IDC InfoBrief
November 2022

The Power of Graph Data Science

Make Better Predictions with Your Existing Data

IDC Doc. #AP241320IB

IDC

Sponsored by neo4j

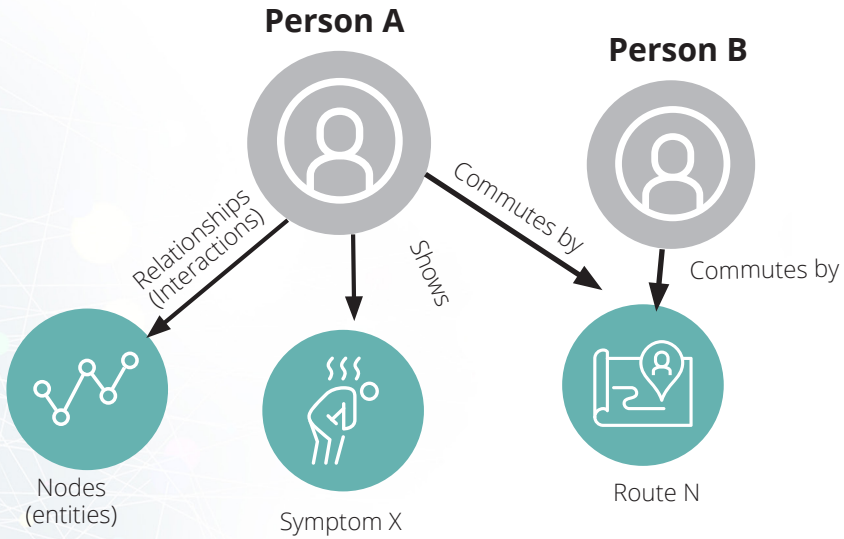
What Is Graph Data?

Graph data science is a scientific approach to gain knowledge from graphs, using the structures of connected data to power predictions, answer questions, and explain outcomes.

What are graphs?

A graph is simply a way of structuring data. Compared with tables in a traditional relational database, which contains rows and columns, graphs contain nodes and relationships as their primary structure, as illustrated here.

For data scientists



What they are *not*

The term "graph", when used in the context of graph data science, or graph analysis, does not refer to data charts, such as line graphs or bar graphs.

The Three Pillars of Graph Data Science

Graph Statistics

Provides basic measures about a graph as a whole, such as the number of nodes and distribution of relationships.

These insights may influence how you configure and execute more complex analysis as well as interpret results.

Graph Analytics

Builds on graph statistics by answering specific questions and gaining insights from connections in existing data.

Graph queries and algorithms are typically applied here as "recipes" during graph analytics.

Graph Learning

Apply graph data and analytics results to train machine learning (ML) models.

In addition, it can support probabilistic decisions within an artificial intelligence (AI) system.

Various Types of Graphs

Directed vs Undirected Graphs

In an undirected graph, connections are reciprocal. In a directed graph, relationships have direction, and may only be unidirectional

Homogeneous vs Heterogeneous Graphs

In a homogenous graph, there is only one type of node and one type of relationship. Heterogeneous graphs have multiple types of nodes and relationships

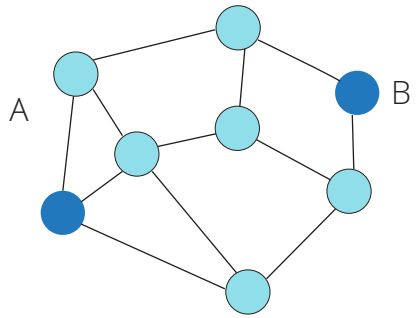
Static vs Dynamic Graphs

A static graph is unchanging. A dynamic graph is continually updated and the structure changes with time.

Why Graph Data Science Matters: The Big Questions It Helps to Answer

PATH FINDING & ANALYSIS

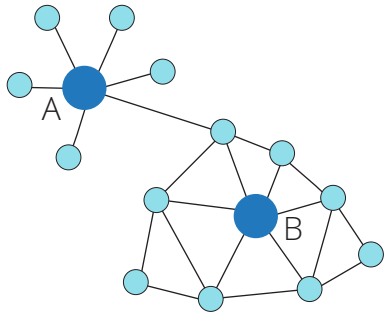
Estimate the relationships between nodes, often used to find the shortest path (connected edges) in logistics and transportations.



What's the optimal route?

CENTRALITY FINDING & ANALYSIS

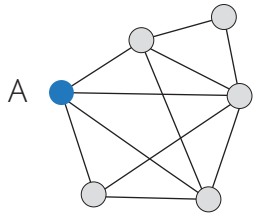
Determine how important a node in the network is. There are different types of centrality – degree centrality, betweenness centrality, etc.



Who is the most influential?

COMMUNITY FINDING & ANALYSIS

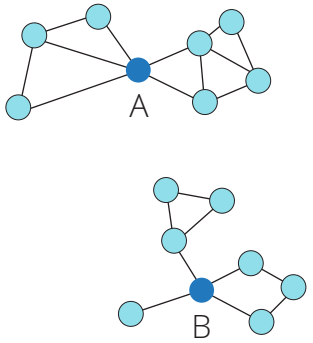
Find groups of people interacting frequently with each other. Evaluate the tendency for such interactions to strengthen or break apart.



What's unusual?

SIMILARITY FINDING & ANALYSIS

Understand the similarity of nodes based on their neighborhoods or their properties.



What should I recommend?

Diving Deeper: Graph Algorithms and Their Uses

Types of Algorithms	Range of Problems: Fraud Detection Banking, insurance, investment services	Recommendations eCommerce, social networks, media & entertainment	Supply Chain Optimization Logistics, transportation, manufacturing	Entity Resolution Customer 360
Path finding			✓	
Centrality	✓	✓		
Topological link	✓		✓	✓
Community	✓	✓	✓	
Similarity scoring	✓	✓		✓
Embeddings	✓	✓		✓

By 2024, 70% of A2000* will expand the use of AI/ML across all business-critical horizontal functions like marketing, legal, HR, procurement, and supply chain logistics.

* A2000 refers to the top 2000 companies in Asia.

Source: IDC FutureScape: Worldwide Artificial Intelligence and Automation 2022 Predictions, APEJ Implications

Why Graph Data Science Matters: Versatility for Data Scientists

Versatility in data types processed

Graph learning works on both structured and unstructured data.

- Structured data: sales transactions, molecular structures, machine logs, and social networks are naturally graphy.
- Unstructured data: images, voices, texts, etc can be represented as nodes, connected to structured data.

Versatility in learning methods applied

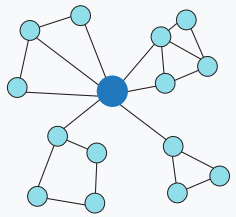
Both supervised and unsupervised techniques can be applied to graph data to address a wide range of questions.

- Unsupervised learning on a graph uses relationships to find important nodes, identify close knit communities, or find outliers and anomalies - all based on connectivity.
- Supervised learning on a graph uses the graph data itself as labels, and can fill in missing data when the graph is only partially labelled.

The Building Blocks of Graph Learning

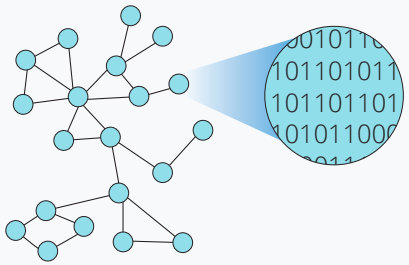
Knowledge Graph

Find patterns in graph data



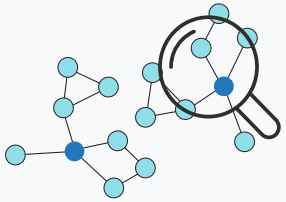
Graph Algorithms

Apply unsupervised learning methods



Graph Native Machine Learning

Learn features and make predictions



Diving Deeper: The Predictive Power of In-graph Machine Learning

Link prediction
Finds where relationships will likely form and predicts changes.

Node Classification
Categorizes the classes of graph nodes and predicts where they belong.

Node Embedding
Translates graphs into low-dimensional vectors, which can be used for exploratory and predictive analyses.

Supervised Learning

Suggests
future connections

Find labels/tags for graph nodes

Unsupervised Learning

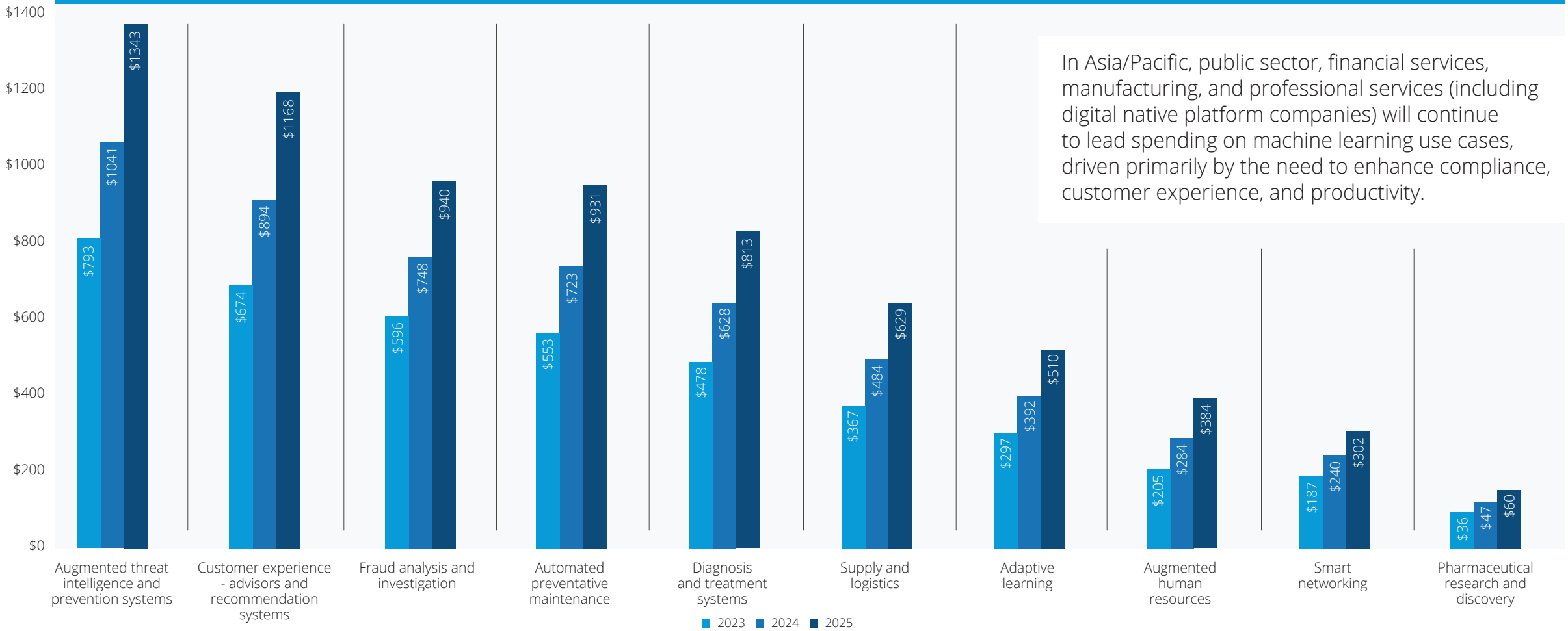
Find important nodes - identify anomalies & communities

IDC sees a new generation of machine learning solutions, underpinned by graph data science, will represent, explore, and make better predictions of the inter-connected world.

Opportunities with Graph-powered Machine Learning

IDC predicts that in Asia/Pacific including Japan (APJ), spending on machine learning with a graph data science focus will grow by an average rate of 30% year-on-year, through 2025.

Projected Spending on Machine Learning – Use Cases with Graph Data Science Focus



In Asia/Pacific, public sector, financial services, manufacturing, and professional services (including digital native platform companies) will continue to lead spending on machine learning use cases, driven primarily by the need to enhance compliance, customer experience, and productivity.

Many Organizations Are Innovating with Graph Data Science

Versatile Maritime Routing Service

Use case type: Supply chain & logistics
Graph algorithms used: Path finding

ORBIT MI uses graph data science to deliver point-to-point maritime routing services that:

- support spatial, linear and tabular data sets
- provide a library of path-finding algorithms;
- incorporate parameters such as distance, weather, and canal and narrow passage constraints.

> *"Test unlimited "what if" scenarios; eliminate the guess work in expense planning; reduce port delays."*

Source: https://neo4j.com/case-studies/orbitmi/

Enriching Customer Profiles

Use case type: Customer 360
Graph algorithms used: Community finding, similarity finding

MEREDITH uses graph data science to power recommendation engines

Transforming billions of page views into millions of pseudonymous identifiers

> *"Instead of "advertising in the dark" we now better understand our customers, which translates into significant revenue gains and better-served consumers."*

Source: https://neo4j.com/case-studies/meredith/

Synergy in the Workforce

Use case type: Predictive maintenance
Graph algorithms used: Similarity finding

BOSTON SCIENTIFIC uses graph data science to investigate failure modes and prevent recalls

Up to 92% reduction
Analytical query times have dropped from 2 minutes to 10-55 seconds

> *"Everyone involved with the project, from business stakeholders to technical implementers, is able to understand one another because they're all speaking a common language."*
>
> *Eric Wespi, Data Scientist, Boston Scientific*

Source: https://neo4j.com/case-studies/boston-scientific/

Message from the Sponsor

Neo4j Graph Data Science

Connecting the Dots in Complex Data for Predictive Insights

Neo4j offers the only graph data science engine built for data scientists to improve predictions and ML models, at scale, with seamless integration across the data stack to get more data science projects to production.

Improve predictions

- Explore billions of data points in seconds and identify hidden connections.
- Discover what's important, what's unusual, and what's next by adding contextual analysis.
- Reuse and share models across teams to support the most accurate predictions.

- Graph native ML Pipelines
- Catalog of over 65 pretuned algorithms across Search, Centrality, Similarity, Link Prediction, Community Detection, Graph Embeddings, etc.
- Reuse and share models

Create and share actionable insights

- Quickly demonstrate the value of data science projects with graph analysis and visualizations.
- Visualize the patterns and relationships in your graph and share them with decision makers.
- Give your business the confidence to answer questions like what's important, what's unusual, and what's next using graph algorithms.

- Native Python Client
- Graph visualization tools & BI connectors
- KNIME, Dataiku, and Google Cloud Vertex A
- Apache Spark & Apache Kafka Connectors

Deploy data science models to production

- Get more data science projects adopted with our flexible deployment options and multiple methods for moving models into production.
- Easily integrates with your data stack and use native connectors to access, store, move, and share data.
- Works on all major clouds and infrastructures to provide stable and scalable production.

- 30+ connectors and extensions
- Native BI Connector
- > Data Warehouse Connector
- > Compatible with all major clouds

Learn More

neo4j

IDC Asia/Pacific
83 Clemenceau Avenue
#17-01 UE Square, West Wing
Singapore 239920
T 65.6226.0330

idc.com

@idc

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications, and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

IDC Custom Solutions

This publication was produced by IDC Custom Solutions. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. A license to distribute IDC content does not imply endorsement of or opinion about the licensee.

IDC Doc. #AP241320IB