Applied Computing and Informatics (2017) xxx, xxx–xxx

Saudi Computer Society, King Saud University

Applied Computing and Informatics

(http://computer.org.sa)
www.ksu.edu.sa
www.sciencedirect.com

Network centrality based team formation: A case study on T-20 cricket

Paramita Dey [a], **Maitreyee Ganguly** [a], **Sarbani Roy** [b,*]

[a] *Department of Information Technology, Government College of Engineering and Ceramic Technology, Kolkata 700010, India*
[b] *Department of Computer Science and Engineering, Jadavpur University, Kolkata 700032, India*

Received 22 August 2016; revised 14 October 2016; accepted 18 November 2016

KEYWORDS

Social Network Analysis (SNA);
Centrality measures;
T-20 cricket;
Small world network;
Clustering coefficient

Abstract This paper proposes and evaluates the novel utilization of small world network properties for the formation of team of players with both best performances and best belongingness within the team network. To verify this concept, this methodology is applied to T-20 cricket teams. The players are treated as nodes of the network, whereas the number of interactions between team members is denoted as the edges between those nodes. All intra country networks form the cricket network for this case study. Analysis of the networks depicts that T-20 cricket network inherits all characteristics of small world network. Making a quantitative measure for an individual performance in the team sports is important with respect to the fact that for team selection of an International match, from pool of best players, only eleven players can be selected for the team. The statistical record of each player considered as a traditional way of quantifying the performance of a player. But the other criteria such as performing against a strong opponent or performance as an effective team member such as fielding, running between the wickets, good partnership deserves more credential. In this paper a revised method based on social networking is presented to quantify the *quality* of team belongingness and efficiency of each player. The application of Social Network Analysis (SNA) is explored to measure performances and the rank of the players. A bidirectional weighted network of players is generated using the information collected from T-20 cricket (2014–2016) and used for network analysis. Thus team was formed based on that ranking and compared with their IPL (Indian Premier League) performances of 2016.

1. Introduction

Watts and Strogatz [1] defined Small-World networks as the class where networks are highly clustered like regular lattices, but with small characteristic path length similar to random graphs. Social network analysis provides analytical information about the interrelationships between the members of the network and the network dynamics. In this work we have generated a small world network [2] of international cricket teams

* Corresponding author.
E-mail addresses: dey.paramita77@gmail.com (P. Dey), maitreyee12aug@gmail.com (M. Ganguly), sarbani.roy@cse.jdvu.ac.in (S. Roy).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.aci.2016.11.001

using the nodes as the players and interactions between them as the edges. The focus of this paper is to apply the social network analysis techniques such as clustering coefficients and centrality measurements to quantify the belongingness of an individual player within the team along with their individual performances. As a result, the quality of the players as a team member reflected in team formation which is very important for team sports. Based on that characterization the key players (in a role based approach) can be selected for a team formation.

Cricket is a bat-and-ball game played between two teams, consisting eleven players in the both team, on a cricket field. One team is the batting team, attempts to score as many runs as possible, while their opponents are the bowling team, which give fielding and bowling and try to take as many as wickets and restrict the runs of the batting team. In short term cricket, the length of each innings ranges from 20 overs of six bowling deliveries per side (T-20 cricket) to 50 over deliveries per side (one day cricket) and in case of Test cricket, there is a maximum of four innings played over five days, and per day maximum 90 overs can be played.

In this paper, a ranking system based on both players' performance statistics and belongingness is proposed. To capture the belongingness of a player, centrality measures and clustering coefficient are considered. The data for network generation for this research work are collected from international T-20 cricket matches for the session 2014–2016. Statistical records of top ten countries Australia, England, South Africa, New Zealand, India, West Indies, Pakistan, Sri Lanka, Bangladesh and Zimbabwe are considered. From late 2000, ICC (International Cricket Council, the highest parenting body of cricket) introduced the shortest format called T-20 cricket which are played for twenty-twenty overs and durations of the matches are approximately three hours [3]. Players are selected based on their centrality measures and clustering coefficient and finally a pool of players are formed from where a team of eleven players can be selected. Best team according to the proposed approach is also compared with IPL 2016 teams.

We are comparing our results with IPL 2016 teams. Most of the countries that played Test Cricket have a domestic cup competition, and in case of India it is Indian Premium League, often abbreviated as IPL. The IPL is the most attracting form of cricket in the world and ranks sixth among all sports leagues, contested every year during April–March by franchise teams representing Indian cities from different states. There are currently 8 teams playing in this tournament, and each member of the team is selected by auction of players done by the franchisees. Each franchise can select a squad of 14 players based on their different performance matrices, and from them 11 of the players are played in the day of match.

The rest of the paper is organized as follows. In Section 2, the related works are discussed. The proposed approach for team formation is discussed in Section 3. This section highlights small world characteristics of T-20 cricket network. Different centrality measures and clustering coefficient technique from the view point of cricket network are also discussed. In Section 4, the effectiveness of the proposed approach is evaluated and compared with IPL T-20 session 2016. In Section 5 a brief discussion is made comparing other best team formation approaches with proposed approach. Finally, the paper concludes in Section 6.

2. Related work

In recent research trends, several works have been done for performance measures of a team, where team belongingness of individual performance along with their individual performances in team sports plays a significant role. Different techniques used in analysing team sports like football can be found in [4–8]. In [9,10], interactions among players are analysed in other ball sports such as baseball and basketball respectively. Another network analysis approach has been done on Spanish Teams performance during FIFA World Cup 2010 [11]. They have done a temporal analysis of the resulting passes network, taken into account the number of passes, length of the chain of passes, and centrality, clustering coefficient measure of the players. Some researchers also consider the basic fact of a team sports and instead of measuring simple statistics as number goals or number assists, they put emphasis on the players ability a team member by generation a flow network [12]. In [13,14] network based approaches are used for performance analysis of players in tennis. In [15,16] a network was generated using the interaction of players for cricket. Performances of water polo also quantified through network-based approach.

Not only in sports, network based approach in team building is more recent trend of analysis along with data mining and web mining tools [17,18]. Though these papers focussed on formation of network, social network characteristics are not used for ranking systems or team formation.

3. Proposed approach for team formation

This work presents a social network analysis based approach to form a team. Both the performances and belongingness of players are considered here. Initially, from the data of T-20 cricket session 2014–2016 a pool of players is created o the basis of their performance. A player is included in the pool as a batsman if he has scored more than 900 runs in T-20 international matches during 2014–2016. Similarly, a player is selected as a bowler if he has taken more than 35 wickets during that period in T-20 international matches. The major steps of the proposed approach are (1) network formation of the T-20 cricket, (2) identification of small world properties in the network, and (3) formation of a pool of players with respect to high centrality and clustering coefficient measures. From this pool, four teams are selected on the basis of players' betweenness centrality, closeness centrality, node degree distribution and clustering coefficient. Players are also assigned ranks according to these measures in the corresponding team. These four teams are compared with present IPL teams, 2016. Fig. 1 shows a flow diagram of the proposed approach.

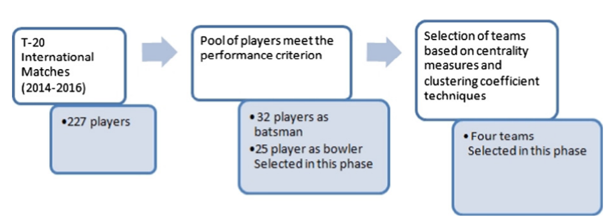

Figure 1 Flow diagram of the proposed approach.

Please cite this article in press as: P. Dey et al., Network centrality based team formation: A case study on T-20 cricket, Applied Computing and Informatics (2017), http://dx.doi.org/10.1016/j.aci.2016.11.001

3.1. Network formation

Cricket database is created from match by match statistics collected from www.espncricinfo.com Web site, and imported on excel sheet as an adjacency matrix as shown in Table 1. This adjacency matrix defines the number of matches played between players. We have formed a player vs. player matrix, where we define the number of matches common between two players. Firstly this was done by individual team basis and the intra network is formed. A small sample network is shown in Fig. 2 Each team member is defined as a node in the graph, and connection between them represents the edges of the graph. The weighted directed edges are created from the ratio of common matches and total number of matches played by that player. From the reference figure it is clear that the thick edges represent profound relationship between two players whereas thin edges represent less interaction. Then a team vs. team matrix was formed for all the countries under consideration and thus network of networks of all countries is formed as shown in Fig. 3. This network is executed in Gephi (https://gephi.org/), which is a free and open source software. Finally, the network parameters are calculated. A partial data table showing the derived parameters is shown in appendix as Table 4.

3.2. Small world characteristics of T-20 cricket network

It is important for the cricket network to be a small world network [1], as our idea is to use the centrality measures and clustering coefficient to detect the key players in the network. However, most of the nodes are not directly connected in a small world, but the distance (i.e., number of hops) between two nodes is significantly less and the nodes connecting other nodes play significant role in the network.

In a network, small world coefficient σ can be defined as.

$$\sigma = (C_{Actual}/C_{Random})/(L_{Actual}/L_{Random}).$$

This value should be greater than 1 for being a small world network [1].

Here C_{Actual} and C_{Random} signify the clustering coefficients of players' network and random network respectively. Similarly, L_{Actual} and L_{Random} respectively signify the average path lengths of players network and random network.

At first, we have generated network for T-20 cricket network. Fig. 4 shows the random graph and T-20 cricket network graph with same number of nodes. Clustering coefficient and average path length for the network are calculated. Clustering coefficient for T-20 is derived as 0.702 and average path length is calculated as 2.147. Now to calculate σ, a random network [19] is generated using random rewiring probability. Clustering coefficients of random network with same number of nodes (as the actual T-20 cricket network) are 0.14 and average Path Length is 2.081. Thus, the small world coefficient, i.e., σ of T-20 cricket network is calculated as 4.86 which is not only greater than 1, but shows higher value than small world coefficients of karate (1.65) and the Internet networks (2.38) [20]. This result clearly depicts that the T-20 cricket network inherits small world phenomenon.

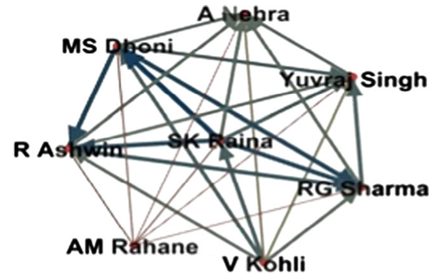

Figure 2 Formation of network.

3.3. Clustering coefficient

Clustering coefficient of a node (player) in the network signifies characteristics of that node (player) forming local cluster that is numbers of nodes (players) those are influenced by that particular node (player). In this context, the dense local cluster signifies that node (player) has great influence to other nodes (players). In social networks, especially in small world networks, generally all nodes are highly connected and the clustering coefficient is also quite high valued than the average clustering coefficient of random network. The local clustering coefficient of a node is also defined as the number of complete graph (clique) that can be formed using the neighbor of that node. This property was first introduced by Watts and Strogatz [1] for defining small world coefficient. The clustering coefficient varies between 0 (no clustering) and 1 (maximum clustering) [20].

Let graph $G = (n, l)$ denotes set of nodes n and a set of links l connecting the nodes. An edge l_{ij} denotes the connection between node n_i with node n_j. Neighborhood N_i for a node n_i can be expressed as node's connected neighbors and can be denoted by following expression [21]:

$$N_i = \{n_j : l_{ij} \in l, l_{ji} \in l\}$$

Here, we consider an undirected graph with identical weight for l_{ij} and l_{ji}. Thus if node n_i has k_i neighbors, $\frac{k_i(k_i-1)}{2}$ edges can be exist between the nodes within local neighborhood. Therefore, local clustering coefficient can be expressed as [21]

Table 1 Formation of adjacency matrix.

Player	Total matches	AM Rahane	S Dhawan	V Kohli	SK Raina	MS Dhoni	RA Jadeja
AM Rahane	17	NA	4	2	5	5	5
S Dhawan	19	4	NA	10	13	13	9
V Kohli	39	2	10	NA	11	11	8
SK Raina	58	5	13	11	NA	13	11
MS Dhoni	64	5	13	11	13	NA	11
RA Jadeja	33	5	9	8	11	11	NA

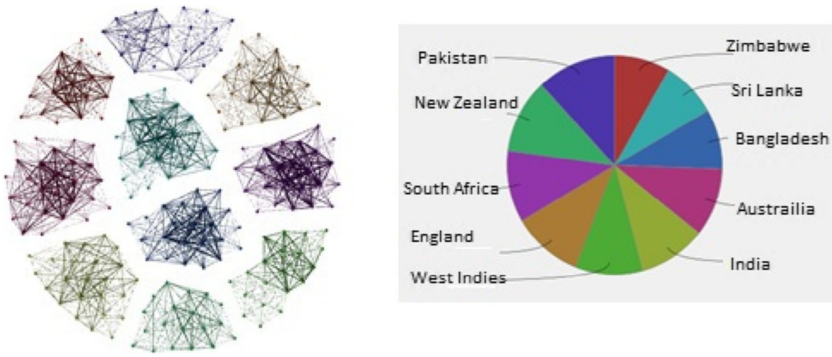

Figure 3 Collection of Intra networks of each country.

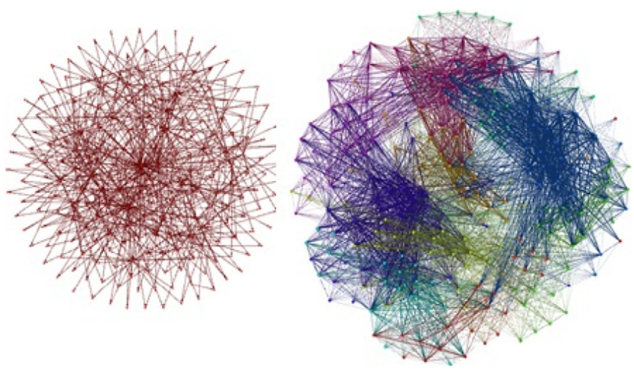

Figure 4 Random graph and T-20 cricket network graph with same number of nodes.

$$LC_i = \frac{2|\{l_{jk} : n_j, n_k \in n, l_{jk} \in l\}|}{k_i(k_i - 1)}$$

If $\lambda_G(n)$ denotes number of triangles on $n \in G$ for an undirected graph G, $\lambda_G(n)$ is total number of subgraphs of G with 3 edges and 3 nodes, $\tau_G(n)$ be the number of triples on $n \in G$. Then clustering coefficient can be defined as below [22].

$$C_i = \frac{\lambda_G(n)}{\tau_G(n)}$$

Clustering can be measured in different ways also. One common procedure for measuring is to find existing triangles, i.e., to check that when two edges share a node, and then in a network with high clustering, it is highly probable that a third edge exists to form a triangles [22]. Small world networks have the characteristics of highly clustered nodes. From the analysis it can be stated that the cricket network is highly clustered as maximum players have high clustering coefficient value. Average clustering coefficient is the mean value of individual clustering coefficients. In this work, a total number of 39,784 triangles are formed in the T-20 network and average clustering coefficient is calculated as 0.702. The maximum clustering coefficient is derived as 1 and distribution graphs of players in the role of batsman and bowler (only for the players who qualified for the performance pool) respectively are shown in Figs. 5 and 6.

3.4. Average path length

In a graph $G = (V, E)$ of order n the average path length is defined as the average distance between any two pair of vertices [21]. It is denoted by the arithmetic mean of distance between any two randomly chosen nodes or players. Thus, average path length can be defined as [21]:

$$\mu(G) = \frac{1}{\binom{n}{2}} \sum_{u,v \subset V} d(u, v)$$

Average path length signifies how well the players in the network are connected to the other players in the player network. The characteristics of small world network are that average path length is comparatively smaller compared to random or scale free network.

3.5. Betweenness centrality

Betweenness centrality is a widely used centrality measure. For each single node present in the network, betweenness centrality of that node can be defined as proportionality of total shortest paths passing through that nodes to all possible shortest paths present in network [23]. A player with high value of betweenness centrality measure has large influence on the other players. Betweenness centrality of node n can be expressed as [23]

$$C_B(n) = \sum_{s \neq n \neq t} \frac{\sigma_{st}(n)}{\sigma_{st}}$$

where σ_{st} defines total number of shortest paths and $\sigma_{st}(n)$ defines the number of shortest paths passing through n. As discussed, betweenness centrality signifies the influence of a particular node in the network, i.e., players with higher

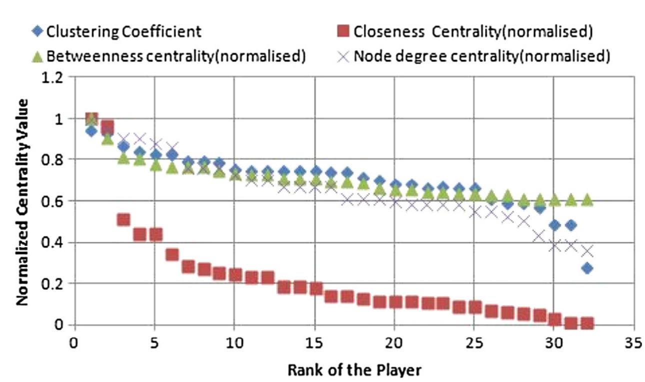

Figure 5 Distribution of different measures of players in the role of batsman.

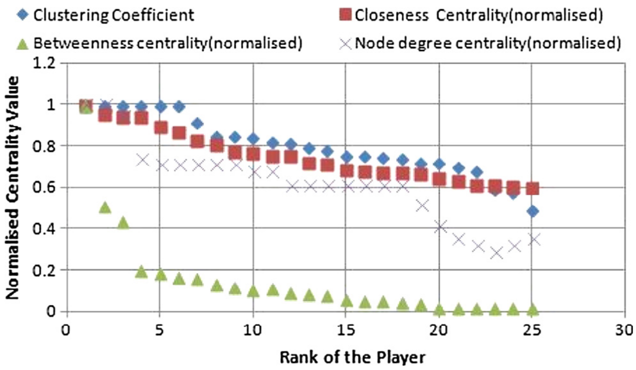

Figure 6 Distribution of different measures of players in the role of bowler.

betweenness centrality play important roles within the network. In the computation, total number of shortest paths is derived as 51,302, for 227 numbers of players. Normalized betweenness centrality distribution graphs for players in the role of batsman and bowler respectively are shown in Figs. 5 and 6.

3.6. Closeness centrality

In a graph representing a small world network, closeness centrality measures how close a node is to others in the network. In other words, the closeness centrality of a node in a network is the inverse of distance between two nodes in that network. By the use of closeness centrality one can determine the efficiency of each node with respect to others for spreading information in the network. The shorter the distance between two nodes, the larger the closeness centrality and thus ensures the more efficient spreading in the network. There are various algorithms for defining closeness centrality measurements by calculating all pair shortest paths. A node is considered as an important node if that node is relatively close to other nodes. Closeness is reciprocal of farness. To calculate closeness centrality of a node the equation can be defined as [21]

$$closeness(u) = \sum_{v} \frac{1}{d_{vu}}$$

where u is the focal node, v is another node in the network, and d_{vu} is the shortest distance between these two nodes. In T-20 cricket network, closeness centrality value ranges between 1 and 52.3. Figs. 5 and 6 shows the normalized closeness centrality distribution of players in the role of batsman and bowler.

3.7. Node degree distribution

Among all centrality measures degree distribution is the simplest one. Degree of a node in a network is determined by the total number of outgoing and ingoing edges from that node. In case of directed graph we have two types of degree distribution namely out-degree and in-degree distribution for each node. Although in case of undirected graph there is no such classification. When analysing the weighted networks, degree has generally expressed as sum of nodes. For calculating the total degree distribution of a node in a directed graph, all the incoming and outgoing edges should be added. The in-degree and out-degree of node u is the total number of connections onto and from node u respectively. Basically degree distribution captures only a small amount of data of a small world network, but that information gives us the important clues about the network. Nodes with highest number of degree have greatest influence on connection the other nodes and working as the hub of the network. In T-20 cricket, the value of node degree centrality ranges from 22 to 68 for the players selected in the performance pool and the distribution (normalized) of players in the role of batsman and bowler is shown in Figs. 5 and 6.

4. Selection of player

From the pool of players who meet the criterion of performance, based on the clustering coefficient, betweenness centrality, closeness centrality and node degree distribution, the key players are selected in a ranked manner. Each cricket team has batsman, bowler, all-rounder and wicketkeeper which can be designated as the role of the players in the team. Based on the role required for a cricket, a team is selected from the pool of players formed using the proposed approach.

In this work, T-20 cricket network is formed for a total of 227 players (session 2014–2016) and total number of shortest paths is derived as 51302. As discussed earlier, performance data (runs, wickets, fielding) are collected for each player. Then those designated players are sorted according to their betweenness centrality, closeness centrality, clustering coefficient and degree in descending order and normalized those values. Now from the pool of players with high performances, players in different roles are selected based on their centrality measures and clustering coefficients. From the analysis, the best teams are shown in Table 2.

For the comparison, we show the first eleven (based on the maximum appearance) from the squad of 25 players of IPL 2016 teams in Table 3. It depicts that almost all players of the best teams are included in the first eleven. Statistics also reflect that performance of these players has a great contribution for their teams. 19 players are appeared in first eleven of eight IPL teams out of 22 players that belong to our best four teams (refer to Tables 3 and 2) Proposed team matching index with IPL is 86.36%, whereas for ICC best rank players (22 numbers) matching index is 14 out of 22, i.e., 63.63% (excluding players from Pakistan). In other words, more than 86% players in the team formed using our proposed approach are included in the first eleven team (excluded the players from Pakistan as they didn't participate in this season of IPL) as compared to 63% of the best ranked players of ICC T-20 ranking.

It is interesting that team Royal challenger Bengaluru have most players in the role of batsman from our proposed pool of players. However, none of their bowlers are from this pool. In the real scenario, the team Royal challenger Bengaluru performed very well and they were the first qualifier for the final of IPL season 2016, however their comparatively poor bowling section was a drawback of their performance. Now the other three teams that include maximum number of players from our proposed best teams are Rising Pune Supergiant, Sunrisers Hyderabad and Gujarat Lions. From these four teams except Rising Pune Supergiant, other teams are top three teams of IPL 2016.

Please cite this article in press as: P. Dey et al., Network centrality based team formation: A case study on T-20 cricket, Applied Computing and Informatics (2017), http://dx.doi.org/10.1016/j.aci.2016.11.001

Table 2 Best team with respect to betweenness centrality, closeness centrality, node degree distribution, clustering coefficient.

Betweenness centrality	Closeness centrality	Clustering coefficient	Degree distribution
AJ Finch (BAT)	CL Breath White (BAT)	CL Breath White (BAT)	Aj Finch (BAT)
DA Warner (BAT)	RG Sharma (BAT)	CH Gayel (BAT)	AD Mathews (BAT)
TM Dilshan (BAT)	CH Gayle (BAT)	HM Amla (BAT)	DA Warner (BAT)
KS Williamson (BAT)	DA Warner (BAT)	F Du Plesis (BAT)	TM Dilshan (BAT)
SR Watson (ALL)	Virat Kohli (BAT)	MJ Guptil (BAT)	SR Watson (ALL)
Shahid Afridi (ALL)	Yuvraj Singh (ALL)	Yuvraj Singh (ALL)	Shahid Afridi (ALL)
Umar Akmal (WC)	MS Dhoni (WC)	AB de Villers (WC)	Umar Akmal (WC)
Mashrafe Mortaza (BALL)	Saeed Ajmal (BALL)	Saeed Ajmal (BALL)	Mashrafe Mortaza (BALL)
DJG Sammy (BALL)	S Badree (BALL)	DW Steyn (BALL)	BB McCulum (BALL)
BB McCulum (BALL)	DW Steyn (BALL)	S Badree (BALL)	Sohail Tanvir (BALL)
James Faulkner (BALL)	R Ashwin (BALL)	WC Parnell (BALL)	D Wiese (BALL)

BAT: Batsman; BALL: Bowler; WC: Wicketkeeper; All: Allrounder.

Table 3 IPL teams of 2016.

Delhi Daredevils	Kings XI Punjab	Gujarat Lions	Kolkata Knight Riders
Zaheer Khan	Murali Vijay	Suresh Raina	Gautam Gambhir
Q de Kock	David Miller	Dwayne Bravo	Piyush Chawla
JP Duminy	Glenn Maxwell	Ravindra Jadeja[c]	Morne Morkel
Carlos Brathwaite[a]	Axar Patel	AJ Finch[a]	Sunil Narine
Amit Mishra	Wriddhiman Saha	Dinesh Karthik	Manish Pandey
Mohammad Sami	Sandeep Sharma	Dhawal Kulkarni[d]	Yusuf Pathan
Chris Morris	Mohit Sharma	Praveen Kumar[d]	Andre Russel
Karun Nair[d]	Hasim Amla[a]	BB McCullum[b]	Shakib Al Hassan
Sanju Samson	Manan Vohra[d]	Dwayne Smith[d]	Robin Uthappa
Rishabh Pant[d]	Shaun Marsh	James Faulkner[a]	Suryakumar Yadav[d]
P Negi[d]	Marcus Stoinis	Pravin Tambe[d]	Umesh Yadav
Mumbai Indians	**Risings Pune Supergiants**	**Royal Challengers Bangalore**	**Sunrisers Hyderabad**
Rohit Sharma[b]	MS Dhoni[b]	Virat Kohli[a]	David Warner[a]
Jasprit Bumrah[c]	R Ashwin[a]	AB de Villiers[b]	Shikhar Dhawan
Jos Buttler	Thisara Perera	Chris Gayle[a]	Mustafizur Rahman
Mitchell McClenaghan	Ajiknya Rahane	SR Watson[a]	KS Williamson[a]
Hardik Pandya	Steve Smith	Stuart Binny	Yuvraj Singh[b]
Krunal Pandya	Murugan Ashwin[d]	Chris Jordan	Eoin Morgan
Ambati Rayadu	Ashoke Dinda[d]	Kane Richardson	Moises Henriques
Tim Southee	F du Plessis[a]	Sachin Baby[d]	Bhuvneshwar Kumar
Kieron Pollard	D. Stein[b]	Lokesh Rahul[d]	Naman Ojha[d]
Martin Guptill[a]	Rajat Bhatia[d]	Yuzvendra Chahal[d]	Ashish Nehra[d]
Nitish Rana[d]	Saurabh Tiwary[d]	Varun Aaron	Deepak Hooda[d]

[a] Appears both in team selected through proposed technique and ICC best ranking list (top 20).
[b] Appears in team selected through proposed technique but does not appear in ICC best ranking list.
[c] Appears in ICC best ranking list but not in team selected through proposed technique.
[d] Represents local players (not considered in input set).

5. Discussion

In this paper, we have presented an approach which includes both performance and bonding with the teammates. It is become important to select the players who have maximum potential and have the capability of playing team game, which builds the success of the team. Moreover for each role/position it requires a proper balance so that the team can play with maximum coordination.

For the ranking of the players, other approaches such as neural network is applied to the selection of players. For example, in the case of annual Australian Football League (AFL) national draft, team formation involves mathematical and statistical approach for extracting knowledge from the neural network [24]. A recent study have been made, where the researchers model a decision-making process of a single sports franchise, which takes a combination of the players estimated value along with the value of the other players currently available, and the position wise analysis required for that particular team [25]. But all these techniques ignore the fact that interaction of the players can play an important role for team performance. We are trying to incorporate this property and find that it is more closure to the real team selection.

Table 4 Partial data table of players with centrality values and clustering coefficients.

Name	Country	Role	Runs scored	Total matches	BC	ClC	CC	Degree
DA Warner	Australia	Batsman	1633	60	17.0675	1.278	0.612	52
SR Watson	Australia	All Rounder	1462	57	50.322	1.056	0.49	62
AJ Finch	Australia	Batsman	974	27	48.046	1.056	0.488	68
CL White	Australia	Batsman	984	47	0.525	1.639	0.936	25
Shakib Al Hasan	Bangladesh	Batsman	1103	54	6.904	1.192	0.743	42
Tamim Iqbal	Bangladesh	Batsman	1154	52	12.683	1.077	0.674	48
Yuvraj Singh	India	All Rounder	1086	52	0.475	1.481	0.945	27
SK Raina	India	Batsman	1163	58	5.649	1.259	0.753	40
RG Sharma	India	All Rounder	1209	56	2.296	1.333	0.284	36
Virat Kohli	India	Batsman	1391	39	5.649	1.259	0.753	40
MS Dhoni	India	WC	982	64	5.649	1.259	0.753	40
Name	Country	Role	Wickets	Total matches	BC	ClC	CC	Degree
SR Watson	Australia	All Rounder	48	57	50.322	1.056	0.49	62
Mashrafe Mortaza	Bangladesh	Bowler	38	49	10.161	1.115	0.7	46
Al-Amin Hossain	Bangladesh	Bowler	39	25	8.295	1.154	0.723	44
R Ashwin	India	Bowler	47	39	4.438	1.296	0.778	38
Mohammad Hafeez	Pakistan	Bowler	46	77	21.938	1.032	0.595	59
Sohail Tanvir	Pakistan	Bowler	47	50	6.179	1.29	0.753	44
Saeed Ajmal	Pakistan	Bowler	85	64	0.1	0.12	.677	20
Umar Gul	Pakistan	Bowler	85	60	2.521	1.387	0.848	38

BC: Betweenness centrality, ClC: Closeness centrality, and CC: Clustering Coefficient.

6. Conclusion

The focus of this paper is team formation based on the properties of small world network in T-20 Cricket. This paper uses three centrality measures betweenness centrality, closeness centrality, node degree distribution and clustering coefficient for evaluation of players. Few key players, that exhibit higher centrality values or clustering coefficient may influence many other players and take a significant role in team formation. In this work, we put an emphasis on quantifying the performance of players from the history of previous years' data about their batting, bowling performances, as well as taking a qualitative measure based on clustering coefficient and centrality measures derived from the network of players. Both the characteristics are used as team formation strategy and a role based team was formed.

Acknowledgement

There is no grant for this research work.

Appendix A.

See Table 4.

References

[1] D.J. Watts, S.H. Strogatz, Collective dynamics of small-world networks, Nature 393 (1998) 440–442.

[2] Qawi K. Telesford, Karen E. Joyce, Satoru Hayasaka, Jonathan H. Burdette, Paul J. Laurienti, The ubiquity of small-world networks, Brain Connect. 1 (2011) 1–5.

[3] M. Amy, The History of Cricket, eSSORTMENT, 2007 <http://www.essortment.com/hobbies/historycricketsngj.html>.

[4] H.V. Ribeiro, R.S. Mendes, L.C. Malacarne, S. Picoli, P.A. Santoro, Dynamics of tournaments: the soccer case, Euro. Phys. J. B (2010) 75–327.

[5] A. Heuer, C. Mller, O. Rubner, Soccer: is scoring goals a predictable Poissonian process?, Europhys Lett. (2010) 89.

[6] Y. Yamamoto, K. Yokoyama, Common and unique network dynamics in football games, PLOS One (2011) 6–12.

[7] P. Passos, K. Davids, D. Araujo, N. Paz, J. Minguens, J. Mendes, Networks as a novel tool for studying team ball sports as complex social systems, J. Sci. Med. Sport (2011) 14–170.

[8] Jos Gama, Micael Couceiro, Gonalo Dias, Vasco Vaz, Small-world networks in professional football: conceptual model and data, Euro. J. Hum. Move. 35 (2015).

[9] C. Sire, S. Redner, Understanding baseball team standings and streaks, Euro. Phys. J. (2009) 473–481.

[10] B. Skinner, The price of anarchy in basketball, J. Quant. Anal. Sports (2010) 3–6.

[11] Carlos Cotta, Antonio M. Mora, Cecilia Merelo-Molina, Juan Julin Merelo, FIFA World Cup 2010: a network analysis of the champion team play, J. Syst. Sci. Complex. 26 (1) (2013) 21–42.

[12] Jordi Duch, Joshua S. Waitzman, Lus A. Nunes Amaral, Quantifying the performance of individual players in a team activity, PLoS One (2010).

[13] D. Lusher, G. Robins, P. Kremer, Measurement in physical education and exercise science, social network analysis, Sport Global. Soc. – Chan. Chall. 14 (2010) 211–224.

[14] F.M.C. Clemente, F. Martins, D. Kalamaras, R. Mendes, Network analysis in basketball: inspecting the prominent players using centrality metrics, J. Phys. Educ. Sport 15 (June) (2015) 212–217.

[15] S. Mukherjee, Identifying the greatest team and captain A complex network approach to cricket matches, Phys. A: Statist. Mech. Appl. 391 (23) (2012) 6066–6076 (Elsevier).

[16] S. Mukherjee, Quantifying individual performance in Cricket – a network analysis of batsmen and bowlers, Phys. A: Statist. Mech. Appl. 393 (2012) 624–637 (Elsevier).

[17] B. Iyer, C.H. Lee, N. Venkatraman, Managing in a small world ecosystem: lessons from the software sector, Calif. Manage. Rev. 48 (3) (2006) 27–47.

Please cite this article in press as: P. Dey et al., Network centrality based team formation: A case study on T-20 cricket, Applied Computing and Informatics (2017), http://dx.doi.org/10.1016/j.aci.2016.11.001

[18] F. Radicchi, Finding statistically significant communities in networks, PLOS One 6 (2011).
[19] P. Erdos, Rnyi, On the evolution of random graphs, Publ. Math. Inst. Hung. Acad. Sci. (1960) 17–61.
[20] Shiu-Wan Hung, An-Pang Wang, A small world in the patent citation network, in: IEEE International Conference on Industrial Engineering and Engineering Management, 2008, pp. 2–4.
[21] G. Wayne, R.O. Oellermann, Distance in Graphs, Structural Analysis of Complex Networks, Springer, 2011.
[22] Matthieu Latapy, Main-Memory triangle computations for very large (sparse (power-law)) graphs, Theoret. Comp. Sci. (TCS) 407 (1–3) (2008) 458–473.
[23] Ulrik Brandes, A faster algorithm for betweenness centrality, J. Math. Sociol. 25 (2) (2001) 163–177.
[24] John McCullagh, Data mining in sport: a neural network approach, Int. J. Sports Sci. Eng. 04 (2010) 131–138.
[25] Michael J. Fry, Andrew Lundberg, Jeffrey W. Ohlmann, A player selection heuristic for a sports league draft, J. Quant. Anal. Sports 3 (2) (2007) (Article 5).

Please cite this article in press as: P. Dey et al., Network centrality based team formation: A case study on T-20 cricket, Applied Computing and Informatics (2017), http://dx.doi.org/10.1016/j.aci.2016.11.001