

Neo4j for Graph Data Science™

Introduction

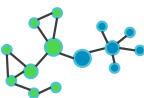


Neo4j - The Graph Company

The Industry's Largest Dedicated Investment in Graphs



Creator of the **Label Property Graph** and **Cypher** language at the core of the **GQL** ISO project



Thousands of Customers
World-Wide



HQ in **Silicon Valley**,
offices include **London**,
Munich, **Paris** & **Malmo**

Industry Leaders use Neo4j

20/25 Top Financial Firms

7/10 Top Retail Firms

7/10 Top Software Vendors



Anyway You Like It

On-Prem



In the Cloud



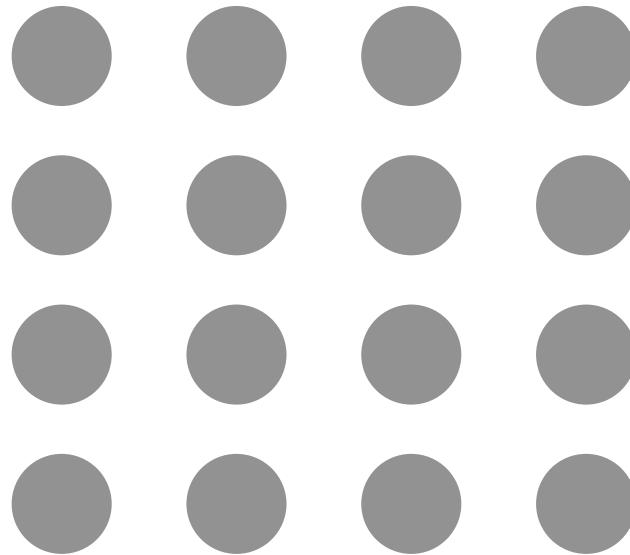
Google Cloud Platform



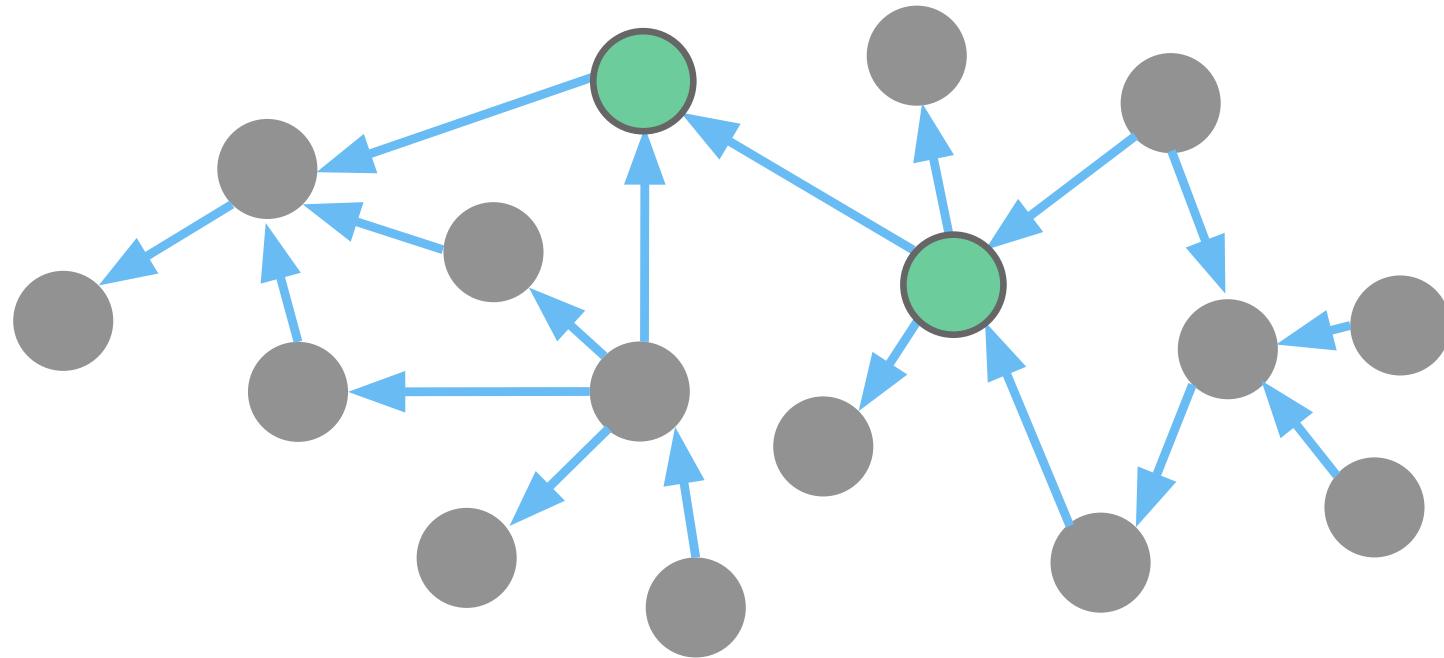
DB-as-a-Service



It's Not What You Know

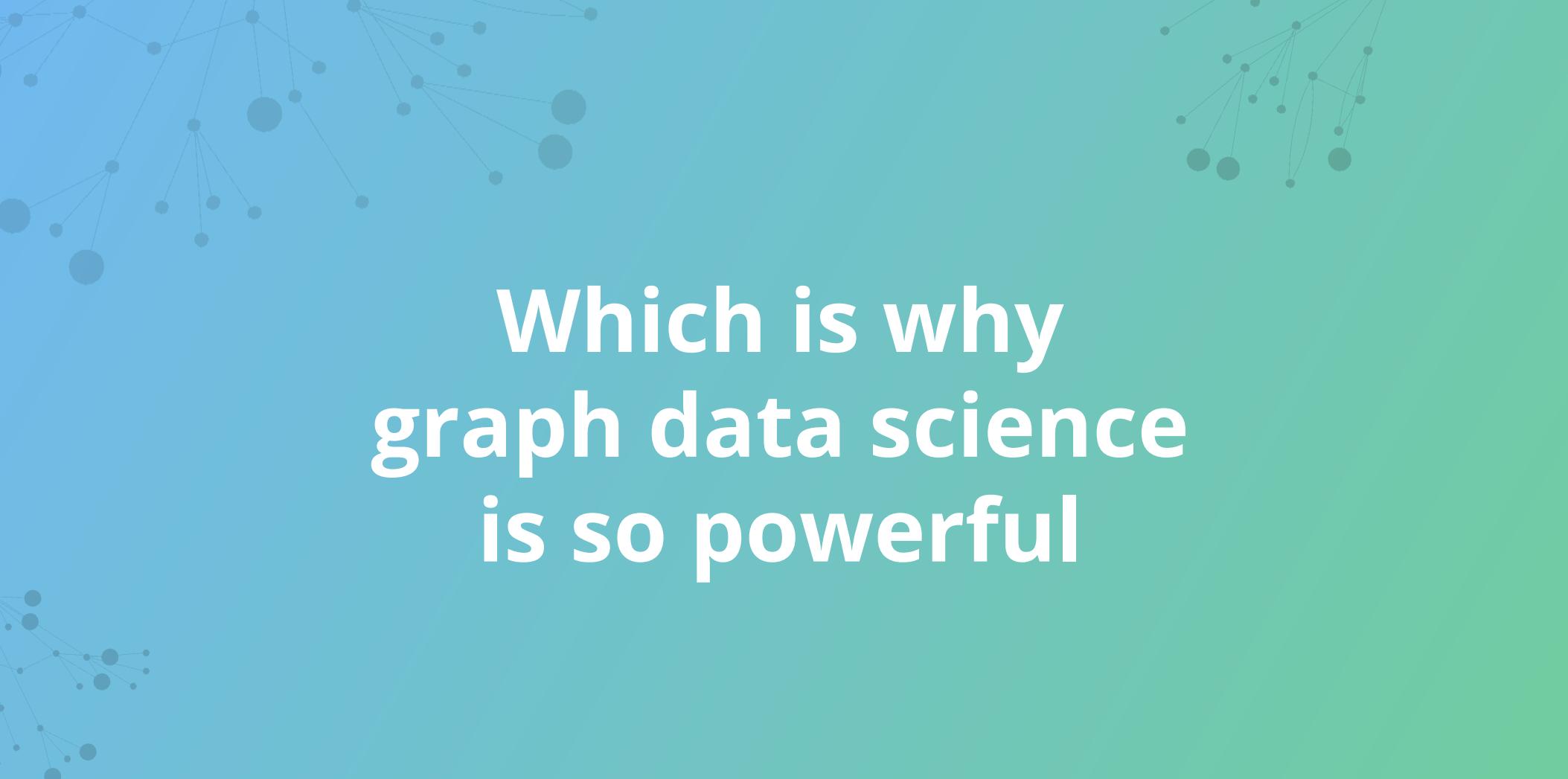


It's Who You Know And Where They Are





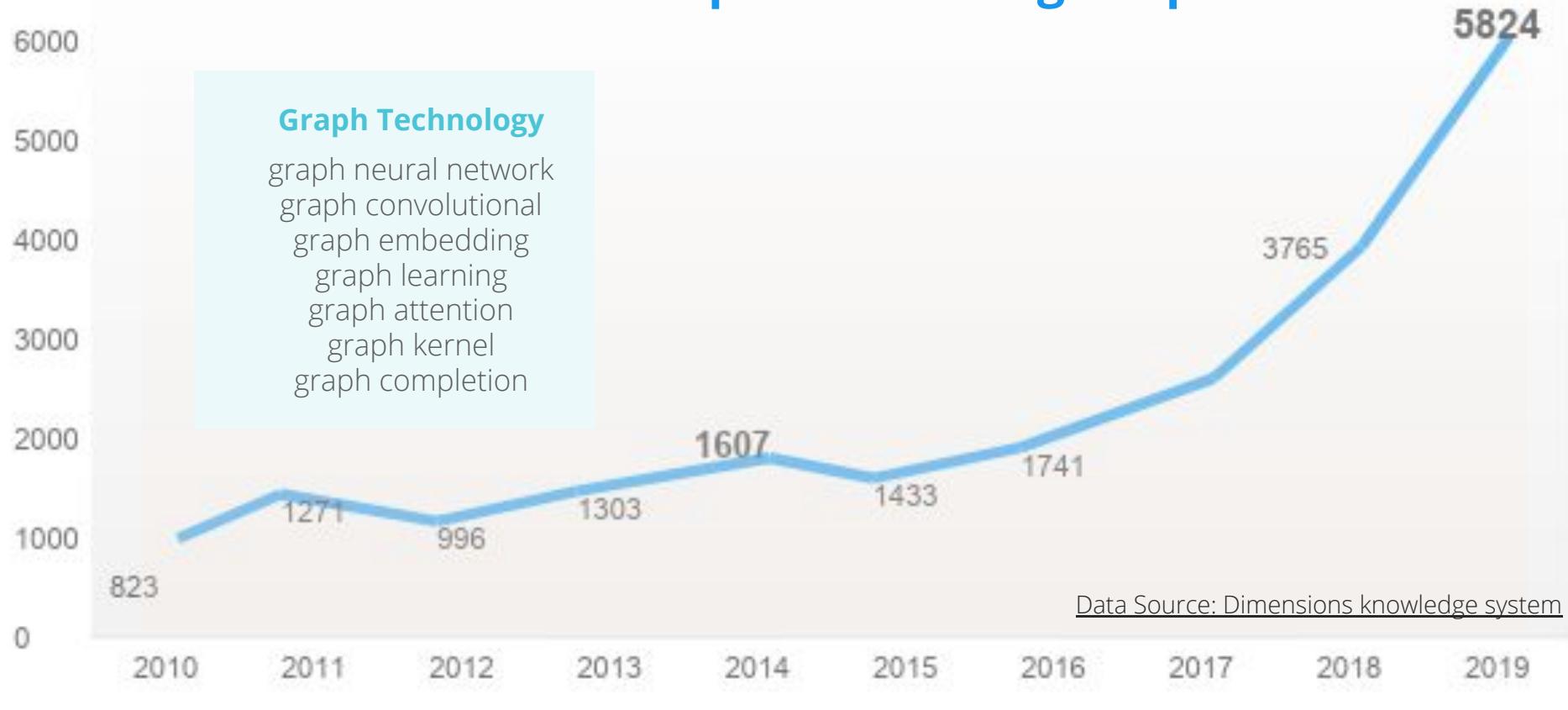
It's a counter-intuitive notion



Which is why
graph data science
is so powerful

Graph Is Accelerating AI Innovation

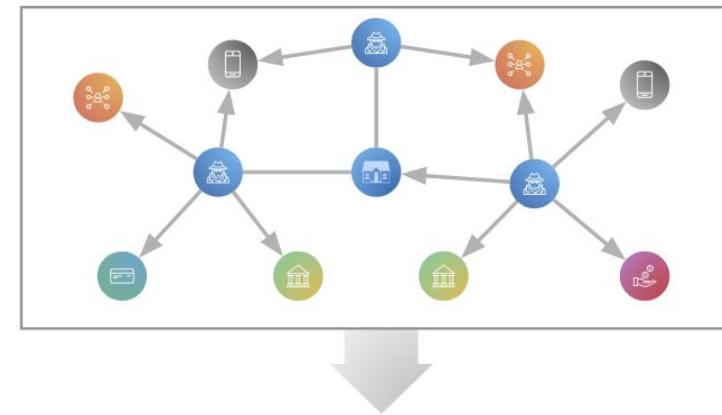
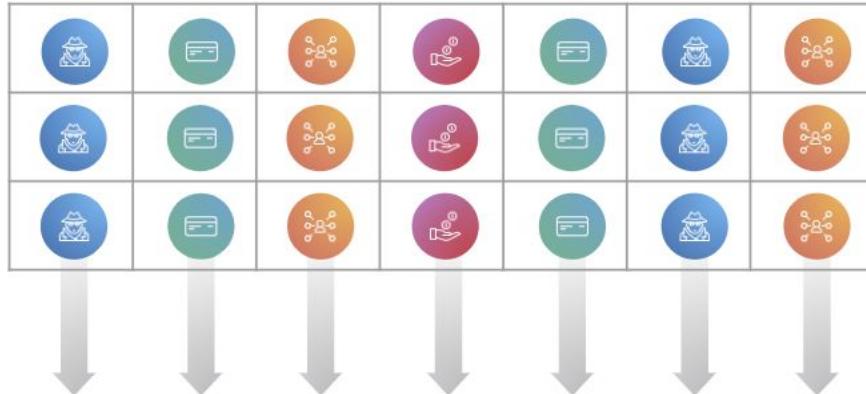
AI Research Papers Featuring Graph



Better Predictions with Graphs

Using the Data You Already Have

- Current data science models ignore network structure
- Graphs add highly predictive features to ML models, increasing accuracy
- Otherwise unattainable predictions based on relationships

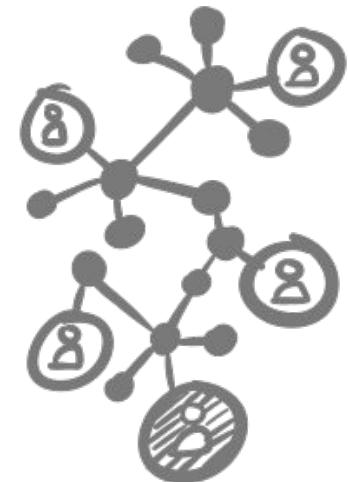


Machine Learning Pipeline



Productionize more accurate,
predictive models

Relationships and
network structures
are highly predictive
and underutilized



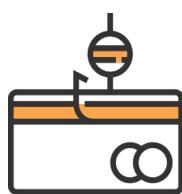
– and already in your data.

Graph are a natural way to
store and use this predictive
information, but different
than what you're doing today.

Graph Data Science Applications

Just a few examples...

Fraud Detection



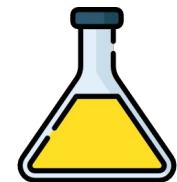
Disambiguation & Segmentation



Personalized Recommendations



Life Sciences



Churn Prediction



Search & Master Data Mgmt.



Predictive Maintenance



Cybersecurity



Top Graph Data Science Applications

in Financial Services and Banking

Fraud
Marketing
Customer Journey



- First party & synthetic identity fraud
- Fraud rings
- Money laundering
- Disambiguation
- Recommendations
- Customer segmentation
- Churn prediction

Top Graph Data Science Applications

in Healthcare and Life Sciences

Discovery

Patient Care

**Regulatory
Compliance**



- Drug repurposing
- Knowledge graph completion
- Risk identification & spread
- Patient journey
- Personalized care
- Contact tracing

Top Graph Data Science Applications

in Marketing and Supply Chain

Market-To
Supply Chain
Logistics



- Disambiguation
- Recommendation
- Customer segmentation
- Logistics and routing
- Predictive fulfillment
- Risk identification

Improving Analytics, ML & AI Across Industries

AstraZeneca Patient Journeys



- Early intervention project with 3 yrs of visits, tests & diagnosis with **10's of Bn of records**
- Finding similarities in patient journeys
- Graph algorithms for identifying communities & **best intervention points**



Meredith Marketing to the Anonymous



- Mostly anonymous users across devices and sites with ever changing cookies
- 4.4 TB: +14 Bn nodes +20Bn relationships
- +160 Mn rich, unique profiles created
- **612% Increase** in visits per profile

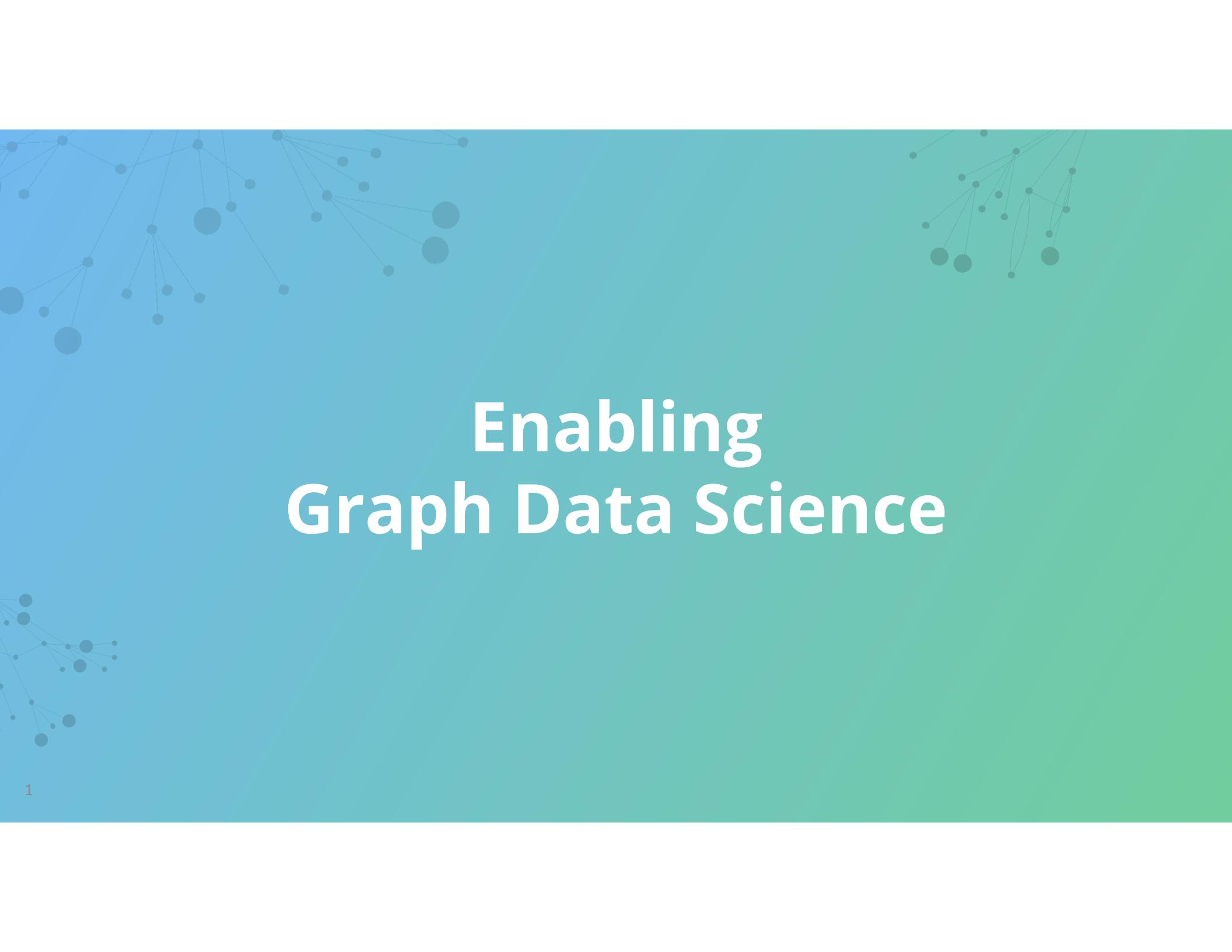


Financial Fraud Detection & Recovery



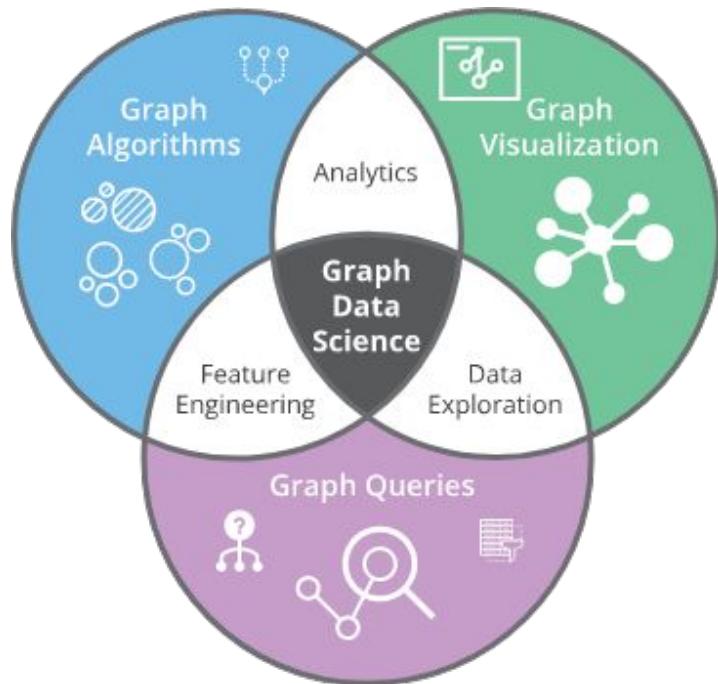
- Almost 70% of Credit Card fraud was missed
- Synthetic Identities were biggest challenge
- +1B Nodes and +1B Relationships to analyse
- Graph analytics with queries & algorithms help **find \$10's of millions of fraud** in 1st year



A large, semi-transparent network graph is visible in the background, consisting of numerous small dark grey nodes connected by thin white lines. In the center, there are two larger, semi-transparent blue nodes. On the left side, there is a cluster of three larger, semi-transparent blue nodes. On the right side, there is another cluster of three larger, semi-transparent blue nodes.

Enabling Graph Data Science

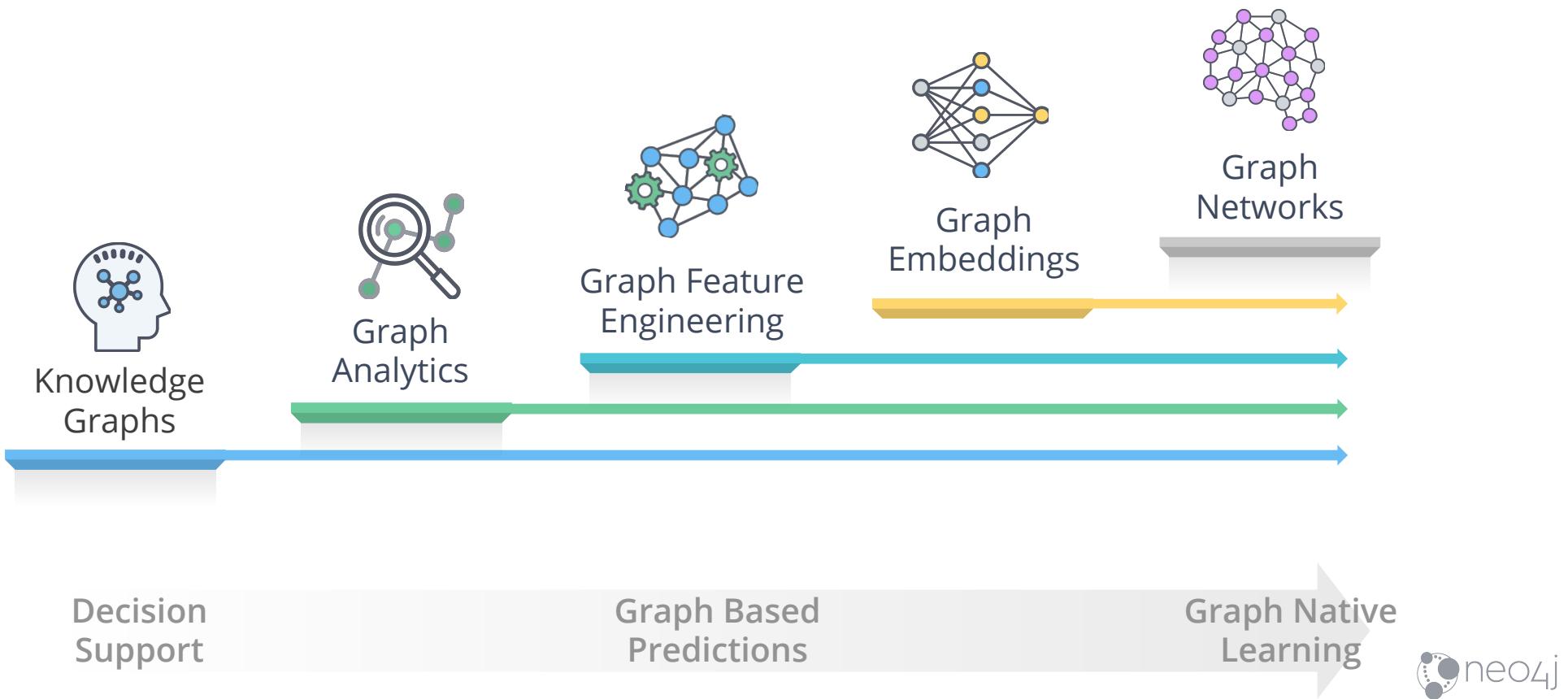
What is *Graph* data science?



Graph Data Science is a science-driven approach to gain knowledge from the relationships and structures in data, typically to power predictions.

Data scientists use relationships to answer questions.

Evolution of Graph Data Science



Evolution of Graph Data Science



Knowledge
Graphs

Graph search
and queries

Support domain
experts



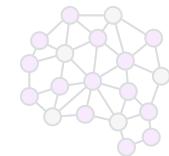
Graph
Analytics



Graph Feature
Engineering



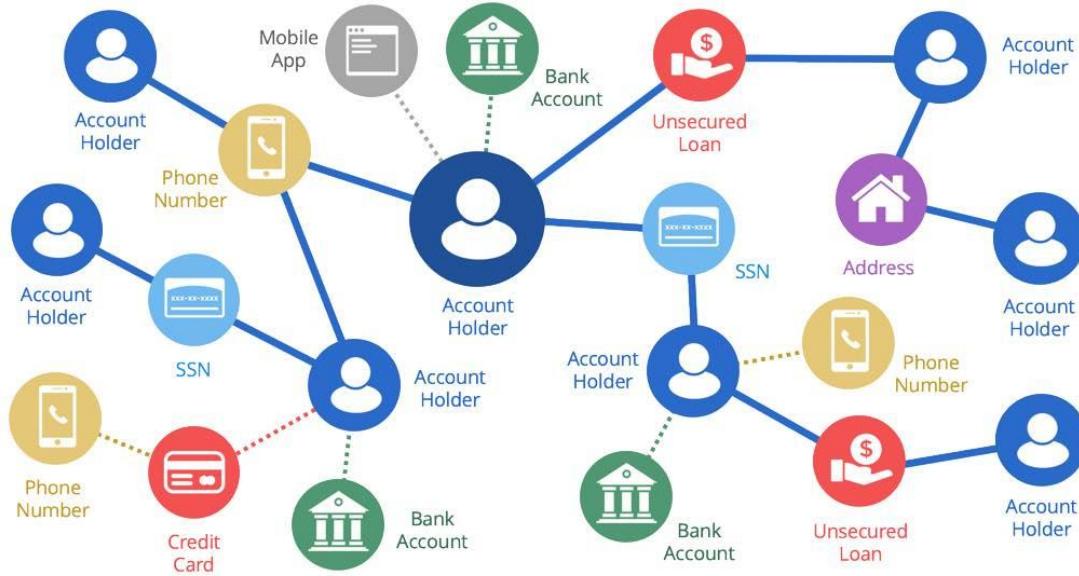
Graph
Embeddings



Graph
Networks

Knowledge Graph Queries in Financial Fraud

Improving existing pipelines to identify fraud via heuristics



Deceptively Simple Queries

How many flagged accounts are in the applicant's network **4+ hops out?**

How many **login / account variables in common?**

Add these metrics to your approval process

Knowledge Graph Queries in Marketing

Improving recommendations and product placement



Deceptively Simple Queries

Collaborative filtering: users who bought X, also bought Y (open-ended pattern matching)

What items make you more likely to buy additional items **in subsequent transactions**?

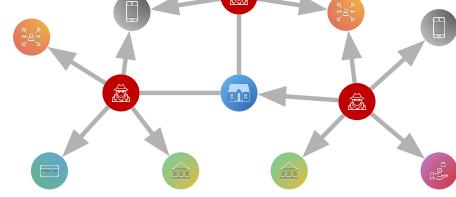
Traverse hierarchies - what items are similar **4+ hops out?**

So, When Do I Need Graph Algorithms?

Query (e.g. Cypher/Python)

Real-time, local decisioning
and pattern matching

**Local
Patterns**

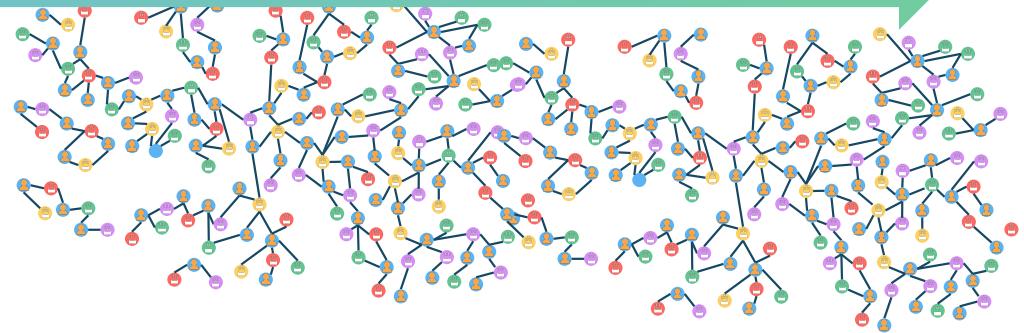


You know what you're looking for and making a decision

Graph Algorithms

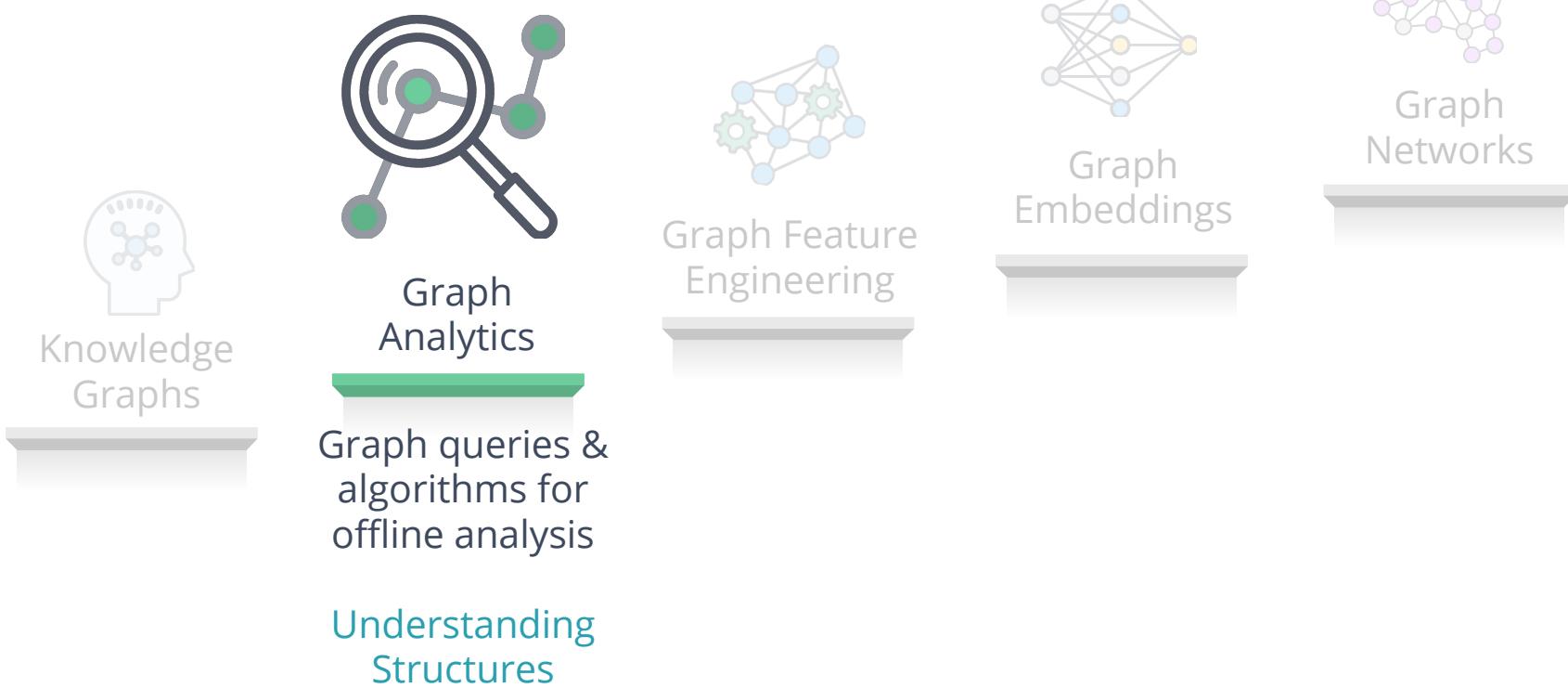
Global analysis
and iterations

**Global
Computation**



You're learning the overall structure of a network, updating data, and predicting

Evolution of Graph Data Science



Graph Algorithms & Functions in Neo4j



Pathfinding & Search

- Shortest Path
- Single-Source Shortest Path
- All Pairs Shortest Path
- A* Shortest Path
- Yen's K Shortest Path
- Minimum Weight Spanning Tree
- K-Spanning Tree (MST)
- Random Walk



Centrality / Importance

- Degree Centrality
- Closeness Centrality
- CC Variations: Harmonic, Dangalchev, Wasserman & Faust
- Betweenness Centrality & Approximate
- PageRank
- Personalized PageRank
- ArticleRank
- Eigenvector Centrality



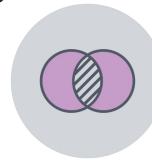
Community Detection

- Triangle Count
- Clustering Coefficients
- Connected Components (Union Find)
- Strongly Connected Components
- Label Propagation
- Louvain Modularity
- K-1 Coloring
- Modularity Optimization



Link Prediction

- Adamic Adar
- Common Neighbors
- Preferential Attachment
- Resource Allocations
- Same Community
- Total Neighbors



Similarity

- Euclidean Distance
- Cosine Similarity
- Node Similarity (Jaccard)
- Overlap Similarity
- Pearson Similarity
- Approximate KNN

...and also Auxiliary Functions:

- Random graph generation
- One hot encoding
- Distributions & metrics

Graph Algorithms for Detecting Fraud

Graph algorithms enable reasoning about **network structure**

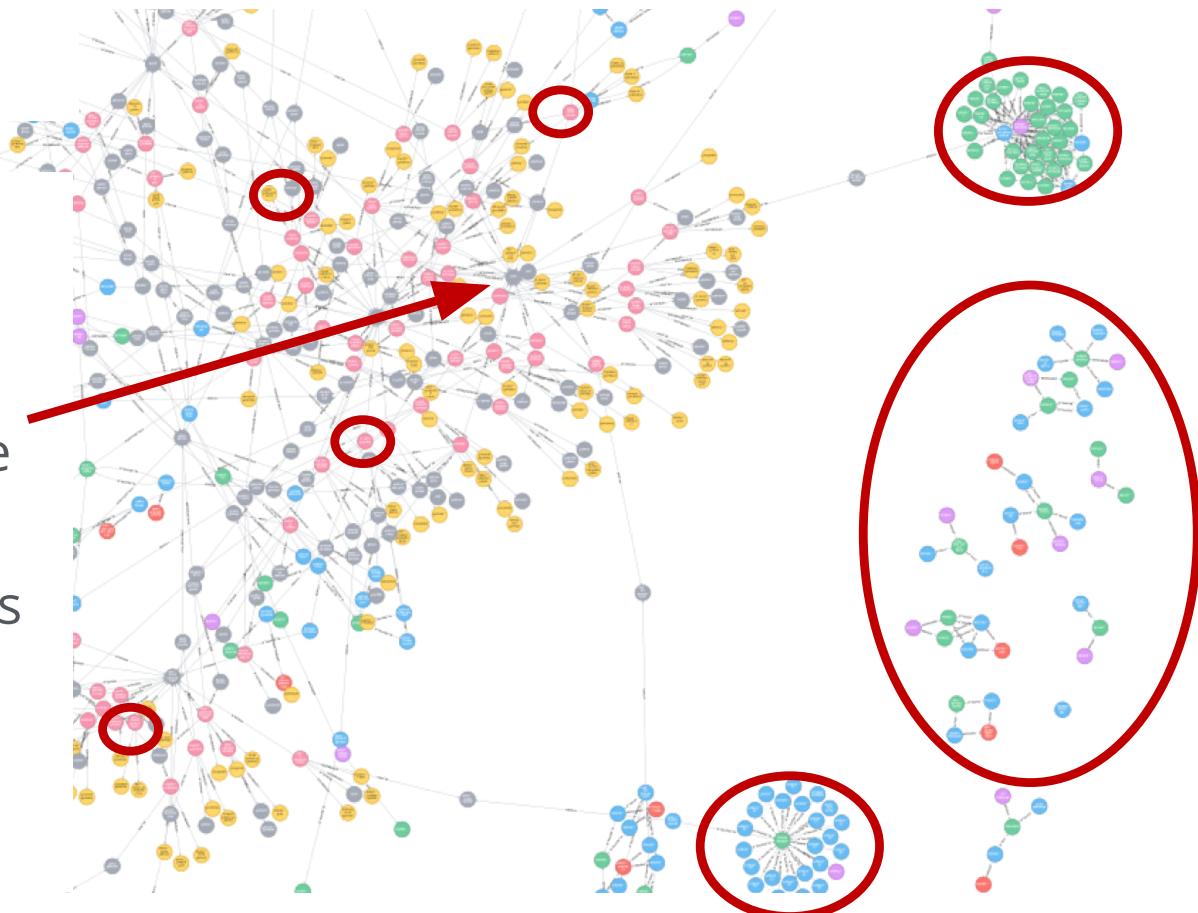


Connected components identify disjointed group sharing identifiers

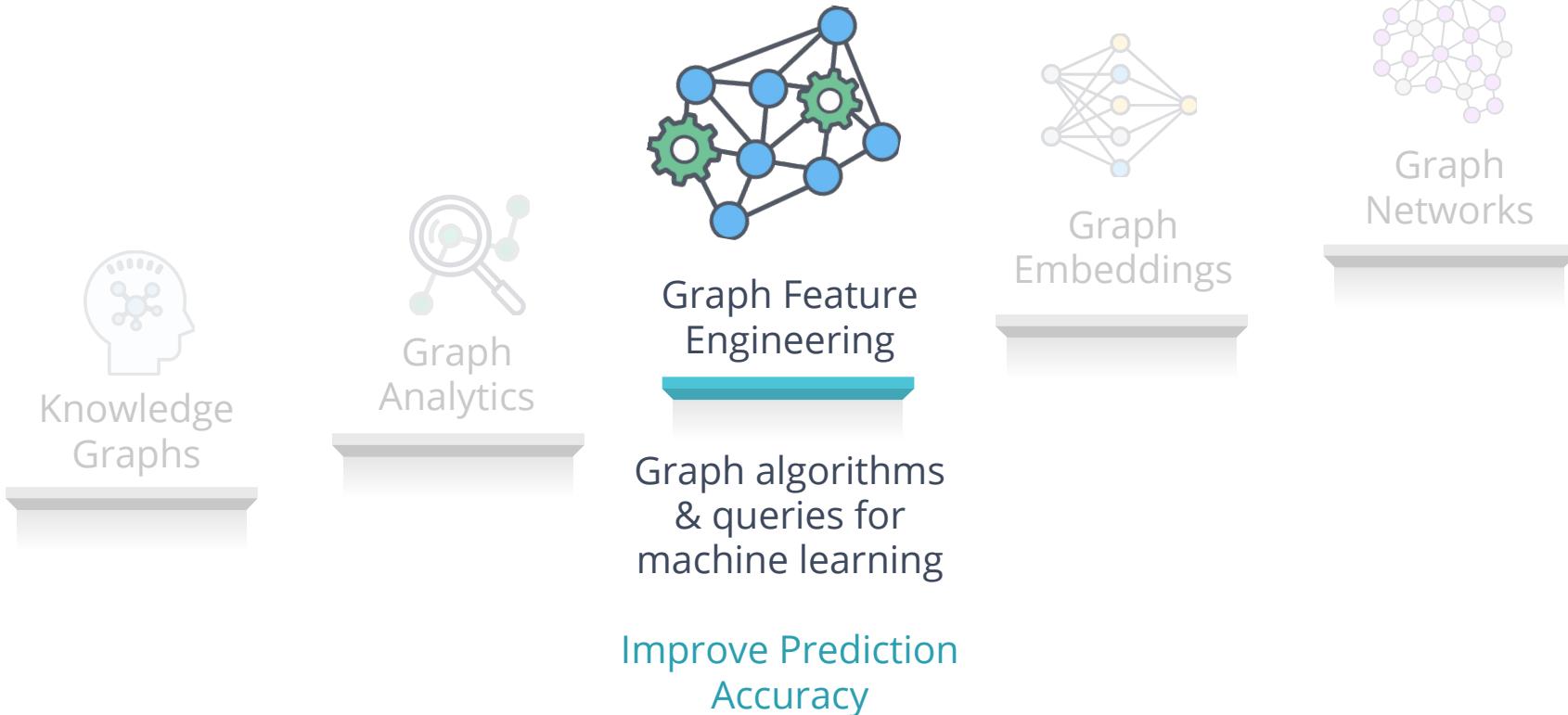
PageRank to measure influence and transaction volumes

Louvain to identify communities that frequently interact

Jaccard to measure account similarity

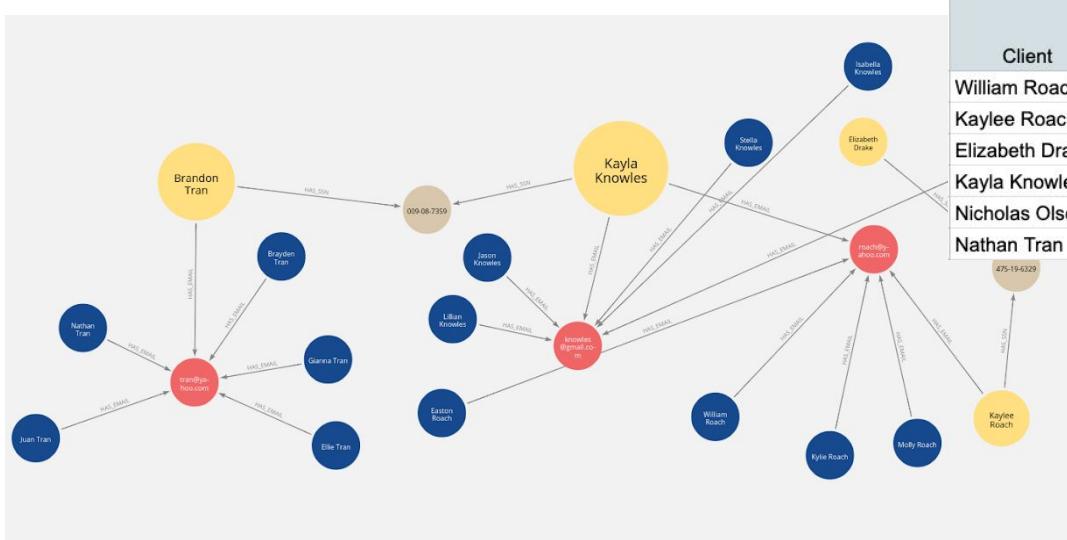


Evolution of Graph Data Science



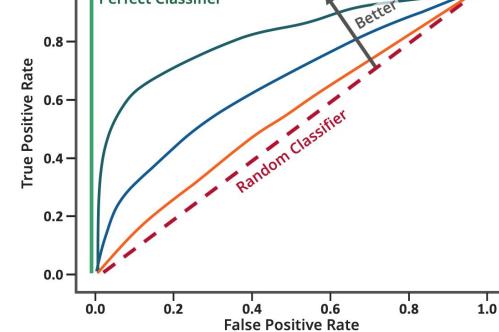
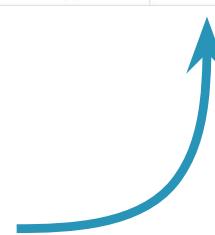
Graph Feature Engineering in Fraud

Feature Engineering is how we combine and process the data to create ***new, more meaningful features***. Using graphs we can base ML on **influential people** that share identifiers.

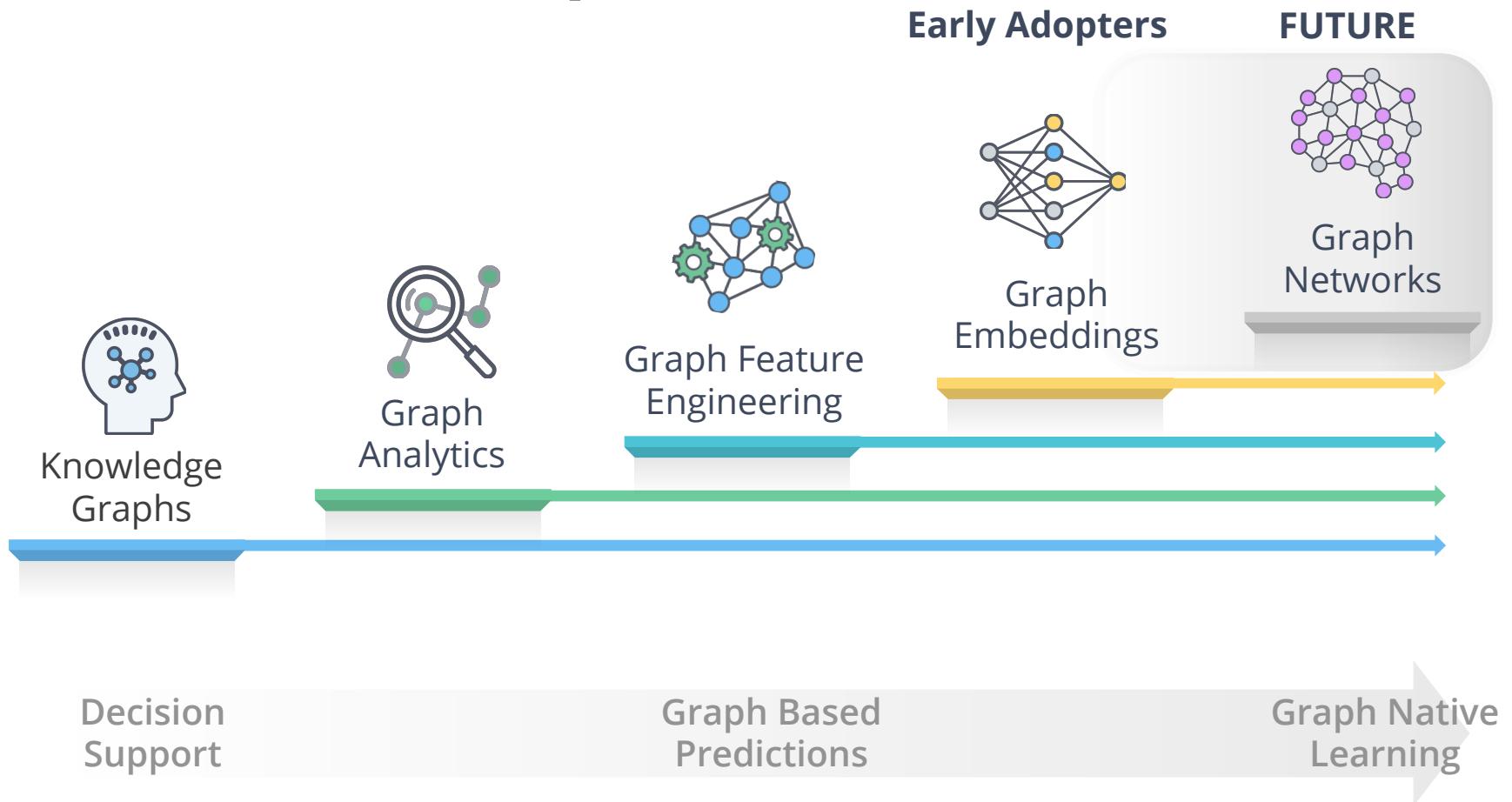


Financial Transaction Data

Client	Betweenness	Unique People Sharing Identifier	Weighted Shared Score	ML Model Prediction
William Roach	0	1	1	Normal
Kaylee Roach	32	2	4	Fraudster
Elizabeth Drake	0	1	20	Fraudster
Kayla Knowles	192	3	3	Fraudster
Nicholas Olsen	0	1	2	Normal
Nathan Tran	0	1	0	Normal



Evolution of Graph Data Science





Neo4j Makes this Evolution Possible

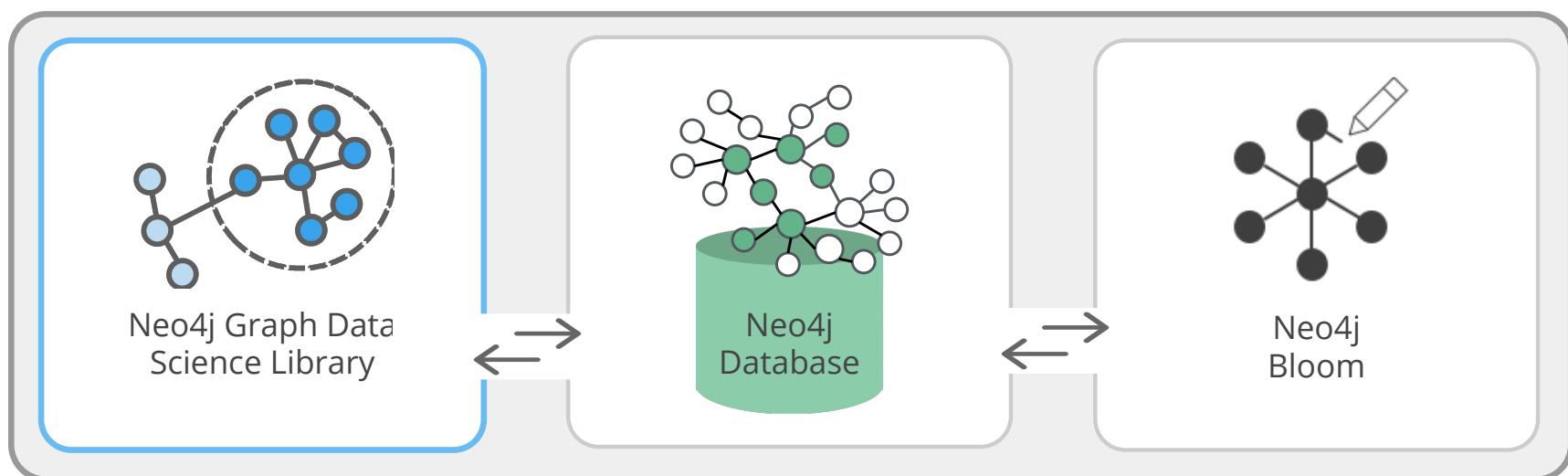


neo4j for Graph Data Science™

Scalable Graph
Algorithms & Analytics
Workspace

Native Graph
Creation & Persistence

Visual Graph
Exploration & Prototyping



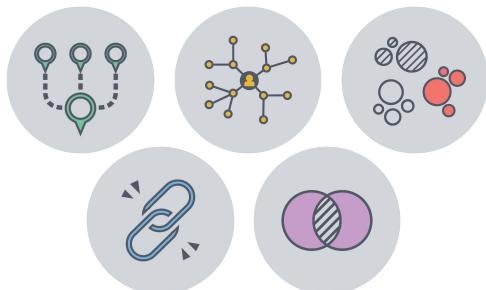
Practical

Integrated

Intuitive



The Neo4j GDS Library

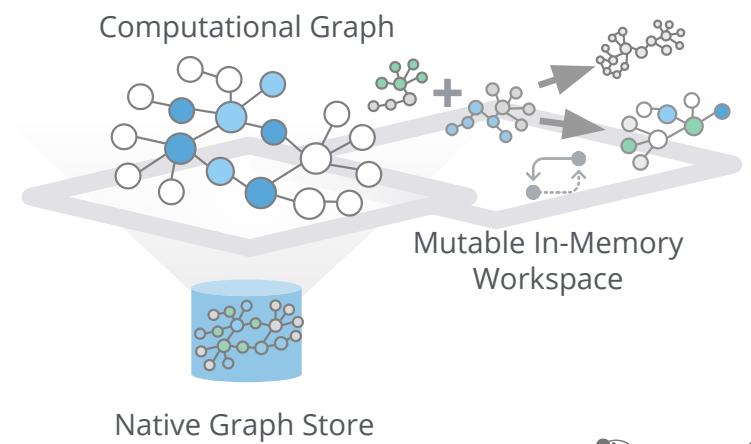


Robust Graph Algorithms

- Run on a loaded graph to compute metrics about the topology and connectivity
- Highly parallelized and scale to 10's of billions of nodes

Efficient & Flexible Analytics Workspace

- Automatically reshapes transactional graphs into an in-memory analytics graph
- Optimized for analytics with global traversals and aggregation
- Create workflows and layer algorithms



Increase your predictive accuracy with Graph Algorithms

Take advantage of hardened, validated graph algorithms that enable reasoning about [network structure](#).



Answer previously intractable questions
[with the data you already have](#)

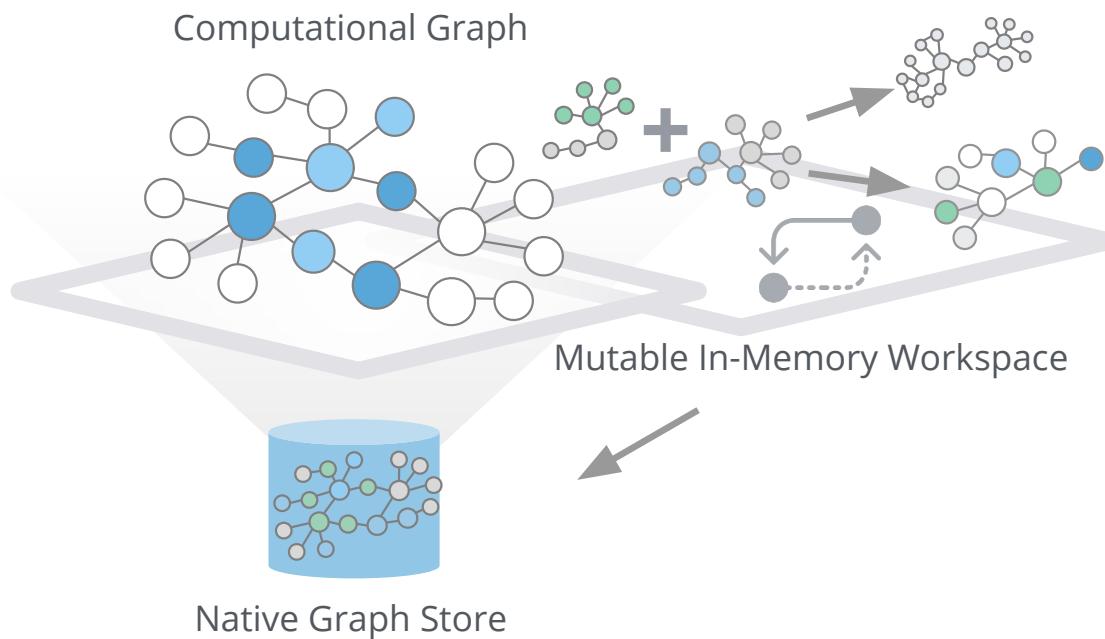
- Deep Path Analytics & Structural Pattern Matching
- Community & Neighbors Detection
- Influencer and Risk Identification
- Disambiguation
- Link and Behavior Prediction

Massive scale to 10's billions of nodes with optimized
algorithms



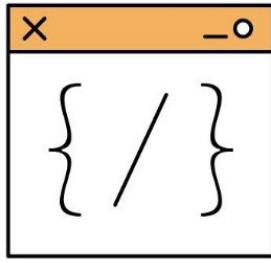
Most efficient way to do graph data science with enterprise-grade features, support and scale

A graph-specific analytics workspace that's mutable – integrated with a native-graph database



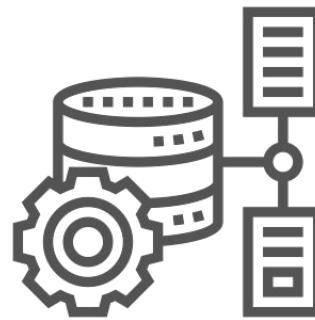
- Neo4j automates data transformations
- Fast iterations & layering
- Production ready features, parallelization & enterprise support

Find Value Faster with Neo4j's practical Graph Data Science framework



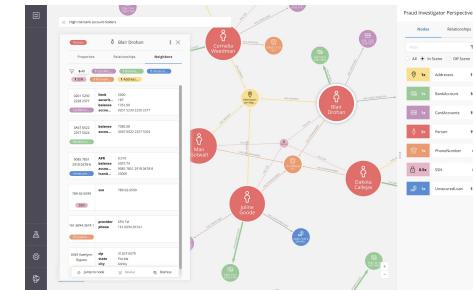
Drastically **simplified** and **standardized** API that enables **custom, flexible configurations**

Documentation, training, and examples so **getting started is simple**



Friendly data science experience with logical guardrails like **memory mgmt.** & **suggestive errors**

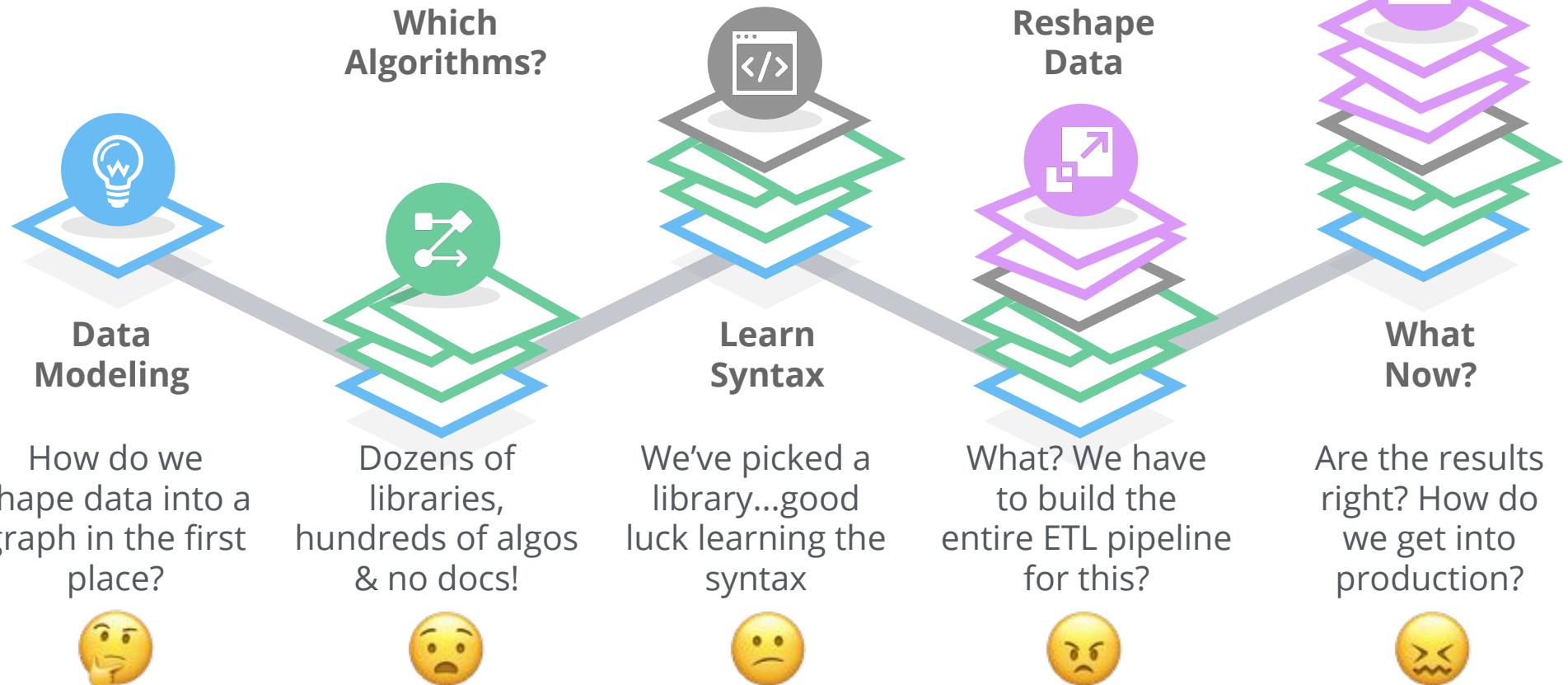
Reshaping, node & relationship aggregation / deduplication and multipartite algos



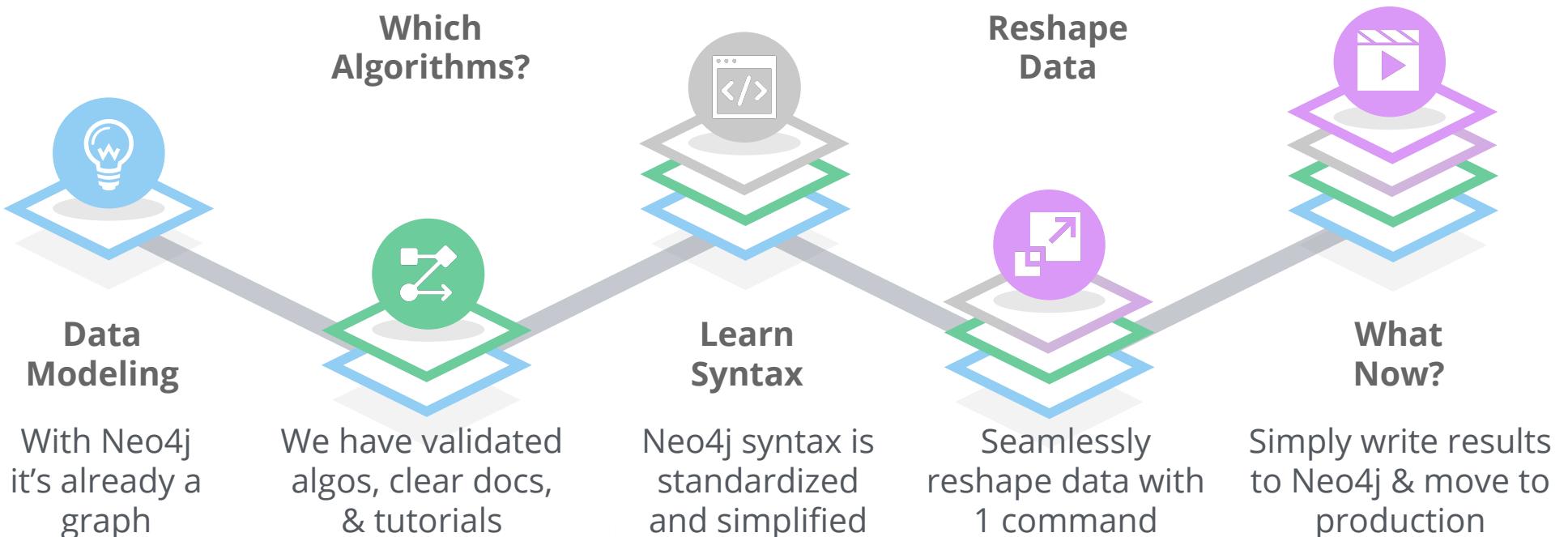
Explore graphs and algorithm results visually with **Bloom**

Share insights across teams for **better collaboration**

Data Science - It's Complicated



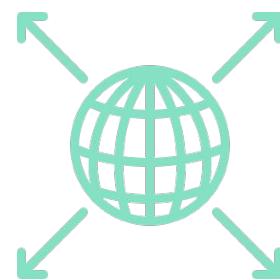
Simplify Your Experience



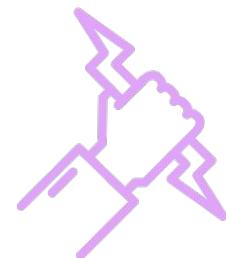
Improve Your Predictions. Get Your Models into Production.



Answer intractable
questions and increase
predictive accuracy - with
existing data



First graph data science
framework with
enterprise features,
scale, and support



Practical, easy-to-use
graph data science and
visual exploration