



EBOOK

# Graph—what your models are missing

Learn how Neo4j Graph Data Science and Amazon SageMaker take machine learning to new heights

Jaime D'Anna - Neo4j

Antony Thevaraj - AWS



# Table of contents

Tap into the power of the graph .....	3
Neo4j Graph Data Science and Amazon SageMaker take ML to new heights.....	4
Graph data science sharpens predictions from ML modeling .....	5
Graph is already shaping real-world use cases .....	6
Answer big questions that power critical decisions .....	8
What will you do with graph data science? .....	9

# Tap into the power of the graph

Across today's business landscape, the field of data science is making it possible for companies to tap into massive datasets and derive value. From marketing to healthcare to finance, data science powers faster, better decision-making. Whether you're genuinely interested in getting insights and solving problems using data or just attracted by what has been called the most promising career, chances are you're familiar with this fast-growing field.



## But what about graph data science?

In a nutshell, graph data science not only helps you answer questions with your data, but also the connection between your data points. Graph data science combines graph analytics and statistics with artificial intelligence (AI) and machine learning (ML) techniques to improve predictive and prescriptive models.

The advantage of graph data science is that it can leverage relationships in data. Most data scientists work with data in tabular formats. However, to get better insights—to answer questions you can't answer without leveraging connections—graphing your data is key.

Neo4j is an Amazon Web Services (AWS) Advanced Technology partner with a competency in Data and Analytics and seamless integration with AWS services such as Redshift, EMR, EKS, Lambda, SageMaker, and more. Keep reading this ebook to find out how Neo4j and AWS are making it easier for companies to tap into the graph. You'll learn how to improve predictions from ML modeling, as well as real-world applications for graph data science.

## Top benefits of graph data science

### Improve predictions

Incorporate the predictive power of relationships in advanced analytics to continually improve predictive accuracy. Graph data science leverages the connections between data points for more accurate and interpretable predictions.

### Answer questions

Enhance understanding of complex systems and answer more complicated questions with graph algorithms designed to analyze network structures.

## With Neo4j plus Amazon SageMaker you can:

- Combine the latest graph data science insights with your ML pipelines
- Answer business critical questions and make predictions
- Accelerate your path from proof of concept to production
- Understand relationships and improve your models

# Neo4j Graph Data Science and Amazon SageMaker take ML to new heights

Leading data scientists and developers will tell you that adding graph analytics to their ML pipelines allows them to quickly gain insights from their data. But there's a challenge: graph datasets are nearly impossible to use in traditional ML tasks.

That's why AWS and Neo4j offer scalable, intelligent tools for making the most of graph data. Neo4j Graph Data Science™ and Amazon SageMaker make building ML models on top of graph data fast and easy.



The **Neo4j Graph Data Platform** is fully integrated with AWS, giving data scientists and developers tools and a managed framework that includes an analytics engine, a robust data science framework, and an insights layer for exploration.



**Neo4j Graph Data Science** gives data scientists a unified workspace to improve models and predictions through the signals and patterns in their data and incorporate them into their Amazon SageMaker workflows.



**Neo4j AuraDB for AWS** is a graph database as a service, enabling you to store, query, analyze, and manage highly connected data.



**Amazon SageMaker** is a fully managed service to prepare data and build, train, and deploy ML models for any use case with fully managed infrastructure, tools, and workflows.

Neo4j AuraDB and Neo4j Graph Data Science integrate seamlessly with Amazon SageMaker so you can quickly and effortlessly prepare, train, and deploy high-quality ML models using agile graph dynamics and build future-proof, scalable ML solutions.

# Graph data science sharpens predictions from ML modeling

## Interest in graph-based ML is exploding

By 2025, Gartner predicts that graph technology will fuel 80 percent of data and analytics innovations, up from 10 percent in 2021.<sup>1</sup> It's gaining popularity because bringing graph data into ML pipelines greatly improves the ML model's ability to rapidly learn predictive features. Most models today aren't exploring the relationship between datapoints or network structures. This leaves insights on the table.

Being able to sharpen predictions also improves the practice of data science. Because data science is inherently iterative, it's essential to use a framework that brings in highly predictive relationships. With better predictions comes a more streamlined process of moving from data to analysis to visualization and back.

Let's take a closer look at three advanced graph data science techniques that improve ML modeling.

**Link prediction** fills in the blanks in your data and predicts missing relationships or relationships that are likely to form in the future. With graph data science, you can train supervised ML models based on the relationships and node properties in your graph to predict the existence—and probability—of relationships.

**Node embedding** transforms the topology and features of your graph into a low-dimensional vector representation of each node. These vectors, also called embeddings, can be used for exploratory data analysis, similarity measurements, and ML. Node embeddings can aggregate information about a node's position in the graph, its local neighbors, its centrality and influence, and, in some cases, other numeric node properties.

**Node classification** models predict the class of nodes in your graph. Node classification can incorporate a broad range of input features, including the network structure of your graph and properties from your source data.

# 80%

of data and analytics innovations will be fueled by graph technology

# Graph is already shaping real-world use cases

The ability to improve predictions and answer tough questions is a fundamental business need, so it's no surprise that graph data science is in demand across industries. From healthcare to financial services, the applications of graph are far reaching.

## Here are a few use cases that are rapidly becoming popular among data scientists:



### Anomaly and fraud detection

Anomaly detection across corporate networks can help identify cybersecurity attacks and prevent data loss. The same strategy used to identify threat actors in a cybersecurity context can be used to detect fraud in banking, insurance, and government programs by analyzing the relationships and behaviors in your graph.



### Customer 360

Across the globe, businesses try to better understand their customers and improve lifetime value. With graph data science, customer knowledge can become more accurate and complete through entity resolution.



### Recommendation engines

The ability to recommend products or services to users based on their preferences and previous activity is an essential component of any digital customer experience strategy. Recommendation engines powered by graph data science drive some of the most personalized experiences for shoppers.





### Supply chain optimization

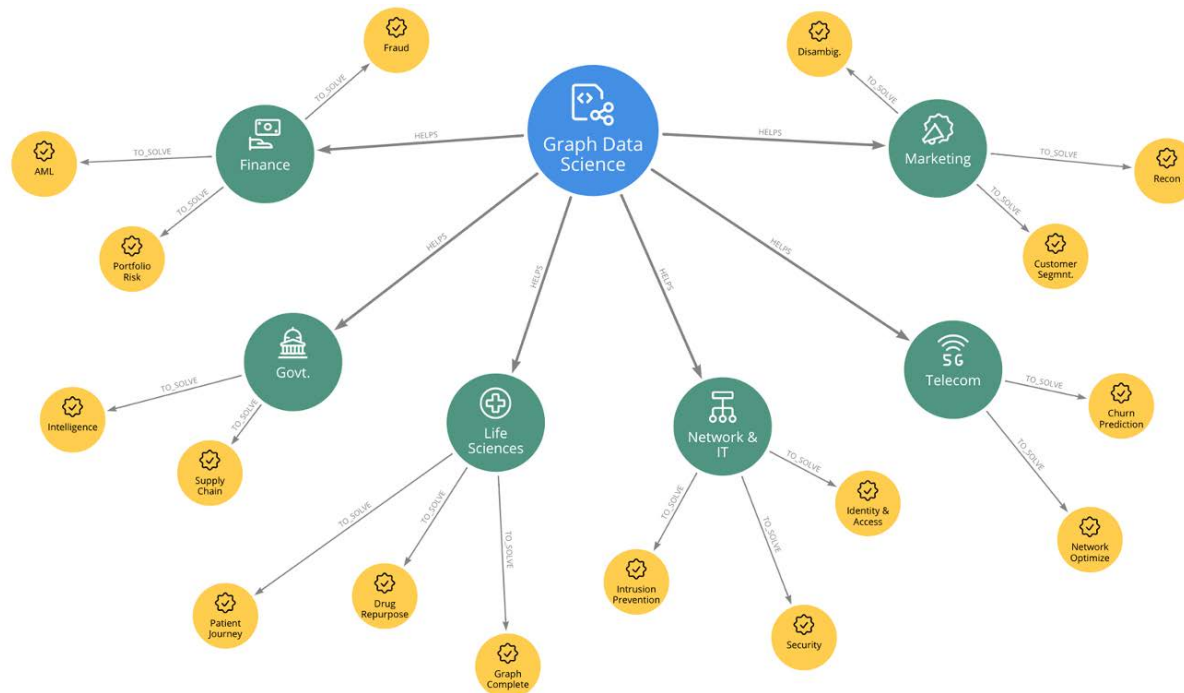
Improving a supply chain leads to savings, not just in dollars, but also in carbon emissions. Graph data science helps optimize supply chain routes by finding the best path, then balancing cost and efficiency with customer satisfaction and sustainability.



### Drug discovery

Finding new drug candidates involves interrogating vast amounts of data and typically takes years. Graph technology enables drug companies to create a system of scalable biological knowledge that consolidates disparate datasets.

This is just a subset of the many use cases that illustrate the effectiveness of deploying graph data science to drive successful business outcomes. You can find more use cases in this [Graph Data Science Selection Guide](#).





## Answer big questions that power critical decisions

Beyond industry-specific use cases, graph data science helps answer big questions that are common starting points for critical business decisions. If your team is asking these questions, it's probably time to use graph data science to answer them.

### What's unusual?

To understand anomalies and hidden patterns, consider using community detection. Communities are clusters within your graph, and community detection algorithms can be used to discover and identify these clusters. Detecting communities helps you uncover unusual patterns, predict similar behavior, find duplicate entities, or simply prepare data for other analyses.

### What's most important?

To understand what's important, who is the most influential, or where is your biggest risk, use centrality algorithms. Centrality algorithms reveal which nodes are important based on their position in the larger network, including their connections. These algorithms are used to infer group dynamics such as credibility, cascading vulnerability, and bridges between groups.

### What should I recommend?

To build a recommendation engine, consider using similarity, which identifies repeating patterns in your graph. Similarity algorithms employ set and distance comparisons to score how alike individual nodes are based on their neighbors or properties. This approach is used in applications such as personalized recommendations and developing categorical hierarchies.

### What is the optimal route?

To understand route optimization, use pathfinding algorithms that find the best routes across your connected data. Pathfinding algorithms are foundational to graph analytics and find the most efficient or shortest paths to traverse between nodes. They can be used to understand complex dependencies and evaluate routes for uses such as physical logistics and least-cost call or IP routing.



# What will you do with graph data science?

Whether it's use cases or big questions, graph data science brings tremendous value to advanced analytics and ML. Its unique ability to sharpen predictions has made graph data science a popular toolset that will underpin modern business analytics as the field continues to evolve.

See what answers your graph holds.

[Visit the Neo4j page on AWS Marketplace to learn more and get started ›](#)

## About Neo4j

Neo4j is a recognized leader in graph database technology across the analyst community, including the Q4 2020 Forrester Wave for Graph Data Platforms. Plus, Neo4j has a community of over 240,000 developers. Fortune 500 organizations trust Neo4j as their graph database vendor, because it has been tested and proven against a broad set of use cases in production and can scale to billions of nodes and relationships on a single machine and extends scalability to trillions of relationships, making it even easier to support AI-powered applications and increased predictability through ML models on AWS.

In collaboration with



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

