3. SAMPLE TWITTER NETWORK AND ITS ANALYSIS

Social networking site is an online platform allows users to create a public profile and interact with other users on the website. Social network brings people together to share ideas and interests or make new friends. Social network analysis helps to understand the structure and properties of network. In this research, twitter network data is extracted from real time network using Twitter API and network analysis is carried out as an initial study. This chapter describes the process of twitter data extraction and network structure creation. Network analysis carried out with four most important properties such as closeness, degree, network density, betweenness is also elucidated in this chapter.

3.1 INTRODUCTION

The increasing prevalence of digital communications technology – the internet and mobile phones provides the possibility of analyzing human behavior at a level of detail previously unimaginable. The internet has generated many information sharing networks, the most well-known of which is the World Wide Web. Recently, a new class of information networks called online social networks has exploded in popularity and now rivals the traditional web in terms of usage. Social networking sites such as MySpace, Facebook, Orkut, and LinkedIn are examples of widely popular networks used to find and organize contacts. Other social networks such as Flickr, YouTube, and Google Video, are utilized to share multimedia content, and others such as LiveJournal and BlogSpot are used to share blogs [75].

Unlike the traditional web, which is largely planned by content, online social networks embody users as first-class entities. Users join a network, publish their own content, and make links to other users in the network called friends. This basic user-to-user link structure facilitates online interaction by providing a mechanism for organizing both real-world and virtual contacts, for finding other users with similar interests, and for locating content and knowledge that has been donated or endorsed by friends.

Blogs, content aggregation sites, internet fora, online social networks, and call data records provide access to data that vary by the second. These data require new tools to acquire process and analyze. These tools are no more difficult to study and utilize than other qualitative and quantitative methods, but they are not commonly taught to social scientists.

Social Network Service

A social network service is a web-based service providing a platform to build social networks or social relations. In social network services, users can register for account with public profile create a connection with other users to interact with other users to share common interest. Social network services also provide some additional services such as photo/video sharing, sending instant messages or emails. Users with common interests or backgrounds in social network services can also gather around in a group to communicate with each other. With the rapid development of social network services, users can communicate with others regardless of the borders of countries at any time.

Social network services are used by users to communicate with others. The registered users of social network services grow in a rapid speed and many commercial companies use social network services to promote their product or services as a kind of advertisement. Also, many researchers utilize the large amount of data created in the platform of social network services for research.

3.2 TWITTER NETWORK DATA

Twitter is one of the largest social networks, with billions of active users from almost every country and over $1 billion of annual revenue. Twitter is a real-time, highly social micro blogging service that allows users to post short status updates, called *tweets* that appear on timelines. Tweets may contain one or more entities in their 140 characters of content and reference one or more places that map to locations in the real world. An understanding of users, tweets, and timelines is particularly essential with effective use of Twitter's API,

Twitter is a public platform with millions of tweets created every day from billions of user accounts, which are of research and commercial values. Twitter developers can access to the data of Twitter via Twitter API. There are mainly three kinds of API in Twitter such as Search API, Streaming API and REST API.

The Twitter Search API allows Twitter developers to access to the data with specific keywords, tweets referencing a specific user or tweets from a specific user. It is important that the Search API focus on relevance and some data may be missing. The Streaming API can provide a dataset with relatively higher level of completeness. Twitter Streaming API provides real-time data with large amount for Twitter developers, which is different from REST API. In order to get access to Streaming API, Twitter developers need to build a long-lived HTTP connection and maintain it. In the process of utilizing Streaming API, developers

can also set up some criteria to collect data related to some keywords or geotagged tweets from a specific region. The major disadvantage of Streaming API is that it only provides a sample of the real-time data. The REST API provides developers with data related to a certain user, such as user timelines, status updates and other information. The REST API provides the history data of a certain user, different from the real-time data of Streaming API. Besides, developers can also interact with the platform by publishing tweets, following other users or updating the profiles of the user via REST API.

Twitter data can be acquired, processed, and analyzed using many programming languages, including R and Python.

A sample twitter network of a sport person is crawled and used in this research for investigations on various community detection approaches. The core data considered here are sports person's friends and followers list. In a directed network like twitter, the friends are the other users added by the sports person and the followers are the other users linked to the sports person. The process of crawling the sports person's network data from twitter is described below. The R packages namely twitter, devtools, Rcurls, igraph, RoAuth, thttr, base64encr are utilized for twitter network data crawling. The data is collected at run time from twitter network using R3.5.1, a statistical tool.

Twitter Network Data Crawling

A Twitter account is created, and an application is created with that account. The user account is then registered with Twitter developer's website. The application is connected to Twitter's API where an account can own multiple applications. Using the account, the credentials such as secret key and access token needed to access the Twitter API are obtained.

Using Python and R scripts authentication with seven-digit code is done to access into to developer's website. The virtual connection is then established between the user and Twitter API enabling the user to search for tweets, download users' tweets, downloads details of accounts' followers and friends. API allows downloading only 1% random sample of the Twitter stream that match language, geographic, account, or keyword parameters [76]. Twitter network data crawling process flow is shown in Fig.3.1.

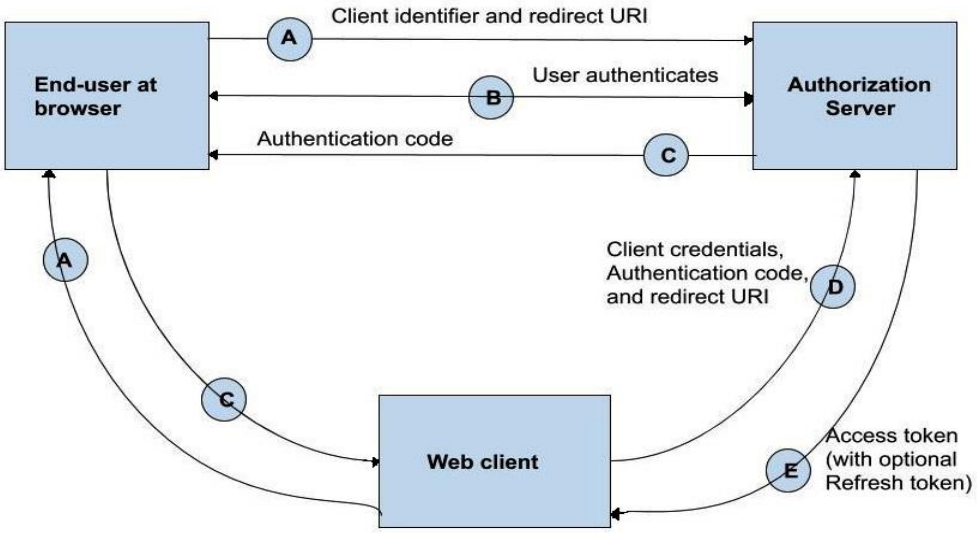

Fig. 3.1 Twitter Network Data Crawling Process

The process followed to crawl sports person's friends and followers is outlined below. Creation of Twitter account: With basic user's details and email account, an account is created at www.twitter.com.

Application creation: With user account and the website www.dev.twitter.com, by choosing the options "Developers" and "Documentation" and then selecting "Manage My Apps" the twitter account is accessed. Next by choosing the option "Create New Application" a new application is created. In the main page of application four navigation tabs are found such as "Details", "Settings", "Keys", "Access Tokens", and "Permissions". By selecting "Keys", "Access Tokens", the access tokens and the secret key are obtained. The screen shot of this process is shown in Fig. 3.2.

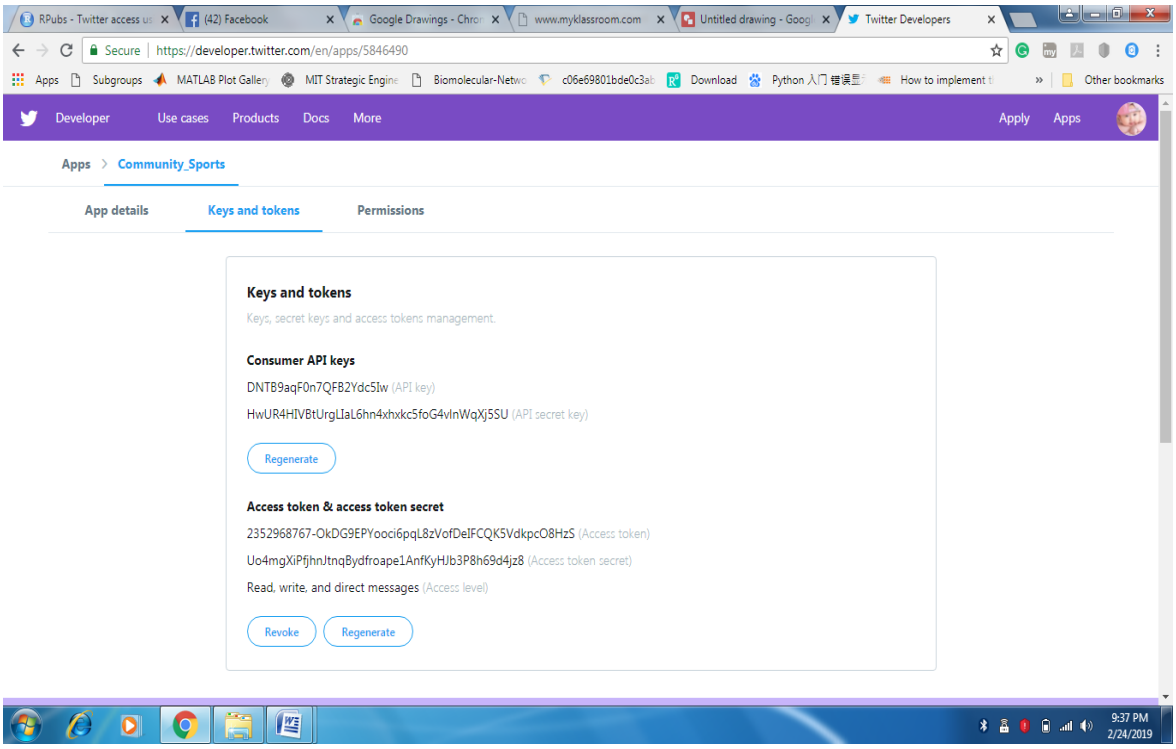

Fig. 3.2 Getting Access Token and Secret Keys in Twitter API

Authentication: The credentials obtained in the previous step are used in R script, which generates a Universal Resource Locator (URL). This URL is used in the browser to generate the seven-digit identification number. The id is then fed into Twitter Developer's website in order to establish a virtual connection between the account and the Twitter API. Now the environment is ready for accessing the data. The screenshots of this process are shown in Fig. 3.3 and Fig. 3.4.

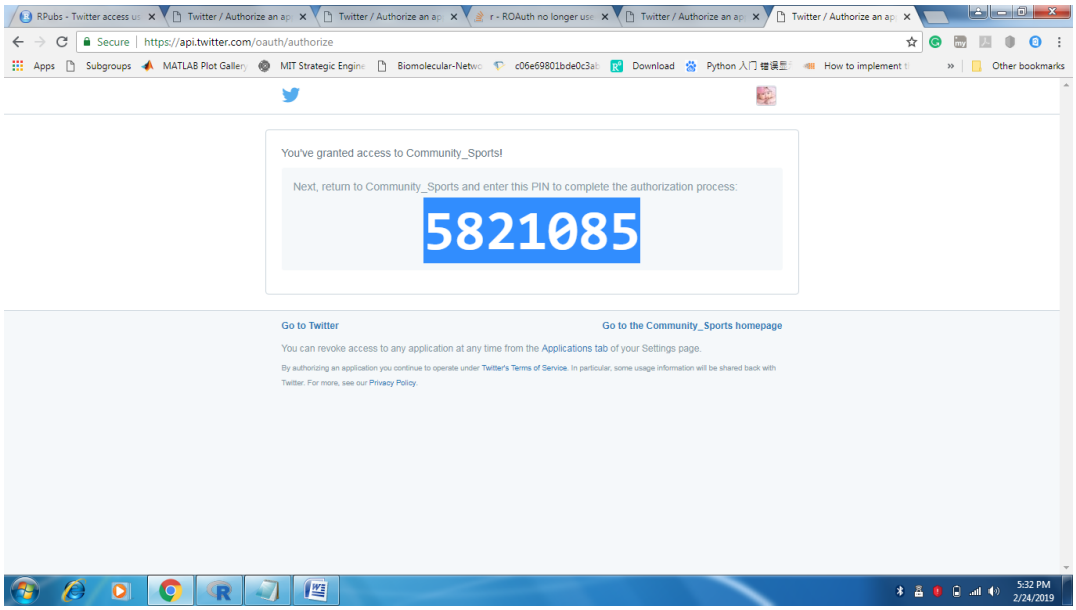

Fig. 3.3 Generation of Seven Digit Identification Number in Twitter API

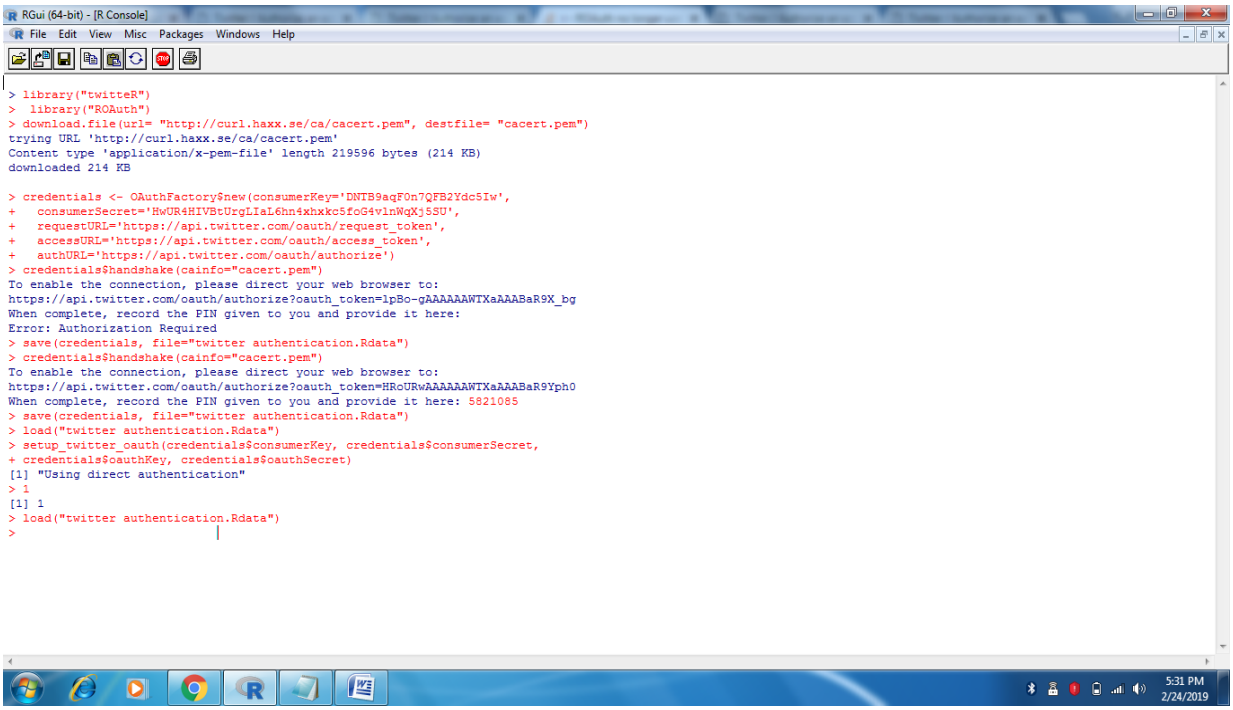

Fig. 3.4 Authorization and Virtual Connection to Twitter

Data extraction: The sports person twitter id is extracted from Twitter API and used in R script to extract friends and followers lists. Two endpoints, GET followers/list and GET followers/ids, provide information about followers. The former provides fully hydrated user objects for each follower at the rate of up to 15 followers per 15 minutes. The latter provides only the identification number of followers, but it does for 75,000 followers per 15 minutes. The required count is then fed to GET users/lookup, from which up to 18,000 completely hydrated user objects are returned every 15 minutes. GET followers/list, therefore, saves one step, but is slower than using GET followers/ids with GET users/lookup. The same logic holds for retrieving who a user follows. It connects to GET friends/ids instead of GET followers/ids, and these friend identification numbers are fed to GET users/lookup. In this work, GET followers/ids and GET friends/ids along with GET users/lookup are used to extract sports person's friends and followers lists. Only active and valid user's friends and followers lists are drawn. 19000 objects including both friends and followers are crawled and stored in .csv file. Twitter API uses JSON to pull data about Twitter users. The sample data crawled from Twitter is shown in Table IV.

Table IV Sample Twitter Data of Sports Person with Friends and Followers

USER	FRIENDS	USER	FOLLOWERS
imVkohli	Aaron Finch	ImVkohli	Ashish Jamwal
imVkohli	AB de Villiers	ImVkohli	ashish kumar
imVkohli	adidas	ImVkohli	ashish kumar raul
imVkohli	adidas Originals	ImVkohli	Ashish mishra
imVkohli	Amitabh Bachchan	ImVkohli	Ashish Tolmare
imVkohli	Aneesh Gautam	ImVkohli	ashok kumar srivasta
imVkohli	Blades of Glory	ImVkohli	Ashok Shejule
imVkohli	Bunty Sajdeh	ImVkohli	Ashok singh
imVkohli	Cristiano Ronaldo	ImVkohli	baba dom
imVkohli	ESPNcricinfo	ImVkohli	babar ali sabir
imVkohli	Gary Kirsten	ImVkohli	Babar Bilal
imVkohli	Harbhajan Singh	ImVkohli	Babasantosh
imVkohli	JP Duminy	ImVkohli	Babli kaparwan nodiy
imVkohli	Justin Timberlake	ImVkohli	Badinenijayanthkumar
imVkohli	KAVI KAVI	ImVkohli	Chaitanya
imVkohli	Kevin Pietersen	ImVkohli	Chanakaprasanna
imVkohli	Mahendra Singh Dhoni	ImVkohli	chandan Sharma
imVkohli	mark boucher	ImVkohli	Chandan Singh
imVkohli	Neha Dhupia	ImVkohli	Daison
imVkohli	praveen kumar	ImVkohli	DALIA ELBASSAL
imVkohli	Ritika	ImVkohli	Damith Indika
imVkohli	Ross Taylor	imVkohli	DAMODARREDDY
imVkohli	sachin tendulkar	imVkohli	danish.p
imVkohli	Ashwinravi99	imVkohli	Gaurav Mishra
imVkohli	Vijay Mallya	imVkohli	Gaurav shrivastava
imVkohli	yuvraj singh	imVkohli	Gaurav#max
imVkohli	Sachin Tendulkar	imVkohli	Rameshpawar

Twitter maintains user profile containing a rich set of information about user. Each Twitter user is identified by either a screen name (for example, imVkohli) or a unique integer

id. In, Twitter network, all the users relative to a particular user are labeled either as followers or friends. The set of friends' IDs are joined with the followers' IDs to determine the total set of unique IDs of Twitter users that either follow or are followed. The crawled data file has three fields labeled as ‘user’, ‘friends’ and ‘followers’. In Table IV, first column contains user, second column has user’s friends list and user’s followers are listed in fourth column. For example, the user ‘imVkohli’ has friends like yuvraj singh, sachintendulkar, Mahendra Singh Dhoni and users like Ashish mishra, Ashish Tolmare, ashok kumar srivasta etc. followers of ‘imVkohli’. Various meaningful representations of this data are portrayed in the following section.

3.3 NETWORK STRUCTURE OF TWITTER NETWORK DATA

A social network consists of a set of nodes or actors or vertices connected via some type of relations, which are called ties, links, arcs, or edges. The nodes usually represent actors, be that individuals, groups, teams, communities, organizations, political parties, or even nation-states. Generally, most of the SNA algorithms accept data in various modes such as network, graph, adjacency matrix, edge list. Visual presentation or visualization of social networks helps to understand network data and interpret the results of analysis.

Network data always consists of two types of datasets. A regular dataset called the nodelist wherein the nodes are the units of observation or rows. Another dataset contains the relationships among the units of observation i.e. nodes which is represented as adjacency matrix or an edge list. In an adjacency matrix or network matrix, the nodes constitute both the rows and the columns, and the cells specify the relationship that exists between the nodes in the row and in the column. An edge list is a dataset containing the edges of the graph.

A real-time Twitter data extracted as described in section 3.2 from Twitter API is represented as network structure and used for analysis. The raw data is converted into the data frame and then it is transformed into (a) graph, (b) adjacency matrix (c) node list (d) edgelist using R script. Except node list all three representations of twitter data have been used in implementation of various community detection approaches explored in this research work.

The network structure for the sports person’s twitter data with friends and followers given in Table IV is depicted in Fig.3.5 and the corresponding adjacency matrix is given in Table V. The sports person’s initial community network shows that the network has 7095 edges and 6831 vertices. The sample twitter data includes 3925 friends and 15075 followers

in the network. The principle node in the network is imVkohli and others are friends and followers. Few nodes are very popular in the network. The most popular node is imVhohli and has largest number of friends and followers in the network. The top influence nodes are imVkohli, Sachin Tendulkar, Suresh raina, ashwinrave99, Virender Sehwag, Mahendra Singh Dhoni, Harbhajan Singh, Gary Kirsten and yuvraj singh in the network. The most interactive nodes, imVkohli, Sachin Tendulkar, Suresh raina, ashwinrave99, Virender Sehwag, Mahendra Singh Dhoni, Harbhajan Singh, Gary Kirsten and yuvraj singh as the corresponding nodes have 1114, 432, 342, 338, 354, 347, 412, 321, 265 of friends and 3966, 2010, 1168, 1543, 1290, 1252, 1289, 1190, 1367 followers in the network. The relationship of users with friends and followers is depicted in Fig. 3.6.

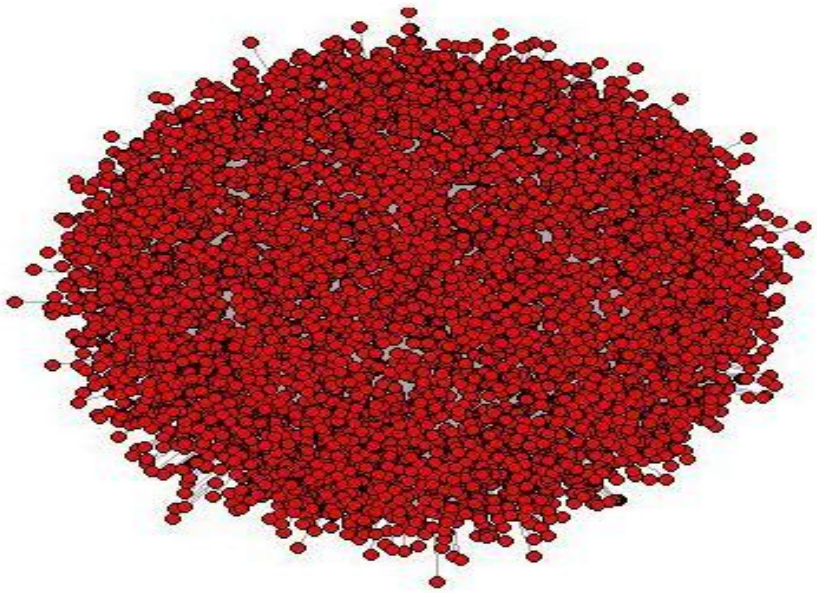

Fig. 3.5 Sports Person's Network with Friends and Followers

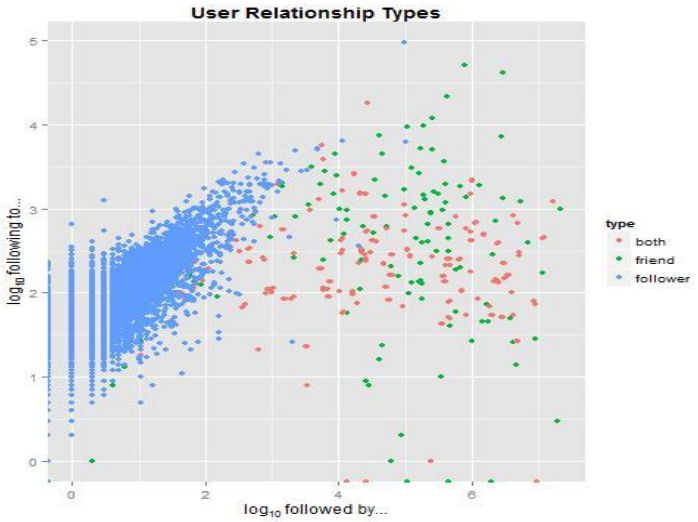

Fig. 3.6 Relationship of Users with Friends and Followers

There is a strong relation between graphs and matrices. An adjacency matrix is a square matrix utilized to represent a finite graph. The elements of the matrix specify whether pairs of vertices are adjacent or not, in the graph. In the unique case of a finite simple graph, the adjacency matrix is a (0, 1) matrix with zeros on its diagonal. Two vertices *i* and *j* of a directed graph are adjacent if there is an edge from *i* to *j* or from *j* and *i*. If such an edge exists, then *i* and *j* are its endpoints. If there is an edge from *i* to *j* then *i* is often entitled tail, while *j* is called head. It is indicated as 1 in the respective cell of the matrix, otherwise 0. Notice that there can be no more than two edges between any two vertices.

Table V Adjacency Matrix of Twitter Network Data

	Suresh Raina	Aaron Finch	AB de Villiers	Mahendra Singh Dhoni	Ishant Sharma	John Abraham	Kartik Murali
imVkohli	1	1	1	1	1	1	1
Aaron Finch	0	1	1	1	0	0	1
AB de Villiers	1	0	1	1	1	0	1
Adidas	1	0	0	0	1	0	0
adidas Originals	1	1	1	0	0	1	1
Amitabh Bachchan	1	0	0	1	0	0	0
Aneesh Gautam	0	0	0	1	0	0	0
Blades Of Glory	0	1	1	0	0	1	0
Bunty Sajdeh	1	1	1	0	0	1	1
Cristiano Ronaldo	1	1	1	0	0	1	1
ESPNcricinfo	1	1	0	1	1	1	1
Gary Kirsten	1	1	1	1	1	1	1
Harbhajan Singh	1	0	0	1	1	1	1
JP Duminy	1	0	0	1	1	1	1
Justin Timberlake	0	1	1	1	1	1	1
sachin tendulkar	1	0	1	1	1	1	1
Ashwinravi99	1	0	1	1	1	1	1
Vijay Mallya	1	0	1	1	1	1	1
yuvraj singh	1	0	1	1	1	1	1
Sachin Tendulkar	1	0	1	1	1	1	1

In Table V, vertices imVkohli and suresh raina are linked because there is an edge from imVkohli to all, while vertices Justin and suresh raina are not linked. There is no edge from node suresh raina to node justin. The node list and the edge list for the same data are given in Table VI and Table VII.

NodeList is an object that consists of a list of all of nodes in a network. NodelList representation of network data is commonly used in most of the graph theory techniques and also used in clique percolation method. The node A in NodeList shown in Table VI has two vertices, such as msdhoni, Amit Mishra and node B has ashwinravi99, abhinav mukund.

Table VI Sample Nodelist of Twitter Data

Node	Friends/Followers	Friends/Followers
A	Msdhoni	Amit Mishra
B	ashwinravi99	abhinav mukund
C	A. R. Rahman	Gautam Gambhir
D	A. T. Rajamani Prabhu	Harbhajan Singh
E	AB de Villiers	Ishant Sharma
F	abhinav mukund	John Abraham
G	Achuuuuu	Kartik Murali
H	ajinkyarahane88	Rudra Pratap Singh
I	aparajith baba	sachin tendulkar
J	Arun Jaitley	Sreesanth
K	yuvraj singh	Ashish Jamwal
L	Sachin Tendulkar	ashish kumar
M	Suresh raina	ashish kumar raul
N	Gary Kirsten	Ashish mishra
O	Harbhajan Singh	Gaurav Mishra
P	JP Duminy	Gaurav shrivastava

An edgelist is a list or array of all of the edges E in a graph. Edgelists are one of the simplest representations of a graph. The entries in the list are of the form of (x, y), when there is an edge from x to y. For example, first row in Table VII shows there are edges from imVkohli to Msdhoni and imVkohli to Amit Mishra.

Table VII Sample EdgeList of Twitter Data

User/Index	Vertex 1	Vertex 2
imVkohli	Msdhoni	Amit Mishra
imVkohli	ashwinravi99	Ashwin Ravichandran
imVkohli	A. R. Rahman	Gautam Gambhir
imVkohli	A. T. Rajamani Prabhu	Harbhajan Singh
imVkohli	AB de Villiers	Ishant Sharma
imVkohli	abhinav mukund	John Abraham
imVkohli	Achuuuuu	Kartik Murali
imVkohli	ajinkyarahane88	Rudra Pratap Singh
imVkohli	aparajith baba	sachin tendulkar
imVkohli	Arun Jaitley	Sreesanth
imVkohli	praveen kumar	DALIA ELBASSAL
imVkohli	Ritika	Damith Indika
imVkohli	Ross Taylor	DAMODARREDDY
imVkohli	sachin tendulkar	danish.p
imVkohli	Ashwinravi99	Gaurav Mishra
imVkohli	Vijay Mallya	Gaurav shrivastava
imVkohli	yuvraj singh	Gaurav#max
imVkohli	Sachin Tendulkar	Rameshpawar

Therefore, for the sample network taken for study, the adjacency matrix is 2-dimensional array which has the size 15653 x 15653. The nodelist contains 14765 instances and edgelist hold 18654 instances.

3.4 NETWORK ANALYSIS OF SAMPLE TWITTER DATA

The study of social networks for behavior analysis of actors involves two aspects; (a) the use of formal theory organized on the root of mathematical conventions and (b) the pragmatic analysis of network data as quantified by various social network analysis metrics. It is understood that social network metrics play an important role in SNA. These metrics have different meanings in different types of networks. This research work investigates various social network metrics for the sample network drawn. This section illustrates the network analysis of the sample twitter network taken into account along with the metrics and the inferences drawn about the network.

The important graph measures such as modularity, graph energy, the number of loops, self-loops, vertex eccentricity, link density and maximal clique lengths are determined as below.

Modularity =	0.083
Graph energy =	177.7514
Number of loops =	8555
Number of self-loop	13
Vertex eccentricity	12
Link density	1.1949
Max-clique	121

Centrality is a measure of the information about the relative importance of nodes and edges in a graph. Centrality measures are used to depict the level of importance of a given node in relation to other nodes in a network or community. Centrality determines like Degree Centrality, Closeness Centrality, Betweenness Centrality, Eigenvector Centrality, Katz Centrality, and Alpha Centrality play a significant role in graph theory and network analysis to measure the importance or prestige of actors or nodes in a network. Definitions of some centrality measures and the corresponding results are stated below.

Degree Centrality: It is the simplest of all the centrality calculates and its value for a given node in the network is the number of links incident on it and is used to identify nodes that have the highest number of connections in the network. It does not take into account the centrality or prestige of the incident nodes. For a graph G=(V,E), the degree of a node vertex $v, (v \in V)$ is expressed using Eq.3.1.

$$C_D(v) = deg(v) \tag{3.1}$$

where deg(v) is the number of edges incident on the vertex *v*. For entire graph *G*, the Degree Centrality is expressed using Eq.3.2.

$$C_D(G) = \frac{\sum_{i=1}^{|V|}[C_D(v*) - C_D(v_i)]}{H} \tag{3.2}$$

where $v*$ is the node in G with highest degree centrality and $H = \sum_{i=1}^{|Y|} C_D(y*) - C_D(y_j)$, where $y*$ be the node with the highest degree centrality in a graph x of G with y nodes. The value of H is maximum when a graph has a star like structure.

Eigenvector Centrality: A more sophisticated edition of degree centrality is eigenvector centrality. It not only depends on the number of occurrence links but also the quality of those links. This means that connections with high prestige nodes contribute to the centrality value of the node [77].

Closeness Centrality: The degree of nearness between any node and the rest of the nodes in the network is represented by closeness centrality. It is the inverse of sum of the shortest distance between a node and rest of all in the network. For a graph *G* with *n* nodes, the closeness centrality of a node *v* is expressed using Eq.3.3.

$$C_C(v) = \frac{n-1}{\sum_{k=i,}^{n} d(u_i, v)} \tag{3.3}$$

where $d(u_i, v)$ denotes the geodesic distance between u_i and v.

Betweenness Centrality: In order to recognize the leaders in the network, the amount of interest in many social network studies is the Betweenness Centrality of an actor. Betweenness Centrality measures the tendency of a node found along the shortest path between two other nodes. It measures the fraction of all shortest paths that pass-through a given node i.e. it quantifies the number of times a node acts as a bridge along the shortest path between two nodes. A node with high BC is important in a network because it serves as an important route for information flow in that network. It means that removal of such node will either collapse the network or weaken the network considerably [77].

Clustering Coefficient: Clustering coefficient is a measure of the ability of a node's neighbor to form a complete graph, also called a clique. The value of the clustering coefficient is directly proportional to the degree of connectedness of the neighbors of that node, the more the connections among the neighbors, the higher the clustering coefficient. The clustering co-efficient of a network is the average of the clustering co-efficient of all the nodes in the network. The average clustering coefficient is expressed using Eq.3.4 as follows:

$$\overline{C} = \frac{1}{n}\sum_{i=1}^{n} C_i \tag{3.4}$$

where $C_i = \lambda_G(v)/T_G(v), \lambda_G$ is the number of subgraphs G having 2 edges and 3 vertices including the vertex v. $T_G(v)$ is a number of subgraphs G having 2 edges and 3 vertices including v such that v is incident on both edges.

Average Degree: The number of vertices adjacent to a vertex v is called as the degree of v or deg(v). Based on this measure a maximum degree, minimum degree or average degree can be

obtained. The average degree of a graph is a network level measure and it is calculated from the value of a degree or all the nodes in the network. For a graph G with V vertices and E edges the average degree of G is expressed using Eq.3.5.

$$D_A(G) = \frac{2 \times |E|}{|V|} \tag{3.5}$$

Density: The Density of a graph quantifies the number of connections between various actors in the network. The graph is considered dense if the number of edges in the graph approaches the maximal number of edges which a graph can have and sparse otherwise [77]. For an undirected or directed graph G with V vertices and E edges, the density of G is expressed using Eq.3.6 as follows:

$$D_G = \frac{2|E|}{|V|(|V|-1)} \tag{3.6}$$

The minimal density is 0 and the maximal density is 1 for complete graphs.

Various centrality measures such as degree, closeness and betweenness are evaluated and analysis of the social network is carried out. These centrality measures are evaluated with various properties like minimum and maximum values of in-degree, out-degree, total degree, in-closeness, out-closeness, total closeness and betweenness. Three centrality measures such as degree, closeness and betweenness are evaluated for the network presented in section 3.3 using R script.

The degree centrality represents the number of connections that a particular node. Since twitter network is a directed graph, a node has both in-degree and out-degree. The out-degree is the number of arcs from a node to other nodes and it is 95 for this network. The in-degree is the number of arcs coming into a node from other nodes and it is 7000 on the same network. The total degree centrality measure is 7095. The histogram representation of in-degree, out-degree and total degree measurement for the cricket player's network is shown in Fig.3.7a and Fig.3.7b. Average degree is determined based on in-degree distribution, out-degree distribution and total degree distribution and average degree of this dataset is found as 7095. The minimum and maximum values of in-degree and out-degree measures are given in Table VIII.

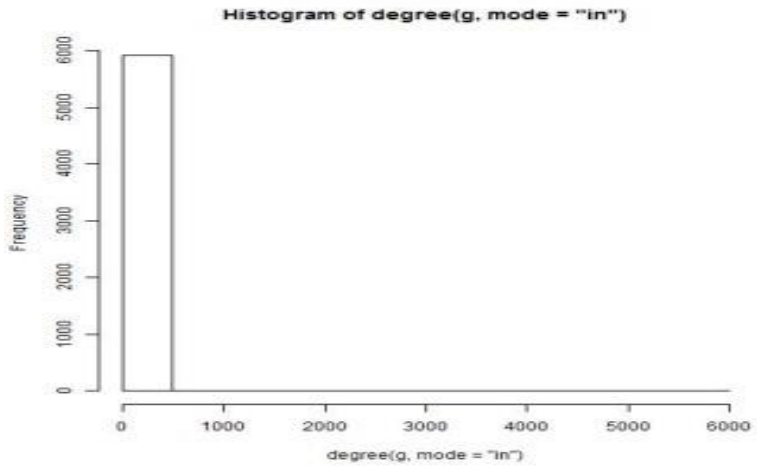

Fig. 3.7a In-Degree of given Network

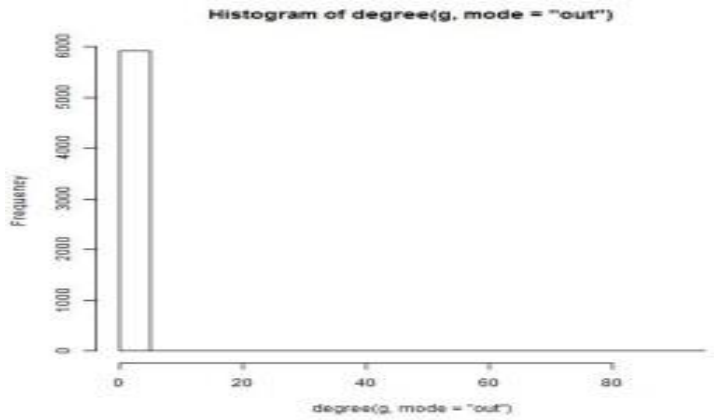

Fig. 3.7b Out-Degree of given Network

Similarly, the closeness centrality measure is evaluated for the same directed graph. The closeness measure represents the shortest path between nodes connected with it. Out-closeness is 2.17E-08 and in-closeness is 1.61E-06 for the same network. The total closeness centrality is 2.17E-08. Fig.3.8a and Fig.3.8b displays the histogram representation of in-closeness and out-closeness of this network. The minimum and maximum values of in-closeness and out-closeness measures are given in Table VIII.

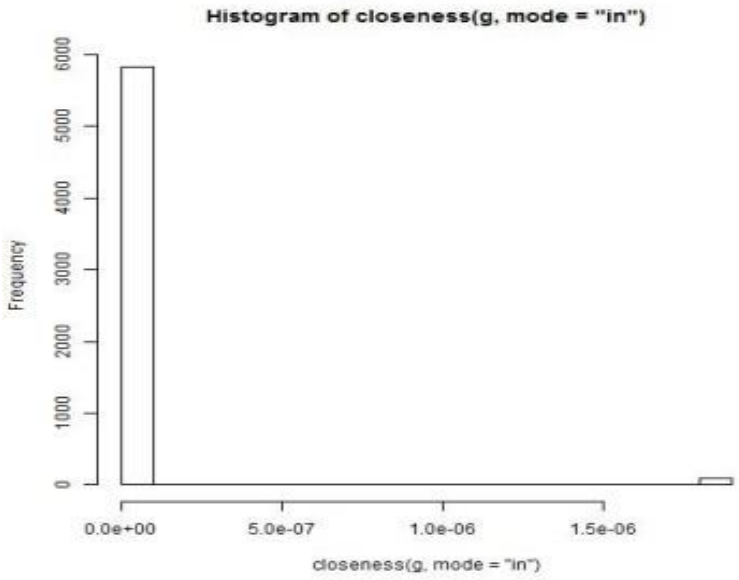

Fig. 3.8a In-Closeness of given Network

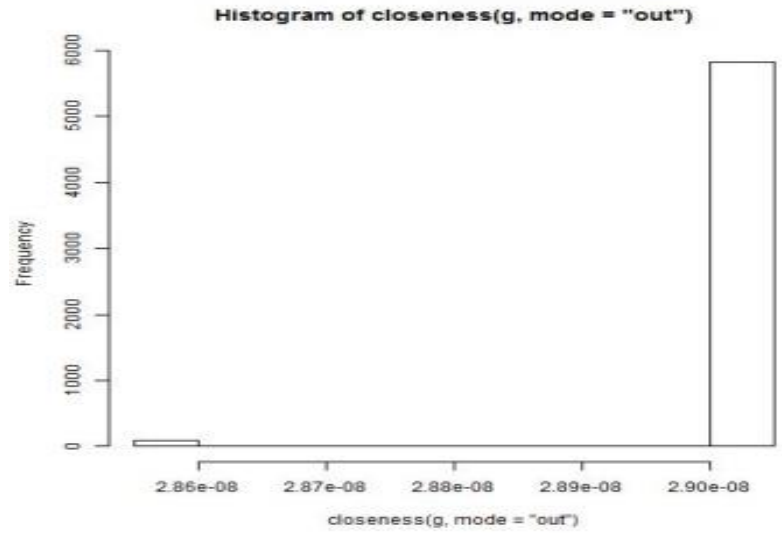

Fig. 3.8b Out-Closeness of given Network

The minimum and maximum values of betweenness measures are computed for this sports person's network in a similar manner and the values are given in Table VIII.

Table VIII Measures of Sample Twitter Network Data

	Community Detection Measures						
Measures / Limitations	**Degree**			**Closeness**			**Betweenness**
	In	**Out**	**Total degree**	**In**	**Out**	**Total Closeness**	
Min	1	1	1	2.14E-08	2.14E-08	2.14E-08	1
Max	7000	95	7095	1.61E-06	2.17E-08	2.17E-08	640295

Eigenvector centrality is calculated as the influence of a node in a network. Eigenvector centrality is an addition of degree centrality. The degree centrality of a node is simply the total number of nodes that are connected, whereas eigenvector centrality not only considers the total number of adjacent nodes but also considers the importance of the adjacent nodes. That is, connections with an influenced person will lend a person more influence than a connection with less influenced persons. Eigenvector centrality is calculated for nodes of the network with the greatest connectivity. Eigenvector centrality scores of a few nodes of the sports person's network are given in Table IX.

From the results, it has been found that user1 has high eigenvector score which indicates that the node has many connections, which in turn have many inter-connections in the network. Graph density of the sample twitter network, a directed graph, is also measured using equation 3.6 and it is observed as.0.786.

Table IX Sample Value of Eigen Centrality

S. No	User	Eigenvector Centrality
1	imVkohli	1.0000000
2	Msdhoni	0.9168980
3	ashwinravi99	0.8358122
4	sachin tendulkar	0.7788376
5	Virender Sehwag	0.6538709

Findings

From this analysis, it is investigated that, the out-degree is 95 and in-degree is 7000 for the sports person's network. An entity or node with high degree centrality shows that the node is an active player, hub and having an advantaged position in the network. Since closeness centrality is low the node has slow interaction to other entities in a network. Also, the node is in a powerful position and a better influence over the other nodes in the network because the betweenness centrality is 640295 for this network. The user with high eigen vector centrality has more inter-connections in the network. The node 1 is an active node with high graph density. Hence it is a prominent node in the network.

SUMMARY

Twitter network data has been drawn for the purpose of the research and the process of crawling network data from Twitter API is illustrated in this chapter. The four most important properties employed in social network analysis such as closeness, network density, degree, and betweenness are studied and the basic network analysis carried out for the sample twitter network data with various metrics is also elucidated with results and charts. From this basic experiment, the thesis proceeds with implementation of various community detection algorithms and are presented in the subsequent chapters 4 to 7.