Mphasis
The Next Applied

Social Network Analysis Approaches for Fraud Analytics

Dr. Archisman Majumdar
Senior Manager, Mphasis NEXTlabs

NEXT labs

Introduction

The impact of fraud on organizations is becoming increasingly costly. Every year financial institutions lose millions of dollars in revenue to systematic fraud. The emergence of new technologies and forms of payments, as well as sophistications in fraud, complicate the challenges faced by organizations in creating effective fraud detection strategies. Many of the existing techniques rely solely on the business rules developed by experts, which require great amount of user inputs, and need to be constantly updated.

However, the ability to link multiple data sources, analyze large volumes of data, and apply newer algorithms on the transactions, provide organizations an opportunity to capture, and sometimes predict, fraud in a more efficient manner. More recent analytics based approaches include the use of descriptive & predictive analytics, machine learning, and social network analysis methods for fraud detection. This paper discusses some of the key approaches and developments in the use of social network analysis in fraud detection and prevention.

FRAUD AND FRAUD ANALYTICS

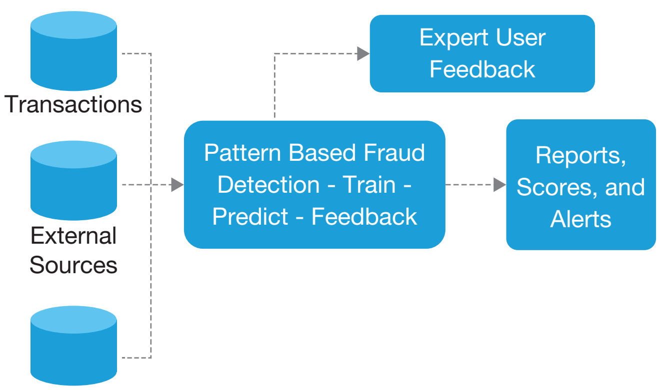

Figure 1 Fraud Analytics System

Fraud is "an uncommon, imperceptibly concealed, time-evolving and often carefully organized crime which appears in many types of forms" (Baesens, Van Vlasselaer, and Verbeke 2015). The definition highlights some of the key elements of fraud and also points out some of the major identification methods.

The classical approach to fraud identification relies on creation of explicit rules (IF-THEN-ELSEIF-…) based on the recommendation of experts. These rules are developed and modified through their collective field experiences. Nevertheless, over time, due to the dynamic and sophisticated nature of the frauds, the rules become complex and difficult to maintain and implement (unless they are very regularly updated). This is also a very labor intensive approach requiring human intervention at every stage of evaluation, identification, and monitoring.

The availability of data from multiple sources, and the ability of present systems to process and analyze this data have provided new opportunities for identifying fraud. As is apparent from *Figure 1 Fraud Analytics System*, the use of multiple data sources to identify patterns is one of the cornerstones of a data mining approach to fraud detection. Fraud analytics also provide a potential to automate multiple stages of the fraud detection, monitoring, and intervention stages of a typical cycle.

HyperGraf™ combines data from multiple sources, including credit scores, enterprise transactional data, and social media to identify and analyze fraud. One of the key methods used in HyperGraf™ is network analysis for fraud detection and the following section highlights some of its key aspects.

Network Analytics in Fraud Detection

Social Network Analysis, one of the emergent data mining methods in fraud analytics, is a technique which represents the entities as nodes and relationships between the entities as links. Representing the relationships reveals a lot more information than simply listing out the properties of the entities. The analysis of links and relationships enables the application of various graph mining algorithms on the data source.

Traditional data mining techniques rely on the statistical patterns used for identifying fraud. Yet, given the uncommon, time-evolving, and carefully concealed nature of fraud, these methods are often unable to detect various types of frauds. Application of a number of graph algorithms can help in identifying such patterns by utilizing relationship information in addition to the user level attribute information.

In fraud detection, the interactions and exchanges can be viewed as heterogeneous networks with multiple participants. The number 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.

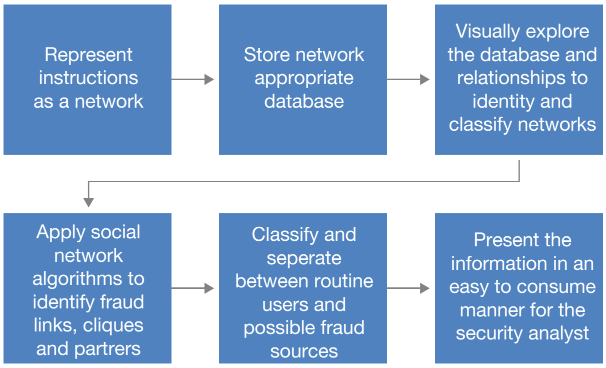

Figure 2 End-to-end fraud analytics approach using social network analysis methods

Some of the popular network analytics methods used and their typical business use cases for fraud detection are listed in the following figure –

Networks visualization and ego-centric analysis
- Displays relationship between selected alert and any other related alerts through link and nodes
- Identify links with known (blacklisted) entities

Entity link analysis / entity resolution
- Detecting rings in two mode networks of people and attributes
- Identify rings in first party fraud collusions

Graph walking to identify rings
- Walk through the graph to identify rings
- Identify paper rings of fraudulent collusions

Centrality measures
- Ranking nodes based on various graph centrality parameters
- Identify leaders in fraud networks

Snowball method
- Identify suspects and recursively expand their connections using snowball method
- Identify linkages to known (blacklisted) entities

Peer group analysis
- This technique detects abnormal behavior of a target by comparing it with its peer group and measuring the deviation of its behavior from that of its peers
- Abnormal changes when compared to the peer group

Network topologies - cliques and stars
- Any quantitative or qualitative features of a user behavior in online social networks that are inconsistent with the rest of users can be considered anomalies
- Anomaly Detection - Identify outliers in networks

Page-rank
- The Page-Rank algorithm can be used to discover the critical accounts of the groups
- Collusive fraud groups

Combining user level and network level features
- Combine user level attributes with network level attributes
- Identify conspired groups

Challenges to Network Analytics in Fraud Detection

Network analysis opens new avenues for fraud detection. These can augment the existing rule-based, and data-mining approaches in the organization. While network analytics techniques promise key breakthroughs in fraud detection, there are certain key challenges which make implementing network analysis in Fraud detection difficult. These include -

- Emergent body of knowledge leading to difficulties in identifying the correct methods, their applications and interpretations.
- Requirements for novel data storage and ware house 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.
- Retro-active nature of social network analysis which makes them less suitable for prediction compared to other methods, such as machine learning based approaches.
- Lack of automation in network analytics in fraud detection and the need for expert analysis and interpretations.

Conclusion

Addressing these challenges require organizations to continually innovate and use new systems with specific capabilities. Some solutions, like HyperGraf™, provide a platform for guided analytics in fraud detection. Nevertheless, the role of domain experts and data scientists in applying these methods are often the key factors in the successful implementation of a fraud detection strategy.

Reference

Baesens, Bart, Véronique Van Vlasselaer, and Wouter Verbeke. 2015. "Fraud Analytics Using Descriptive, Predictive & Social Network Techniques."
https://lirias.kuleuven.be/handle/123456789/500346

Dr. Archisman Majumdar
Senior Manager, Mphasis NEXTlabs

Archisman is a Senior Manager at Mphasis NEXTlabs. At Mphasis, he conceptualizes, develops, and leads multiple products in the analytics R&D space. He has extensive experience in the IT industry at various project management, research, and engineering roles.

He holds a PhD from the Indian Institute of Management Bangalore (IIMB) in the Quantitative Methods and Information Systems area, and was a visiting researcher at the IT University of Copenhagen during his PhD. His areas of expertise are business analytics, social media, product management, and information systems research.

About Mphasis

Mphasis (BSE: 526299; NSE: MPHASIS) applies next-generation technology to help enterprises transform businesses globally. Customer centricity is foundational to Mphasis and is reflected in the Mphasis' Front2Back™ Transformation approach. Front2Back™ uses the exponential power of cloud and cognitive to provide hyper-personalized ($C = X2C^{2}_{TM} = 1$) digital experience to clients and their end customers. Mphasis' Service Transformation approach helps 'shrink the core' through the application of digital technologies across legacy environments within an enterprise, enabling businesses to stay ahead in a changing world. Mphasis' core reference architectures and tools, speed and innovation with domain expertise and specialization are key to building strong relationships with marquee clients. To know more, please visit www.mphasis.com

For more information, contact: Nextlabs@mphasis.com