Analítica de redes, instrumento para la detección de fraude en seguros

El fraude, una actividad delictiva generalmente organizada, requiere la interacción de varios actores que crean las condiciones o explotan vulnerabilidades de los sistemas para conseguir la suplantación, la falsificación o la modificación de documentos y la simulación de accidentes, entre otras manifestaciones en el sector asegurador.

José Manuel Pérez
Profesional de Gestión Institucional Contra el Fraude de Fasecolda

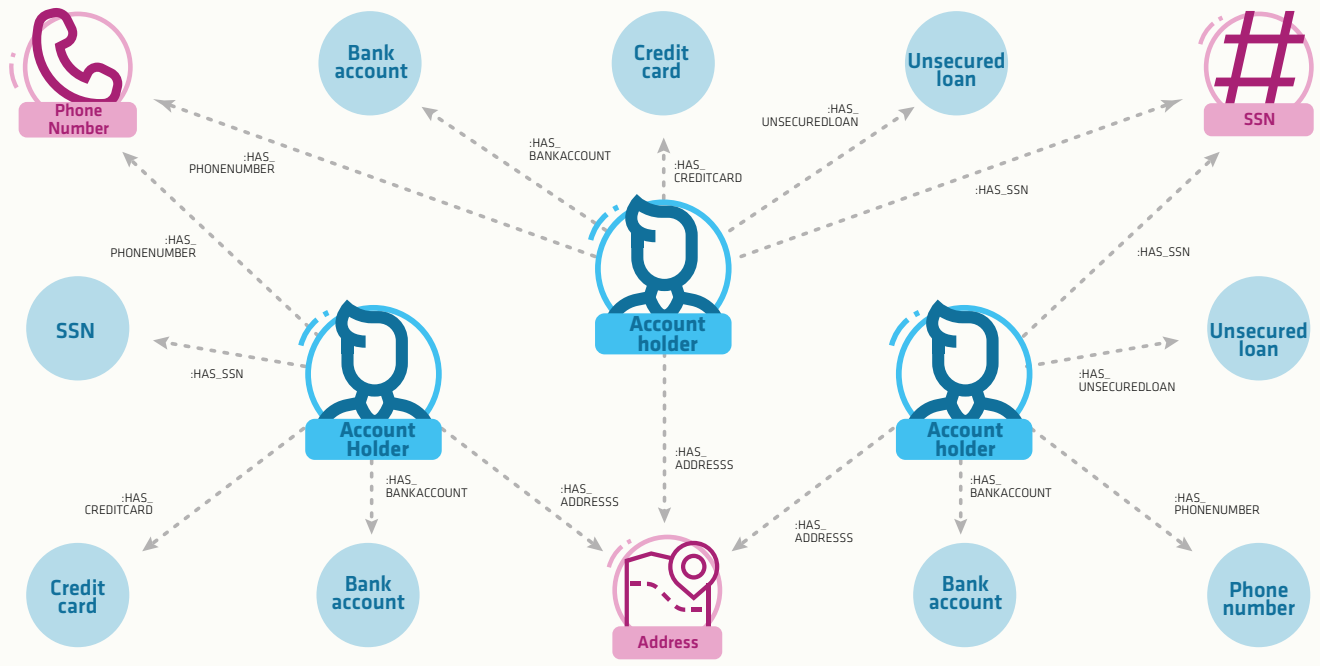
Estas redes van mutando de acuerdo con los controles y las nuevas vulnerabilidades que son detectadas, tanto por los perpetradores como por las autoridades, y presentan ciclos e incidencias geográficas particulares. Las herramientas de analítica, comúnmente utilizadas en la detección del fraude, identifican atipicidades y validan reglas de negocio y reportes de casos confirmados de fraude luego de ser investigados; sin embargo, estas dependen de la calidad y completitud de la información, y del proceso de detección que hacen las compañías y las autoridades.

El uso de la analítica de redes puede complementar y potenciar los procesos de detección de fraude y el entendimiento de las relaciones que permiten estos actos delictivos; además, permite ampliar el alcance de herramientas y métodos tradicionales de alertas y modelos supervisados, que ya enfrentan algunas limitaciones, como bajas tasas de detección, el desbalance de los datos para el aprendizaje automático y el procesamiento estadístico.

La analítica de redes: una arquitectura diferente para aprender cómo se genera el fraude

El uso de analítica de redes es la aplicación de un paradigma complementario al tradicional, es decir, en vez de tener grandes bases de datos estructuradas que pueden ser cruzadas y alimentar modelos de detección, esta información se puede mapear por medio de un grafo o red de actores y mostrar sus relaciones en las reclamaciones que se consolidan.

Ilustración 1:
Ejemplo de red de actores o grafo



Fuente: *Financial Fraud Detection with Graph Data Science. How graph Algorithms and Visualization Better Predict Emerging Fraud Patterns. Neo4j. P. 2*

Formas de uso de la analítica de redes

Usar grafos permite realizar búsquedas específicas dentro de la red (en algunos casos, de manera más eficiente y potente que en tablas regulares), algunas consultas se hacen para explorar comportamientos y relaciones con patrones locales; además, se emplean algoritmos para aprender sobre la estructura y topología general de la red, incluyendo patrones generales y anomalías.

El portal Neo4j publicó el documento *Financial Fraud Detection with Graph Data Science. How Graph Algorithms & Visualization Better Predict Emerging Fraud Patterns* en 2022, en el cual explica que existen diferentes tipos de algoritmos que se pueden implementar para predecir los patrones de fraude, como los de detección de agrupaciones o vecindarios, otros que encuentran similitudes o niveles de centralidad (como el algoritmo PageRank), aquellos que predicen enlaces o relaciones y los que permiten buscar nodos o trayectorias específicas. Cada uno busca entender la red y presenta fortalezas y debilidades específicas.

En el documento se presentan ejemplos de fraudes y su detección:

> *Ejemplo: Identificación de fraude de primera persona*
>
> *En este fraude, una persona (o un grupo de personas) adultera su identidad o brinda información falsa al solicitar un producto o servicio financiero.*
>
> *Según McKinsey, el tipo de fraude de primera persona que crece más rápidamente es el de identidad sintética, en el cual el estafador suele combinar información falsa y real para establecer un historial crediticio bajo una nueva identidad (sintética). Este tipo de fraude genera importantes detrimentos para las instituciones financieras; se estima que el 80% de todas las pérdidas por fraude de tarjetas de crédito se deben a esta modalidad.*
>
> *(Traducción realizada por el autor)*

Ilustración 2. Tipos de algoritmos que se pueden implementar en un grafo

Type of Graph Algorithm	Examples Algorithms
 **Community Detection**	Weakly Connected Components (Union Find), Louvain Modularity, Label Propagation
 **Similarity**	Node Similarity using Jaccard
 **Centrality**	PageRank
 **Heuristic Link Prediction**	Common Neighbors
 **Pathfinding & Search**	Shortest Path

Fuente: *Financial Fraud Detection with Graph Data Science. How graph Algorithms and Visualization Better Predict Emerging Fraud Patterns. Neo4j.*

El ejemplo detalla varios pasos a seguir: el primero es crear un grafo con las relaciones de los individuos involucrados, como números de cuentas, nombres, direcciones IP y de correo electrónico, etc.; el segundo es consultar con los investigadores para definir lo que se debe buscar (atributos comunes, varias personas usando una misma cuenta, patrones en transacciones, entre otros); el tercero es hacer consultas sobre los atributos o aplicar algoritmos de similitud para determinar conexiones y agrupaciones e incluir los resultados en el grafo; finalmente, los puntajes de los algoritmos se convierten en características que se agregan a un modelo de aprendizaje automático para que se puedan identificar los casos fraudulentos con mayor rapidez.

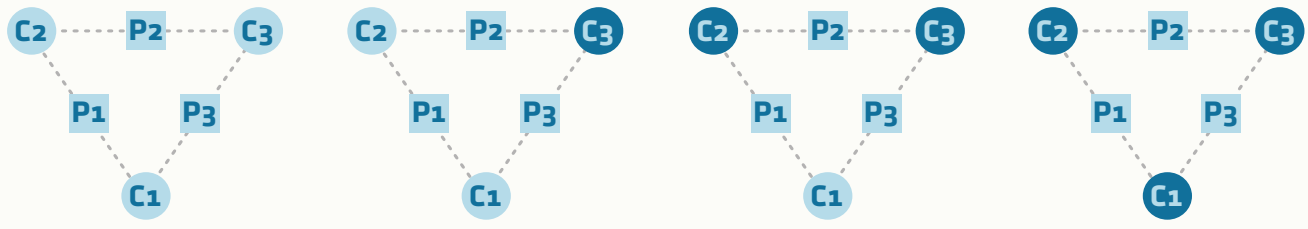
Un caso aplicado a los seguros en automóviles

En el trabajo de Óskarsdóttir et al. (2022) se usa una red social para incorporarla en un modelo supervisado de detección de fraude en seguros. El modelamiento se realizó como fenómeno social en la red y se utilizó el algoritmo BiRank con un vector de consulta específico para calcular una puntuación para cada siniestro. A partir de la red, se extrajeron características relacionadas con las puntuaciones de fraude, así como la estructura de vecindad de las reclamaciones. Las características de la red se combinaron con aquellas específicas de los siniestros y se construyó un modelo supervisado para detectar el fraude en el seguro de automóviles, como variable objetivo; aunque se hizo solo para este ramo, la red incluye siniestros de todas las líneas de negocio disponibles. Los resultados muestran que los modelos con características derivadas de la red funcionan bien a la hora de detectar fraudes e incluso superan a los que utilizan únicamente las particularidades específicas de los siniestros.

> El uso de la analítica de redes permite ampliar el alcance de herramientas y métodos tradicionales de alertas y modelos supervisados

Dentro de los actores modelados en la red se encuentran asegurados, corredores, peritos y talleres, también se incluye a terceras personas, ya sean víctimas directas o no del siniestro; todos ellos se vinculan a las reclamaciones de estudio. Para esta investigación se creó una red bipartita, entre reclamaciones (los círculos) y actores (los cuadrados) (ilustración 3)

Ilustración 3.
Ejemplos de ciclos con reclamaciones fraudulentas



Fuente: Óskarsdóttir M et al. *Risk Analysis*, Vol 42. No. 8, 20

Una vez construida la red es posible detectar ciclos de distintos tamaños que se pueden relacionar con estructuras fraudulentas; en la ilustración 3 estas se encuentran en los nodos resaltados en gris. En el análisis se determinó que los siniestros fraudulentos se agrupan en seis ciclos y que hay una frecuencia relativa con vecinos de segundo y de cuarto orden en el grafo.

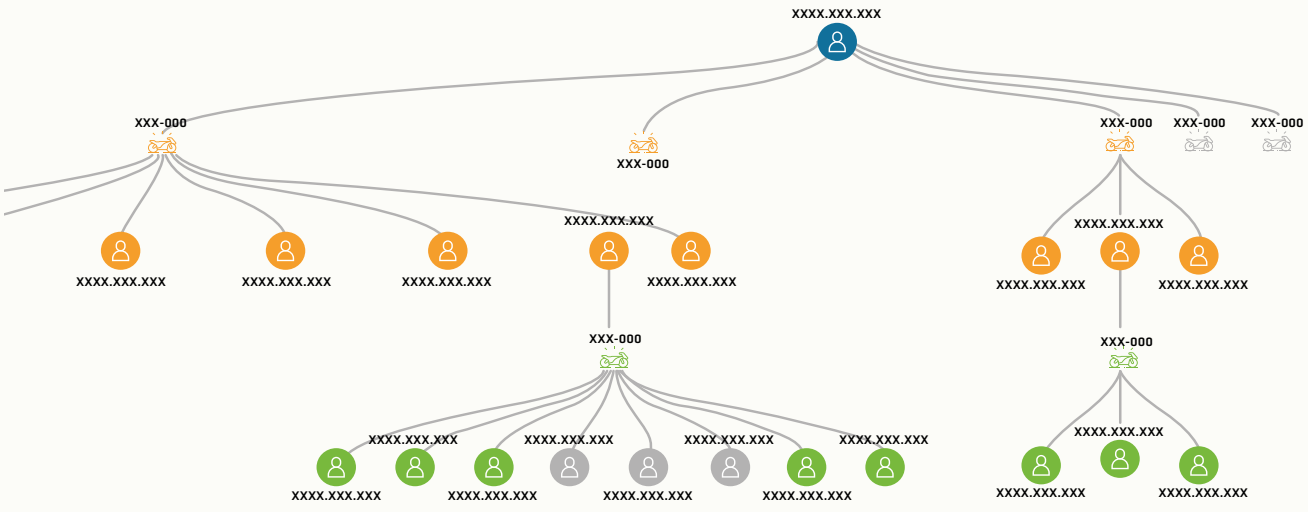
Finalmente, los investigadores usan un algoritmo PageRank personalizado para detectar centralidad de los nodos expuestos a reclamaciones fraudulentas marcadas e incluir estas características como la cantidad de nodos que están conectados en primer y segundo orden o el número de reclamaciones etiquetadas como fraude, conectadas en segundo orden. Luego, estas características se incorporan al modelo predictivo tradicional, que en este caso fue una regresión logística y un modelo de Random Forest o bosque aleatorio, y se comparan sus métricas de rendimiento respecto a los modelos iniciales.

La Dirección de Gestión Institucional Contra el Fraude opera un modelo de redes dentro del Sistema para la Detección Inteligente de Atipicidades en el SOAT (DIAS).

El uso de redes de atipicidad en el sistema DIAS

La Dirección de Gestión Institucional Contra el Fraude opera un modelo de redes dentro del Sistema para la Detección Inteligente de Atipicidades en el SOAT (DIAS), en el cual se conforma una red de placas de vehículos y víctimas, a partir de un actor

Ilustración 4. Ejemplo de una red en el sistema DIAS



Fuente: Óskarsdóttir M et al. *Risk Analysis*, Vol 42. No. 8, 20

inicial que el usuario define. El algoritmo busca a este actor en todas las bases de datos actualizadas del sistema y genera una red de relaciones entre placas y víctimas; además, se puede descargar una tabla resumen que incluye a los prestadores de servicios de salud relacionados.

Este servicio de redes de atipicidad requiere consultas sobre las fuentes de información, que demandan gran capacidad de procesamiento. Para escalar y potenciar esta herramienta se debe usar una infraestructura dedicada a redes (Amazon Neptune, AllegroGraph, Neo4j, entre otros) y mapear los actores en esta nueva arquitectura; incluso es posible construirla con información de varios ramos que pueden estar relacionados, como autos y SOAT, para realizar exploración de actores y reclamaciones atípicas. Finalmente, como se hizo en el estudio mencionado, esto puede mejorar los modelos predictivos que ya están funcionando en los diversos sistemas de Fasecolda.

Referencias

Neo4j. *Financial Fraud Detection with Graph Data Science. How graph Algorithms and Visualization Better Predict Emerging Fraud Patterns.*

Óskarsdóttir M, Ahmed W, Antonio K, Baesens B, Dendievel R, Donas T, Reynkens T. Social Network Analytics for Supervised Fraud Detection in Insurance. *Risk Analysis.* 2022 Aug;42(8):1872-1890. doi: 10.1111/risa.13693. Epub 2021 Feb 6. PMID: 33547691.