International Journal of Engineering Research ISSN:2319-6890)(online),2347-5013(print)
Volume No.5 Issue: Special 4, pp: 790-991 20 May 2016

Forest fire model proposal for match fixing in cricket based on criminal network analysis

Dr.K.Udayakumar[1] K.Ananthapadmanabha[2]

[1]Principal and Professor, Department of Computer Science and Engineering
Adarsha Institute of Technology, Bangalore, India
[2,]Associate Professor, Department of Computer Science and Engineering
Adarsha Institute of Technology, Bangalore, India
siridevikap@yahoo.co.in

***Abstract*—In cricket, match fixing occurs when the match is placed to completely or partially to predetermine the result of the game for favors and kickbacks received by violating the rules of the game and law of the land. Also match fixing is possible due to unlawful nexus between players, officials, bookmakers, middlemen and gang leaders. This nexus leads to a small criminal social network. In this paper to analyze the match fixing in cricket explore of exemplar areas of mining on social networks using link mining technique. Key tasks and challenges faced in this area are highlighted. A graph model is used to represent the network and its related features.**

Keywords—match fixing, social networks, link mining, cricket

I. INTRODUCTION

Dark moments in world cricket include tours to south Africa during apartheid era, body line series of 1930's, pyjama cricket called packer series in Australia, ball tampering rows and the recent one being that of match fixing in cricket

In organized sports like cricket match fixing is done by violating sports man spirit of the game. It is possible due to unholy relationship between cricketers, umpires, officials, bookies, gangsters and middle man. This network is represented as[1]

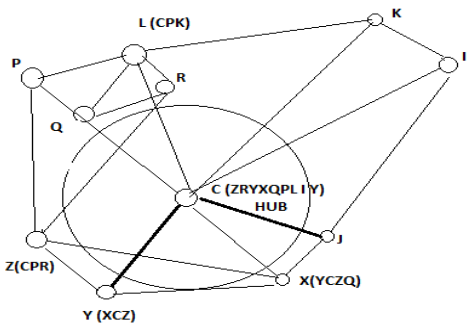


Figure 1: Match fixing network analysis

At the heart of cricket worst problem is the status of gambling in Indian subcontinent. Cricket being a religion in India with more than one billion fans following and billions of dollars invested on cricket in India. Without the option of legitimate betting, gamblers in cricket call on the services of illicit bookmakers who are in the pockets of gangsters.

II. MATCH FIXING NETWORK ANALYSIS[3]

Human investigations can analyze match fixing data and can predict perfectly but this takes more time effort and cost. Also it is error prone. It provides criminals enough time and space to destroy evidence. Therefore there is a need to analyze match fixing in cricket based on data mining technique like association, classification, clustering, prediction, outlier analysis, correlation analysis and link analysis to improve efficiency of prediction and reduce analysis time.

Investigators can use techniques like CART, CHAID, ID3, Neural network, genetic algorithm and decision tree to predict match fixing by analyzing relevant factors.

A. N Dimensional space

N dimensional space in clustering and nearest neighbor is a space defined to indicate what is near and what is for away and here distance can be calculated. It is similar to three dimensional spaces in which distance between objects is defined by Euclidean distance given by the formula

$$D_{(x,y)} = \sum ((x_i - y_i)^2)^{1/2}$$

Real world problems like match fixing consists of n dimension where each predictor that is used can be considered to be a new dimension. In cricket different predictors are batting average, bowling average, batting strike rate, bowling strike rate, number of over's bowled, number of runs scored, average number of balls bowled to take a wicket.

ISSN:2319-6890)(online),2347-5013(print)
20 May 2016

B. *Link analysis*[6]

Link mining deals with research in social networks, link analysis, hyper text and web mining, graph mining, relational learning and inductive logic programming. It includes tasks like link based object classification, object type prediction, link type prediction, link existence prediction, link cardinality estimation, object reconciliation, sub graph identification and meta data mining.

C. *Task and challenges in link mining*

- Link based object classification

  It is used to predict category of object based on attributes of objects, links and attributes of linked objects.

- Object type prediction

  It is used to predict type of the object based on its attributes and its links and on the attributes of of objects linked to it.

- Link type prediction

  It predicts the type or purpose of the link based on properties of the objects. In our application given cricket match data to predict whether a player is match fixer or not.

- Predicting link existence

  Here we try to predict whether a link exists between two objects. Example if a captain of a cricket team is involved in match fixing we try to predict the links that exists between captain and other players.

- Link cardinality estimation

  Here the link cardinality estimation is the number of links to an object. This is called in links. Example if captain is an object, the number of links he has will tell number of players who may be suspected of match fixing along with the captain

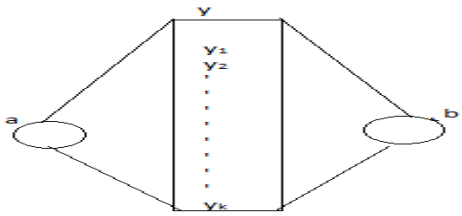
III. GRAPH MINING

Graph represents a more general class of data structures than sets, sequences, lattices and trees. They are one of the most suitable data structures to represent networks. Graph mining is used to mine frequent graph patterns and perform characterization, discrimination, classification and cluster analysis over large data sets. They can also be used to mine sub graph. There are two methods for sub graph merging procedure namely vertex growing and edge growing procedures. In vertex growing methods new candidates are generated by adding a new vertex to an existing frequent sub graph. In edge growing procedure new candidates are generated by adding a new edge to an existing frequent sub graph provided the edge or link exists.

IV. SOCIAL NETWORKS[4]

It is a heterogeneous and multi relational data sets represented by a graph. The graph is very large with nodes of the graph corresponding to objects and edges corresponding to links representing relationship or interaction between objects. Both nodes and links have attributes. Small world social networks reflect the concept of small world which is focused on network among individuals.

A criminal network is created due to nexus between cricketers, umpires, officials, middlemen, bookies and gangsters for match fixing in cricket. Here there will be lot of indirect relationship or association between people involved in the network.



V. INDIRECT ASSOCIATION FOR MATCH FIXING IN CRICKET[4]

Figure 2: Indirect association for match fixing in cricket

An indirect association between a pair of items a, b is indirectly associated via a mediator set y if the following conditions holds good.

1. $S(a,b) < t_s$ This is called itempair support condition
2. $\exists\ y \neq 0$ such that
   a) $S(a) \cup y >= t_f$ and $S(b) \cup y >= t_f$ this is called mediator support condition.
   b) $d(a,y) >= t_d$, $d(b,y) >= t_d$ where $d(x,z)$ is an objective measure of the association between x and z. This is called mediator dependence condition.

An algorithm called indirect association is as follows:

doi : 10.17950/ijer/v5i4/016

ALGORITHM indirect association( )[4]

1. Generate F_k, the set of frequent item set
2. For K=2 to K_{max} do
3. C_k = { a,b,y}/{a}∪y ∈ F_k, {b}∪ y∈F_k, a≠b
4. For each candidate(a,b,y) ∈C_k do
5. If S(a,b)< t_s ^ d([a},y)>=t_d ^ d({b}y) >= t_d then
6. I_k = I_k ∪{(a,b,y)}
7. Endif
8. End for
9. End for
10. Result = ∪I_k

Research on social networks highlight key features like

1. There appears to be a universal "six degrees of separation" between any two individuals in the world.
2. Many social, physical, human designed networks exhibit small world characteristics.
3. Research interest in mining social networks is to understand their structure and topology which helps in identifying their mode of operandi and behavior.
4. Criminal small world match fixing networks are not social in network. They are criminal in the context. The aim of those networks is to fix matches to suit their requirement. Hence there is a need to understand

- The network structure and topology
- The individual nodes that constitute the network, their mode of operandi, their relationship with other nodes and type of relationship like direct relationship or indirect relationship.
- The links of the networks represented by the edges of the graph indicates their contacts and behavior.
- Searching criminal social networks can help to understand how to reach other people in the network involved in match fixing. Also it helps to design smarter search agents on the network which can find relevant information about networks and nodes of networks all of which are within smaller number of degrees of separation.

VI. CHARACTERISTICS OF SOCIAL NETWORKS[4]

1. The nodes degrees which indicate number of edges incident to each node.
2. The distance between a pair of nodes representing shortest path length.
3. Individuals on the network are linked via short chains.
4. Network diameter indicates maximum distance between pair of nodes.
5. Effective distance indicates the minimum distance for at least 90% of reachable node pairs.
6. Dynamic nature of social networks indicates nodes or edges can be added or deleted with time. They are not static.
7. They follow densification power law. Which states that e(f) ∞ n (t)[a]

Here e(f) indicates number of edges at time t. n(t) indicates number of nodes at time t. Exponent a strictly lies between 1 and 2. If a=1, this corresponds to constant average degree over time. If a=2, it indicates it is an extremely dense graph.

VII. FOREST FIRE MODEL PROPOSAL[6]

A forest fire model for graph generation is proposed to capture the characteristics of graph evolution overtime. Here new note attached to the network by burning through existing edges. The parameters used are

1. Forward burning probability P
2. Backward burning ratio q

Suppose a new node v arrives at time t, it attaches to graph G through the following steps.

Step 1: it chooses an ambassador node W at random and form a link to w.

Step 2: It selects a link incident to w where x is a random number with binomial distribution with mean $(1-p)^{-1}$. It chooses from out links and in links of w but selects in links with probability r times lower than out link. Let w_1…….w_z denote the nodes at the other end of the selected edges.

Step 3: The new node v forms out link to w_1…….w_z and then applies step 2 recursively to each w_1……w_z. Nodes cannot be visited second time for preventing cycle formation. The process continues until it dies down.

VIII. CONCLUSION

Match fixing in cricket can be represented by a social network with criminal intensions. In this work future trends in match fixing are described using in degree and out degree concepts. Social network includes indirect

ISSN:2319-6890)(online),2347-5013(print)
20 May 2016

association. This issue has been addressed thoroughly. Special types of match fixing like spot fixing and session fixing can be considered for future work. Also community mining and multi relational mining for large data bases can be applied for match fixing in cricket using graph mining concepts.

ACKNOWLEDGMENT

The authors wish to thank the management of Adarsha institute of technology for constant support and encouragement. Special thanks to Dr.P.V.Krupakara for his constant support.

REFERENCES

[1] G.K.Gupta, "Introduction to Data Mining with case studies", PHI Learning Private Limited.
[1] Crickinfo.com
[2] Crickbuzz.com
[3] Espnstarsports.com
[2] Alex Berson , Stephan J. Smith, "Data Ware Housing, Data Mining and OLAP", Tata McGraw Hill Edition publication
[3] R.V.Hauck et.al, "Using coplink to analyze criminal justice data", Computer, March, 2002,30-37
[4] Pang-Ning Tan, Michael Steinback and Vipin Kumar, "introduction to Data Mining", Pearson education. 2015
[5] Arun K pujari, "Data Mining Techniques", 2nd edition, Universities press, 2009
[6] Jaiwei Han and Micheline Kamber, "Data Mining Concepts and techniques" 2nd edition, Morgan Kaufman Publishers, 2006
[7] Web site visited