (11,12).

The tools of complex network analysis have previously been applied to sports. Such as a network approach was developed to quantify the performance of individual players in soccer (13) and football (14) Network analysis tools have been applied to football (15), Brazilian soccer players (16) and Asian Go players (17). Successful and un-successful performance in water polo has been quantified using a network-based approach (18). Head-to-head matchups between Major League Baseball pitchers and batters was studied as a bipartite network (19). More recently a network-based approach was developed to rank US college football teams (20), tennis players (21), Cricket teams and captains (22) and Cricket players (23).

Social network analysis (SNA) allows researchers to explore the intra-group and inter-group relations between players, thus providing an informal relation between various players. Such an analysis provides insight about the pattern of interaction among players and how it affects the success of a team. This article points out that topological relations

between players need to be explored in order to better understand individuals who play for their teams. SNA is well suited to investigate the complex relations between team members (24). Such an approach to cricket at the microscopic level, form a basis of elucidating the individual importance and impact of a player.

Cricket is a game played in most of the Commonwealth countries. There are three versions of the game - Test, One Day International (ODI) and Twenty20 (T20) formats. Test cricket is the longest format of the game dating back to 1877. Usually it lasts for five days involving 30 - 35 hours. Shorter formats, lasting almost 8 hours like ODI started in 1971 and during late 2000 ICC introduced the shortest format called T20 Cricket which last approximately 3 hours (22).

In the recent years there has been a growth in research interests on Cricket. Multiple linear regression techniques were used to determine relative batting and bowling strengths in ODI and Test cricket (25) Again, the distribution of runs scored in Test cricket followed a negative binomial distribution (26). Recently it was shown that the score dynamics in cricket is an anomalous diffusive process (27). However these studies do not focus on the style of play adopted by different teams. In this paper, we apply tools from SNA to understand a Cricket team's style of play, relative importance of each player within a team and effect of removing a player from the team. We also compare the individual performance based on SNA with Man of the Match, judged by Cricket experts.

1.1 Cricket as a Complex Network

The central goal of network analysis is to capture the interactions of individuals within a team. Teams are defined as groups of individuals collaborating with each other with a common goal of winning a game. Within teams, every team member co-ordinate across different roles and the subsequently influence the success of a team. In the game of Cricket, two teams compete with each other. Although the success or failure of a team depends on the combined effort of the team members, the performance or interactive role enacted by individuals in the team is an area of interest for ICC officials and fans alike. We apply network analysis to capture the importance of individuals in the team. The game of Cricket is based on a series of interactions between batsmen when they bat in partnership or when a batsman is facing a bowler. Thus a connected network among batsmen arises from these interactions.

Cricket is a bat-and-ball game played between two teams of *11* players each. The complex nature of this sport demarcates it from other sports like soccer or baseball. For example, although baseball appears to be similar to Cricket, they share notable differences in the manner in which they are played. In cricket there are many factors that determine the outcome of a game. For example in a cricket line-up, the openers lay the foundation of an innings, by seeing-off the new ball and playing a sheet anchor role. The lower-middle depending on the score either tries of make as many runs as possible or try to save their wickets. From the perspectives of network theory, while baseball pitchers and batters could be investigated as a bipartite network (19), in Cricket one cannot represent batsmen and bowlers as a bipartite graph. This is due to the fact that a bowler has to bat once the top order batsmen are dismissed. Again, sometimes a batsman could be used as a part time bowler.

The team batting first scores as many runs as possible, while the other team bowls and fields, trying to dismiss the batsmen. At the end of an innings, the teams switch between batting and fielding. In Cricket two batsmen always bat in partnership. Usually the

opening partnerships are responsible to face the 'new ball' and score runs at the same time. Middle order partnerships are entrusted with consolidation of the innings. Lower order partnerships are much smaller than the opening partnerships. The outcome of a match depends on the batting partnerships between batsmen. Large partnerships not only add runs on the team's score, it may also serve to exhaust the tactics of the fielding team. Again, the concept of partnership become vital if only one recognized batsman remains. It is therefore important to identify the key players in a team by constructing network of batting partners (28). We analyze the data of batting partnership of teams in Indian Premier League (IPL) 2014. Two batsmen are connected if they form a batting partnership in the match. A directed and weighted batting partnership network is generated for each team.

2. Performance Index

We generate weighted and directed networks of batting partnership for all teams, where the weight of a link is equal to the fraction of runs scored by a batsman to the total runs scored in a partnership with another batsman. Thus if two batsmen *A* and *B* score *n* runs between them where the individual contributions are n_A and n_B, then a directed link of weight n_A/n from *B* to *A*. In Figure 1 we show an example of weighted and directed batting partnership network for two teams – Kings XI Punjab and Kolkata Knight Riders (KKR). We quantify the batting performance of individual players within a team analyzing the centrality scores - in-strength, PageRank score, betweenness centrality and closeness centrality.

2.1 In-strength

For the weighted network the in-strength is given by

$$s_i^{in} = \sum_{j \neq i} w_{ji}$$

where w_{ji} is given by the weight of the directed link.

2.1 PageRank

We quantify the importance or `popularity' of a player with the use of a complex network approach and evaluating the PageRank score. Mathematically, the process is described by the system of coupled equations

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

where w_{ij} is the weight of a link and s_j^{out} is the out-strength of a link, p_i is the PageRank score assigned to team *i* and represents the fraction of the overall ``influence'' sitting in the steady state of the diffusion process on vertex *i* (21) , *q* is a control parameter that awards a `free' popularity to each player and *N* is the total number of players in the network.

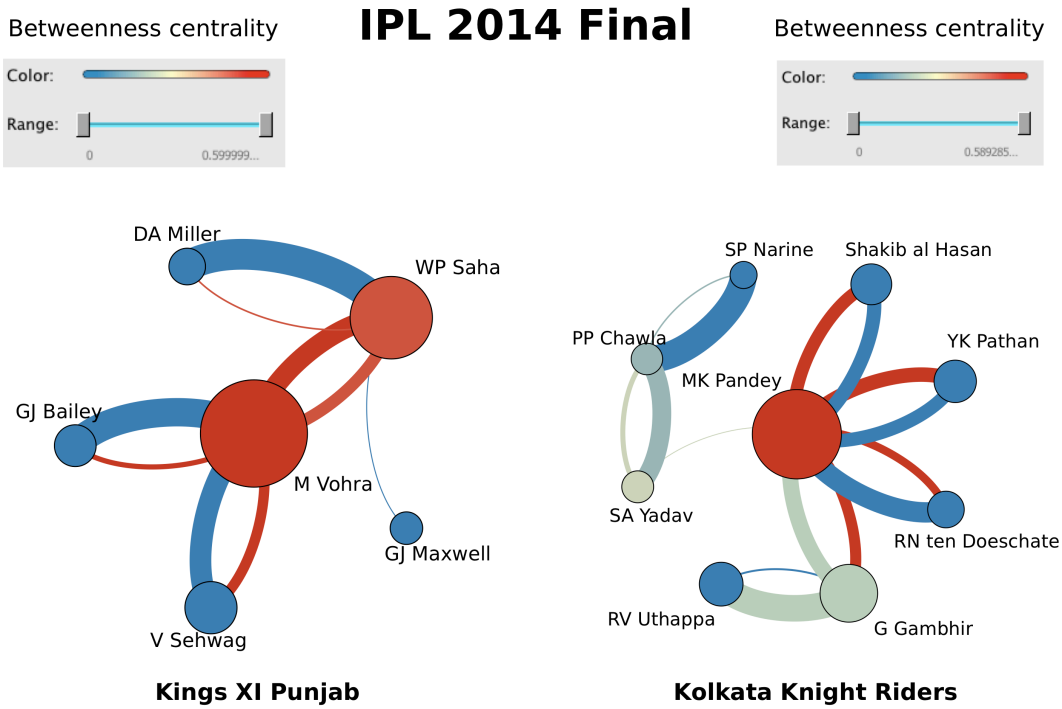

Figure1: Batting partnership network of Kings XI and KKR The colour of nodes are according to the betweenness centrality. The size of the nodes is proportional to the PageRank. The color of the edge is according to the color of the source node.For example if we see the connection between RV Uthappa and G Gambhir, Gambhir has a larger share of the runs they scored (5/6). For KKR, MK Pandey has the highest betweenness, PageRank and In-strength. It is to be noted that the color and size of the nodes are all with respect to the individual team and not normalized according to absolute values of teams.

The first term on the Right Hand Side (RHS) of the equation represents the portion of the score received by node i in the diffusion process obeying the hypothesis that nodes redistribute their entire credit to neighbouring nodes. The second term on the RHS stands for a uniform redistribution of credit among all nodes. The thirds term on the RHS of the equation serves as a correction in the case of the existence nodes with null out-degree, which otherwise would behave as sinks in the diffusion process. It is to be noted that the PageRank score of a player depends on the scores of all other players and needs to be evaluated at the same time. To implement the PageRank algorithm in the directed and weighted network, we start with a uniform probability density equal to $1/N$ at each node of the network. Next we iterate through Equation2 and obtain a steady-state set of PageRank scores for each node of the network. Finally, the values of the PageRank score are sorted to determine the rank of each player. According to tradition, we use a uniform value of $q=0.15$. This choice of q ensures a higher value of PageRank scores (21).

Another performance index is *betweenness centrality*, which measures the extent to which a node lies on a path to other nodes. In cricketing terms, *betweenness centrality* measures how the run scoring by a player during a batting partnership depends on another player. Batsmen with high *betweenness centrality* are crucial for the team for scoring runs without losing his wicket. These batsmen are important because their dismissal has a huge impact on the structure of the network. So a single player with a high *betweenness*

centrality is also a weakness, since the entire team is vulnerable to the loss of his wicket. In an ideal case, every team would seek a combination of players where *betweenness* scores are uniformly distributed among players. Similarly the opponent team would seek to remove the player with higher *betweenness centrality*.
Closeness centrality measures how easy it is to reach a given node in the network. In cricketing terms, it measures how well connected a player is in the team. Batsmen with high closeness allow the option for changing the batting order depending on the nature of the pitch or match situation.

3. Results

Match-Teams	PageRank	In-strength	Betweenness	Closeness	Man of Match
Kolkata	YK Pathan	**JH Kallis**	**JH Kallis**	**JH Kallis**	**JH Kallis**
	SA Yadav	YK Pathan	YK Pathan	G Gambhir	
	JH Kallis	SA Yadav	Shakib Hasan	RV Uthappa	
Mumbai	AT Rayudu	AT Rayudu	AT Rayudu	KA Pollard	
	RG Sharma	AP Tare	KA Pollard	CM Gautam	
	AP Tare	KA Pollard	AP Tare	CJ Anderson	
Match-Teams	**PageRank**	**In-strength**	**Betweenness**	**Closeness**	**Man of Match**
Bangalore	PA Patel	PA Patel	V Kohli	V Kohli	YS Chahal
	V Kohli	V Kohli	PA Patel	PA Patel	
	NJ Maddinson	NJ Maddinson	Yuvraj Singh	Yuvraj Singh	
Delhi	M Vijay	M Vijay	M Vijay	M Vijay	
	JP Duminy	MK Tiwary	JP Duminy	JP Duminy	
	MK Tiwary	JP Duminy	MK Tiwary	KD Karthik	
Match-Teams	**PageRank**	**In-strength**	**Betweenness**	**Closeness**	**Man of Match**
Punjab	**GJ Maxwell**	**GJ Maxwell**	**GJ Maxwell**	**GJ Maxwell**	**GJ Maxwell**
	DA Miller	DA Miller	DA Miller	DA Miller	
	CA Pujara	V Sehwag	CA Pujara	CA Pujara	
Chennai	DR Smith	MS Dhoni	SK Raina	SK Raina	
	MS Dhoni	DR Smith	MS Dhoni	MS Dhoni	
	SK Raina	SK Raina	DR Smith	DR Smith	

Table 1: Results for first three IPL 2014 games The red color columns are for the winning team and green ones for the losing team. I have added the top 3 performers as per PageRank, In-strength, Betweenness and Closeness and also compare it with Man of Match. In some cases the Man of Match (MOM) went to the bowler, as seen in 2nd match where YS Chahal received MOM.

We present our results for first three matches in IPL 2014 in Table1. The red color columns are for the winning team and green ones for the losing team. We present the top three performers as per PageRank, In-strength, Betweenness and Closeness and also compare it with Man of Match (MOM). The players marked in bold font emerge as the players on which there is full consensus: they won the Man of Match (MOM) award,

judged by Cricket experts and are the best movies according to all four ranking schemes. In Figure1 we show the performance visualization of the final match in IPL 2014 played between KKR and KXIP. For KKR, *MK Pandey* has the highest betweenness, PageRank and In-strength. The nodes are colored according to the betweenness centrality. The size of the nodes is proportional to the PageRank. The color of the edge is according to the color of the source node. It is interesting to note that the MOM award in the IPL 2014 final was awarded to *MK Pandey* for his match winning performance. We observe that in approximately 80% of the times, there is full consensus between centrality measures and expert opinion (MOM award)

4.Discussions and Conclusion

To summarize, we investigated the structural properties of batsmen partnership network (BPN) during the Indian Premier League 2014. Our study reveals that network analysis is able to examine individual level network properties among players in cricket.
The batting partnership networks not only provides a visual summary of proceedings of matches for various teams, they are also used to analyze the importance or popularity of a player in the team. It is reasonable to expect players with higher centrality scores be rewarded with the Man of the Match (MOM). We observe that in majority of cases, the MOM matches well with the top three performers, as elucidated from centrality measures. Identifying the `central' player in a batting line up is always crucial for the home team as well as the opponent team. Again, all renowned Cricket teams have different approach to their game-play, batting line-up and bowling order. Batting order of teams like England is much different from teams like India. Indian team depends on performance of star players, whereas for English players distribute their performance. Our analysis can be extended to identify the pattern of play for different teams and potential weakness in batting line-up.

Some additional features could be applied in our analysis. The networks in our study are static and we assumed all the batsmen are equally athletic in the field. One could add an ``athletic index" as an attribute to each batsman. Also adding the fielders as additional nodes in the networks could provide us with a true picture of the difficulty faced by a batsman while scoring. Moreover, in real life many networks display community structure: subsets of nodes having dense node-node connections, but between which few links exist. Identifying community structure in real world networks have could help us to understand and exploit these networks more effectively. Potentially our study leaves a wide range of open questions, which could stimulate further research in other team sports as well.

Acknowledgements

The author thanks the Cricinfo website for public availability of data. The author also thanks members of Northwestern Institute on Complex Systems for their comments that helped in improving the manuscript. Financial support for this research provided by the Kellogg Graduate School of Management, Northwestern University.

References

1. Ben-Naim, E., F. Vazquez, and S. Redner *J. Korean Phys. Soc*. 50, 124. (2007)
2. Bittner E, Nussbaumer A, Janke W, Weigel M *European Physical Journal B* 67, 459 (2009)
3. Petersen, A. M., W. S. Jung, and H. E. Stanley *EPL* 83, 50010. (2008)
4. Sire, C. and S. Redner, *Eur. Phys. J. B* 67, 473. (2009)
5. Ben-Naim, E., S. Redner, and F. Vazquez *EPL* 77, 30005. (2007)
6. Skinner, B. *Journal of Quantitative Analysis in Sports* 6, 3. (2010)
7. Guerra YD, Gonzalez JMM, Montesdeoca SS, Ruiz DR, Garcia-Rodriguez A, GarciaManso JM *Physica A* 391, 2997 (2012)
8. Heuer A, Rubner O *European Physical Journal B* 67, 445 (2009)
9. Ribeiro HV, Mendes RS, Malacarne LC, Picoli S, Santoro PA *European Physical Journal B* 75, 327 (2010)
10. Heuer, A., C. Muller, and O. Rubner *EPL* 89, 38007. (2010)
11. R. Albert and A. Barabasi, *Reviews of Modern Physics* 74, (2002)
12. Freeman, L. C., Borgatti, S. P., White, D. R., *Social Networks* 13(2), 141 (1991)
13. Duch, J., J. S. Waitzman, and L. A. N. Amaral *PLoS ONE* 5, e10937. (2010)
14. J. L. Pena and H. Touchette *arXiv:1206.6904v1* (2012)
15. Girvan, M. and M. E. J. Newman *Proc. Natl. Acad. Sci. USA* 99, 7821. (2002)
16. Onody, R. N. and P. A. de Castro *Phys. Rev. E 70* 037103. (2004)
17. Xinping Xu, Junhui Hu and Feng Liu *Chaos* 17, 023111 (2007)
18. P. Passos, K. Davids, D. Araujo, N. Paz, J. Minguens and J. Mendes *Journal of Science and Medicine in Sport* 14, 170 (2011)
19. Saavedra, S., S. Powers, T. McCotter, M. A. Porter, and P. J. Mucha *Physica A* 389, 1131 (2009)
20. Park, J. and M. E. J. Newman *Journal of Statistical Mechanics: Theory and Experiment* 10. (2005)
21. Radicchi, F. *PloS ONE* 6, e17249. (2011)
22. S. Mukherjee *Physica A* 391, 6066 (2012)
23. S. Mukherjee *Physica A* 393, 624 (2013)
24. D. Lusher, G. Robins and P. Kremer *Measurement in Physical Education and Exercise Science* 14, 211 (2010)
25. P. E. Allsopp and Stephen R. Clarke *Journal of the Royal Statistical Society: Series A* 167, Part 4, 657 (2004)
26. Philip Scarf, Xin Shi and Sohail Akhtar *Journal of the Royal Statistical Society: Series A* 174, Part 2, 471 (2011)
27. H. V. Ribeiro, S. Mukherjee and X. H. T. Zeng *Physical Review E* 86, 022102 (2012)
28. S. Mukherjee *Advances in Complex Systems* 16, 3 (2013)