SOMAIYA VIDYAVIHAR
K J Somaiya Institute of Engineering and Information Technology
An Autonomous Institute Permanently Affiliated to the University of Mumbai

DEPARTMENT OF INFORMATION TECHNOLOGY

B.Tech - IT - Semester VI - PBL Presentation

Graph Data Mining for Fraud Detection

Team Members:

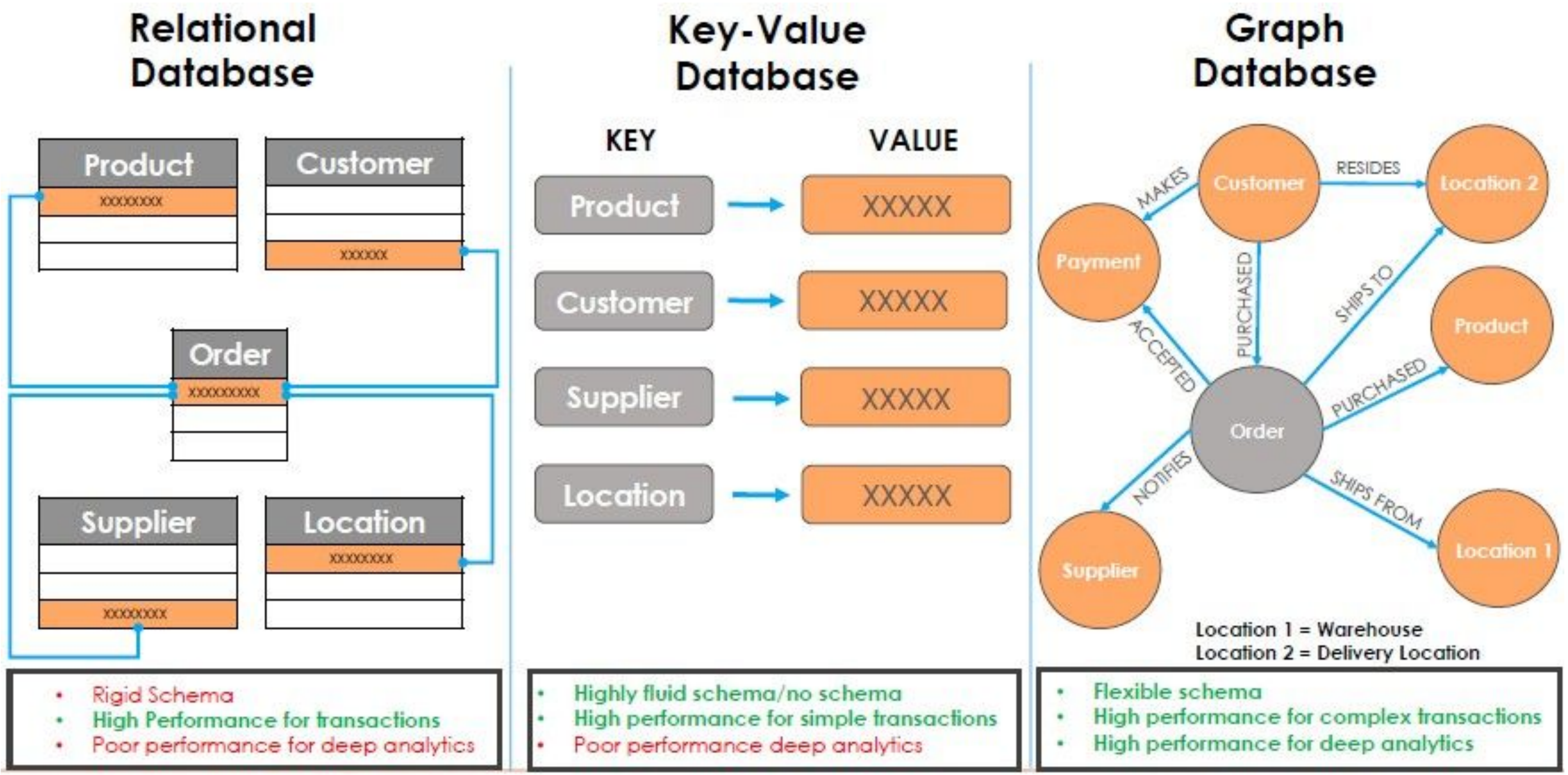
Ananya Kura (Roll No.: 28)
Shaili Mamnia (Roll No.: 32)
Peetamber Pancharia (Roll No.: 42)

Guide:

Dr. Radhika Kotecha
Head – Department of Information Technology

Contents

Rationale


Relational Database
Product
XXXXXXXX
Customer
XXXXXX
Order
XXXXXXXXX
Supplier
XXXXXXXX
Location
XXXXXXXX
Key-Value Database
KEY
VALUE
Product
XXXXX
Customer
XXXXX
Supplier
XXXXX
Location
XXXXX
Graph Database
Customer
MAKES
RESIDES
Location 2
Payment
PURCHASED
SHIPS TO
Product
ACCEPTED
Order
PURCHASED
NOTIFIES
SHIPS FROM
Supplier
Location 1
Location 1 = Warehouse
Location 2 = Delivery Location
Rigid Schema
High Performance for transactions
Poor performance for deep analytics
Highly fluid schema/no schema
High performance for simple transactions
Poor performance deep analytics
Flexible schema
High performance for complex transactions
High performance for deep analytics

The Challenge: By the Numbers

$800 billion - $2 trillion

Estimated amount of money laundered every year

$26 billion

AML and KYC fines by financial services firms from 2008-2018

91%

Percentage of fines paid by U.S. financial service firm

$25 billion

Amount spent annually by financial service firms for AML compliance

$90 million

Average cost of an AML/KYC fine

Literature Review

Ref. No.	Focus / Goal	Approach	Findings	Open Issues
[1]	A case study on the application of various Data Mining techniques for Anti-money Laundering Detection.	Uses multiple techniques such as clustering, neural networks, genetics algorithm, heuristic, etc together to provide a comparative analysis and the most efficient solution for anti-money laundering in banks.	Efficient for real-life data unlike other techniques. Less research available regarding knowledge-based solutions.	Applying Graph Analysis to tackle the problem of very large datasets.
[2]	Applying graph mining approach to identify or detect the suspicious transactions for investigation.	Considers individual transaction dependencies with graph mining method to detect suspicious illegal transactions.	Created a subgraph model using hierarchical patterns and fuzzy numbers.	Accuracy of detection, computational time is high/time-consuming,

Literature Review

Ref. No.	Focus / Goal	Approach	Findings	Open Issues
[3]	Financial Crime & Fraud Detection Using Graph Computing: Application, Considerations & Outlook	Highlights difficulties faced in graph based solutions by organisations for financial processing systems.	Data mining, anomaly detection, sub-graph analysis show effective utilization and exhibit promising results.	Complexity of various techniques arise due to diversity of transactions processing systems.
[4]	Money Laundering Detection using Data Mining and graph analysis with rule based engine.	Used hash based association mining to generate dataset and implement graph theoretical approach to chain accounts in dataset and used rule based AI to find suspicious accounts.	Idea results shown via a graph plotted using the aforementioned algorithm and approach.	Lack of parameters.
[5]	FRAUDRE: Fraud Detection Dual-Resistant to Graph Inconsistency and Imbalance. Ge Zhang, Jia Wu	To detect frauds by analysing features, topological and relational graph inconsistencies or imbalances and build a new model FRAUDRE based on graph neural networks.	Fraudre outperforms all comparative relational graphs, detects camouflaged fraud behaviour unlike other detection algorithms.	Complex unified model.

Problem Statement

"To develop a Fraud Detection System that identifies suspicious transactions and accounts by applying Data Mining techniques on the data present in Graph Database"

Objectives

01

To use Neo4j GDS library to detect and label two types of fraudsters - First party fraudsters and Money Mules

02

To use graph databases to uncover hidden patterns for fraud detection using various Data Mining Techniques.

03

To detect suspicious fraud rings hidden in mobile financial transactions that may exhibit fraud activities.

Scope of Work

- In-scope:

  Transaction parameters, queries in Neo4j to visualize and understand networking.

  To use graphical database such as Neo4j.

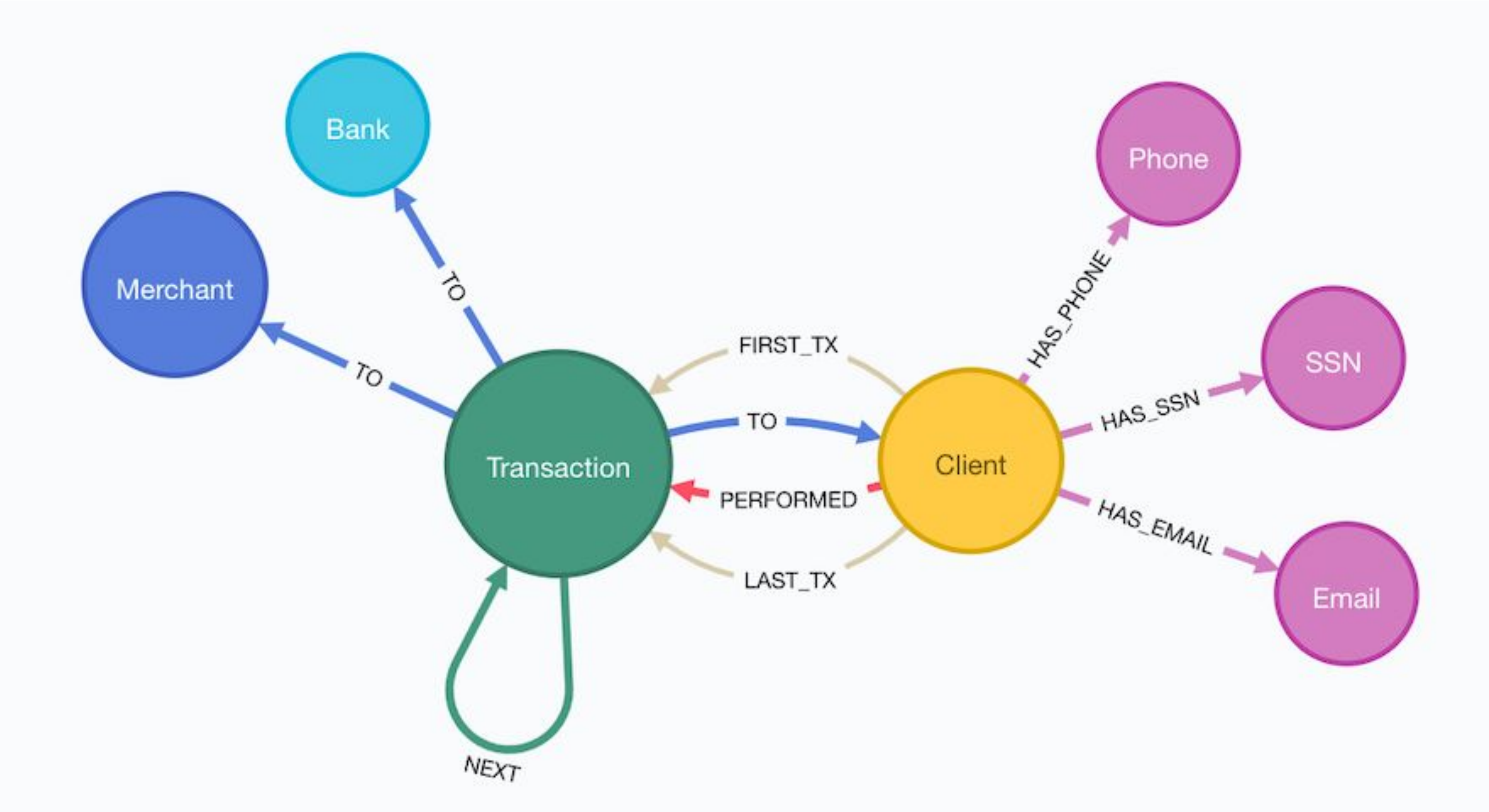
  To provide simple semantics for visualization.

- Out-of-scope:

  Amalgamation of influencing real-time parameters onto this such as rules, etc.

  To Increase accuracy of frauds detected.

Implementation



Database Schema

Fraud Categories

- **First-party Fraud**
  An individual, or group of individuals, misrepresent their identity or give false information when applying for a product or services to receive more favourable rates or when have no intention of repayment.

- **Second-party Fraud**
  An individual knowingly gives their identity or personal information to another individual to commit fraud or someone is perpetrating fraud in his behalf.

- **Third-party Fraud**
  An individual, or a group of individuals, create or use another person's identity, or personal details, to open or takeover an account.

First-Class Frauds

Identify clients that share identifiers and create a new relationship between clients that share identifiers

Identify clusters of clients sharing PII using a community detection algorithm (Weakly Connected Components)

Find similar clients within the clusters using pairwise similarity algorithms (Node Similarity)

Use computed fraud scores to label clients as potential First-Party Fraudsters

Calculate and assign fraud score to clients using centrality algorithms (Degree Centrality)

Second-Class Frauds

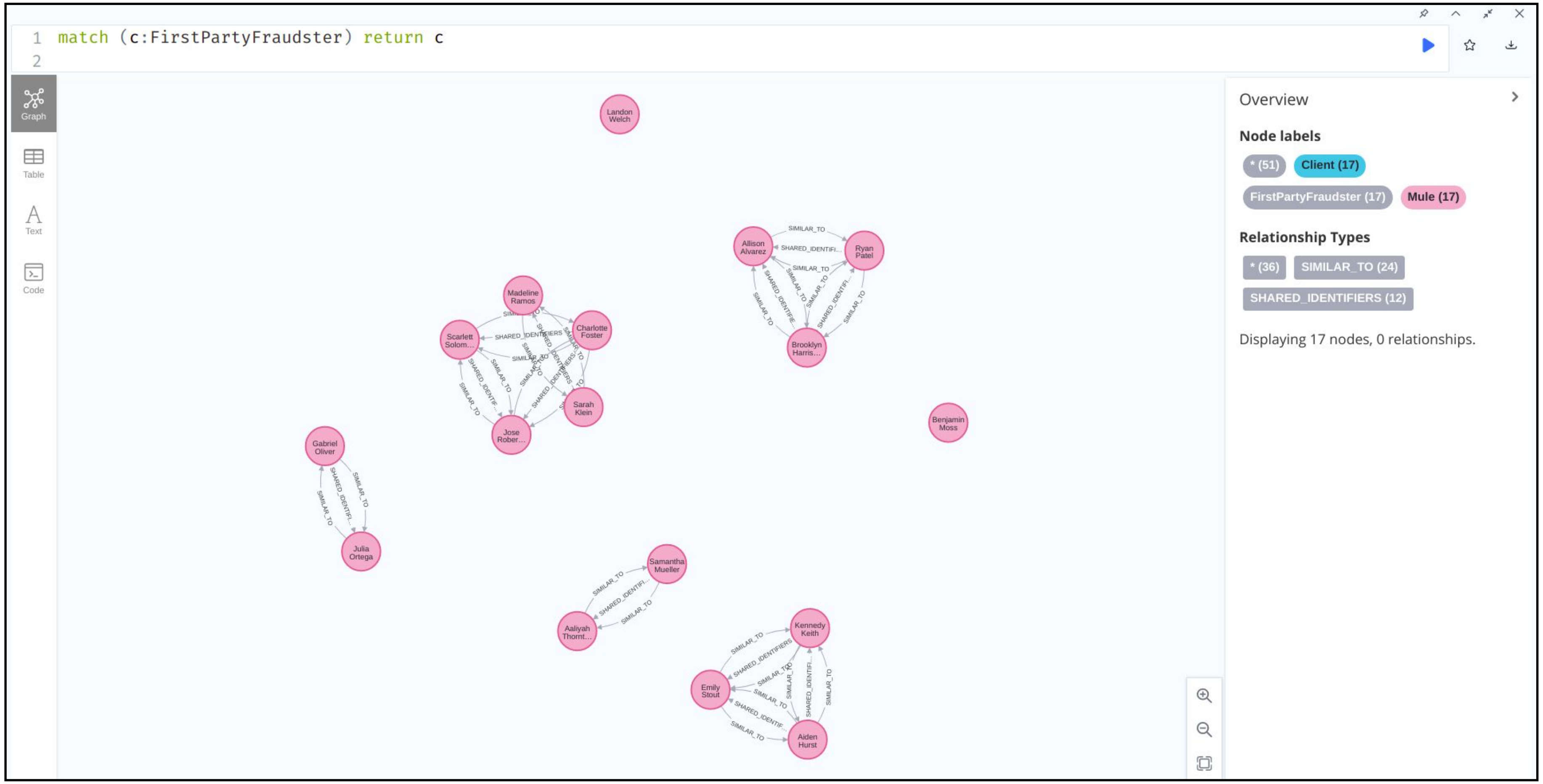
Use WCC (community detection) to identify networks of clients who are connected to first party fraudsters

Use PageRank (centrality) to score clients based on their influence in terms of the amount of money transferred to/from fraudsters

Assign risk score to these clients

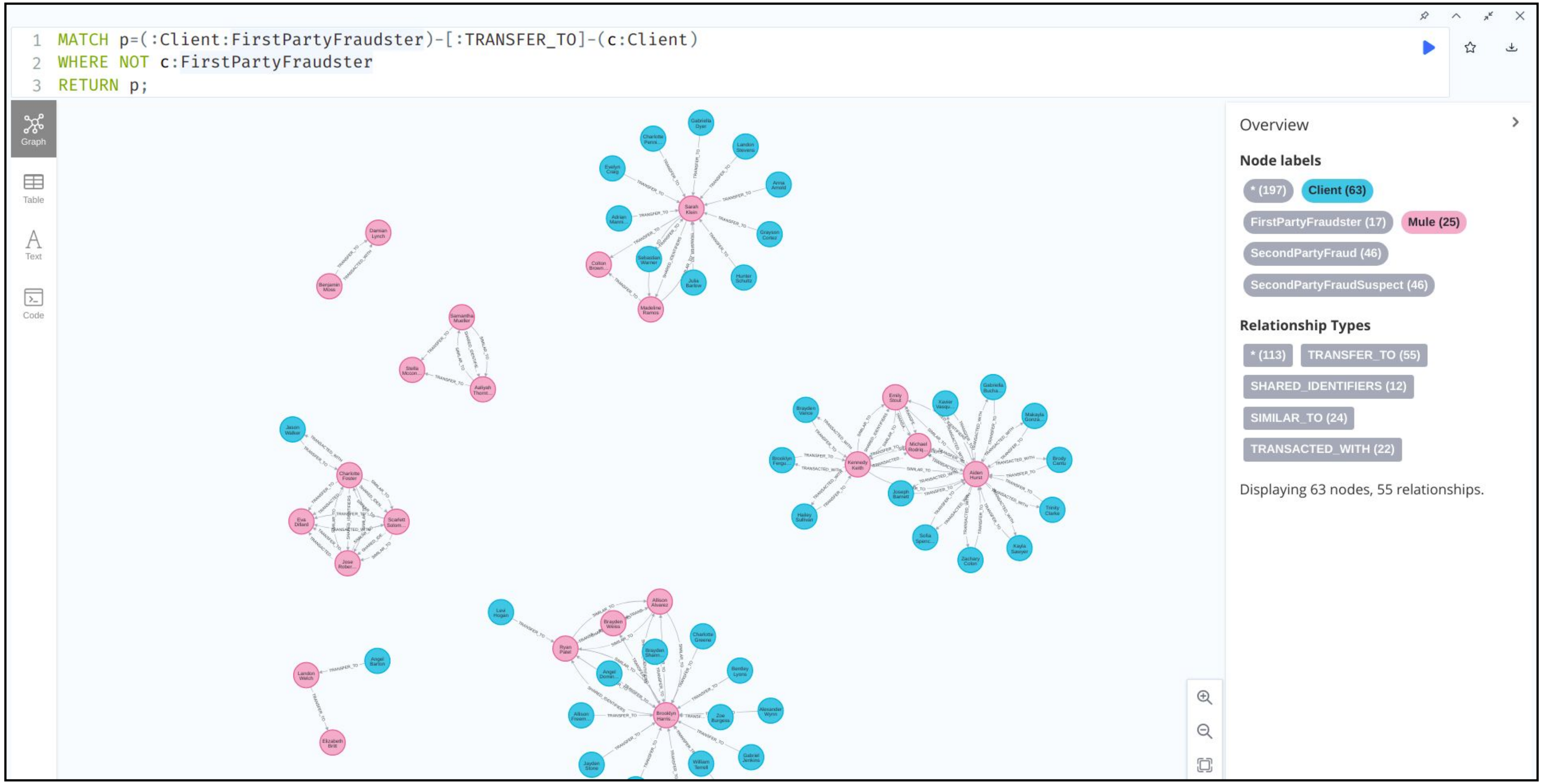
Find 2nd - Party Fraudsters using the risk score.

Results



First-Class Fraudsters

Results



Second-Class Fraudsters

Technology Stack

neo4j

Plan of Execution

Conclusion

Fraud is a connected data problem. Graph data science enables you to uncover more fraud and shut it down quickly. The Neo4j Graph Data Science Library offers an enterprise-ready toolset for running sophisticated graph algorithms on connected data at scale. Graph analytics and feature engineering both add highly predictive relationships to your machine learning for better results.

Future Scope

The current implementation can be improved by analyzing more information like account numbers, IP addresses, etcs,. Graph technology is the ideal enabler for efficient and manageable fraud detection solutions, hence with time, its algorithms and functionalities will improve, giving us more methods to experiment on.

References

[1] Le-Khac, Nhien-An & Kechadi, Tahar. (2010). *Application of Data Mining for Anti-money Laundering Detection: A Case Study*. Proceedings - IEEE International Conference on Data Mining, ICDM. 577-584. 10.1109/ICDMW.2010.66.

[2] Michalak, Krzysztof & Korczak, Jerzy. (2011). *Graph Mining Approach to Suspicious Transaction Detection.* 2011 Federated Conference on Computer Science and Information Systems, FedCSIS 2011. 69-75.

[3] Kurshan, E., Shen, H., & Yu, H. (2020). *Financial Crime & Fraud Detection Using Graph Computing: Application Considerations & Outlook*. 2020 Second International Conference on Transdisciplinary AI (TransAI), 125-130.

[4] Rohini Mohite, Jayraj Ghelani ,Harshal Bhitre, Nikit Sawant , Anita .A. Lahane (2018). *Money Laundering Detection using Data Mining.* IJARCCE ISO 3297:2007. Vol. 7, Issue 4, April 2018. ISSN 2319-5940

[5] Zhang, Ge & Wu, Jia & Yang, Jian & Beheshti, Amin & Xue, Shan & Zhou, Chuan & Sheng, Quan. (2022). *FRAUDRE: Fraud Detection Dual-Resistant to Graph Inconsistency and Imbalance.* DOI-10.1109/ICDM51629.2021.00098.

[6] Nuha khalid Balkir and Faraj A. El-Mouadib. 2021. *Money Laundering Detection System Using Data Mining Functionalities*. The 7th International Conference on Engineering & MIS 2021 (ICEMIS'21). Association for Computing Machinery, New York, NY, USA, Article 12, 1–7. DOI:https://doi.org/10.1145/3492547.3492583

[7] Gorka Sadowksi & Philip Rathle. *Fraud Detection: Discovering Connections Using Graph Databases*, White paper, January 2015.

Thank You!