

# Analysis of Stock Crime using Graph Mining

Jigyasha Arora  
CSE Dept  
UTU, Dehradun

Pawan Kumar Mishra  
CSE Dept  
UTU, Dehradun

## ABSTRACT

Graph mining is the use of most important structure of graph to obtain frequent patterns of information. It has board range of applications. This technique can be used to find the probability of persons doing crime in the stock market .Some case studies of people involved in stock crime were studied to obtain the attributes such as persons involved in crime, whether they are educated or not, style of crime, earning from the particular threat. These attribute lead to the construction of graph database and an algorithm has been proposed which replaces previously existing graph mining algorithm which was unable to generate subgraph for non immediate nodes. Proposed algorithm deals with large database including the features which captures the properties of graph in few parameters and establish relationship among nodes. It then adopts depth first search approach to traverse the graph in top down manner which finds the common node in the graph & establishing subgraph for both immediate & non immediate nodes. Supervised approach has been used that provides correct segmentation of training data. For the prediction of future crime, a new framework as been designed which determines the suspicious behaviour if similar crime has been occurred before based on property such as style of crime which therefore can prevent many crimes that can occur in future. We have used the NetBeans for implementing the algorithm, Neo4j for graph database and NeoEclipse for analysis of graph.

## Keywords

Graph database, Graph mining, Common Node, Sub graph.

## 1.INTRODUCTION

Graph mining is a technique used to extract characteristic patterns from a variety of graph structured data. A graph is set of nodes, pairs of which might be connected by edges. Graph is important data source as any kind of relationship can be relationship can be represented with the help of graphs. Its application is very important in terms of web and internet. Graphical representation is used for the visual representation of concepts. An easy and effective way is provided by graph to represent relation between data and entities. Graph mining is an easy representation of data in the form of relation among the nodes .In Graph mining, we split relational data to extract some pattern to get new knowledge. In this knowledge discovery process the knowledge can be a sub graph, relation to other nodes and this has been used in various fields. This paper provides an algorithm and we compare our algorithm with previous algorithm. The remainder of this paper is organized as follows. Section 2 introduces the overview of existing graph mining technique; Section 3 describes our proposed algorithm; Section 4 comparative study of our proposed technique with other existing technique; Conclusion and future scope in Section 5 and all the used references are given in Section 6.

## 2.OVERVIEW OF EXISTING ALGORITHM

### 2.1 Inference Algorithm

Inference algorithm also called belief propagation uses heuristic approach for traversing. In this each node gathers incoming message, multiplies them into one vector and set the vector as node new belief has been mentioned in paper [13]. Finally node generates new message for its neighbor using updated belief. The process continues until al belief changes. It then finds the subgraph for all immediate neighborhood and all edges in the neighborhood. It further uses unsupervised algorithm that correlate pair of features and pin points nodes whose features deviate from other nodes. The short coming of inference algorithm is that it is able to generate the subgraph for immediate nodes only.

### 2.2 Prediction of Crime

Prediction of crime done earlier was based on geographical location or hot spots in which analysis was done based on high sensitive area or low sensitive area. It was not based on behavioral pattern.

## 3.PROPOSED ALGORITHM

3.1 The proposed algorithm overcomes the previous algorithm by generating subgraph for both immediate nodes as well as non immediate nodes in the neighborhood based on ant colony optimization. It also traverses all the edges in the neighborhood. It further finds the common factor based on cost estimation which has high probability of crime using supervised approach which provides correct segmentation of training data. It then introduces a new transactional framework which finds the suspicious behaviour of crime based on DEMAT number .Crime is detected if similar pattern of crime has been occurred before. User interface has been created so that user can enter more queries and create graph for retrieval of subgraph. These case studies are stored in MySQL and graph database has been created. Furthermore, Subgraph retrieved based on properties. The steps are listed below:

1. It involves creation of graph database and nodes.
2. Properties nodes are set and relationship among them is established.
3. Further property is searched based on selection and sorting is done with the help of relationship.
4. Traversing to a particular node takes place in which property gets matched to node property. If it matches then the graph gets spilt and traverses according to depth first search. Relationship of first match also gets compared to node relationship. If it matches then it displays all the properties of the nodes and continues to traverse down until all nodes has been traversed.
5. Subgraph generation is done based on cost estimation which has high probability of crime and common node is identified based on ant colony optimization.
6. For future prediction of crime, a transactional framework has been created which determines the suspicious behaviour which previously that style of crime has been used.

Several types of crime that has occurred include style of crime like insider trading, fake loans, fake press release, high growth of price of stock, more number of employees than actual, snooping etc.

#### ALGORITHM FOR TRAVERSING

```

Setup
Step 1 Create Graph Database
Step 2 Create Node
Step 3 Set Properties of nodes
Step 4 Create Relationship
Step 5 Select property /* Property to be searched */
Step 6 while Node position traverse
Step 7 if p =node.property // if property matches
Step 8 Split Graph
Step 9 traverse <- depth
Step 10 r <- node-relationship // store relationship of first match
Step 11 if node.relationship equals r
Step 12 display properties
Step 13 continue traverse down
Step 14 else
Step 15 traverse <- down next node
Step 16 end if
Step 17 end if
Step 18 if p = node.property
Step 19 repeat Step 8 through14
Step 20 end if
Step 21 end if
Step 22 end loop
    
```

#### ALGORITHM FOR FINDING COMMON NODE & SUBGRAPH

```

String kw=null, int c=0,int i, int j, int nc=0; cost=5; /*cost=5 is related to fake */
Step1 String node[][]=new String[100][12] /*All the nodes get stored in a matrix.*/
Step 2 If the node has more elements then go to step 3, else exit.
Step3 Get the nodes over the set of elements then go to step 4 else exception is thrown./* Step 2 & 3 are implemented using iterator interface*/
Step4 Store the keywords to be searched in an array as String keyword= {"fake","insider","trading"}
Step5 for (i=5, i<n, i++) /* i is set to 5 as four main header nodes are placed above before traversing*/
    for (j=0, j<9,j++) /* limit of j is 9 as 9 column having different properties are set*/
        if node [i][j].contains (keyword[nc]) then set c=c+1
Step6 if(c>cost)
    kw=keyword[nc]
Step7 Display kw
Step 8 int ii=2;
Step 9 for (i=5;i<n;i++)
    Set ii=ii+1;
Step10 if (node[i][8] contains property
Step11 Display properties
Step 12 Relationship are set
Step13 else if (node[i][8] contains another property
Step14 Display properties
Step15 Relationships are set
    
```

#### ALGORITHM FOR PREDICTION OF CRIME

```

String prop, int flag;
String promote="" */ /*entered by user*/
Step1 String node [][]=new String[25][9] /*All the nodes get stored in a matrix.*/
Step 2 If the node has more elements then go to step 3, else exit.
Step3 Get the nodes over the set of elements then go to step 4 else exception is thrown./* Step 2 & 3 are implemented using iterator interface*/
Step4Element retrieved from graph db get stored in prop and type casted to String data type
Step5 if (prop. contains (promote)
Step6 set flag=1 and fraud is suspected
Step7 else no crime suspected & exit.
    
```

## 4.COMPARITIVE STUDY AND DISCUSSION

The proposed algorithm when compared with above existing algorithms works remarkably as it generates sub graph both for immediate nodes as well as non immediate nodes. Also the common factor gets generated. When compared with prediction method, it predicts the crime based on behavioral pattern such as style of crime rather than geographical location which replaces the previous techniques. Neo4j is the graph database used for graph mining. NeoEclipse is used for analysis of crime. Fig1 represents overview of stock crime, fig2 represents user interface through which user enter query, fig3 subgraph of all persons involved in crime at NSE (immediate nodes only), fig4 represents Subgraph of criminals involved in money making through fake press release (both immediate and non immediate nodes), fig5 represents user interface for checking suspicious behaviour of crime and fig 6 represents result of query showing suspicious behaviour.

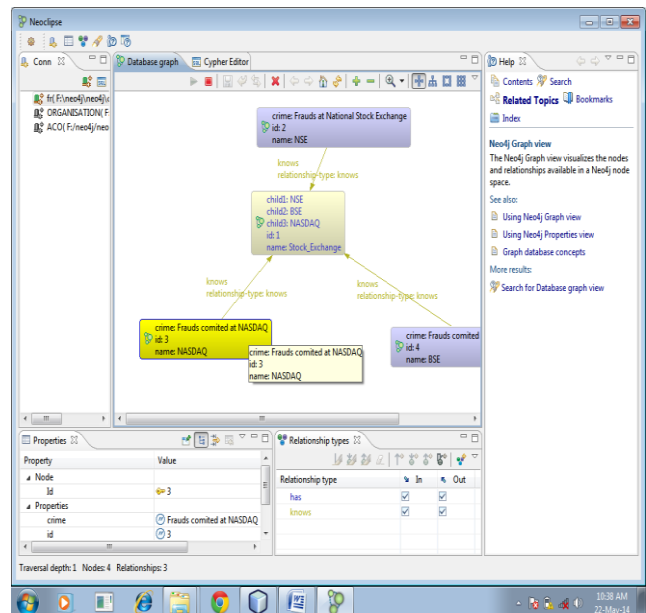


Fig1. Overview of Stock Crime Exchange

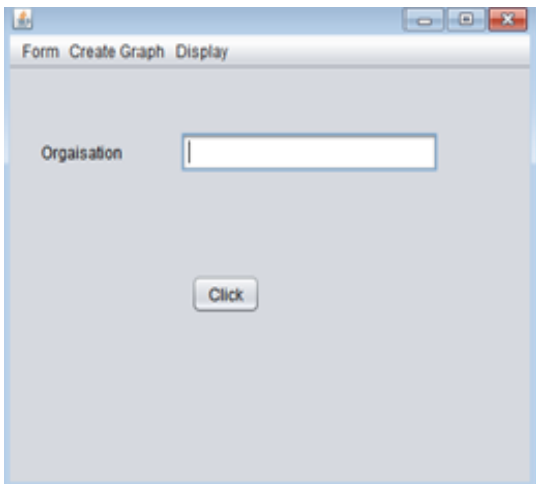


Fig2. User interface through which user enter query

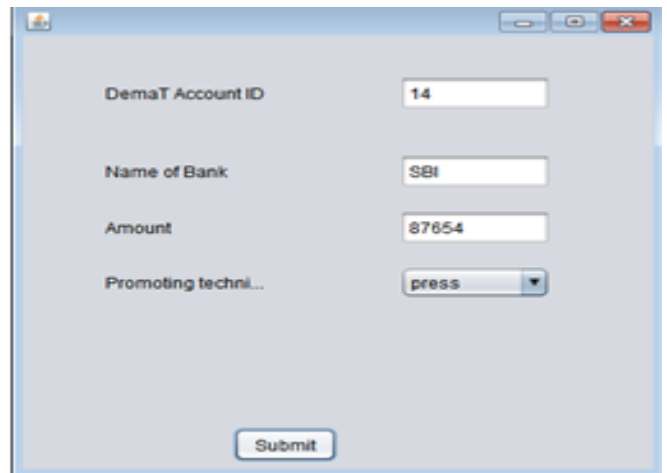


Fig 5. User interface for checking suspicious behaviour of crime

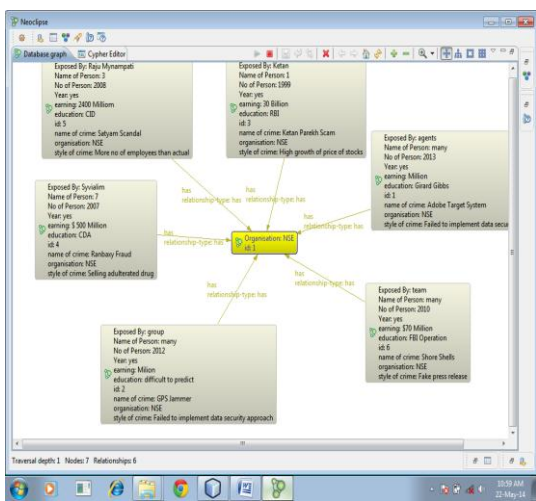


Fig3. Subgraph of all persons involved in crime at NSE (Immediate nodes only)

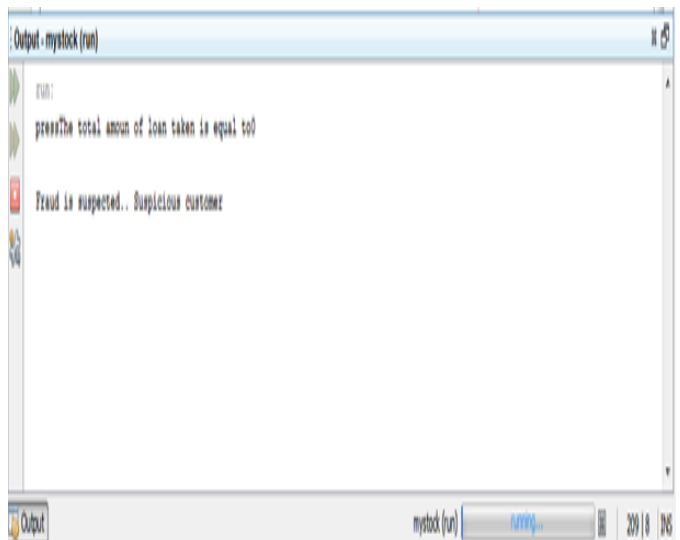


Fig 6. Result of query showing suspicious behaviour

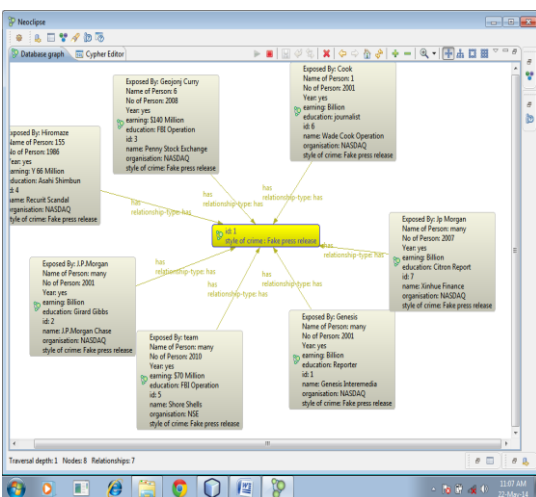


Fig4. Subgraph of criminals involved in money making through fake press release (both immediate and non immediate nodes)

## 5. CONCLUSION & FUTURE SCOPE

The key insight behind the proposed method is to identify the hotspots by searching, utilizing and presenting patterns in the Graph database by processing the crime related datasets into a transaction based dataset. In our technique firstly the graph has been created and then by traversing that according to the property entered by the user a subgraph has been created that will show the entire related data on the basis of query entered by user with their relationships. Then this introduces a similarity method to summarize the identified patterns. Based on these patterns, a relevant report of crime and related variables is visually presented for domain experts. The system utilizes some new approaches regarding the crime detection, which use related variables in crime mapping with the effective of primary direction. Although, the current algorithm already performs quite well, it can be implemented in real time systems to trace the pattern of stock rise and fall in the share market and we can compare the current pattern of stock fluctuations with the pattern present in graph database, so that if it finds any resemblance in the pattern it can impose a security check over that particular stock and predict the future steps. This can be helpful in planning the prevention of several crimes which can contribute to the people who gets affected due to this share

price manipulation. Graph mining is a currently very active research field. The application areas of graph mining are widespread ranging from biology & chemistry to internet applications.

## 6. REFERENCES

- [1] Xifeng Yan and Jiawei Han, "Span: Graph-Based Substructure Pattern Mining", 2002.
- [2] Xifeng Yan, Philip S. Yu and Jiawei Han: Graph Indexing, A Frequent Structure Based Approach, 2004.
- [3] Ingrid Fischer and Thorsten Meinl, "Graph Based Molecular Data Mining - An Overview", IEEE 2004.
- [4] Deepayan Chakrabarti, Yiping Zhan and Christos Faloutsos, "R-MAT: A Recursive Model for Graph Mining", School of Computer Science, CMU, 2004.
- [5] Hsinchun Chen, Wingyan Chung, Jennifer Jie Xu, Gang Wang Yi Qin and Michael Chau, "Crime Data Mining: A General Framework and Some Examples", IEEE 2004.
- [6] Katsutoshi Yada, Hiroshi Motoda, Takashi Washio, and Asuka Miyawaki, "Consumer Behavior Analysis by Graph Mining Technique", M. G. h. Negoita et al. (Eds.): KES 2004, LNAI 3214, pp. 800–806, Springer 2004.
- [7] Lisa Singh, Mitchell Beard and Lis Getoor: Visual Mining of Multi-Modal Social Networks at Different Abstraction Levels, 2006.
- [8] Junmei Wang, Wynne Hsu Mong and Li Lee Chang Sheng, "A Partition-Based Approach to Graph Mining", Proceedings of the 22nd International Conference on Data Engineering (ICDE'06), IEEE 2006.
- [9] Kamrul Abedin Tarafder, Shah Mostafa Khaled, mohammad Ashraf Islam, "Reverse Apriori algorithm for frequent pattern mining", Medwell journals, 2008.
- [10] Frank Eichinger, Klemens Böhm and Matthias Huber, "Improved Software Fault Detection with Graph Mining", Appearing in the 6th International Workshop on Mining and Learning with Graphs, Helsinki, Finland, 2008.
- [11] William Eberle: Detecting Insider Threats Using a Graph Based Approach, 2010.
- [12] Anton Dries and Siegfried Nijssen: Analyzing graph database by aggregate queries, 2010.
- [13] Polo Chau: Catching Bad Guys with Graph Mining, Springer, 2011.
- [14] Ciro Cattuto, André Panisson, Marco Quaghiotto and Alex Averbuch, "Time-varying Social Networks in a Graph Database", 2012.
- [15] G. Kishore Kumar, Dr. V. K. Jayaraman "Clustering of Complex Networks and Community Detection Using Group Search Optimization", 2012.
- [16] Garima Jaiswal and Arun Prakash Agrawal "Comparative analysis of Relational and Graph databases", IOSR Journal of Engineering (IOSRJEN), 2013.
- [17] Quist-Aphetsi Kester, "Criminal Geographical Profiling: Using FCA for Visualization and Analysis of Crime Data", IEEE 2013.
- [18] Justin J. Miller, "Graph Database Applications and Concepts with Neo4j", Proceedings of the Southern Association for Information Systems Conference, Atlanta, GA, USA March 23rd-24th, 2013.