Network Security

Course No. : SS G513

Fraud Detection using Graph DB: Connecting the dots

Submitted To

Dr. Ashutosh Bhatia

Assistant Professor

Department of CSIS, BITS Pilani

Submitted By

Umang Dhiman - 2016H112151P

Sarita Sharma - 2016H112163P

April 30, 2017

Abstract

The Indian Banking Industry has undergone a major transformation since 1991, the year of Liberalization. In the year of 2010-11, there had been a significant increase in people switching to Online Banking. Online Banking, though having its multiple benefits, like, location independent banking, time saving, also had a dark side. With the increasing number of users opting for Online Banking, online frauds had also had a boost. The fraudsters became smarter and it became more and more difficult to trace them.

Network analytics is an effective tool to amplify traditional fraud detection approaches. The fraud detection world, basically, deals in rich details and confronts constantly emerging and evolving techniques. This is where network graph analysis is of central value– it offers a method for capturing the rich context of fraud in a standard, machine readable and transferable format. Once captured in such a format, deep pattern and statistical analysis can be conducted on existing datasets. Network analytics is thus a complementary approach which enhances the traditional fraud detection approaches. Our study basically deals with how the problem of banking frauds in general can be solved using network graph analysis.

Keywords: Banking Frauds, RBI, Graph DB, pagerank

1. Introduction

The Indian Banking Industry has undergone a major transformation since 1991, the year of Liberalization. Though banking domain is generally highly regulated, it still faces the issues related to frauds, NPA (Non-Performing Assets) etc. Post liberalization the frequency, complexity and cost of banking frauds have increased manifold resulting in a very serious cause of concern for regulators, such as the Reserve Bank of India (RBI). The year 2010-11, witnessed a significant increase in people switching to Online Banking. Online Banking, though having its multiple benefits, like, location independent banking, time saving, also had a dark side. With the increase in users opting for Online Banking, online frauds had also had a boost. The fraudsters, now, have become smarter and it has become more and more difficult to trace them. During the last three years, public sector banks (PSBs) in India have lost a total of Rs. 22,743 crore, due to various banking frauds. The number of banking frauds have decreased due to the various measures have been initiated by the Reserve Bank of India (RBI), but amount of money lost has increased in these years. The robustness of a country's banking and financial system is a direct indicator of the well-being and living standards of its citizens. Therefore, if the banking system is overshadowed with high levels of frauds then it is a cause of worry, as it reflects financial distress of borrower clients, or inefficiencies in transmission mechanisms.

Indian economy suffers to a great extent from these problems, and this served as the prime motivation for us to carry out this detailed study of frauds in the Indian banking system and examining frauds from

different angles with the sole aim to design a framework to reduce the number of frauds and also the money lost in frauds effectively. We propose a novel approach to map the financial transactions to graphs and further perform graph analysis to detect suspicious accounts.

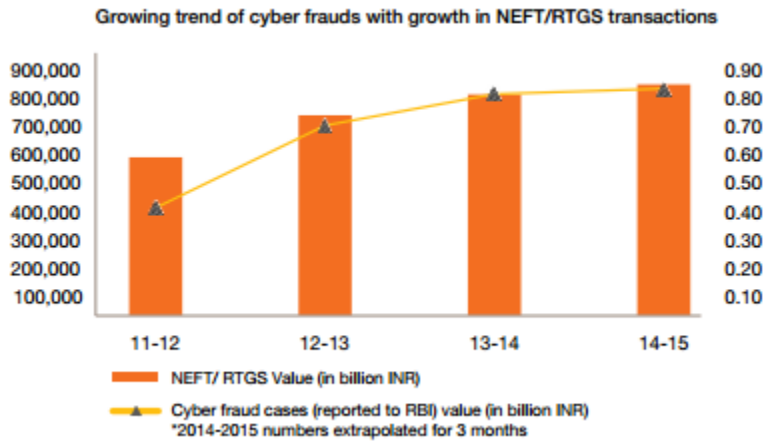

Fig 1: Growing trend of cyber frauds with growth in NEFT/RTGS transactions

(Source: Economic Times Mar 4, 2015)

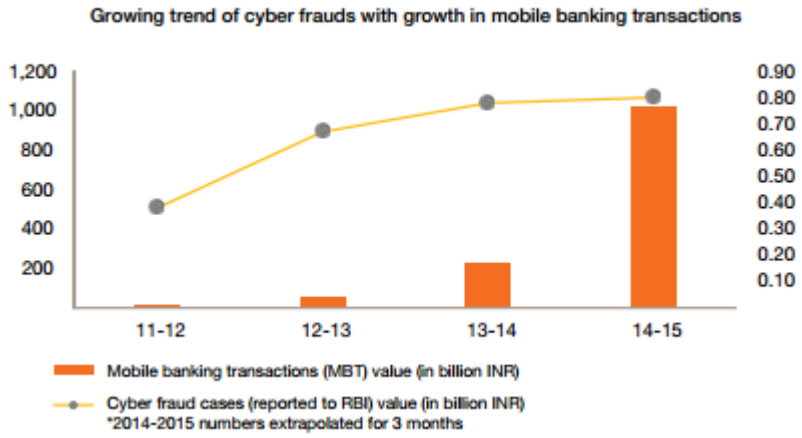

Fig 2: Growing trend of cyber frauds with growth in mobile banking transactions

(Source: Economic Times Mar 4, 2015)

1.1 What is fraud in banking domain?

RBI, the regulator of banks in India, defines fraud as,

"*A deliberate act of omission or commission by any person, carried out in the course of a banking transaction or in the books of accounts maintained manually or under computer system in banks,*

resulting into wrongful gain to any person for a temporary period or otherwise, with or without any monetary loss to the bank". [1]

1.2 Some common frauds in Indian Banks

Indian banking industry is crippled by various types of frauds. Some of the most prevalent ones being:

1.2.1 Identity Theft: Identity theft basically refers to the intentional use of someone else's identity, either of a fake person or that of a third person, usually as a means to gain a financial advantage which may in some cases lead to the other person's loss. There are three popular types of identity thefts:

a. **First party fraud** - First party fraud refers to fraud that is committed by an individual or group of individuals on their own account by opening an account with no intention of repayment. The fraudsters may use either their own identity or a fictitious identity.
b. **Third party fraud** - Third party fraud refers to fraud that is committed without the knowledge of a person whose identity is used to commit the fraud. In such cases, the third person whose identity is misused is often the victim.
c. **Insider fraud** – The deliberate misuse of bank's assets by the employees with the intention to compromise data, operations or security is known as insider fraud.

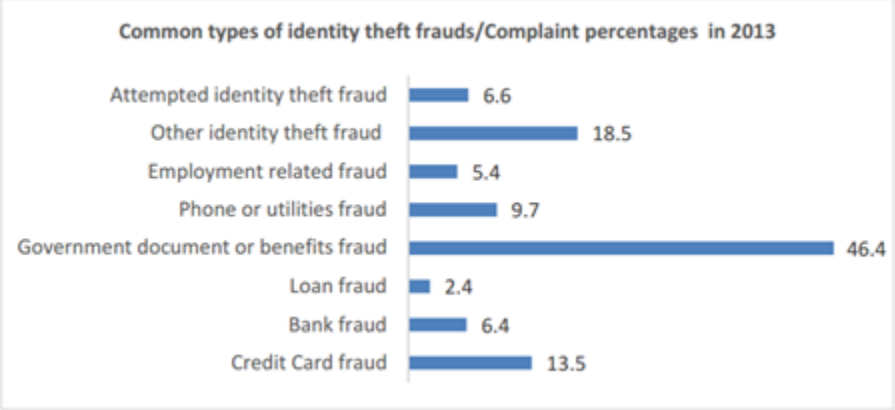

Fig 3: Identity theft frauds (Source: PWC India and ASSOCHAM(2014))

1.2.2 Money Laundering: Money Laundering is basically the process of creating the appearance that the money acquired through illegal sources, such as, drugs trafficking and terrorist activity, originated from a legitimate source. India is extensively gripped in the clutches of money laundering. It is usually used by criminals to hide money made through illegal act. It is designated as the source of illegally obtained funds covered through a series of transfers and deals in order that those same funds can eventually be made to appear as legitimate income.

1.2.3 Cybercrime: A majority of the Indian banks are now offering online and mobile banking services. Most of the transactions are conducted via payment cards, debit and credit cards. Consequently, both private and public banks as well as other financial institutions in India are becoming increasingly vulnerable to sophisticated cyberattacks.

1.2.4 Black money: Black money refers to the act of illegal accumulating of money and not declaring this income to the government and thus, not paying any taxes associated with the same. According to the Global Financial Integrity Report, the total amount of illicit money moving out of India rose to 439.59 billion USD (28 lakh crore INR) from 2003 to 2012. In 2012, India ranked third globally, with an estimated 94.76 billion USD (nearly 6 lakh crore INR) in illicit wealth outflows.

1.2.5 Loan loss: The risk of loan loss is high in India. Due to lack of appropriate due diligence and monitoring of loans, the number of loan defaults has increased in recent years. The non-performing assets are growing in last few years while the GDP has been declining.

2. Literature Review

Krzysztof Michalak, Jerzy Korczak [8] [year 2011] proposed a Graph Mining Approach to detect Suspicious Transactions. They basically dealt with detection of money laundering activities, i.e., activities aimed at obscuring the origin of illegally-obtained money. The money laundering process is divided into three stages: placement, layering and integration. Each stage involves various schemes of transferring money between bank accounts. These schemes can be identified as subgraphs in the transaction graph. Money laundering is difficult to detect because the occurrences are very infrequent in the graph of all transactions and individual bank transfers involved in a money laundering scheme are structured so that they appear as legitimate. These obstacles are especially difficult to overcome if the detection process is based only on features of individual transactions. Therefore, graph mining methods seem to be interesting, because they are able to detect complex dependencies between transactions. It is also possible to take into account properties and relations of entities involved in transactions.

Ali Ahmadian Ramaki, Reza Asgari and Reza Ebrahimi Atani [9], in the year 2012, proposed the application of Ontology Graph to detect Credit Card fraud. In this approach, the ontology graph was created for every user's transaction behavior and then stored in the system. In the event of abnormality detection only those transactions from registered history of transactions were selected to perform computation which were highly similar to entry transactions. The method used to detect fraud in credit cards transactions is outlier detection. Outliers were applied in abnormality detection for determining detrimental behaviors. The most salient advantage of this method in fraud detection was its being real-time and high accuracy. By this method in addition to real-time detection of fraud occurrences through running transactions, storage capacity of various data patterns is low because of no maintenance of similar patterns.

Richard Colven [10], in the year 2011, proposed a social networking approach to detect First Party Frauds. These frauds occur when customers apply for credit cards, loans, overdrafts or other unsecured banking credit lines with no intention of paying them back. These frauds are difficult to trace as they give the impression of legitimate activity in the beginning. These fraudsters typically use a technique known as 'cash cycling'; where an amount of money is circulated amongst many fraudulent accounts creating the illusion of legitimate activity. These false payments never leave the network and the fraudster, on the back

of a good CIBIL score, would have accumulated several cards with significant levels of credit available, cheque books, cashcards and invitations to take out loans.

As per Richard [8], these fraudsters always tried to cover their tracks by having no obvious connection to other members of their network. However, by pooling data from multiple sources — phone numbers, addresses, known relationships, transactions, historic data and third party data —fragments of these records can be used to link groups of people with suspicious transactions. Thus, these footprints can be used to detect whether a person is a fraudster or not.

In the year 2015, [11] proposed Social Network Analysis (SNA) approach to detect first-party frauds. Fraud detection models were basically based on patterns in credit card transactions, which often involved fake schemes based on the creation of a synthetic identity to obtain credit cards, overdrafts or loans. They argued that patterns often exist, however, in the networks of individuals perpetrating these crimes.

Dr. Archisman Majumdar [12] [year 2016], also advocated the network analytics approach to detect fraud in financial industry. As per him, in fraud detection, the interactions and exchanges could be viewed as heterogeneous networks with multiple participants. In a real world scenario, the numbers of participants are generally huge, but the kind of interactions among the individuals is generally limited and known. Graph analysis techniques can be used further to identify suspicious individuals, groups, relationships, unusual changes over time/geography, and anomalous networks within the overall graph structure.
He also highlighted some of challenges in using network analysis for fraud detection which include –

- emergent body of knowledge, leading to difficulties in identifying the correct methods, their applications and interpretations
- requirement for novel data storage and warehouse methods: Traditional databases are not optimized or designed for network analysis and operations. New NoSQL and graph databases are often more suitable for these operations
- high volume and variety of data which needs to be processed
- many graph algorithms are ‘computationally intractable’, i.e., even though the problems can be solved in finite time, the amount of processing required make them infeasible
- lack of automation in network analysis in fraud detection

Trung Le, Ba Quy Tran, Hanh Dang Thi My, Thanh Hung Ngo [13] [year 2015] focused on using Rules Set Technique and Graph Model to evaluate the reliability of online payment transactions. The decision of abnormal values in a transaction was represented by fraud detection rules. Fraud Detection Rules were set up via system programming language combined with Cypher query language, based on vertices and relations among the vertices on the graph database. According to the Rules set, Inference Engine was assigned status for transaction. As per their algorithm, the transaction would be processed only if its status is safety. If transaction status was doubt or fraud, it would be reported to admin to confirm or reject.

3. Proposed Algorithm

3.1 Hypotheses

Hypothesis 1: There exists 2 types of transactions from account perspective, i.e., incoming and outgoing transactions. Incoming transactions refers to the transactions wherein an account receives money from another account, in outgoing transactions amount is debited from the account. If an account has a healthy incoming and outgoing transactions then it can be marked as regular account, whereas an account which has higher probability of becoming a sink can be red flagged. Red flagged accounts represent accounts that have high probability of being a fraud account.

Hypothesis 2: If an account pays another account, then there exists a trust relationship among those accounts. If an account transacts with a set of accounts regularly then it can be inferred that there is high trust among these accounts. An account's trust score can be calculated based on the incoming and outgoing transactions of an account. An account can be red flagged if it has low trust score.

3.2 Graph Concepts

While banking systems record numerous attributes about transactions, we restrict ourselves to the four main attributes: source account, destination account, transaction initiation timestamp, and amount of money transferred (in a common currency).
Using this data, we define a quiver (or multidigraph) as follows:

Definition 1: A quiver (or multidigraph) G = (V, E, s, d) where V is a set of vertices of G, E is a set of edges, and s and t are mappings s: E → V and d : E → V that returns the source and destination vertices for an edge, respectively. Each edge has two properties: the timestamp and the amount involved in the transaction.

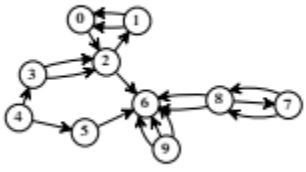

Fig 4: Quiver

Definition 2: Source Vertex: A vertex v1 is termed as the source vertex if indegree of the vertex is 0.

Definition 3: Sink Vertex: A vertex v1 is termed as the source vertex if outdegree of the vertex is 0.

3.3 Graph Creation

This section analyses how the transactions can be viewed as edges in a multigraphs. The transactions transferring money from one account to another can be represented as edges between the nodes. The weight of the edges in the multigraph includes the amount of money involved in the transaction along with the timestamp for the transaction.

Thus, we propose to build a graph G(V,E) where V is the set of vertices in the graph and E is the set of edges between the nodes. The set V represents the account details; for example, account number. The set E consists of a pair of vertices, (v1, v2), which have an edge between them in G. The edge basically denotes that v1 has done a transaction between v2. Each edge contains details regarding the amount involved in the transaction along with its timestamp.

3.4 Page Rank Algorithm

Page rank [6] algorithm was initially developed to compute the importance of webpages. The concept of pagerank can be used to identify the importance or trust among the accounts. In our graphs PageRank algorithm is used to compute the reputation of an account in terms of the payments made to it. Our hypothesis is that the account with a high value of trust computed using PageRank is less likely to be fraudulent. In the PageRank algorithm, an account evenly distributes its own PageRank to the accounts it pays, and the algorithm iterates until convergence:

$$PR(u) = (1-d)/ N + d \sum_{v \in P(u)} PR(v) / |P(v)| \quad (1)$$

Where, P(u) is the set of accounts u pays, d is a damping factor and N is the total number of nodes in the graph.

Initially, the damping factor modeled the probability a random web surfer stops on a page. In the financial scenario, the damping factor can be used to model an account saving. We use a default constant damping factor of 0.5; future work will explore using a per-account damping factor based on past spending behavior.

PageRank algorithm has 2 forms unweighted and weighted. In the weighted form, the distribution of an account's PageRank to its neighbors can be made proportional to the transaction amount. Alternatively, we could weight edges based on the number or frequency of the transactions. The weight of edge considered in the algorithm is a function of amount and timestamp of the transaction.
If the trust score is below a threshold, then the account is marked as suspicious. These suspicious accounts should be reported to bank officials for special monitoring.

3.5 Problem of Scalability in Real Life Data

While analyzing the transactions in real time, there exists a problem of scalability. In real life scenario, there exists the problem of big data. If the above stated algorithm is to be implemented in real life scenario, then it should be able to handle big data efficiently. The transactional data should be developed incrementally rather than constructing the entire graph each time. Incremental construction of graph means only the new transactions should be added to the existing graph.

Also, to be effective in practice, the fraud detection algorithm must be able to compute the trust score of the accounts in real time. The trust score can be incrementally calculated to minimize the cost of computation. In each level, only the trust score of the previous level is used and only those trust scores are recomputed only for those accounts which are involved in the new transactions. The trust scores are then incrementally calculated until convergence. Thus, the complexity of computation of trust score is reduced.

4. Conclusion

We have proposed an algorithm to analyse the transactions based on the network graph analytics. Our proposed algorithm includes the concept of pagerank to compute the trust score of a bank account. The trust score (page rank) can be used to give an indication about the fraudulent level of an account. A low trust score indicates that it is less trusted by its peers and hence, could be suspicious. These accounts can be red flagged and then specially monitored to analyse whether the account is truly fraudulent. The proposed approach is channel independent and could be applied in any class of banking frauds.

References

[1] https://www.rbi.org.in/scripts/BS_ViewBulletin.aspx?Id=14351
[2] http://sbioacc.com/downloads/GSrepo25thgb.pdf
[3] Reserve Bank of India (2014a), “Master Circular on Frauds- Classification and Reporting”, RBI Circulars.
[4] Reserve Bank of India (2015a), “Framework for dealing with loan frauds”, RBI Circulars.
[5] Current Fraud trends in the Financial Sector ASSOCHAM June, 2015
[6] http://www.experian.com/assets/decision-analytics/white-papers/bust-out-fraud-white-paper.pdf
[7] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. Computer Networks and {ISDN} Systems, 30(17):107 – 117, 1998. Proceedings of the Seventh International World Wide Web Conference.
[8] “Graph Mining Approach to Suspicious Transaction Detection ” - https://fedcsis.org/proceedings/2011/pliks/218.pdf
[9]“ Credit Card Fraud Detection Based On Ontology Graph ” - http://airccse.org/journal/ijsptm/papers/1512ijsptm01.pdf
[10] “How To Use Social Networks In The Fight Against First Party Fraud” - http://www.businessinsider.com/how-to-use-social-networks-in-the-fight-against-first-party-fraud-2011-3?IR=T
[11] “Social network analysis for fraud detection in payments” - http://banking.com/analysis/social-network-analysis-for-fraud-detection-in-payments/
[12] “Social Network Analysis Approaches for Fraud Analytics” - http://www.mphasis.com/nextlabs/wp-content/uploads/2016/08/Social-Network-Analytics-for-Fraud-Detection.pdf
[13] “ Evaluating Online Payment Transaction Reliability using Rules Set Technique and Graph Model ” - https://www.irjet.net/archives/V2/i9/IRJET-V2I902.pdf

[14] Junghanns, Martin; Petermann, Andre; Neumann, Martin; Rahm, Erhard Management and Analysis of Big Graph Data: Current Systems and Open Challenges In Handbook of Big Data Technologies (eds.: A. Zomaya, S. Sakr) , Springer 2017, to appear 2017-02