What’s New in Neo4j Graph Data Science 2.4

June 2023

What is Neo4j Graph Data Science?

Neo4j Graph Data Science is an analytics and modeling engine that uses the relationships in your data to improve predictions. It plugs into enterprise data ecosystems so you can get more data science projects into production quickly. Using pretuned graph algorithms, data scientists explore billions of data points in seconds to identify hidden connections and generate compelling visualizations that lead to better stakeholder decision making.

Areas of Investment

Neo4j offers the only Graph Data Science solution built for data scientists to improve their predictions and ML models, at scale, with seamless integration across the data stack. We continue to build on the momentum of our 2.0, 2.1, 2.2, and 2.3 releases and focus on delivering the most comprehensive Graph Data Science solution on the market.

We're investing in four key areas:

1. Make Better Predictions: Easily identify and interpret meaningful relationships in your data to discover what's important, what's unusual, and what's next.
2. Built for Data Scientists: Work in a familiar environment and quickly demonstrate practical business value.
3. Integrate with your Data Ecosystem: Integrate Graph Data Science with the existing tools across your technology stack and data pipeline using native connectors.
4. Production Ready: Trusted, Scalable, and Robust: Deployment flexibility and options for moving models into production to get more data science projects adopted.

What’s new?

Highlights from this release include:

Maker Better Predictions

Features that enable you to easily identify and interpret meaningful relationships hidden in your data include:

- **New** Community Detection Algorithm: K-Core Decomposition: Partitions a graph into layers from peripheral to more central nodes. This is accomplished by recursively finding subgraphs where each node has at least k connections, k going from 0 up to the max degree of the graph. Use this algorithm to group nodes according to their overall centrality and influence in your data. Useful for analytics, feature engineering, and visualization; popular use cases include:
    - **Social networks:** community detection and user influence
    - **Drug discovery:** protein-protein interactions
    - **Network design:** optimize information spreading

- **New** Shortest Path Algorithm: Bellman-Ford shortest path algorithm: Compute shortest paths in graphs with negative weights. A high-performance adaptation of the original Bellman-Ford algorithm called **Shortest-Path Faster Algorithm (SPFA)** significantly reduces computation time via sampling and parallelization techniques. Use cases for this improved Bellman-Ford alternative include:
    - **IT Networking:** routing, load balancing, network topology, and cost optimization
    - **Supply Chain Optimization:** resource allocation
    - **Finance:** routing transactions

- **New** Common Neighbour Aware Random Walk (CNARW): Sample graphs with dense clusters. Graph sampling helps you scale advanced analytics and machine learning to large graphs by providing representative subgraphs that preserve graph structure and attributes. It uses common neighbour awareness to decrease the odds of sampling duplicate nodes and increase the odds of selecting nodes across different clusters. It improves the convergence speed of algorithms, increasing the quality of the samples by avoiding loops and sampling efficiently clustered graphs that are common in online social networks. This has a broad range of use cases including:

    - **Recommendation engines:** suggest items that are likely to be of interest based on common neighbors of users or items
    - **Anomaly and Fraud Detection:** identify nodes that deviate significantly from other nodes
    - **Market Segmentation:** effectively identify densely connected groups of nodes that often correspond to communities

- **Improved** Performance for finding multiple shortest paths (Yen’s Algorithm): Parallelized Yen’s algorithm for faster computation on large complex graphs.

Built for Data Scientists

Work in a familiar environment and quickly demonstrate practical business value with:

- **New** Share ML Models across Neo4j Graph Data Science instances: Save time and prevent ML model replication across Graph Data Science instances. Transfer all node classification and link prediction models and pipelines including multi-layer perceptrons (MLP) and Random Forest **[Enterprise Only]**.

Integrate with your Data Ecosystem

New and improved connectors, extensions, and integrations across the data pipeline ecosystem include:

- **New AuraDS Enterprise on AWS Early Access Program:** AuraDS Enterprise, a fully managed graph analytics and ML solution, is now available on AWS through the Early Access Program. With AuraDS Enterprise on AWS, you can integrate, analyze, and manage your graph data science pipelines using Amazon SageMaker and Amazon EMR. As an AWS Data and Analytics Competency Partner, we help organizations make the most of their data connections and relationships. If you're interested in joining the Early Access Program for AuraDS Enterprise on AWS, please contact us.

- **New AuraDS Enterprise on Azure:** AuraDS Enterprise, a fully managed graph analytics and ML solution, is now available on Azure through the Early Access Program. With Neo4j AuraDS Enterprise on Microsoft Azure, you can integrate, analyze, and manage your graph data science pipelines using Azure Machine Learning and Azure Synapse Analytics. As a Microsoft Azure Partner Network member, we help organizations make the most of their data connections and relationships. If you're interested in joining the Early Access Program for AuraDS Enterprise on Azure, please contact us.

- **New BigQuery integration:** The native Neo4j integration to Google BigQuery allows you to extend SQL analysis with graph-native data science and machine learning by working seamlessly between BigQuery and Neo4j Graph Data Science whether using BigQuery SQL or notebooks. Data science teams can improve and enrich existing analysis and ML using the graph-native data science capabilities within Neo4j by running graph analysis-in memory- directly from BigQuery.
    - Conduct deep graph analysis of connected data, in-memory, and return results to BigQuery for use in downstream models and analysis.
    - Build a graph data model in BigQuery and move it directly into Neo4j Graph Data Science in simple steps. (No additional software or drivers to download, install, and use).
    - Simplify data movement between BigQuery and Neo4j by eliminating the need for extra software or data staged in a bucket or queuing system.
    - Write graph features and algorithm results back to BigQuery
    - Data Engineers work entirely within BigQuery to build models (without new software or tool switching) for deep graph analysis in Neo4j Graph Data Science.

Production Ready: Trusted, Scalable, and Robust

Features that improve deployment flexibility and move models in production faster:

- **New Fine-grained access control** in AuraDS Enterprise grants administrators permission to manage user access to graph elements like node labels, relationship types, and properties, based on user roles and privileges. Access control is achieved through catalogs, where users can only read from their own catalogs, barring explicit publishing of graphs/models or for admin users, ensuring a higher level of data security and user autonomy.

- **Improved High-degree node graph performance:** Users can more efficiently analyze graphs with high-degree nodes, meaning nodes with an above-average quantity of edges. Augmented degree-based partitioning for higher-performance parallelization shows significant improvements on FastRP, HashGNN, Leiden, Max K-Cut, Conductance, Link Prediction (training), and the graph projection: ToUndirected algorithms.

Resources

Want to learn more about the 2.4 Graph Data Science release? Check out the Graph Data Science Manual and Graph Data Science Changelog to learn more.

Previous Announcements

Neo4j Graph Data Science 2.3 - February 2023

Neo4j Graph Data Science 2.2 - October 2022

Neo4j Graph Data Science 2.1 - June 2022

Neo4j Graph Data Science 2.0 and AuraDS - April 2022