



5

FUNDAMENTOS DE  
LA CIENCIA DE  
DATOS GRAFOS

QUE TODOS  
DEBERÍAN SABER

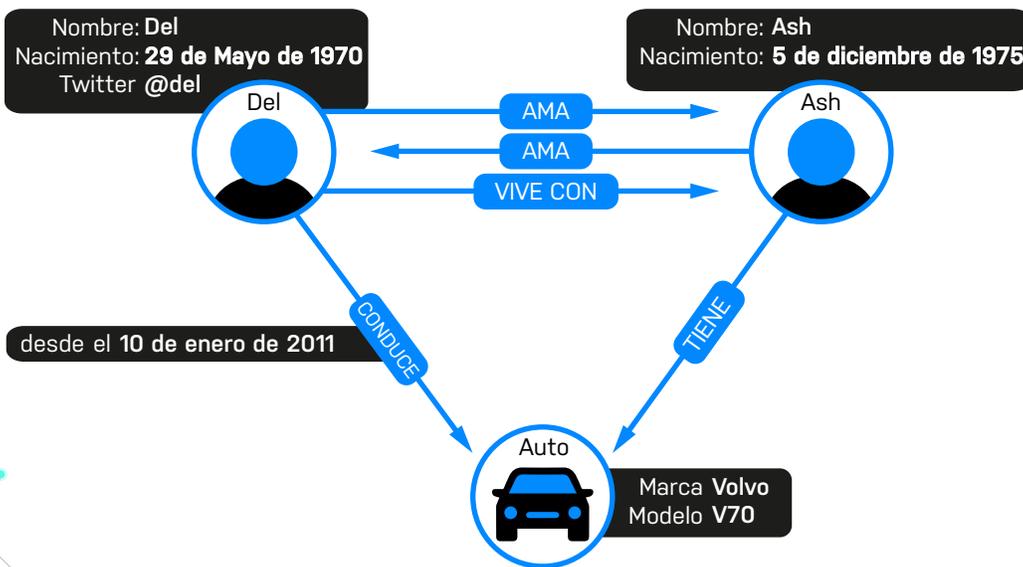
# ¿Por qué la ciencia de datos grafos y por qué ahora?

Las aplicaciones comerciales de la ciencia de datos grafos son nuevas y la mayoría de los expertos en datos todavía se están poniendo al día sobre cómo utilizarlos mejor en sus organizaciones. Mientras que algunos expertos en datos estudiaron la teoría de grafos, otros no han sabido mucho de ella. Esta ciencia de datos reúne análisis de grafos, estadísticas y técnicas de IA y ML para mejorar sus modelos predictivos y prescriptivos. Este documento es una guía a través de los fundamentos de la ciencia de datos grafos para que sientas la seguridad de saber cuándo usarlo en el trabajo diario.

## 1. ¿Qué es un Grafo?

Antes de que puedas entender la ciencia de datos grafos, necesitas entender los grafos. En su forma más fundamental, un grafo es simplemente una forma diferente de estructurar los datos. En lugar de filas y columnas, como en una tabla de base de datos relacional tradicional, los grafos usan nodos (sustantivos) y relaciones (verbos) como su estructura primaria.

En un grafo, los sustantivos -personas, lugares, cosas, organizaciones- son nodos, y las relaciones entre ellos son verbos: es amigo de, trabaja para, le gusta, etc.



Los grafos representan datos a través de relaciones.



Los nodos representan una entidad en el grafo.



Las relaciones (bordes, enlaces) conectan los nodos entre sí.



Las propiedades describen un nodo o una relación, nombre, edad, altura.

## 2. ¿Cómo difiere la ciencia de datos grafos del análisis tradicional?

Cuando analizas los datos en forma tabular, como en una base de datos relacional, intentas dar sentido a los puntos de datos sin una forma coherente de analizar sus conexiones. Ben Squire, un científico de datos, comparó los métodos tradicionales con "tratar de resolver un cubo de Rubik mirando solo un lado." Un grafo proporciona a los expertos en datos la capacidad de observar, comprender y analizar las conexiones entre cada punto de datos. Esto da un contexto que es imposible obtener de un modelo de datos tabular, ya que puedes ver cuán fuertes son las conexiones, dónde se forman grupos de conexiones, cuán importante es cada conexión y cómo las conexiones se influyen entre sí.

La ciencia de datos grafos reúne estadísticas de grafos, análisis y ML para poner los datos en contexto y responder preguntas urgentes.

**Las estadísticas de grafos, las consultas y la visualización** impulsan la exploración y los conocimientos. Estas estadísticas proporcionan medidas básicas sobre los grafos, como el número de nodos y la distribución de las relaciones. Las consultas del grafo responden a cualquier pregunta, no importa cuán profunda sea, ya sean 6 o 600 grados de separación. La visualización de los mismos permite a los expertos en datos ver sus datos y explorar.

El **análisis de grafos** se basa en las estadísticas al responder preguntas específicas y obtener información de las conexiones en datos históricos o existentes. Las consultas de grafos y los algoritmos se aplican típicamente juntos en "recetas" durante su análisis, y los resultados se utilizan directamente para éste.

**Grafo mejorado ML** es la aplicación de datos grafos y resultados de análisis para entrenar modelos de ML o apoyar decisiones probabilísticas dentro de un sistema de IA. Las estadísticas y análisis de grafos a menudo se utilizan en conjunto para responder a ciertos tipos de preguntas sobre sistemas complejos y los conocimientos posteriores, aplicados para mejorar el ML.

# Vista rápida: Técnicas de ML en grafo que hacen el trabajo por tí

El interés en el aprendizaje automático basado en grafos está explotando. Para 2025, Gartner proyecta que la tecnología de grafos impulsará el 80% de las innovaciones de datos y análisis, frente al 10% en 2021. Aquí hay algunas técnicas avanzadas de ciencia de datos grafos que tu equipo puede explorar.

La **predicción de enlaces (Link prediction)** llena los espacios en blanco en tus datos y predice cambios en la estructura de los grafos. Es una tarea común de aprendizaje automático aplicada a los grafos: entrenar un modelo para aprender dónde deben existir relaciones entre pares de nodos en un grafo. Puedes pensar en la predicción de enlaces como la construcción de un modelo para predecir las relaciones que faltan en su conjunto de datos o las relaciones que es probable que se formen en el futuro. Con la ciencia de datos grafos, puedes entrenar modelos de ML supervisados basados en las relaciones y propiedades de nodos en tu grafo para predecir la existencia -y la probabilidad- de las relaciones.

La **incrustación de nodos (Node embedding)** transforma la topología y las características de tu grafo en una representación vectorial de baja dimensión de cada nodo. Estos vectores, también llamados incrustaciones, se pueden utilizar para el análisis de datos exploratorios, las mediciones de similitud y el ML. Las incrustaciones de nodos pueden agregar información sobre la posición de un nodo en el grafo, sus vecinos locales, su centralidad e influencia, y en algunos casos, otras propiedades de nodos numéricos.

Los **modelos de clasificación de nodos (node classification)** de nodos predicen la clase de nodos en el grafo. Una clase puede ser un indicador binario, como por ejemplo si una cuenta de usuario está involucrada en fraude o un indicador multivalor, y también a qué segmento de mercado pertenece un cliente. Estos modelos pueden entrenarse para predecir a qué nodos de clase (incluyendo cualquier nodo nuevo) pertenecen. La clasificación de nodos puede incorporar una amplia gama de características de entrada, incluyendo la estructura de red del grafo y de sus datos de origen.



### 3. ¿Por qué importan los grafos?

En toda una organización, cada departamento puede beneficiarse de la ciencia de datos grafos para responder preguntas como quién o qué es importante, qué debo hacer a continuación y qué es inusual sobre esto.

Si bien los casos de uso para la ciencia de datos grafos abarcan industrias y líneas de negocio -desde las ciencias de la vida hasta la fabricación- algunos de ellos se están convirtiendo rápidamente en los más populares entre los científicos de datos.



#### Detección de anomalías y fraudes

La detección de anomalías en las redes corporativas puede ayudar a identificar los ataques de ciberseguridad y prevenir la pérdida de datos. La misma estrategia utilizada para identificar a los actores de amenazas en un contexto de ciberseguridad puede ser utilizada para detectar el fraude en la banca, seguros, y programas del gobierno, analizando las relaciones y los comportamientos en su grafo.

*Aquellos que utilizan la ciencia de datos grafos para frenar el fraude han visto mejoras de detección de más del 300%, ahorrando millones cada año.*

## Cientes 360

En todo el mundo, las empresas tratan de entender mejor a sus clientes y mejorar el valor de vida (LTV). Con la ciencia de datos grafos, el conocimiento del cliente puede ser más preciso y completo a través de la resolución de entidades. Este proceso examina todas las entradas de la base de datos e identifica los duplicados. Crear una entrada de base de datos completa y maestra para cada cliente en lugar de tener múltiples entradas incompletas, mejora LTV y profundiza el conocimiento del cliente, lo que permite optimizar programas y ofertas de marketing.

*Aquellos que utilizan la ciencia de datos grafos para el cliente 360 pueden aumentar el conocimiento del cliente en un 30%.*

## Motores de recomendación

Los motores de recomendación se hicieron conocidos a través de Netflix y de experiencias de compra en línea. Sin embargo, tienen usos en todos los sectores, desde el desarrollo de productos, hasta los recursos humanos para retener a los empleados a través de la formación de perfeccionamiento; e impulsan algunas de las partes más importantes de un negocio.

## Gestión de la cadena de suministro

Mejorar una cadena de suministro conlleva ahorros, no solo en dólares, sino también en emisiones de carbono. Cada ruta optimizada, la sincronización perfecta y la entrega perfecta significan clientes más felices y menos desperdicio del tiempo, la infraestructura y las emisiones. La ciencia de datos grafos ayuda a optimizar las rutas de la cadena de suministro al encontrar el mejor camino, equilibrando el costo y la eficiencia con la satisfacción del cliente y la sostenibilidad.

*Un ROI de decenas de millones para una organización que utiliza la ciencia de datos grafos para la optimización de rutas, no es nada comparado con las 60.000 toneladas de emisiones de carbono eliminadas mediante el uso de esas rutas optimizadas.*

## 4. ¿Cuáles son las grandes preguntas que la ciencia de datos grafos ayuda a responder?

La ciencia de datos grafos ayuda a responder preguntas importantes para tomar decisiones empresariales críticas. Si escuchas preguntas como éstas, puedes beneficiarte de usar la ciencia de datos grafos para responderlas:

### ¿Qué es lo inusual?

Para entender las anomalías y los patrones ocultos en los grafos, considera usar la **detección comunitaria**. Las comunidades son grupos dentro del grafo, y [los algoritmos de detección de la comunidad](#) se pueden utilizar para descubrir e identificar estos grupos. Detectar comunidades te ayuda a descubrir patrones inusuales, predecir comportamientos similares, encontrar entidades duplicadas o simplemente preparar datos para otros análisis.

### ¿Qué te puedo recomendar?

Para construir un motor de recomendación, considera usar similitudes. La **similitud** identifica patrones repetitivos en el grafo. Los [algoritmos de similitud](#) emplean comparaciones de conjunto y distancia para calificar cómo los nodos individuales se basan en sus vecinos o propiedades. Este enfoque se utiliza en aplicaciones como las recomendaciones personalizadas y el desarrollo de jerarquías categóricas.

### ¿Qué contenido es el más importante? ¿Quién es el más influyente? ¿Dónde está nuestro mayor riesgo?

Para entender lo que es importante, usa algoritmos de **centralidad**. Las métricas de centralidad como [PageRank](#) te ayudan a identificarlo y estos algoritmos revelan qué nodos son valiosos en base a la topología de grafos. Identifican, a su vez, nodos influyentes en función de su posición en la red más grande, incluidas sus conexiones. Estos algoritmos se utilizan para inferir dinámicas de grupo tales como credibilidad, vulnerabilidad en cascada, y puentes entre grupos.

### ¿Cuál es la ruta óptima?

Para entender la optimización de rutas, utiliza el **pathfinding**. Los [algoritmos de pathfinding](#) encuentran las mejores rutas a través de sus datos conectados. Estos algoritmos son fundamentales para el análisis de grafos y encuentran las rutas más eficientes o más cortas para atravesar entre nodos. Se pueden usar para comprender dependencias complejas y evaluar rutas para usos como la logística física y llamadas de menor costo o enrutamiento IP.

## 5. ¿Quién utiliza herramientas de ciencia de datos grafos?

Cualquiera que trabaje con datos puede aprovechar la ciencia de los datos grafos para encontrar respuestas a preguntas difíciles. Aquí están algunos de los roles que actualmente utilizan la ciencia de datos grafos.



Los **científicos de datos** suelen ser los principales usuarios de las herramientas de ciencia de datos grafos porque son profesionales de la ciencia de datos con un profundo conocimiento de algoritmos y modelos.



Los **ingenieros de Machine Learning** trabajan para escalar, mejorar, integrar u operar modelos de aprendizaje automático desarrollados por científicos de datos.



Los **científicos de datos ciudadanos** aplican técnicas de ciencia de datos grafos sin necesidad de análisis profundo y experiencia en ML, utilizando una experiencia de código bajo/ sin código.



Los **analistas de datos empresariales** visualizan, desarrollan e informan sobre el análisis realizado por un científico de datos para usuarios empresariales.



Los **analistas de datos comerciales** visualizan, desarrollan e informan sobre el análisis realizado por un científico de datos para usuarios comerciales.



Los **desarrolladores de aplicaciones y arquitectos de software** que están aprendiendo sobre bases de datos grafos, a menudo usan consultas de grafos, típicamente usando [lenguaje de consultas Cypher](#). Aprenden algoritmos de grafos mientras tratan de analizar patrones a través de sus grafos y usan esas técnicas en sus aplicaciones (la ciencia de datos grafos es a menudo la salsa secreta en las aplicaciones de diferenciación).

## Mantente a la vanguardia

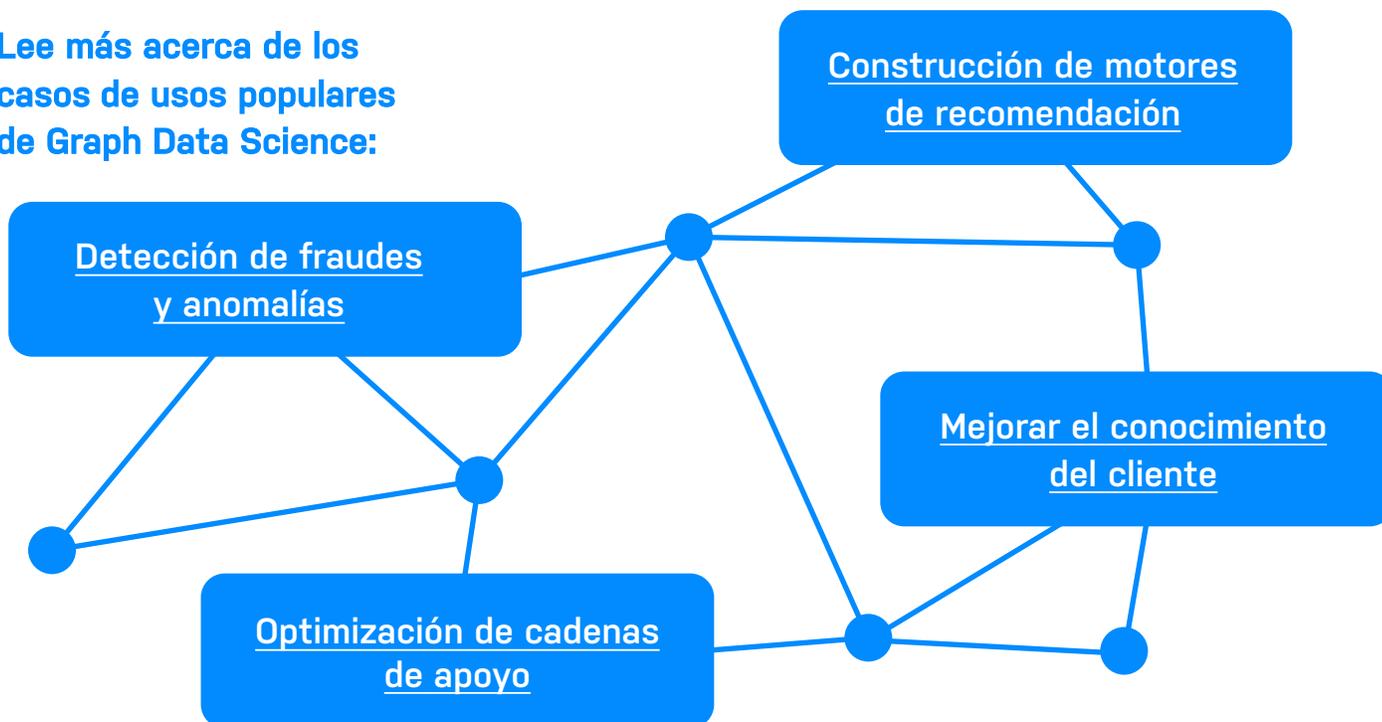
Las aplicaciones comerciales de la ciencia de datos grafos son nuevas y los expertos en datos todavía están tratando de comprender cómo utilizar mejor la ciencia de datos grafos en sus organizaciones. Al considerar casos de uso, los expertos en datos y los científicos de datos deben tener en cuenta que la ciencia de datos grafos ayuda a responder preguntas esenciales como qué es importante, qué es inusual y qué es lo siguiente. Usando este encuadre, es más fácil identificar oportunidades para utilizar la ciencia de datos grafos para mejorar los modelos y hacer predicciones.

Organizaciones de todos los tamaños, todas las industrias y dentro de todos los departamentos están utilizando la ciencia de datos grafos para hacer recomendaciones, identificar anomalías y encontrar estafadores, mejorar el conocimiento del cliente y optimizar las cadenas de suministro.

### Entonces, ¿qué harás con la ciencia de datos grafos?

¿Listo para probarla? Accede [al sandbox gratuito Neo4j Graph Data Science](#) con datos pre poblados para escenarios comunes.

### Lee más acerca de los casos de usos populares de Graph Data Science:



Neo4j es la plataforma de datos grafos líder en el mundo. Ayudamos a las organizaciones, incluidas **Comcast**, **ICIJ**, **NASA**, **UBS**, y **Volvo Cars** – a capturar el rico contexto del mundo real que existe en sus datos para resolver desafíos de cualquier tamaño y escala. Nuestros clientes transforman sus industrias al frenar el fraude financiero y el ciberdelito, optimizar las redes globales, acelerar la investigación innovadora y brindar mejores recomendaciones. Neo4 ofrece en tiempo real procesamiento de transacciones, AI/ML avanzado, visualización intuitiva de datos, y más. Encuentrenos en [neo4j.com](#) y síguenos en [@Neo4j](#).

Preguntas acerca de Neo4j? Contactenos alrededor del mundo:

[info@neo4j.com](mailto:info@neo4j.com)  
[neo4j.com/contact-us](https://neo4j.com/contact-us)