

# Region-based Teams Ranking in the Game of Cricket using PageRank Algorithm

Akbar Hussain

College of Information and  
Computer

Taiyuan University of Technology,  
China

Yan Qiang

College of Information and  
Computer

Taiyuan University of Technology,  
China

M. Abdul Qadoos Bilal

College of Information and  
Computer

Taiyuan University of Technology,  
China

Inam Ullah

School of Software Engineering  
Shandong University, Jinan, China

Najeeb Ullah

Department of Computer Science  
Government Post Graduate Jahanzeb College  
Swat, Pakistan

## ABSTRACT

Cricket sport is a favorite source of entertainment in commonwealth countries. Traditionally in cricket, the performance of the team measures after winning or losing a match. The international cricket council (ICC) is a governing body that ranks cricket teams globally without considering the strength and weakness of the region. Region-based team's strength and weakness is an essential factor for the teams' selection. In this paper, we proposed the Region-wise Team Rank (RWTR), an extension of the PageRank algorithm to rank cricket teams across the region. The intuition is to get more points to a team that wins a match from a stronger. The proposed ranking reflects more the flow of the game compared to the traditional ranking.

## General Terms

Information systems~ Content ranking • Information systems  
Combination, fusion & federated search • Computing  
methodologies Ranking.

## Keywords

PageRank, Principal Component Analysis (PCA), region  
strength, team's strength.

## 1. INTRODUCTION

Cricket is a bat and ball game played between two teams having 11 players each side. For each phase of play, a batting side tries to score as many runs as possible after striking the ball thrown at the wicket with the bat, while a bowling/fielding side tries to prevent the score/runs and get them out or dismiss all the players. The batting side gives a target score in the first innings. In the second innings, the first inning's batting side starts bowling, and the first inning's bowling side starts batting, and they are going towards the target. Either they complete the target in the specific overs, or they lose the match<sup>1</sup>.

Ranking in sports is a significant measure for the performance and comparison between different teams. The International Cricket Council (ICC) is the governing body that ranks cricket teams based on their current performance. Traditionally, the ICC rating system entirely based on the number of wins or

loses in a match to assign a rating score<sup>2</sup>.

There are three formats in international cricket; Twenty20, One Day International (ODI) and Test cricket format. T-20 cricket is the latest and fastest international cricket format, in which each team has permitted to play 20 over the maximum. ODI cricket is a 50-50 over a match, it can be played during the day or day-night, and it is faster than test cricket. Test cricket matches can last up to five continues days with each day split into three sessions interrupted by lunch and tea breaks<sup>3</sup>. The focus of our research is to rank ODI, cricket teams.

In cricket, the region plays an important role in measuring players' and teams' performance because every region has different playing conditions. Region-based ranking classifies the team's strengths and weaknesses across the region/venue. Consistent performance of the players in batting, bowling and fielding automatically improve their team's strength. There are five regions in which cricket are played internationally. These regions are classified as Asia, Africa, the Americas, Europe, and Oceania Region. The teams with higher strength are usually considered higher in rank. If a team visits to play a match or a series of matches against the opponent team, having a regional ranking enables a more equitable regional distribution. The team's management, coach and the captain should be aware of their teams' strength against the opposite team from the visiting region so that the selectors select a strong combination from his team

In this article, we extend the approach of Daud et al. [1] and Hussain et al. "in press" and computed the region-wise teams' ranking. Daud et al. rank cricket teams using the extension of the PageRank algorithm, but they do not clearly define the strengths and weaknesses of the team across the region. Hussain et al. proposed a statistical procedure (Principal Component Analysis) to find out the region-based team strength for Twenty20, ODI and Test cricket teams. In our model, we applied the results of the prior method for ODI cricket teams using the PageRank algorithm. The proposed method shows a strong correlation with the traditional rating system.

<sup>2</sup><https://www.liveabout.com/how-do-icc-rankings-work-962225>

<sup>3</sup> <https://www.icc-cricket.com/about/cricket/game-formats/the-three-formats>

<sup>1</sup><http://www.espnricinfo.com/ci/content/story/239757.html>

The major contributions of the work are as follows, (1) Consideration of graph-based (RWTR) weighting methods derived from Pagerank algorithms to rank region-based cricket teams, (2) Estimations of region-based teams' strength by investigating the players & teams statistics and consequently aggregating their joint impact, (3) Hybridization of (RWTR) to provide a unified aggregated impact of the teams (URWTR).

In the remaining parts of the paper, a literature review is presented in Section 2. In Section 3, the proposed method is explained. Section 4 represents the experimental results and their evaluation, and, finally, the conclusion and future work are explained in Section 5.

## 2. RELATED WORK

Research Cricket is a second popular sport around the world originated from England and New Zealand and now is the most popular game in Asian countries. Several researchers' interest is attracted by their popularity. There are limited articles found related to region-wise teams 'ranking; however, a comprehensive review revealed the following. The author developed a new performance measure to quantify the performance of the players [2]. Scoring runs against a strong bowling lineup or taking wickets against the strong batting lineup deserves more credit to a player. The advent of the recent format of the cricket sport, the combined bowling rate quantify the performance of the player using multiple linear regressions [3]. The multiple linear regression model is also applied to predict the match outcome during the match [4]. Several variables are used for training and testing the model, such as the past and current form of the team, home region performance, home ground advantage to predict the match outcome. Logistic regression is applied to predict the best team after the knock-out phase in ODI cricket [5]. Fielding is also an essential parameter in cricket sport. Parag Shah has developed to measure the aggregated fielding performance of each player in ODI and T20 cricket matches [6]. The application of machine learning, the artificial neural network is applied to predict the match results [7]. Daud et al. proposed a Region-wise Players Link fusion (RPLF) algorithm to rank cricket players using a region-wise intra-type and inter-type relation-based features to rank the players in cricket in ODI cricket [8].

PageRank [9] was a very useful algorithm for ranking web pages or finding important nodes in the graphs. It has been applied to many domains such as authors ranking, institutions ranking, and sports ranking. A weighted directed model for co-authors network was introduced in reference [10] to measure the impact and popularity of the individual's author in the network. They identified the degree, closeness, and centrality measure to compare the performance of the author. Biological network analysis gained popularity in the area of bioinformatics. The author proposed a personalized PageRank algorithm [11] to analyze protein interaction networks. Important proteins are found between directed links due to several chemical reactions. Google PageRank is proposed for ranking football team. Several statistics, such as goal scored and matches own are combined in different metrics to assign weights to a graph [12].

Recently PageRank was applied to rank cricket teams and their captains [13], but the team's strength and weaknesses were not considered. Later on, Daud et al. considered the extensions of the h-index and PageRank algorithm to rank cricket teams [1]. Besides, they also considered the matches win by runs and wickets parameters for both graph and non-graph based methods. Unfortunately, they ignored teams 'strengths and weaknesses parameters across the region. Hussain et al. (in

press) find out the region-based teams 'batting, bowling and fielding strength for Test, ODI and Twenty20 cricket using Principal Component Analysis (PCA). We proposed the extension of the PageRank algorithm with the region-based teams 'strength and computed the ODI cricket teams 'ranking across the region.

This research is the first generalized approach to rank ODI cricket teams. The information retrieval technique is applied to build a model that can effectively measure the performance of a team across the region.

## 3. RESEARCH METHODOLOGY

### 3.1 Principal Component Analysis

Principal Component Analysis (PCA) is a nonparametric variable reduction approach that uses an orthogonal transformation, and it is well-suited for correlated data [14]. The objective of PCA is to collapse a set of correlated variables into fewer uncorrelated variables as linear combinations of the original variables called principal components.

### 3.2 PageRank

The PageRank is one of the most widely used ranking algorithms, states that if a page has important links to it, its links to other pages also become important [9]. Therefore, PageRank takes the backlinks into account and propagates the ranking through links: a page has a high rank if the sum of the ranks of its backlinks is high [15]. Google PageRank (PR) was created as the backbone to what is now the most influential search engine ever created. Its purpose is to rank web pages when a user makes a query. The PR of the pages should be high if they link to important web pages. PR is based on the number of incoming and outgoing links to determine their rank. In general, the PR value for a page A can be determined by using the following generalized equation:

$$PR(A) = c \sum_{v \in B_A} \frac{PR(v)}{L(v)} \quad (1)$$

Where, PR(A) is the value of PageRank of a page A for each page v contained in the set  $B_A$  (the set containing all the pages linking to page A), divided by the number of outgoing links  $L(v)$  from page v. and c is a normalization factor.

### 3.3 List of Abbreviations

Table 1 show the complete abbreviations which are used though out the article.

**Table 1: List of Abbreviations and Description**

Abbreviations	Description
ICC	International Cricket Council
ODI	One Day International
PCA	Principal Component Analysis
PR	PageRank
RTAR	Region-based Teams Average Runs
RTSR	Region-based Teams Strike Rate
RBAR	Region-based Batsman Average Runs
RBER	Region-based Batsman Economy Rate
RTBR	Region-based Teams Batting Runs
RBMRA	Region-based Batting Milestone Reaching Ability

RTERC	Region-based Teams Economy Rate Conceded
RTARC	Region-based Teams Average Runs Conceded
RbARC	Region-based bowler Economy Rate Conceded
RTbERC	Region-based bowler Economy Rate Conceded
RbSRC	Region-based bowler Strike Rate Conceded
RbMRA	Region-based bowler Milestone Reaching Ability
RFC	Region-based Fielder Catches
RWC	Region-based Wicketkeeper Catches
RRO	Region-based Run Out
RWS	Region-based Wicketkeeper Stumps
RWTR	Region-wise Team Rank
URWTR	Unified Region-wise Team Rank

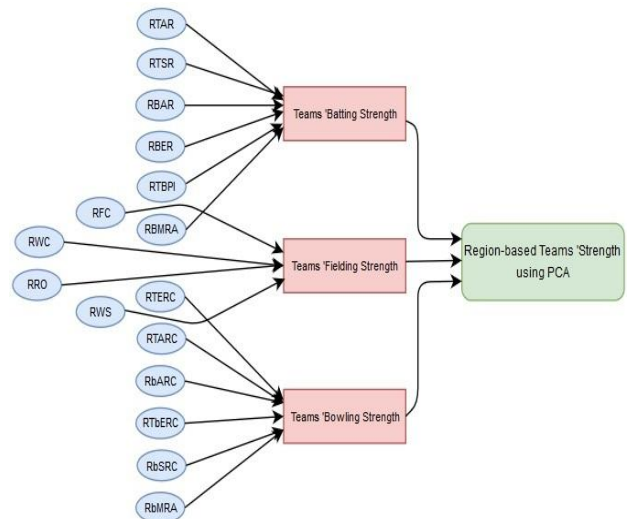


Figure 1: Representation of Region-based Teams Strength using Principal Component Analysis (PCA)

### 3.4 Region-based Team Strength

In cricket sport, the teams 'strength is the combination of several parameters such as bowling, batting and fielding strengths.---deeper analysis and statistics required to figure out the teams 'strength across the region.

In the proposed method Hussain et al. "in press" has created three datasets from the online website<sup>4</sup> using scraping tools (i.e. parsehub and import.in) then created different features for each of the datasets of batting, bowling and fielding respectively. In the experimental setup the authors used six features to each batting and bowling parameters and four features to fielding parameter. By applying the Principal Component Analysis (PCA), it is clearly shown to reduce the dimensions of the features and found out the strengths and weaknesses of each parameter across the region. In conclusion, the author combined the strengths of all the regions and find out the aggregated score of each team called a Unified Team's Strengths. Figure 1 explains the graphical representation of team strength. The author has computed the region-based teams 'strength for ODI cricket teams using equation (2).

$$\sum_{t,r=i}^n RT_{Strength} =$$

$$RTAR*0.23+RTSR*0.26+RBAR*0.12+RBER*0.51+RBPI*0.48+RbMAR*0.11+RERC*0.04+RTARC*0.19+RbARC*0.23+RbERC*0.55-RbSRC*0.52-RbMR*0.56+RFC* 0.40+RWC*0.49+RRO*0.21+ RWS *0.19(2)$$

### 3.4 Ranking Cricket Teams

In the proposed method we created a network of a directed graph having different ODI cricket teams played head-to-head matches. The region-based rank of team t is high if the team wins many matches against opposite teams that have a higher team strength and win-loss ratio. The proposed method shows the network of a team played head-to-head matches in which nodes represent the teams and edges between them are the matches played. If two teams,  $A_i$  and  $B_j$ , played a match in a specific region, and if team  $A_i$  wins from team  $B_j$  then, a directed edge is generated from team  $B_j$  to team  $A_i$ . Consequently, a directed graph is generated iteratively to rank the teams by considering the number of matches played and the team strengths across the region. The graphical representation of different teams which played head-to-head matches is shown in Figure 2.

The RTR is determined by using the following equation.

$$RTR(T) = \frac{1-d}{N} + d \left[ \frac{RTR(T_i)}{CT_i} * (\sum_{t,r=i}^n RT_{Strength}) +, \dots, + \frac{RTR(T_n)}{CT_n} * (\sum_{t,r=i}^n RT_{Strength}) \right] \quad (3)$$

Where,

- $RTR(T)$  is the Region-based TeamRank (RTR) of team T,
- $RTR(T_i)$  is the region-based team rank of Team  $T_i$  which links (lost matches) to team T,
- $CT_i$  is the number of out-links (matches lost) by team  $T_i$ ,
- $\sum_{t,r=i}^n RT_{Strength}$  is the region-based team strength (calculated from equation 2) of Team  $T_i$  in region r,
- d is a damping factor (it is included to prevent sinks i.e. Teams with no lost matches) that is usually set between 0 and 1 (in our experiment we set  $d=0.85$ ), and
- N is the total number of teams played in region r.

In (equation 3) incoming and outgoing links show the number of matches won and lost by Team  $t_i$  in region r.

<sup>4</sup> www.espnricinfo.com

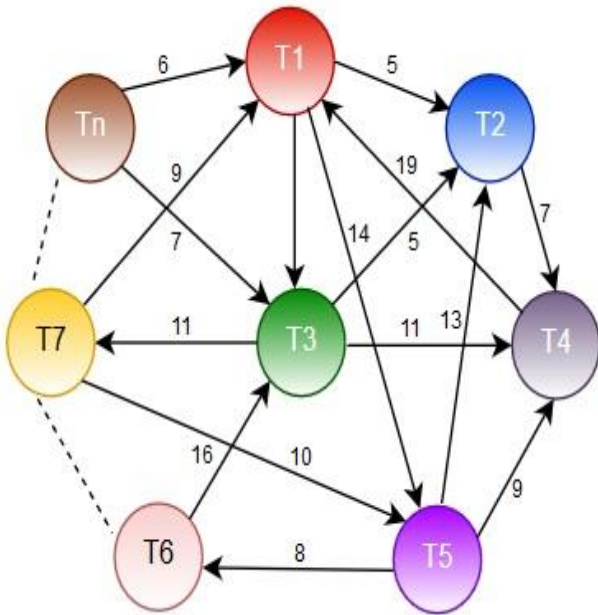


Figure 2: A Graph of teams played head-to-head matches

#### 4. EXPERIMENTAL RESULTS

This section explains the dataset, performance evaluation and comparison of the ICC official teams ranking versus the proposed ranking.

##### 4.1 Dataset

The dataset is taken from the ESPNcricinfo website<sup>4</sup> of ODI cricket matches played between January 2001 and August 2019. The ICC grouped cricket teams into three categories, i.e. full member countries, associate member countries, and affiliate

member countries. In our case, we considered only full member countries; they have the right to play Twenty20, One-Day and Test cricket matches and follow the ICC rules. The authors considered only those teams for a ranking which had played at least ten matches across the region.

##### 4.2 Results and Discussions

The proposed ranking determined the strengths and weaknesses of a team across the region. If a team visits to play a match against the opponent team, having a regional ranking enables a more equitable regional distribution. The captain, the coach, and the team’s management should be aware of their teams’ strength against the opposite team from the visiting region so that the selectors select a strong combination from his team.

Consequently, we compare the results of the ICC ranking with the proposed ranking using Spearman’s rho, Pearson’s r and Kendall’s tau correlation to assess the validity and reliability of the proposed method. The region-based ranking and the URWTR scores are shown in Table 4 while the ICC current ranking is shown in Table 3. Table 2 shows that there is a strong correlation when using Spearman’s rho and Pearson’s r (77%) while using Kendall’s tau only 60% correlation between the traditional method and the proposed method. The results and discussions prove that the proposed methods are highly correlated and very useful in terms of ranking cricket teams across the region. The International Cricket Council (ICC) only presents a general ranking based on the most recent performance of the teams. However, an ODI rating system from the ICC is used as a baseline for comparing the region-wise and a unified team’s ranking. In addition, the outcomes of the proposed methods are also compared subjectively with the existing ICC ranking.

Table 2: Correlation of ICC ranking with URWTR Ranking using Spearman’s rho

Teams	ICC Rank	URWTR Rank	Correlation between ICC and UTR Ranking		
			Pearson’s r ( $\rho$ )	Spearman’s rs ( $\rho$ )	Kendall’s tau ( $\tau$ )
England	1	4	77.3%	77.3%	60%
India	2	3			
New Zealand	3	6			
Australia	4	1			
South Africa	5	2			
Pakistan	6	5			
Bangladesh	7	10			
Sri Lanka	8	7			
West Indies	9	8			
Afghanistan	10	9			
Zimbabwe	11	11			

**Table 3: ICC Current Ranking for ODI Cricket Teams**

Teams	Matches	Points	Ratings	Rank
England	54	6,745	125	1
India	58	7,071	122	2
New Zealand	43	4,837	112	3
Australia	50	5,543	111	4
South Africa	47	5,193	110	5
Pakistan	49	4,756	97	6
Bangladesh	46	3,963	86	7
Sri Lanka	54	4,425	82	8
West Indies	49	3,740	76	9
Afghanistan	40	2,359	59	10
Zimbabwe	35	1,538	44	12

Table 4 represents the region-wise teams 'ranking and unified team ranking of ODI cricket teams. In most of the region Australian, Indian and South African cricket teams having better ranking because of their higher win/loss ratio and greater teams 'strength. Considering the overall ranking, the Australian cricket team is ever best because their win/loss ratio is higher than the other teams across the region. The highlighted number shows the top three teams in every region. In the region-wise ranking, we considered the overall wins and losses and teams 'strength and weakness at the specific region while the traditional method only considered the recent performance of the teams. Even though the Australian cricket team has the fourth rank in the ICC ranking but their region-wise ranking is excellent, and that is the reason that they are first in the URWTR ranking.

Pakistani, Indian and Sri Lankan cricket teams are the top three teams in Asia region while they are 6th, second and 8th at the ICC ranking. The Australian team has retained the same rank (4th) as in the ICC rank. The English cricket team has the fifth rank in Asia region because of lower teams 'strength while first at the ICC ranking. The recent two Year performance of the English team is excellent so that they gain the top position. The

Afghanistan cricket team is the newly cricket team they recently played a very good game and won a lot of matches in Asia region. Their rank in the Asia region is eighth and ninth respectively in the ICC rank.

South African, Australian and, Indian teams are the top three teams in Africa region while they are 5th, fourth and 2nd at the ICC ranking. The Australian and South African team has gained a better rank while considering this region because of their region-wise team strength. The English cricket team has the seventh rank in Africa region because of lower batting 'strength while first at the ICC ranking. Pakistan and Sri Lankan teams are fourth and fifth respectively in Africa region while sixth and eighth respectively in the ICC rank. West Indies cricket team has the eighth rank in the African region while 9th in the ICC rank.

Australia, West Indies and South African teams are the top three teams in America region while they are 4th, ninth and 5th at the ICC ranking. The Australian team has retained the same rank again as in the ICC rank. The English cricket team has the fifth rank in America region because of lower teams 'strength while first at the ICC ranking. South African team has the same rank in both the Europe and ICC ranking. The Zimbabwean cricket team is 10th in this region while 12th at the ICC ranking. The Afghan cricket team is the newly cricket team they did not play many matches at this region, so we had not considered their rank in this region.

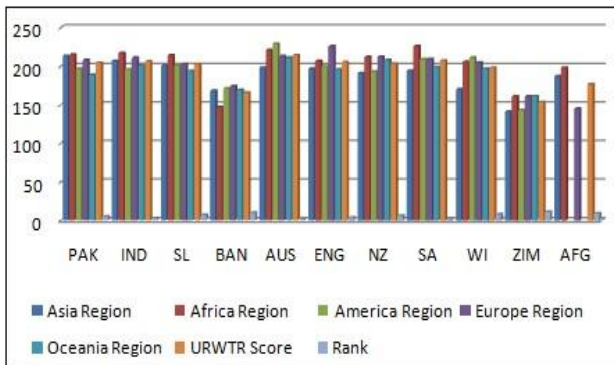
England, Australia and New Zealand cricket teams are the top three teams in the Europe region while they are 1st, 4th and, 3rd at the ICC ranking. The English team has retained the same rank as in the ICC rank. Their batting and bowling strength is better as compared to the other teams in this region. The Afghan cricket team has the last rank in this region because their batting and fielding strength is very weak.

Australia, New Zealand and, Indian teams are the top three teams in Oceania region while they are 4th, third and 2nd at the ICC ranking. The New Zealand team has gained one rank in this region because their winning percentage is better in this region. The English cricket team has the sixth rank in this region because of lower batting strength while first at the ICC ranking. The Afghanistan cricket team is the newly cricket team they recently played a very good game and won a lot of matches in Asia region. They did not play many matches in the Oceania region, so they did not consider in this region.

**Table 4: Region-wise ranking for One-Day International (ODI) cricket teams**

Teams	Asia Region	Africa Region	America Region	Europe Region	Oceania Region	URWTR	Rank
PAK	213	215	197	208	189	204.4	5
IND	207	217	196	211	201	206.4	3
SL	201	214	201	203	194	202.6	7
BAN	168	147	171	174	169	165.8	10
AUS	198	221	229	213	211	214.4	1
ENG	197	207	200	226	196	205.2	4
NZ	191	212	193	212	208	203.2	6
SA	194	226	209	209	199	207.4	2
WI	170	206	211	205	197	197.8	8
ZIM	141	161	143	161	161	153.4	11
AFG	187	198	N/A	145	N/A	176.7	9

The graphical representation of the region-wise and URWTR ranking for ODI cricket teams is shown in Figure 3. The Australian cricket team has ever best ranking across the region and in the URWTR ranking.



**Figure 3: A graphical representation of region-wise teams ranking**

## 5. CONCLUSIONS & FUTURE WORK

In this paper, we have proposed a new method for the region-based teams ranking using the PageRank algorithm. A region-based teams 'strength is presented to identify stronger and weaker teams across the region. The proposed method has used several features for the teams batting, bowling and fielding strengths. The results of the work have been compared with the traditional ranking method. Experimental results showed that the proposed method generates better results across the regions.

In the future, we plan to propose a cluster-based prediction for a team. If a team is moved from his region to the other region, what will be the effect on their ranking? Similar studies can also be utilized to rank other sports as well.

## 6. ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (61872261).

## 7. REFERENCES

[1] A. Daud, F. Muhammad, H. Dawood, and H. Dawood, "Ranking Cricket Teams," pp. 1–17, 2015.

[2] P. Shah, "New performance measure in Cricket," vol. 4, no. 3, pp. 28–30, 2017.

[3] D. Bhattacharjee and D. G. Pahinkar, "Analysis of Performance of Bowlers using Combined Bowling Rate," vol. 06, no. 03, pp. 184–192, 2012.

[4] M. Bailey and S. R. Clarke, "Predicting the match outcome in one day international cricket matches, while the game is in progress.," *J. Sports Sci. Med.*, vol. 5, no. 4, pp. 480–487, 2006.

[5] P. Taylor and H. H. Lemmer, "Predicting the names of the best teams after the knock-out phase of a cricket series," *Eur. J. Sport Sci.*, no. September, pp. 37–41, 2014.

[6] P. Shah, "Measuring Fielding Performance in Cricket," vol. 23, no. 2, pp. 113–114, 2016.

[7] R. P. Bunker and F. Thabtah, "Applied Computing and Informatics A machine learning framework for sport result prediction," *Appl. Comput. Informatics*, 2017.

[8] H. Daud, A., Hussain, A., Abbasi, R. A., Aljohani, N. R., Amjad, T., & Dawood, "Region-wise Ranking of Sports Players based on Link Fusion," in *Companion of the web conference 2018*, 2018, no. 1, pp. 259–266.

[9] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank Citation Ranking: Bringing Order to the Web," *World Wide Web Internet Web Inf. Syst.*, vol. 54, no. 1999–66, pp. 1–17, 1998.

[10] X. Liu, J. Bollen, M. L. Nelson, and H. Van De Sompel, "Co-authorship networks in the digital library research community," *Inf. Process. Manag.*, vol. 41, no. 6, pp. 1462–1480, 2005.

[11] G. Iván and V. Grolmusz, "When the web meets the cell: Using personalized PageRank for analyzing protein interaction networks," *Bioinformatics*, vol. 27, no. 3, pp. 405–407, 2011.

[12] A. Y. Govan and C. D. Meyer, "Ranking National Football League Teams Using Google PageRank," *Work. Pap.*, vol. 8205, pp. 1–10, 2006.

[13] S. Mukherjee and B. Engineering, "Identifying the greatest team and captain – A complex network approach to cricket matches," vol. 391, no. 23, 2012.

[14] P. C. A. P, "Principal component analysis: Principal component analysis," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 2, no. 4, pp. 433–459, 2010.

[15] S. K. Madria, S. S. Bhowmick, W. K. Ng, and E. Lim, "Proceedings of the First International Conference DataWarehousing and Knowledge Discovery (DaWaK'99)," vol. 1676, no. May 2013, 1999.