

Detección de fraude digital mediante el uso de grafos

Digital Fraud Detection Using Graphs

Diego Saldaña Ulloa*

ROR  Benemérita Universidad Autónoma de Puebla, México

diego.ulloa13@hotmail.com

 <https://orcid.org/0009-0004-1428-6625>

Recepción: 8 de mayo de 2024

Aprobación: 14 de marzo de 2025



Guillermo De Ita Luna

ROR  Benemérita Universidad Autónoma de Puebla, México

guillermo.deita@correo.buap.mx

 <https://orcid.org/0000-0001-7948-8253>

RESUMEN

Se describen algunas características del fraude digital, así como los principales métodos que se han utilizado para la detección de fraude. A su vez, la transición que ha existido hacia un enfoque que combina grafos con aprendizaje automático ha propuesto las llamadas *redes neuronales de grafos*. Finalmente, se mencionan algunos retos del área relacionados con las características de los datos necesarios para operar este tipo de algoritmos. En este sentido, el fraude transaccional en plataformas digitales ha incrementado en años recientes. Los defraudadores recurren a diferentes técnicas para apropiarse de los recursos económicos de usuarios y sacar provecho de algunas vulnerabilidades. Por esta razón, se han implementado diferentes soluciones basadas en el aprendizaje automático para abordar el problema.

PALABRAS CLAVE: redes neuronales de grafos, detección de fraude, aprendizaje máquina

ABSTRACT

This paper describes some characteristics of digital fraud and the main methods used for fraud detection. It also describes the transition to an approach that combines graphs with machine learning, which has led to the development of so-called graph neural networks. Finally, some challenges in the field related to the characteristics of the data required to operate these algorithms are mentioned. Transactional fraud on digital platforms has increased in recent years. Fraudsters use various techniques to steal users' financial resources and exploit specific vulnerabilities. For this reason, various machine learning-based solutions have been implemented to address the problem.

KEYWORDS: graph neural networks, fraud detection, machine learning.

INTRODUCCIÓN

Una definición de *fraude* consiste en aquellas acciones que hacen uso de diferentes tipos de técnicas para ir en contra de la verdad y perjudicar a una persona o conjunto de personas a costa de un beneficio ilícito (Stripe, 2024). El fraude se llega a cometer en diferentes ámbitos como el legal, el económico, el político o el bancario, por mencionar algunos. Independientemente del tipo, provoca numerosos efectos negativos que involucran daños morales, legales o económicos para las entidades que son objetivo del hecho.

Con el avance tecnológico y el continuo desarrollo de la era digital, el fraude ha encontrado formas de adaptarse a la actualidad a través del uso del internet y de las *apps* digitales. En particular, el fraude digital se define como el conjunto de técnicas empleadas en plataformas digitales y que abarcan desde el fraude bancario, el robo de información personal o el acceso no autorizado a cuentas digitales (cuentas de comercios digitales) (Saldaña-Ulloa *et al.*, 2024). Este tipo de fraude ocasiona múltiples pérdidas económicas que afectan a los usuarios de las plataformas y entidades como comercios o bancos. Es común que se cometa a través del uso no

*AUTOR PARA CORRESPONDENCIA

diego.ulloa13@hotmail.com

autorizado de tarjetas bancarias (en la plataforma digital) o también por el acceso no autorizado a las cuentas digitales de los usuarios (Saporta y Maraney, 2022). Lo anterior se traduce en un fraude transaccional que termina impactando negativamente al comercio y al usuario final.

Debido a esta situación, a lo largo de los años se han usado herramientas con el objetivo de detectar este tipo de fenómeno con fines preventivos, que van desde el uso de reglas estratégicamente establecidas hasta el despliegue de algoritmos de aprendizaje automático que manejan grandes conjuntos de información histórica para lograr las tareas de detección. Por otro lado, durante los últimos años se ha propuesto el uso de técnicas que involucran información de redes (grafos) para develar patrones estructurales intrínsecos que ayuden a dicho propósito (Liu *et al.*, 2020; Dou *et al.*, 2020; Saldaña-Ulloa *et al.*, 2024). Este artículo se centra en exponer una reseña sobre aspectos fundamentales de la detección de fraude digital, los métodos utilizados y los que en la actualidad están basados en el uso de redes neuronales con grafos.

1. FRAUDE DIGITAL EN PLATAFORMAS EN LÍNEA

Con el continuo avance tecnológico existen cada vez más aspectos de la vida cotidiana que de una u otra forma se van incorporando al mundo digital. Desde el pago de servicios básicos del hogar, la administración de cuentas de ahorro, aspectos educativos relacionados al aprendizaje, la movilidad mediante el uso de transporte, actividades de ocio orientadas al entretenimiento e inclusive la comunicación entre personas. Cada uno de estos aspectos involucra por parte de los usuarios, el acceso de información a las plataformas proveedoras de estos servicios. Debido a esta dinámica, existen riesgos que comprometen su información e involucran procesos de fraude dentro de dichas plataformas en línea.

Los casos más comunes se relacionan con transacciones monetarias fraudulentas que involucran el uso no autorizado de medios electrónicos de pago, como tarjetas bancarias o fondos monetarios existentes en las cuentas de los usuarios dentro de las plataformas (Jendruszak, 2024). A menudo, este último punto está relacionado (o no) con una apropiación de cuentas (*account takeover*), producto de campañas de hackeo enfocadas a robar la información del acceso a cuentas del usuario.

Esta situación va escalando y en algunos casos implica acciones por parte de grupos delictivos (defraudadores) dedicados exclusivamente a explotar vulnerabilidades del lado de los usuarios y de las plataformas digitales (FAFT-Interpol, 2023); en otros casos, el fraude lo cometen grupos más pequeños e inclusive una sola persona. En cualquiera de los casos, los defraudadores cometen transacciones fraudulentas con el uso de tarjetas que son obtenidas de forma ilegal en foros de internet de la red oscura o *dark web* (J. P. Morgan Chase, 2021).

Las transacciones fraudulentas a partir del uso de tarjetas no propias es el tipo de fraude más común dentro de una plataforma digital y consiste en un robo de los fondos de tarjetas que han sido comprometidas, frecuentemente con hackeos masivos a las bases de datos de los bancos. Después, la información de las tarjetas robadas es puesta en venta de manera ilegal y adquirida por defraudadores (BBC, 2024). Las transacciones fraudulentas concretadas representan una enorme pérdida económica para usuarios, comercios y bancos. Se estima que para 2028 el monto acumulado de pérdidas debido a fraude digital sea de 362 000 millones de dólares (Malone, 2023).

Otro tipo de fraude que se traduce en transacciones fraudulentas en plataformas en línea es el ya mencionado *account takeover*, donde el acceso a la cuenta de un usuario se ve comprometido y el defraudador tiene la capacidad de realizar diferentes acciones, por lo regular el robo de fondos monetarios. Esta situación se debe al robo explícito del dinero dentro de la cuenta del usuario o al uso de las tarjetas de pago registradas (Seon, 2024).

Los defraudadores suelen conseguir el acceso a las cuentas de los usuarios mediante campañas de *phishing* (suplantación de identidad), que se define como el conjunto de técnicas aplicadas para engañar al usuario y robarle información al simular un comportamiento legítimo, por ejemplo con correos electrónicos ficticios de entidades bancarias, campañas publicitarias u ofertas muy llamativas (Seon, 2024). En la figura 1 se presenta un diagrama de cómo es la operación de campañas fraudulentas para obtener acceso de manera ilícita a la información de los usuarios.

Ambos tipos de fraude –por uso no autorizado de tarjeta o por *account takeover*– representan un peligro continuo para todos los actores involucrados en el uso de plataformas digitales. Por esta razón, se tienen que tomar algunas medidas para frenar este comportamiento. Recientemente, el uso de algoritmos basados en aprendizaje de máquina (*machine learning*) ha sido la herramienta más usada para dar solución a este problema.

Con la información generada en las plataformas digitales, los usuarios quedan identificados en cuanto a su comportamiento, por ejemplo el correo electrónico para acceder a la cuenta, las acciones realizadas dentro de cada plataforma, el tipo de dispositivo para conectarse, etc. Este tipo de datos ayudan a generalizar y diferenciar entre un comportamiento legítimo o uno fraudulento (Ravelin, 2024).

Con esta información también surgen otro tipo de perspectivas útiles en un proceso de detección de fraude. Esto se logra traduciendo los datos a una representación del conocimiento basada en grafos (redes), gracias a la cual se accede a capas ocultas que de otra manera no se obtendrían, o que probablemente serían más difíciles de notar con el uso de algoritmos tradicionales.

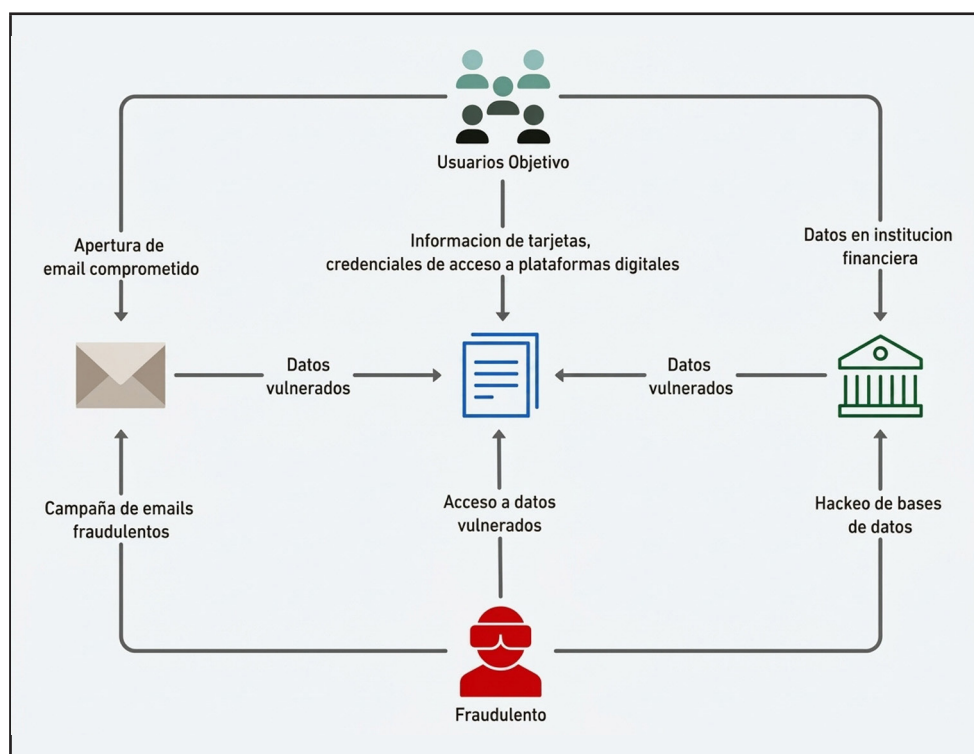


FIGURA 1

Diagrama del robo de información del usuario

Fuente: elaboración propia.

Nota: los fraudulentos organizan campañas de correos para obtener acceso a los datos de los usuarios y, de la misma forma, datos de tarjetas a través del hackeo a bases de datos de entidades financieras

2. BREVE RESEÑA DE LOS MÉTODOS DE DETECCIÓN DE FRAUDE

Históricamente, la detección de fraude digital comenzó con la implementación de reglas construidas por analistas de fraude. Estas reglas se basan en patrones explícitos sobre el comportamiento de usuarios fraudulentos dentro de la plataforma (Saporta y Maraney, 2022). Sin embargo, el uso de reglas supone una alta tasa de falsos positivos (casos no fraudulentos detectados como fraude), ya que se basan en intervalos muy específicos relacionados con un comportamiento general.

A principios de este siglo, y con el continuo desarrollo del área de aprendizaje automático, comenzaron a implementarse algoritmos más complejos para dar solución al problema del fraude digital. Unos de los primeros métodos fueron los basados en perceptrones y redes bayesianas (Maes *et al.*, 2002). Los perceptrones son arquitecturas sencillas de las llamadas *redes neuronales*, y se basan en aplicar algoritmos con el objetivo de hacer una aproximación real sobre un conjunto de datos de entrada. El proceso involucra el uso de un gran conjunto de datos que representen observaciones del fenómeno a modelar.

Las redes bayesianas son modelos probabilísticos que trabajan mediante dependencias condicionales sobre los datos de entrada para otorgar un resultado. De manera similar, algunos algoritmos del área de minería de datos han sido empleados para analizar datos financieros y encontrar patrones de riesgo (Yue *et al.*, 2007). A pesar de que estos enfoques resultaban novedosos, debido a su complejidad era difícil aplicarlos en procesos de detección en tiempo real, por lo que más bien se dejaron para tareas de análisis.

Con el paso del tiempo y la implementación de nuevos paradigmas de aprendizaje automático, la detección de fraude se ha estandarizado hacia el uso de metodologías usadas en la minería de datos. Por esta razón, diferentes tipos de algoritmos se comenzaron a aplicar para dar solución a este problema (Meng *et al.*, 2020), entre ellos el uso de redes neuronales.

Considerando únicamente una metodología enfocada en modelos de aprendizaje automático, el problema se centra en encontrar una serie de características que categoricen mejor las observaciones correspondientes al fraude digital. De este modo, los algoritmos tienen la capacidad de separar observaciones fraudulentas de las que no lo son. Parte de esta solución se centra en la llamada *ingeniería de características* o *feature engineering*, que se refiere al proceso de modelado o descubrimiento de características en las transacciones. Este proceso se realiza de forma manual y a veces toma tiempo en su implementación. Entre los tipos de características más utilizados en un proceso de detección de fraude se encuentran los relacionados con las transacciones, los dispositivos y las tarjetas, así como el comportamiento general dentro de la plataforma digital, solo por mencionar algunas características.

En los algoritmos que usan redes neuronales, el proceso de *feature engineering* se lleva a cabo de forma parcial dentro del propio algoritmo (Fu *et al.*, 2016) y no es necesario una preselección de las características. Este enfoque forma parte del marco del aprendizaje profundo que se centra en extraer patrones intrínsecos de los datos de entrenamiento (datos utilizados para la creación del algoritmo).

En la actualidad, las soluciones de aprendizaje automático aplicadas a la detección de fraude digital tienen un problema relacionado a la poca disponibilidad de conjuntos de datos para realizar investigaciones (Sulaiman *et al.*, 2022). Esto se debe a que la información contiene datos sensibles relacionados con el usuario y su interacción con la plataforma digital, además de las políticas de privacidad relativas a la no divulgación de cómo operan los procesos de detección con el objetivo de que terceros no hagan un mal uso de esta información (Saporta y Maraney, 2022).

Hasta mediados de la primera década del siglo XXI, los diferentes algoritmos de detección de fraude utilizaban principalmente datos relacionados a las transacciones de usuarios, pero en la última década, y gracias al enfoque de aplicar teoría de grafos en el área de aprendizaje automático, surgieron nuevos modelos basados en grafos que mejoran los resultados de los algoritmos de clasificación (Zhao *et al.*, 2019).

Los grafos son modelos abstractos que representan conocimiento mediante vértices y aristas que, entre otras, permiten la construcción de redes, en nuestro caso, redes relacionadas con la interacción de individuos en plataformas digitales. Sin embargo, los grafos también tienen diferentes aplicaciones en diversas áreas como las ciencias sociales, química, física, por mencionar algunos. En general, un grafo puede ser dirigido (o no dirigido) dependiendo de si las aristas poseen una dirección definida (o no marcan una dirección). De la misma forma, también se obtienen particiones del grafo llamados *subgrafos* que son útiles en conceptos como el vecindario local de un vértice, es decir, los elementos más cercanos a un vértice (Hamilton *et al.*, 2017).

Por otro lado, se ha encontrado que el uso de características provenientes de un modelo mediante un grafo mejora las métricas de clasificación (verdaderos/falsos positivos), por lo que el enfoque actual en el diseño de algoritmos de aprendizaje automático se centra en la información sobre la red (grafo) de los usuarios y sobre las características transaccionales (Lebichot *et al.*, 2017).

3. ¿COMO DETECTAR FRAUDE USANDO GRAFOS?

Como se ha mencionado, la vía más directa de detectar transacciones fraudulentas utilizando un enfoque de grafos es con la incorporación directa de características (extraídas del grafo); sin embargo, este tipo de enfoque no toma en cuenta parte de la información que de manera implícita está contenida. En la figura 2 se muestra un ejemplo de cómo la información tabular se representa con base en un grafo y así extraer características para su posterior procesamiento.

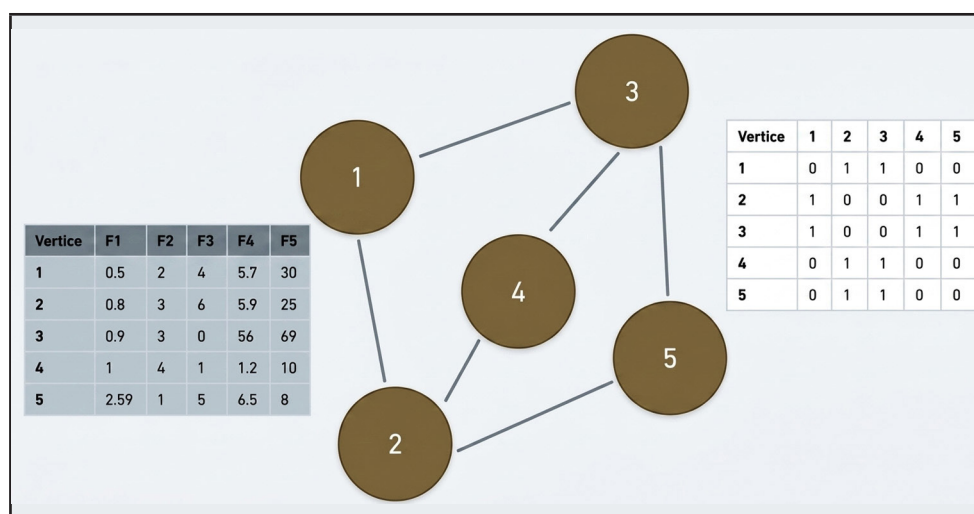


FIGURA 2
Representación tabular y mediante grafos de información

Fuente: elaboración propia.

Nota: en la parte derecha se muestra una tabla que representa si dos vértices están unidos con una arista. Los "0" indican que no existe unión entre vértices, mientras que los "1" indican que hay una unión entre los vértices (de renglón con columna). La imagen central es la representación mediante grafos de la tabla de la derecha y la tabla de la izquierda es una capa extra de información que muestra que cada vértice puede estar asociado a un conjunto de características (*features*). Los nombres de las columnas F1, F2,..., F5 corresponden a 5 diferentes *features*. En este ejemplo cada vértice (fila) posee valores de *features* diferentes.

Como una forma de representación del conocimiento, un grafo posee más información relacionada con el contexto de vértices y aristas que se infieren del número de conexiones que tiene un vértice, de los tipos de vértices que se conectan entre sí e, inclusive, por la distancia que existe entre un par de vértices. Es difícil que este tipo de contexto sea empleado por algoritmos tradicionales como árboles de decisión o clasificaciones por grupo (*clusterings*), ya que la forma en que operan estos métodos se basa en una ingesta de datos tabulares (información representada mediante arreglos de valores) (Saldaña-Ulloa *et al.*, 2024).

3. 1. Metodología para la detección de fraude con grafos

Para tener la capacidad de aprovechar información contextual, se necesita un método que combine los enfoques tradicionales del aprendizaje automático y el uso de grafos. La manera más directa es una combinación de algoritmos, por ejemplo redes neuronales y grafos. Este paradigma se refiere a las llamadas *redes neuronales de*

grafos (*graph neural networks* o GNN) que funcionan gracias a una agregación de información en el grafo y un procesamiento mediante una red neuronal (Saldaña-Ulloa *et al.*, 2024).

Los algoritmos GNN pueden aplicar el llamado *paso de mensaje*, que consiste en obtener información a través de vértices vecinos. La figura 3 muestra un esquema general del proceso del paso de mensaje entre los vértices de un grafo.

Con el paso de mensajes se logra tener contexto sobre el vecindario local respecto de un vértice objetivo. Los datos que se agregan (combinan) corresponden a las características obtenidas para cada uno de los vértices; estas características son las que definen el comportamiento de los vértices en el grafo. Una vez que la información es combinada, se introduce a un algoritmo tradicional de redes neuronales, en donde el objetivo se centra en clasificar si el vértice pertenece a una determinada clase (Saldaña-Ulloa *et al.*, 2024). La clasificación con este tipo de método se basa en el contexto del grafo construido, es decir, los vértices y aristas cercanos o la disposición estructural de cada uno de ellos.

Para detectar transacciones fraudulentas mediante un enfoque de GNN primero se necesita construir un grafo que modele la forma en la que los usuarios interactúan y transaccionan en una plataforma digital. Así, los vértices del grafo corresponderían a usuarios, dispositivos, tarjetas, transacciones o cuentas de banco, por mencionar algunas entidades involucradas. Estos vértices pueden ser clasificados para determinar si corresponden (o no) a una observación fraudulenta, por ejemplo reconocer los vértices que se involucran con una transacción que se cataloga como fraude o si un usuario es predicho como fraudulento (Saldaña-Ulloa *et al.*, 2024).

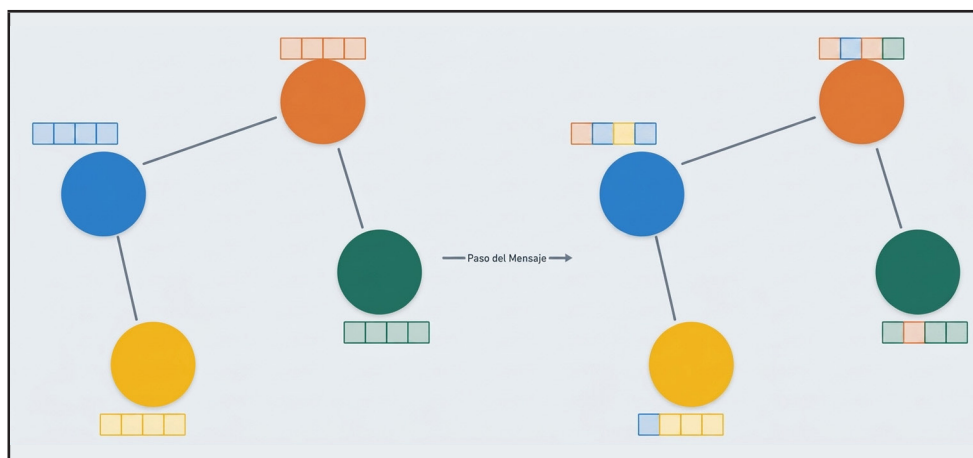


FIGURA 3

Proceso de paso del mensaje entre vértices de un grafo

Fuente: elaboración propia.

Nota: cada vértice posee un conjunto de características (*features*) definidas por las cuadrículas de colores; el paso de mensaje permite un intercambio de información entre vértices. Por esta razón, en la parte derecha de la imagen se observa que ahora los vértices tienen información de sus vecinos (cuadrículas de diferentes colores).

3. 2. Tipos de algoritmos de grafos utilizados para detectar fraude

Adicionalmente, dentro del paradigma GNN, existen algoritmos que trabajan sobre diferentes tipos de grafos, esto es, grafos homogéneos o heterogéneos. Los grafos homogéneos son aquellos que poseen un único tipo de arista y un único tipo de vértice (figura 4).

Por otro lado, los grafos heterogéneos poseen diferentes tipos de aristas y vértices. En una aplicación de detección de transacciones fraudulentas resultan más útiles métodos con grafos heterogéneos, ya que otorgan mayor cantidad de información y contexto sobre lo que se va a resolver (ejemplo de grafo heterogéneo en la figura 5).

A grandes rasgos, los algoritmos de redes neuronales con grafos heterogéneos (HGNN) trabajan con base en diferentes tipos de ponderaciones en las aristas existentes con la finalidad de otorgar mayor relevancia a cierto tipo de aristas (Xiao *et al.*, 2020). Usar diferentes tipos de aristas es muy útil en el proceso de detección de fraude, ya

que modela diferentes tipos de relaciones entre vértices y de transacciones en el sistema de una plataforma digital.

Y puesto que el problema del fraude es un fenómeno dinámico, es decir presenta cambios a lo largo del tiempo, resulta útil incorporar este tipo de información dinámica en el mismo grafo. Por ello, también existen algoritmos GNN que consideran el aspecto temporal (Rossi *et al.*, 2020), ya sea por medio de grafos homogéneos o heterogéneos.

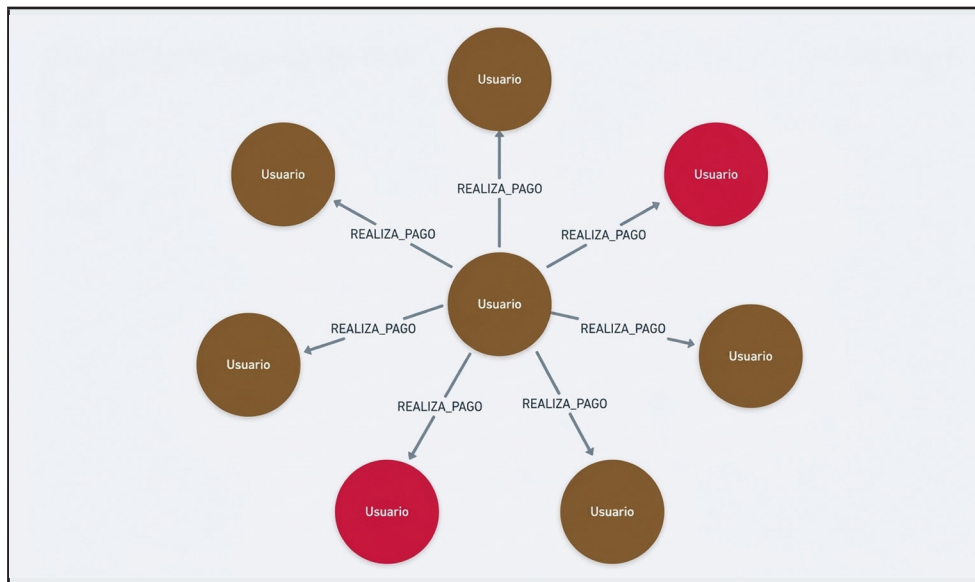


FIGURA 4

Representación de un grafo homogéneo

Fuente: elaboración propia.

Nota: los vértices rojos representan a usuarios fraudulentos.

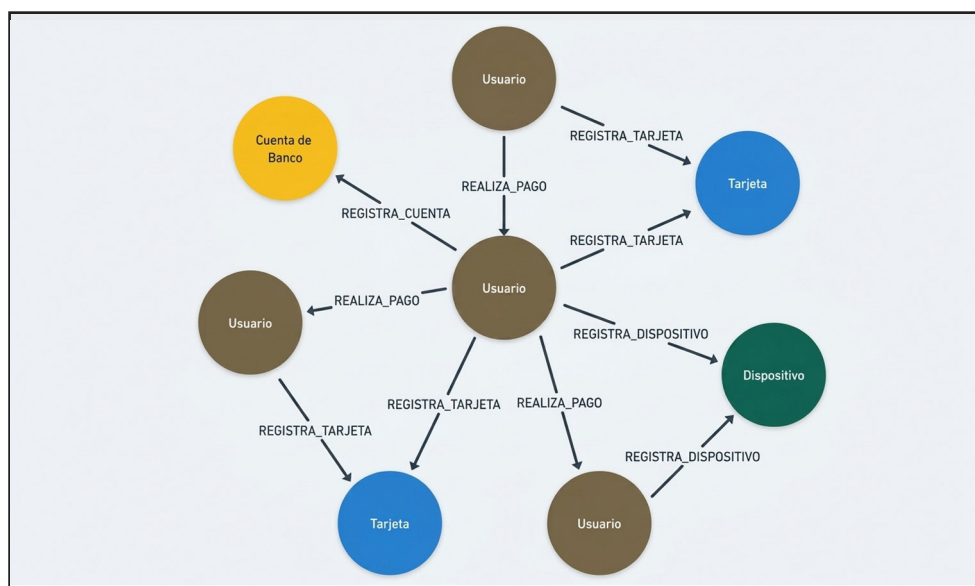


FIGURA 5

Representación de un grafo heterogéneo

Fuente: elaboración propia.

Nota: en este grafo los usuarios completan varios tipos de registros con otras entidades como tarjetas, dispositivos, cuentas de banco u otros usuarios.

La perspectiva dinámica permite obtener una capa adicional de información que ayuda a revelar patrones ocultos del fenómeno, y así mejorar el proceso de detección de fraude. En la figura 6 se ilustra un caso de un grafo heterogéneo que modela procesos dinámicos al agregarse el aspecto temporal de cuando se realizan las interacciones entre los vértices del grafo.

3.3. Perspectiva general sobre detección de fraude usando grafos

El enfoque del uso de grafos en la detección de fraude es relativamente nuevo y existen algunos trabajos que han recurrido a estas herramientas para abordar el problema (Dou *et al.*, 2020; Saldaña-Ulloa *et al.*, 2024). La mayoría de estas propuestas se centran en un paradigma de grafos homogéneos o heterogéneos. Sin embargo, estos últimos son los que se prefieren, ya que la problemática del fraude se ataca con una gran variedad de información contextual procedente de diferentes fuentes (Liu *et al.*, 2020).

En la detección de fraude se deben combinar aspectos dinámicos del problema, que ya han sido expuestos debido, sobre todo, a que la heterogeneidad y la temporalidad de los datos sirven como fuentes extra de información para los algoritmos enfocados a la prevención del fraude. Al respecto, los algoritmos de este tipo que se encargan de combinar diferentes metodologías hacen uso de grandes cantidades de información y es por ello que otro de los aspectos relevantes para este tipo de trabajo recae en la calidad y la cantidad de información, así como en el orden en que se ejecutan los pasos de preprocesamiento y desarrollo algorítmico (Rao *et al.*, 2022).

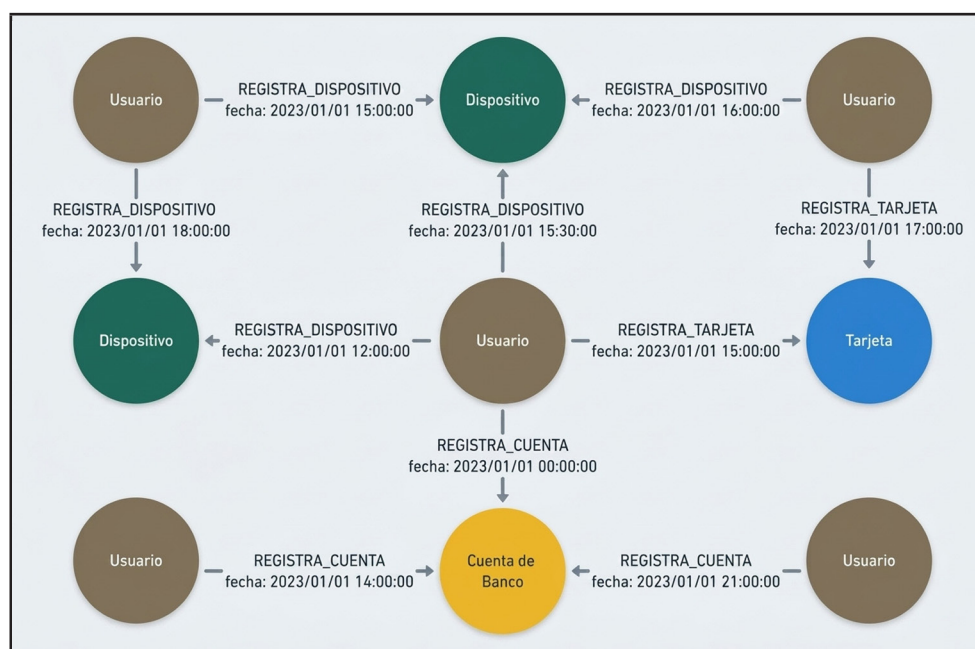


FIGURA 6
Ejemplo de un grafo heterogéneo dinámico

Fuente: elaboración propia.

Nota: cada arista tiene asociada una marca de tiempo que indica el momento en el que se realizó un evento. Los algoritmos que utilizan este tipo de representación tienen que tomar en cuenta las posibles correlaciones relacionadas en el tiempo. En esta figura todos los vértices son diferentes y solo se están agrupando por tipos de vértice (Dispositivo, Usuario, Cuenta de Banco y Tarjeta) con el uso de colores diferentes.

Cabe destacar que en los modelos donde se usan grafos para la detección de fraude, el proceso de paso de mensaje sirve como pilar para considerar el aspecto estructural de la información. Por medio de este enfoque el comportamiento de entidades fraudulentas queda expuesto, ya que el vecindario local ayuda a distinguir obser-

vaciones irregulares de observaciones normales. Más aún, existen ocasiones donde es más relevante considerar ciertos elementos de un grafo. En este caso, al proceso de selección de partes claves del grafo se le llama *mecanismo de atención* y se utiliza para ponderar la información del grafo con la finalidad de encontrar los patrones más relevantes que ayuden a predecir un comportamiento fraudulento (Dou *et al.*, 2020).

Por otro lado, el procesamiento temporal de la información de un grafo puede llevarse a cabo por medio de un enfoque discreto o dinámico. Un enfoque discreto consiste en una representación de varios grafos divididos en intervalos discretos de tiempo en la cual las operaciones fraudulentas se evalúan con mayor contexto. El enfoque dinámico procesa la información del grafo a partir de una secuencia de eventos, en donde cada unión de vértices mediante aristas está asociada a una marca de tiempo. Los grafos dinámicos continuos basados en eventos contienen un mayor contexto para detectar fraude debido a los patrones temporales que surgen del comportamiento de usuarios.

4. RETOS Y LIMITACIONES DEL USO DE GRAFOS EN LA DETECCIÓN DE FRAUDE

A pesar de las ventajas que posee un enfoque de detección de fraude basado en grafos, existen ciertas consideraciones que se deben tomar en cuenta para su implementación. Esto tiene que ver con la cantidad de información necesaria para entrenar (desarrollar) el modelo. En un uso real este tipo de algoritmos se tienen que alimentar de un gran conjunto de datos provenientes de alguna plataforma digital (Ravelin, 2024). Sin embargo, es imposible procesar toda la información en un solo paso debido a las limitaciones respecto al *hardware* de los sistemas.

Por esta razón, comenzaron a desarrollarse algoritmos orientados a los métodos relacionados al entrenamiento de GNN. Estos enfoques realizan un muestreo (Hamilton *et al.*, 2017) sobre el grafo de modo que en cada iteración del entrenamiento del algoritmo solo se procesa una porción; por consiguiente, se optimizan los tiempos de ejecución en el desarrollo. Debido a lo anterior, uno de los objetivos actuales del área se centra en el desarrollo de nuevos métodos para procesar la información, sobre todo para el caso de información proveniente de plataformas digitales (Rao *et al.*, 2022).

Adicionalmente, es necesario considerar aspectos relacionados a la inferencia (predicción) de observaciones fraudulentas, puesto que para generar un resultado en algunas ocasiones es necesario recolectar la información del vértice objetivo y de su vecindario (Hamilton *et al.*, 2017), lo cual agrega una capa adicional de complejidad.

Para finalizar, se debe hacer énfasis en el desarrollo de trabajos que utilicen datos reales para la verificación de algoritmos (provenientes de plataformas digitales) (Saldaña-Ulloa *et al.*, 2024). Sin el uso de información real, el proceso de validación de un método se hace más difícil. Algunos de los trabajos actuales para detección de fraude con grafos son resultado de colaboraciones con plataformas digitales que cuentan con información real (Dou *et al.*, 2020; Liu *et al.*, 2020; Rao *et al.*, 2022; Saldaña-Ulloa *et al.*, 2024).

PROSPECTIVA

La detección de fraude representa un problema que perjudica a usuarios y empresas principalmente por los recursos económicos que se ven comprometidos. El fraude es un problema dinámico que requiere tomar en consideración diferentes aspectos ya mencionados, incluido el uso de algoritmos de aprendizaje profundo. Los aspectos fundamentales recaen en algoritmos computacionales que sean útiles al momento de procesar la información.

En relación con los datos, se requiere que estos sean representativos del problema, por esa razón resulta de mayor utilidad enfoques con datos de transacciones reales y con millones o cientos de millones de registros. En este trabajo se hace una recapitulación de los aspectos fundamentales que se involucran en la detección de fraude de transacciones digitales, en cómo se ha abordado a lo largo de los años y qué técnicas son las que resultan más

útiles en la actualidad para dar solución al problema. El avance en el área de *graph neural networks* representa una gran oportunidad para el desarrollo de métodos que consideren aspectos más fundamentales sobre la detección de fraude. El desarrollo de este tipo de trabajos abre la puerta para el continuo mejoramiento de este tipo de procesos de detección y, a su vez, para la aportación al conocimiento del área de *graph neural networks*.

CONCLUSIONES

Este artículo aborda aspectos generales de la detección de fraude mediante técnicas de aprendizaje automático (*machine learning*) habituales y aquellas que hacen uso de información de grafos. Esta perspectiva ofrece algunas ventajas que versan sobre consideraciones intrínsecas del problema y que ayudan a mejorar la detección. Aspectos fundamentales como la estructura del grafo ayudan a determinar más propiedades que permiten diferenciar entre fraude y un comportamiento normal en las transacciones digitales.

En el mundo actual los diferentes procesos digitales son vulnerables a las operaciones fraudulentas. Diversas plataformas en línea que operan sobre diferentes propósitos son objeto de este tipo de fenómeno que termina por impactar a usuarios y empresas. Como se ha visto a lo largo de este artículo, para detectar este comportamiento se han desarrollado diferentes métodos, algunos de los más novedosos están relacionados con algoritmos de aprendizaje automático. Históricamente, este tipo de enfoques a veces se ha combinado con el uso de reglas para conseguir un proceso de detección más cerrado. A pesar de esto, la información recolectada por plataformas en línea tiene diferentes tipos de representaciones, entre ellas destaca la construcción de grafos (redes) que revelan información implícita dentro del conjunto de datos.

A partir de una combinación de enfoques basados en el aprendizaje automático y en el uso de grafos, se han desarrollado procesos que permiten utilizar este tipo de información con la finalidad de mejorar los procesos de detección de fraude en las plataformas en línea. Sin embargo, este tipo de implementaciones requieren consideraciones especiales debido a la gran cantidad y a la calidad de los datos que se requiere para su entrenamiento (desarrollo). A pesar de ello, estos enfoques representan una de las mejores opciones para la detección de fraude digital debido a la información implícita que contiene el paradigma basado en grafos.

Finalmente, la cantidad y calidad de la información es de suma importancia, ya que este tipo de problema depende mucho del contexto de un comportamiento real en diferentes plataformas. Por esta razón, los trabajos actuales que se enfocan en detección de fraude mediante el uso de grafos utilizan datos provenientes de instancias de pago reales.

AGRADECIMIENTOS

Agradecemos a los árbitros por las mejoras en la estructura del artículo, así como también a Moneypool por proporcionar información útil para desarrollarlo.

REFERENCIAS

- BBC. (2024). *Santander staff and '30 million' customers hacked*. <https://www.bbc.com/news/articles/c6ppv06e-3n8o>
- Dou, Y., Liu, Z., Sun, L., Deng, Y., Peng, H., & Yu, P. S. (2020). *Enhancing graph neural network-based fraud detectors against camouflaged fraudsters*. Arxiv. <https://doi.org/10.1145/3340531.3411903>
- FATF-Interpol-Egmont Group (2023). *Illicit Financial Flows from Cyber-Enabled Fraud*. FATF. Paris. <https://www.fatf-gafi.org/content/dam/fatf-gafi/reports/Illicit-financial-flows-cyber-enabled-fraud.pdf.core-download.inline.pdf>

- Fu, K., Cheng, D., Tu, Y., & Zhang, L. (2016). Credit card fraud detection using convolutional neural networks. *Neural Information Processing*, 9949, 483-490. Cham: Springer. https://doi.org/10.1007/978-3-319-46675-0_53
- Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. In *Proceedings of the 31st international conference on neural information processing systems* (pp. 1025-1035). Long Beach: Curran Associates Inc.
- Jendruszak, B. (2024, 26 de marzo). *Credit Card Fraud Detection: The Complete Guide*. Seon. <https://seon.io/resources/credit-card-fraud-detection>
- J. P. Morgan Chase (2021). *A merchant's guide to preventing card testing attacks*. <https://www.jpmorgan.com/content/dam/jpm/securities/documents/card-testing-white-paper-ada-final.pdf>
- Lebichot, B., Braun, F., Caelen, O., & Saelens, M. (2017). A graph-based, semi-supervised, credit card fraud detection system. In H. Cherifi, S. Gaito, W. Quattrociocchi, A. Sala. (eds.), *Complex Networks & Their Applications V. COMPLEX NETWORKS 2016 2016. Studies in Computational Intelligence, vol 693*. Cham: Springer. https://doi.org/10.1007/978-3-319-50901-3_57
- Liu, Z., Chen, C., & Yang, X., & Zhou, J., & Li, X., & Song, L. (2020). Heterogeneous graph neural networks for malicious account detection. *ArXiv*. <https://doi.org/10.48550/arXiv.2002.12307>
- Maes, S., Tuyls, K., Vanschoenwinkel, B., & Manderick, B. (2002). Credit card fraud detection using bayesian and neural networks. *Proceedings of the First International NAISO Congress on NEURO FUZZY TECHNOLOGIES January 16-19, 2002 (Havana, Cuba)*. <http://como.vub.ac.be/Members/bram.htm>
- Malone, C. (2023). *Online payment fraud: Market forecasts, emerging threats and segment analysis 2023-2028*. <https://www.juniperresearch.com/researchstore/fintechpayments/online%5C%5C-payment-fraud-research-report>
- Meng, C., Zhou, L., & Liu, B. (2020). A case study in credit fraud detection with smote and xgboost. *Journal of Physics: Conference Series*, 1601(5), 052016. <https://doi.org/10.1088/1742-6596/1601/5/052016>
- Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., Monti, F., & Bronstein, M. (2020). Temporal graph networks for deep learning on dynamic graphs, *Arxiv*. <https://doi.org/10.48550/arXiv.2006.10637>
- Rao, S. X., Lanfranchi, C., Zhang, S., Han, Z., Zhang, Z., Min, W., & Zhang, C. (2022). Modelling graph dynamics in fraud detection with "attention". *ArXiv*. <https://doi.org/10.48550/arXiv.2006.10637>
- Ravelin. (2024). *Ravelin Insights. Machine learning for fraud detection*. <https://www.ravelin.com/insights/machine-learning-for-fraud-detection/>
- Saporta, G., & Maraney, S. (2022). *Fraud prevention evaluation and investment in practical fraud prevention*. O'Reilly Media.
- Saldaña-Ulloa, D., De Ita Luna, G., & Marcial-Romero, J. R. (2024). A temporal graph network algorithm for detecting fraudulent transactions on online payment platforms. *Algorithms*, 17(12), 552. <https://doi.org/10.3390/a17120552>
- Seon. (2024). Bank Account Takeover Attacks: How to Keep Safe. <https://seon.io/resources/bank-account-takeover-attacks/>
- Stripe. (2024). *Fraudulent transactions 101: What causes them and how businesses can act*. <https://stripe.com/in/resources/more/fraudulent-transactions-101>
- Sulaiman, R., Schetin, V., & Sant, P. (2022). Review of machine learning approach on credit card fraud detection. *Human-Centric Intelligent Systems*, 2(1-2), 55-68.

- Xiao, W., Houye J., Chuan S., Bai W., Yanfang Y., Peng C., & Philip, S. Y. (2019). *Heterogeneous graph attention network*. New York: Association for Computing Machinery. <https://doi.org/10.1145/3308558.3313562>
- Yue, D., Wu, X., Wang, Y., Li, Y., & Chu, C.-H. (2007). A review of data mining-based financial fraud detection research. 2007 International Conference on Wireless Communications, Networking and Mobile Computing (pp. 5519-5522). <https://doi.org/10.1109/WICOM.2007.1352>
- Zhao, P., Fu, X., Wu, W., Li, D., & Li, J. (2019). Network-based feature extraction method for fraud detection via label propagation. *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)* (pp. 1-6). 10.1109/BIGCOMP.2019.8679414

CC BY NC ND



Disponible en: <https://cienciaergosum.uaemex.mx/article/view/23217>

Cómo citar el artículo:

Saldaña Ulloa, D. y De Ita Luna, G. (2026). Detección de fraude digital mediante el uso de grafos. *CIENCIA ergo-sum*, 33. <https://doi.org/10.30878/ces.v33n0a49>

DOI: <https://doi.org/10.30878/ces.v33n0a49>



Licencia Creative Commons Atribución-NoComercial-SinDerivar 4.0 Internacional.

Declaración de autoría (CRediT):

Diego Saldaña Ulloa: conceptualización, análisis formal, investigación, metodología, administración del proyecto, visualización, redacción-borrador original, redacción-revisión y edición
Guillermo De Ita Luna: Investigación, Administración del proyecto, Supervisión, validación, Redacción-revisión y edición