HELPING THE WORLD MAKE SENSE OF DATA

# GRAPH DATA SCIENCE USE CASES: FRAUD AND ANOMALY DETECTION

*Jaimie Chung*

Product Manager, Graph Data Science

## Problem

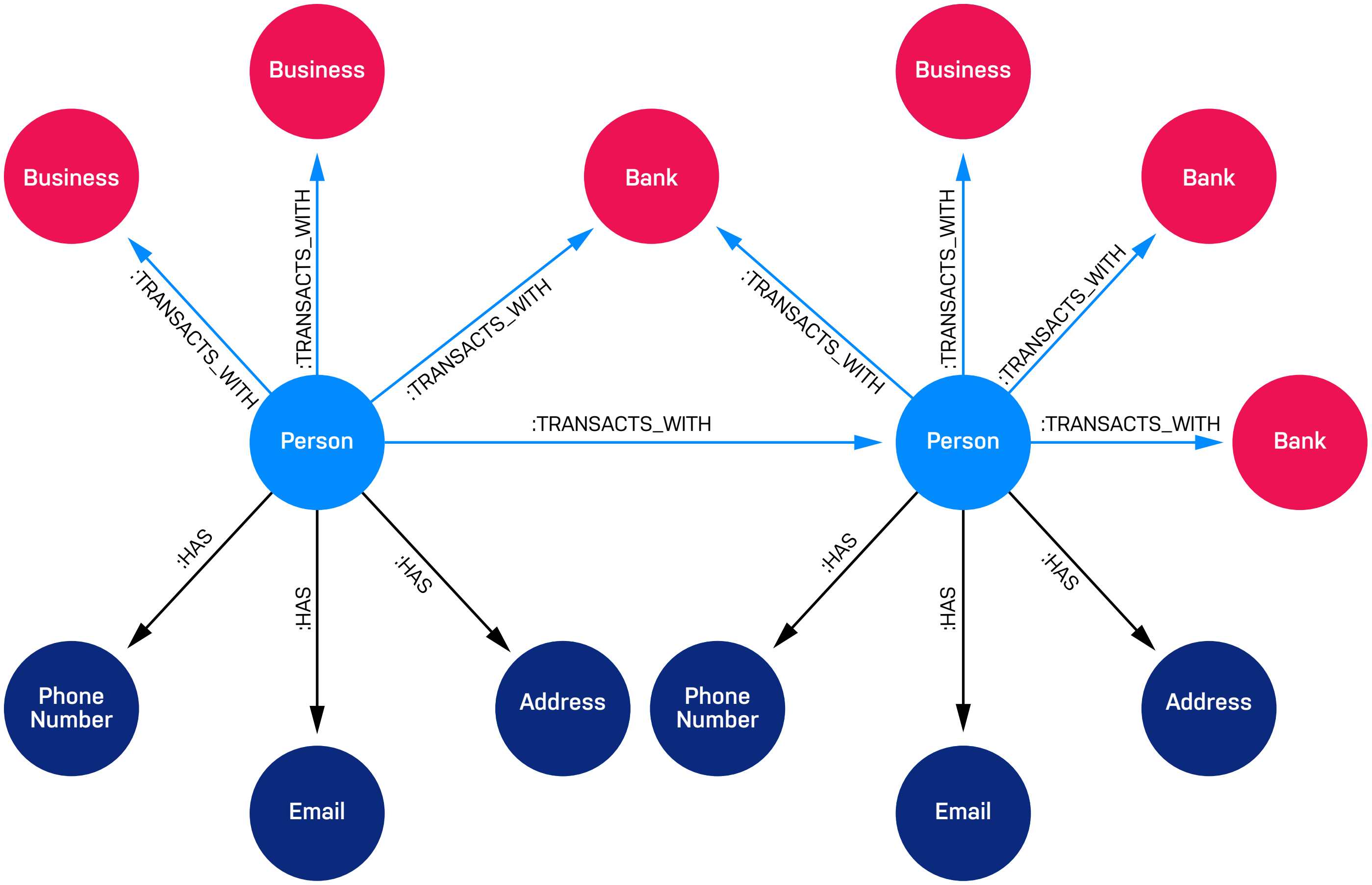
Fraud affects both consumers and businesses, whether it's having your identity stolen, losing money from your bottom line, or exposing nefarious criminal activities such as human trafficking. And criminals are only becoming more prolific as transactions shift toward digital. TransUnion reported a global 149% increase in digital fraud attempts in the first four months of 2021 compared to the last four months of 2020. Fraud targeting businesses is also at an all-time high; a PwC survey of 5,000 businesses in 99 territories showed that 47% had experienced fraud in the last 24 months. This amounted to $42 billion in fraud in 2020.

Without fraud detection practices in place, it's only a matter of time before criminals learn how to steal your money, your identity, or worse. And if you already have fraud analysts manually combing through data, you will always be constrained by human capacity. Many companies are now turning to AI to identify and prevent fraud, using machine learning to identify anomalous or suspicious behaviors. Using machine learning in this way, analysts review only the most likely cases, rather than sifting through billions of data points.

Neo4j's Graph Data Science framework combats fraud by enabling you to identify and predict fraud and anomalies, at scale, using the connections that already exist between your data points. By identifying patterns and creating a feedback loop of insights, consumers can keep their hard-earned dollars, businesses can see direct return on investment, and the world could even become a little bit safer.

## Data Model

A common data model for fraud detection connects people, institutions, and transactions. Mapping financial behaviors in a graph instead of a relational database more accurately represents real-world behavior. Fraudsters aren't analyzed in isolation, but in the context of other fraudsters and their transactions. Algorithms and machine learning techniques can be used to identify fraudsters based on similarity to known fraudsters or to reveal tight communities that may be fraud rings or money launderers.



**A graph data model for fraud and anomaly detection**

This data model demonstrates how entities could be connected to identify fraudulent behavior. For anomaly detection, the nodes could easily be users performing actions on websites who turn out to be bots. This behavior negatively affects your ability to make accurate recommendations to users.

# Solution

Neo4j's Graph Data Science framework offers a variety of analytical approaches to identify fraud and anomalies, ranging from localized query patterns to machine learning–based insights.

### Queries

Cypher is a powerful, graph-optimized query language that lets you find patterns you know exist in your data. For example, people transacting with a known fraudster, or people two or three hops away from a known fraudster, could be flagged for review by a fraud analyst. Cypher makes it simple and efficient to identify people and nodes of interest.

### Graph Algorithms

Neo4j Graph Data Science offers out-of-the-box graph algorithms for similarity and community detection. These algorithms are useful on larger datasets where it's hard to know exactly what you're looking for. These techniques often identify potential fraudsters and anomalies that are not obvious using queries alone. Some examples of useful algorithms include:

- **Community detection algorithms** like Weakly Connected Components can be run to identify first-party fraud where users share identifiers such as IP addresses or social security numbers. Algorithms like Louvain are widely used to identify fraud rings by finding suspicious transaction patterns
- **Similarity algorithms** like KNN use graph structure to find similar profiles, which could find fraudsters based on activities of known fraudsters.
- **Centrality algorithms** like PageRank can be used to find anomalies by scoring accounts based on their transaction behaviors; outliers with oversized impact on your transaction networks may be fraud kingpins.

Together, these techniques lead not only to higher detection rates for fraud and anomalies, but also to higher accuracy, reducing the rate of false positives that require manual review or customer intervention.

### Supervised Machine Learning

Machine learning on a graph encompasses two aspects: how you represent your data and the predictions you want to make with it.

**Graph embeddings** are a powerful tool that take all of the rich information from the connectivity of your graph and encode it into each node, creating a vector (the embedding) that other machine learning techniques build upon. These embeddings enable you to predict changes to the structure of your graph. You can then train a link prediction model to predict duplicate accounts, or use node classification to identify fraudsters and anomalies directly (predicting labels).

# Results

In terms of impact, graph data science affects the bottom line in multiple ways: detecting fraud that evades traditional approaches, achieving higher accuracy in escalating cases for manual review, and, of course, preventing instances of fraud. Although the value of anomaly detection in existing machine learning workflows is more nuanced, comparing results before and after the removal of anomalies readily shows the effectiveness of these techniques.

Graphs are the most effective way to track fraud, crime, and anomalies due to the connected nature of these behaviors. While queries and graph visualizations are a powerful way to get started, graph algorithms and graph machine learning enable organizations to stay ahead of new forms of fraud and crime by leveraging connections to detect the undetectable.

## Customer Spotlight: Banking Circle

Banking Circle relies on Neo4j's Graph Data Science framework to power their fraud detection and anti-money laundering initiatives. With over 150 customers including Stripe, Alibaba, and Paysafe, Banking Circle processed EUR 155bn of payments in 2020 and approached 100 million annual bank transfers by the end of 2021. As a result of using Neo4j GDS, Banking Circle has been able to reduce false negatives by 25% and has halved the numbers of overall alerts escalated for manual review.

The fintech company previously used traditional rule-based strategies, such as assigning risk scores by searching for certain words, accounts, and locations. These alerts were then sent to fraud analysts for manual review. However, Banking Circle realized this solution was slow, expensive, and inflexible and turned to graph and machine learning techniques to scale their operations and impact their bottom line.

Banking Circle's solution involves a graph of accounts connected to payments, and they use community detection to generate features for their machine learning pipeline that detects high-risk clusters. This technique, combined with graph features such as risk scores of near neighbors and distances to tax havens and known fraudsters, has significantly impacted their fraud detection efforts. With GDS, they were able to turn a slow and manual process into a scalable, flexible solution. Banking Circle continues to experiment with GDS algorithms and plans to add more graph features to further tune their model.

## Conclusion

Neo4j's Graph Data Science framework enables you to make sense of your data – at scale.

Whether your data challenge is fine-tuning your machine learning models or combating increasingly advanced fraud, using GDS for fraud and anomaly detection provides fast and accurate results. It's just one of many use cases enabled by graph data science.

Learn more about Neo4j's Graph Data Science framework at neo4j.com/graph-data-science and read about customer use cases. Or get started right away with a Neo4j GDS Sandbox.

Neo4j is the world's leading graph data platform. We help organizations – including **Comcast**, **ICIJ**, **NASA**, **UBS**, and **Volvo Cars** – capture the rich context of the real world that exists in their data to solve challenges of any size and scale. Our customers transform their industries by curbing financial fraud and cybercrime, optimizing global networks, accelerating breakthrough research, and providing better recommendations. Neo4j delivers real-time transaction processing, advanced AI/ML, intuitive data visualization, and more. Find us at **neo4j.com** and follow us at **@Neo4j**.

Questions about Neo4j? Contact us around the globe:

**info@neo4j.com**
**neo4j.com/contact-us**