Fraud Detection in Stock Market Using Collusion Clustering Algorithm

[1]Archana K, [2]Iris T and [3]Mercy W

[1,2,3]Department of CSE, KCG College of Technology, Chennai, India.
[1]archanaa.2923@gmail.com, [2]iristhomas200@gmail.com,
[3]mercy.cse@kcgcollege.com

Article Info
Jenitta J and Swetha Rani L (eds.), *International Conference on VLSI, Communications and Computer Communication*, Advances in Intelligent Systems and Technologies,
Doi: https://doi.org/10.53759/aist/978-9914-9946-1-2_3

Abstract-Fraud is one of the major problems around the world today. Example: Stock market. In this paper, circular trading is detected using graphs. We detect the fraud using graph and machine learning. To resolve this problem, different fraud detection algorithm and approaches were used. In our proposed system, we are using an algorithm called collusion clustering, which is particularly designed for finding fraudulent sets. This algorithm will detect the fraud more efficiently convey the result.

Keywords-Fraud, Collusion clustering, Machine learning, collusion sets, Circular trading, stock market.

I. INTRODUCTION

There will be a set of traders who perform illegal tradingamong themselves than with others. They tend to increase the stock price by spreading counterfeit information which will generate false demand in the market for the purpose of gaining profit. In circular trading, the fraud can beappropriately detected when the transactions are apposite aggregated together. The main objective of this project is to detect the collusion sets in the stock market using graph clustering [1]. Collusion is a strictly confidential concurrencebetween the interested traders for an illegal purpose. Collusion takes place when the traders combine together forthe purpose of their own gain. A set of traders create an artificial demand in any trading platforms like stock market and trade among themselves and this process is known as circular trading [13]. A stock portrays the proprietorship of an affliction. A stock of a product is hiked when there is a high demand for the product. There are a number of misconducts that occurs in stock market. There are many frauds that takeplace in stock market like price propping, price hammering, price manipulation. But in this paper, we focus on specific malpractice called circular trading [12].

II. LITERATURE SURVEY

A set of traders create an artificial demand in any trading platforms like stock market and trade among themselves which is known as circular trading. The stock market defines a secureand ordinance mart where the traders reach an agreement on prices and investments. A class of single or association strive to contrive or strike the movement on exchanging goods for the purpose of gaining benefit through illicit activities. This is an important part of a problem that is to be viewed seriously [2].

Circular trading itself is a collusion based mal practice. This is because, thus incompetence call for class of buyers and traders who combine together and trade to attain an obvious target. Several collusion-based artifices vary with each other in the trading scheme during collusion.

To detect this malpractice, an algorithm called collusion clustering is used. It focuses on several trading related frauds and also involves different types of fraudulent pricingtechniques [11]. Different algorithm used which is more beneficial and productive in aiming at Collusion sets and upgrade the precision of detection. The main difficulty faced inCollusion based malpractice is that, the substantiation for such frauds can be properly found only when the transactions are combined the substantiation for such frauds can be properly found only when the transactions are combined together. Otherwise, the data are concealed in the trading summary.

We develop an algorithm to detect the group of colluding dealers who do heavy illegitimate trading among themselves. Though we have few algorithms for collusion set detection, weare implementing collusion clustering algorithm. We also haveapproaches like Shared nearest neighbor algorithm and mutualnearest neighbor algorithm- but these approaches don't detect effectively [10]. So, we are implementing collusion clustering algorithm. This is implemented to overcome the shared and mutual nearest neighbor algorithm [9].

III. PROPOSED SOLUTION

To detect this circular trading malpractice, this system uses the approach called collusion clustering algorithm. We have improvised the collusion clustering algorithm with modern graph analytics which helped to identify and interpret the fraudulent connections and transactions on the real SEBI dataset, the improvised algorithm is scalable on big data and reduce the execution time. Our proposed system uses the collusion clustering algorithm. Using this approach, we can identify the frauds and malpractices that occurs in the stock market. We have collected transaction data from SEBI [8].

The trading summary in the .csv format has all the data of traders i.e the transactions between all the traders in the stock market. We have taken the trading summary file and start reading the transactions [3]. We have large collection of data available or stored in the data frame but we trim the data and choose only few top data. After reading transactions from the data frame, graph building and visualization takes place with the same data. We use matplotlib package for graph visualization and network for building graph [4]. The directional graph indicates that nodes and edges based on the transactions and are iterated every time during transactions. The link will becreated and finally we add edges firm the link attributes. Then, the graph attributes are all set in graph visualization [14-16]. Nodes indicates all the traders including buyers and sellers. Links are transactions between them and the size of the node indicates how many transactions they are involved. Next, an important method takes place i.e., Machine learning graph clustering. In this step, the algorithm which is used for detecting this collusion based malpractice- circular trading takes place. We use an algorithm called collusion clustering for detecting fraudulent traders. After applying collusion clustering algorithm, fraudulent traders are detected [1]. In Fig 1.

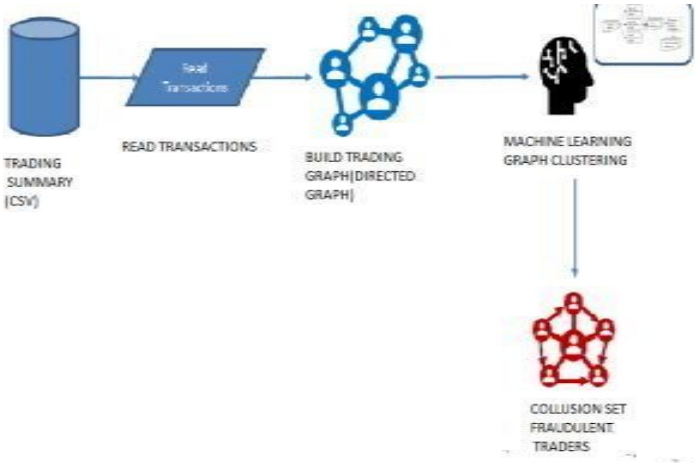

Fig 1. Detection of Collusion Sets

IV. COLLUSION CLUSTERING

After Internal trading refers to the transaction between any particular set of traders. Internal trading is denoted by I(C). External trading refers to the transactions happening among traders who are not involved in the internal trading. When there is no external trading takes place, the value of external trading becomes 1 instead of 0 [5]. It is denoted by E(C). Collusion index can be found by dividing internal trading by external trading. [φ(C) = I(C)/E(C)] in Fig 2.

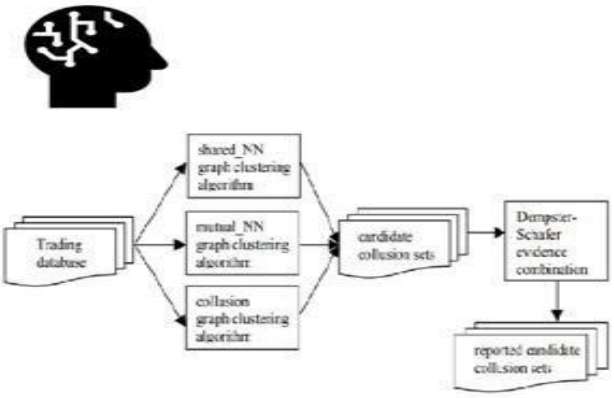

Fig 2. Detailed Machine Learning

The collusion level means the higher values of candidate level L(c) indicate better chances of a candidate collusion set. Based on the outgoing transactions / edges from a node, we need to find top-k neighbors based on the 'amount' of transaction [6].

We created a 'Directed Graph' with the given dataset. Graph visualization of Traders/Transactions (random sample: 100 traders) each 'node' indicates a Trader (buyer U seller)
each 'edge' indicates a Transaction [7].
'direction of the edge' indicates seller--->buyer.
'size' of the node indicates degree of the node (in & out).
Number of transactions a trader is involved.

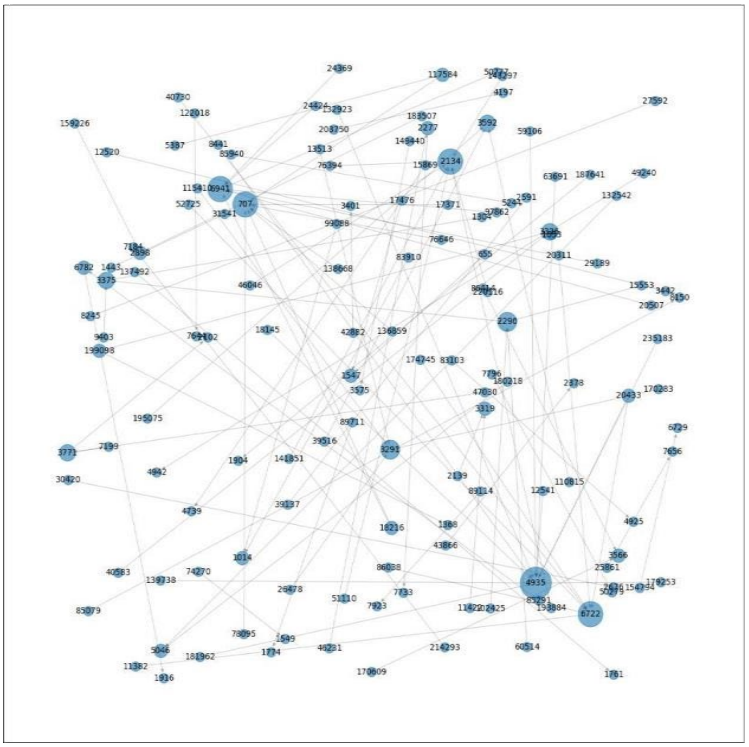

Fig 3. Graph Ingestion

V. RESULTS AND DISCUSSION

In this paper, the proposed system for detecting the circulartrading-fraudulent traders using collusion clustering algorithm was done. In this paper, we have defined the most significant practical problem of collusion set detection among the traders in stock market. The existing method performs well for detecting cross trading but it cannot detect purely circular trading. Based on the prediction calculated using the algorithmof collusion clustering, we can detect the malpractice more efficiently than other approaches. Experimentation conducted with different values of k, m and h. in Fig 3 and Fig 4.

Table 1. Execution time andCollusion Cluster size

k	m	h	Epochs (number of iteration)	ExecutionTime (time to converge)	Largest Collusion cluster size (number of traders involved in collusion)
3	1	0.6	7	23sec	6
4	1	0.6	12	1min 12sec	9
5	1	0.6	18	4min 1sec	10
3	1	0.7	7	23sec	6
4	1	0.7	10	56sec	8
5	1	0.7	16	3min 30sec	10

Observations:

- k value is proportional to Epochs, Execution time andCollusion Cluster size
- When kvalue increases, Epochs (the number of iterations), Execution time (time to converge) and Collusion

cluster size (number of traders involved in the collusion) are increasing.

- On top of Collusion index Lc the compatibility measure of (k,m,h) controls the collusion
- h value changes the collusion set in the initial iterations but give the same collusion set if the k value increases.

Table 2. Collusion Size

k=3, m=1, h=0.6	collusion_size
[4000, 25315, 30159, 20378, 134414, 11599]	6
[92739, 169748, 57413, 45322, 126443]	5
[98740, 30429, 27700, 22959]	4
[33449, 40115, 63691, 9310]	4
[59584, 1442, 22131, 86872]	4
[67665, 15869, 31713, 36342]	4
[231704, 135138, 76730, 136068]	4
[51632, 59106, 24407, 12647]	4
[14048, 205793, 21612, 23593]	4
[17698, 49699, 229202, 118755]	4

Collusion Index 36369.0

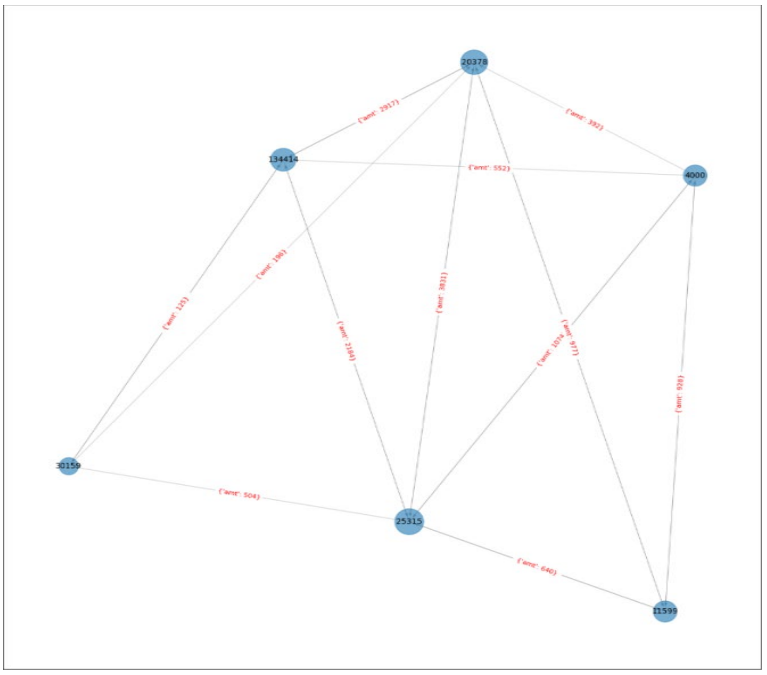

Fig 4. Result values

VI. CONCLUSION AND FUTURE WORK

Several algorithms used in the project is much beneficial and constructive in aiming at Collusion sets and improving the detecting accuracy. As further work, its needed to inquire whether there are additional ways of accommodating mutual nearest neighbour and shared nearest neighbour graph clustering algorithm for the problem of candidate collusion set detection. And also investigating whether these techniques can be extended to detect occurrences of specific mal-practices such as price manipulation and circular trading. It is essential to explore whether classical statistical inference theory can be used to combine results of various experiments. As futurescope, it is investigated whether there are few more effective ways of occurring the mutual nearest neighbor and sharednearest neighbor graph clustering algorithms to the problem of collusion set detection. It is also investigated whether these approaches can be enlarged to detect the occurrence of frauds

in circular trading. We have improvised the collusion clustering algorithm with modern graph analytics which helped to identify and interpret the fraudulent connections and transactions on the real SEBI dataset, the improvised algorithm is scalable on big data and reduce the execution time.

Acknowledgment
We would like to express my gratitude to my supervisor for the valuable and constructive suggestions during the development of this work. We also would like to acknowledge the extended support and guidance provided by KCG College of Technology that made this work possible.

References

[1]. Girish Keshav Palshikar, Manoj M. Apte (2005)," Collusion Set Detection Using Graph", Published online: 16 June 2007Springer Science+Business Media, LLC 2007.
[2]. Kaji Masudul Alam, Md. Nazrullslam, Md.Tarikuzzaman,S. M. Rafizul Haque (2009)," An Approach to Improve Collusion Set Detection Using MCl Algorithm", Proceedingsof 2009
[3]. 12th International Conference on Computer and Information Technology(ICCIT 2009) 21-23 December,2009, Dhaka, Bangladesh.
[4]. K. C. Gowda, G. Krishna, Agglomerative clustering using the concept ofmutual nearest neighborhood, Pattern Recognition, Vol. 10, pp.105-112, 1978.
[5]. S. Le Hegarat-Mascle, D. Richard, C. Ottle, Multi-scale data fusion using Dempster-Shafer evidence theory, Integrated Computer-Aided Engineering, 10(2003), pp. 9–22.
[6]. A.K. Jain, R.P.W.Duin, J. Mao, Statistical pattern recognition: a review, IEEE Trans. Pattern Analysis Machine Intelligence, vol. 22, no. 1, Jan. 2000, pp. 4– 37.
[7]. A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, ACM Computing Surveys, vol. 31, no. 3, Sep. 1999, pp. 264-323.
[8]. R. A. Jarvis, E. A. Patrick, Clustering using a similarity measure based onshared nearest neighbors, IEEE Transactions on Computers, Vol. C-22, No. 11, November 1973.
[9]. G. K. Palshikar, Arun Bahulkar, Fuzzy temporal patterns for analysing stock market databases, in Proc. Int. Conference on Advances in Data Management (COMAD-2000), 2000, Pune, India, Tata-McGraw Hill, pp.135-142
[10]. V. V. Bapeswara Rao, K. Sankara Rao, Enumeration of Hamiltonian circuits in digraphs, Proc. IEEE, vol. 73, pp. 1524-1525, Oct. 1985.
[11]. F. Rubin, A search procedure for Hamilton paths and circuits, Journal of ACM, vol. 21, no. 4, Oct. 1974, pp. 576 – 580.
[12]. G. Shafer, A mathematical theory of evidence, Princeton University Press,1976.
[13]. R. E. Tarjan, R. C. Read, Bounds on backtrack algorithm for listing cycles, paths and spanning trees, Networks, 5, 1975. pp. 237 – 252.
[14]. http://www.sebi.gov.in, SEBI Order against DSQ Holdings dated 10th December 2004. Order no.CO/109/ISD/12/2004.
[15]. Anand R, Fazlul Kareem S, Mohamed Arshad Mubeen R, Ramesh S, Analysis Of Heart Risk Detection In Machine Learning Using Blockchain, Volume 2021- October, 2021 Pages 685689.
[16]. Kavinnila J, Hemalatha E, Jacob M.S, Dhanalakshmi R, Stock Price Prediction Based On LSTM Deep Learning Model (conference paper), 30 July 2021.