A faint, abstract network graph is visible in the background, composed of numerous small, semi-transparent grey dots connected by thin white lines, creating a sense of complex connectivity.

Enabling Graph Data Science

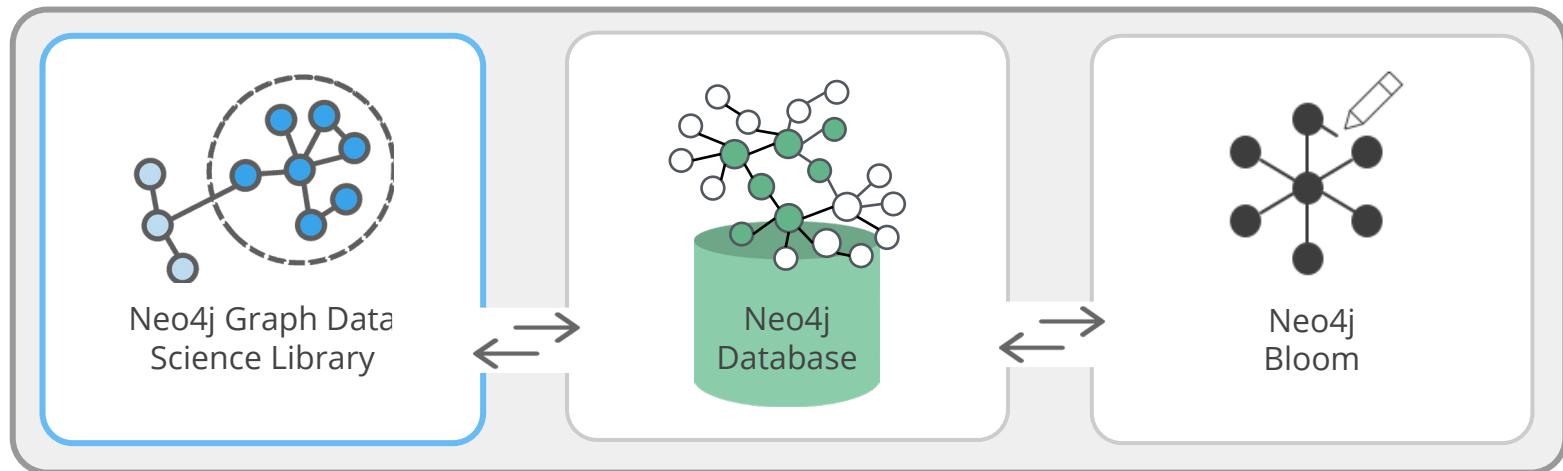


neo4j for Graph Data Science™

Scalable Graph
Algorithms & Analytics
Workspace

Native Graph
Creation & Persistence

Visual Graph
Exploration
& Prototyping

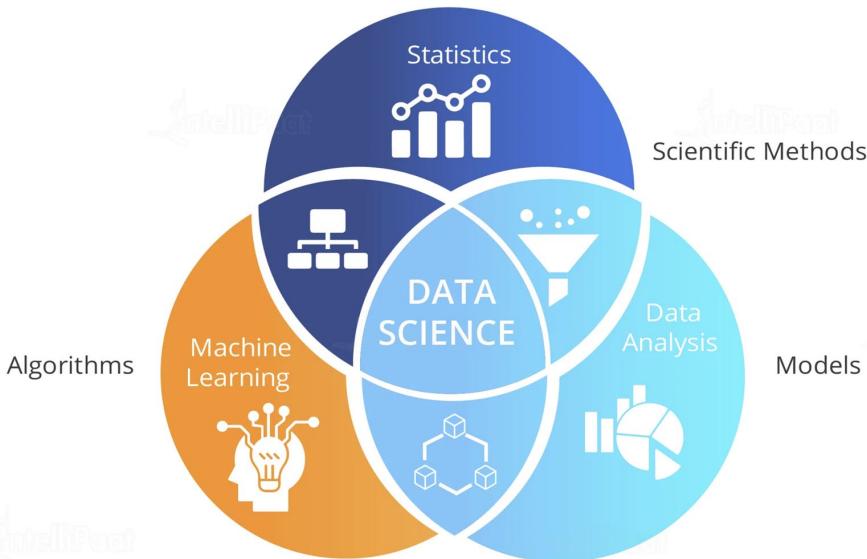


Practical

Integrated

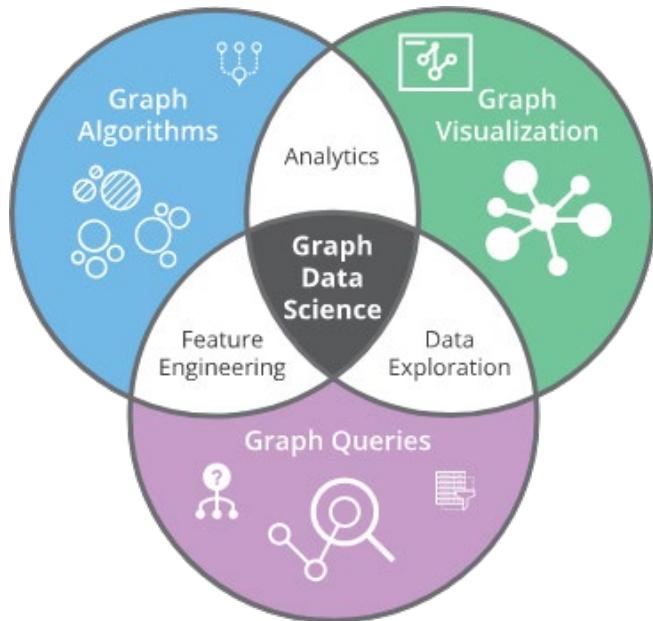
Intuitive

What is data science?



“Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data.” - Wikipedia

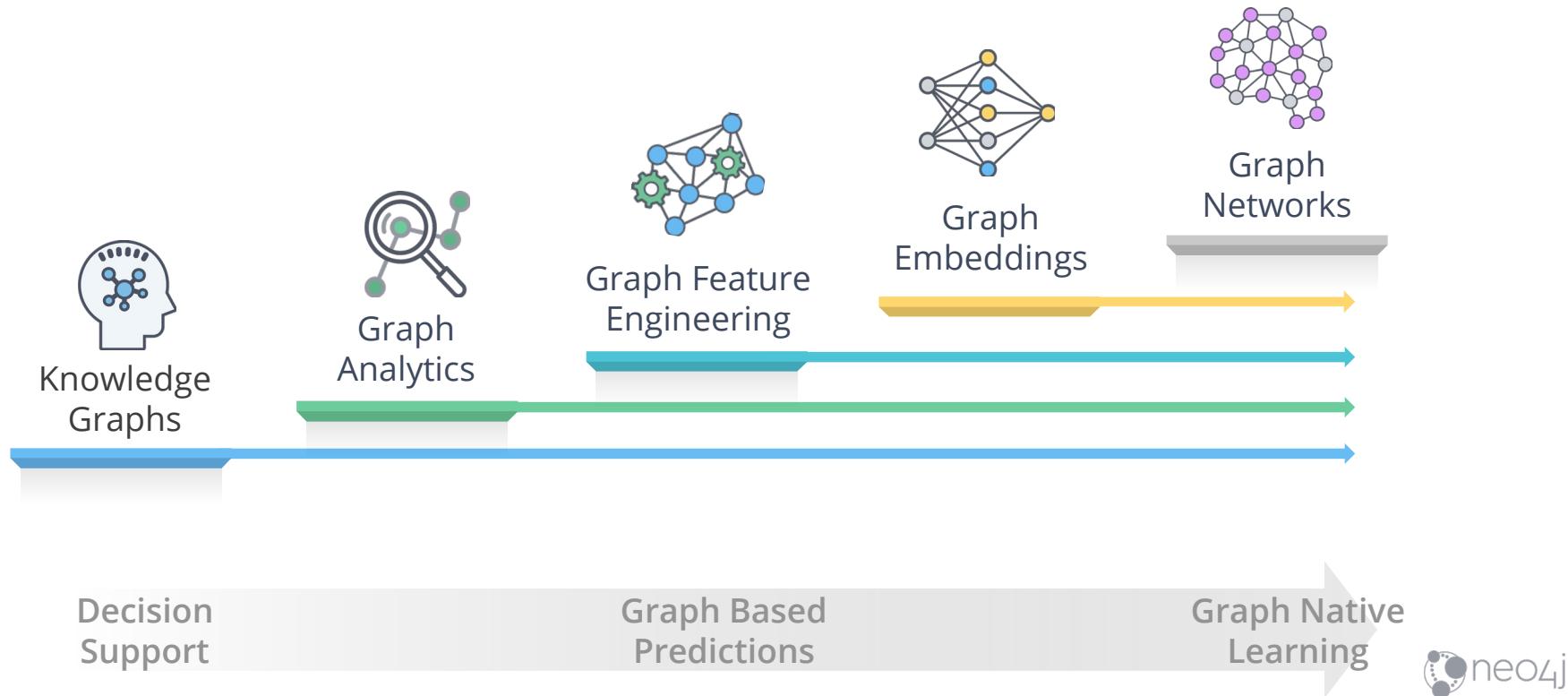
What is *Graph* data science?



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



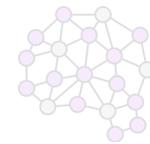
Graph
Analytics



Graph Feature
Engineering



Graph
Embeddings



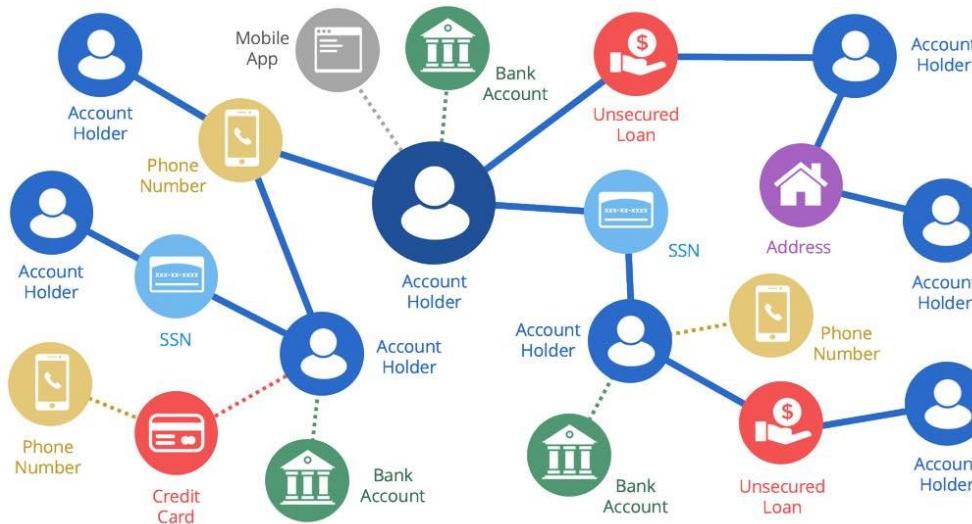
Graph
Networks

Graph search
and queries

Support domain
experts

Knowledge Graph Queries in Financial Fraud

Improving existing pipelines to identify fraud via heuristics



Deceptively Simple Queries

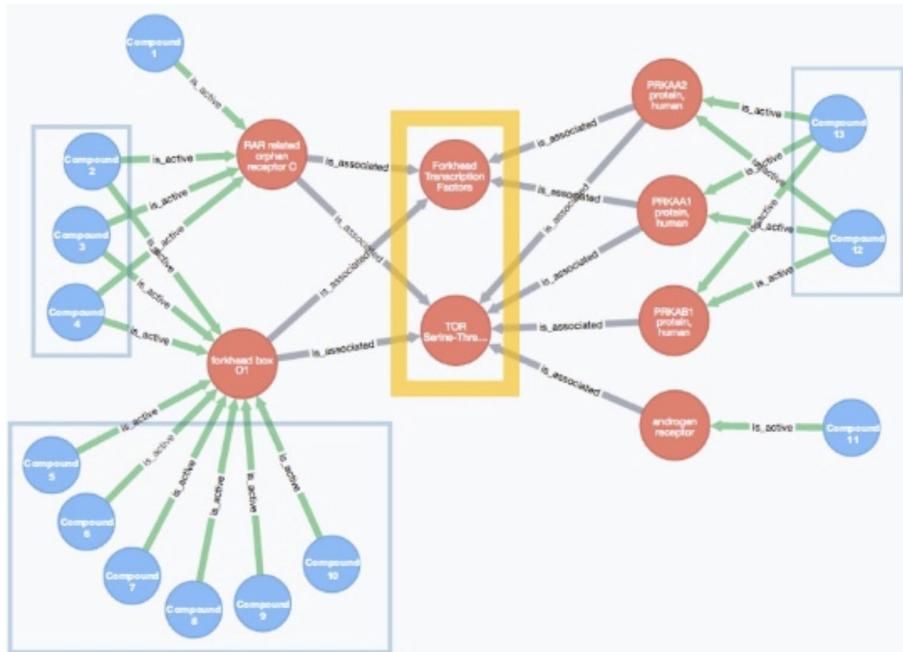
How many flagged accounts are in the applicant's network **4+ hops out?**

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

Add these metrics to your approval process

Knowledge Graph Queries in Drug Discovery

Companies like Novartis and the Chan Zuckerberg Foundation leverage knowledge graphs to mine medical research for insights



Deceptively Simple Queries

What **completes the connections** from genes to diseases to targets?

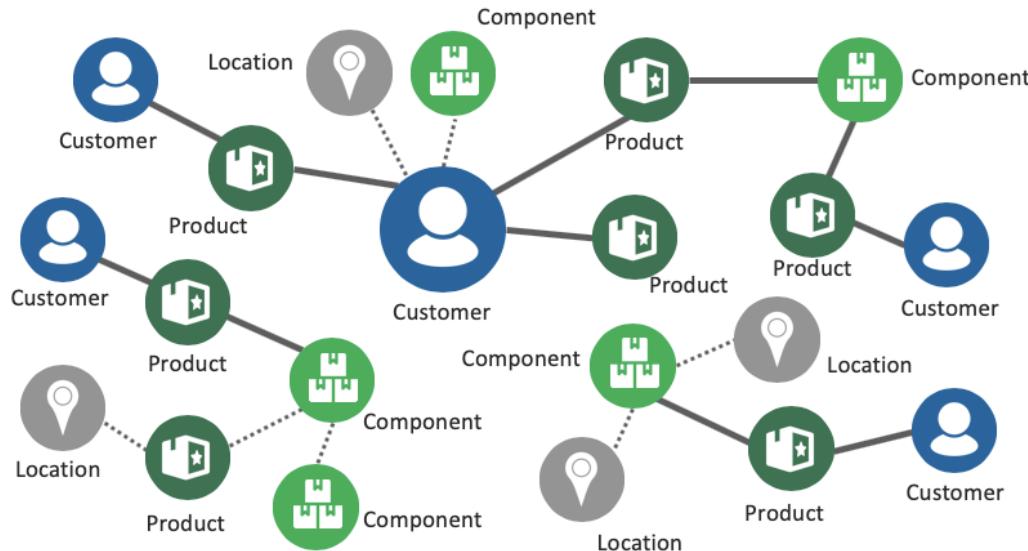
What genes can be reached **+4 hops out** from a known drug target?

What **mechanisms in common** are there between two drugs?

Which **biological pathways** are regulated by these genes?

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?](#)

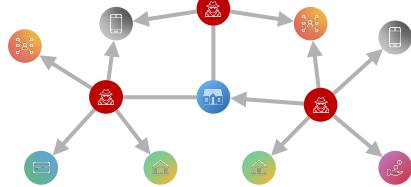
Traverse hierarchies - what items are similar [4+ hops out?](#)

So, When Do I Need Graph Algorithms?

Query (e.g. Cypher/Python)

Real-time, local decisioning
and pattern matching

Local Patterns

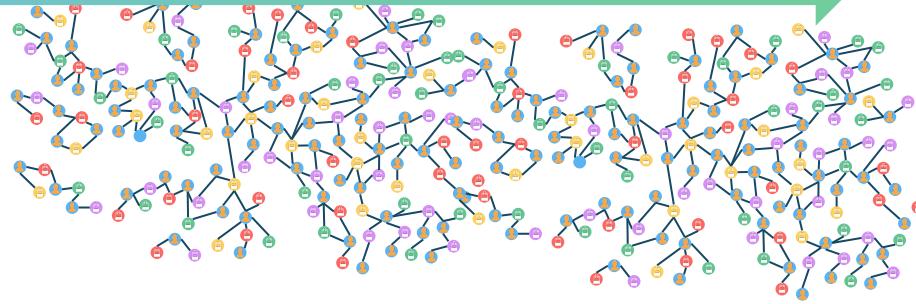


You know what you're looking for and making a decision

Graph Algorithms

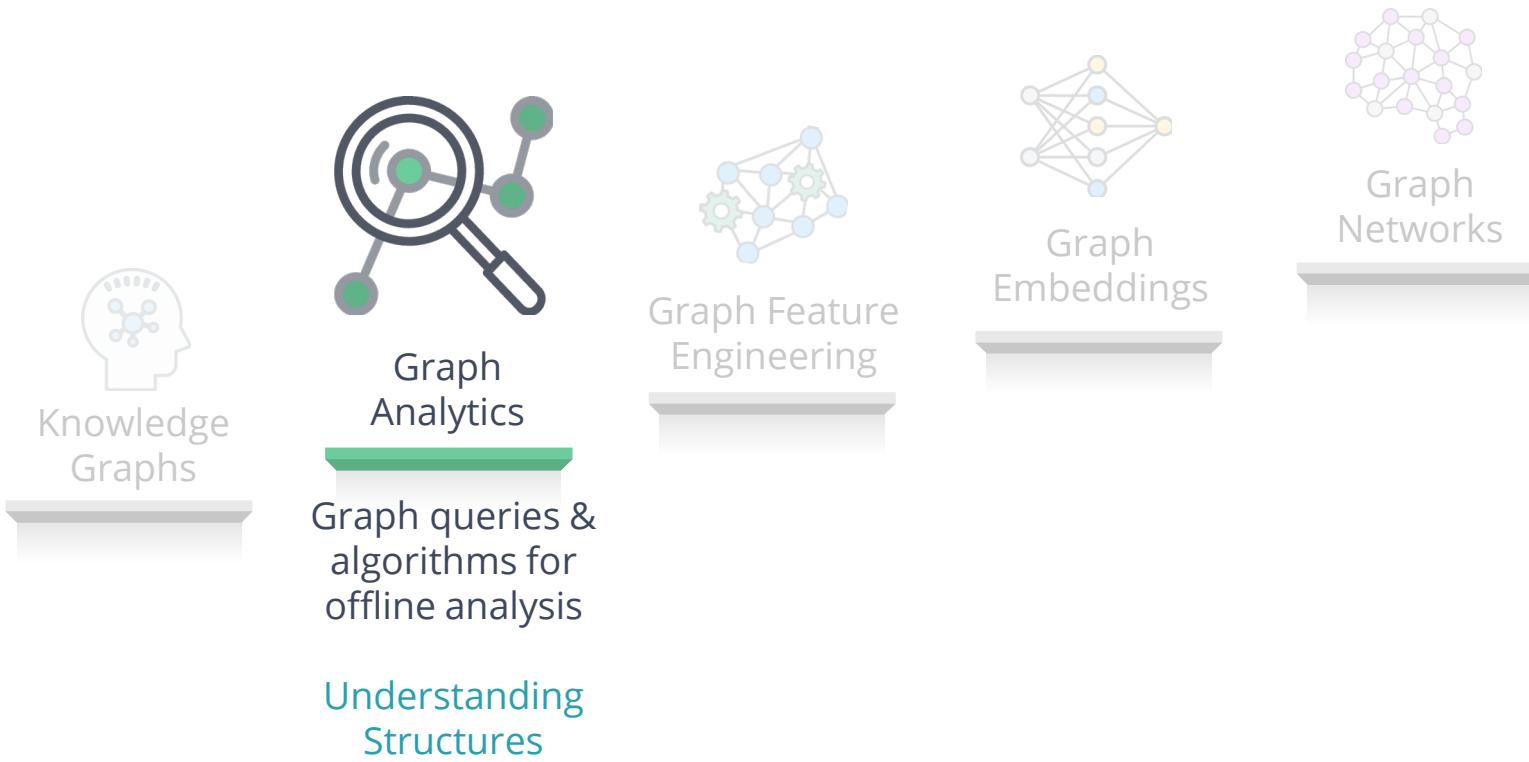
Global analysis
and iterations

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



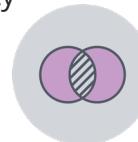
Community Detection

- 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**

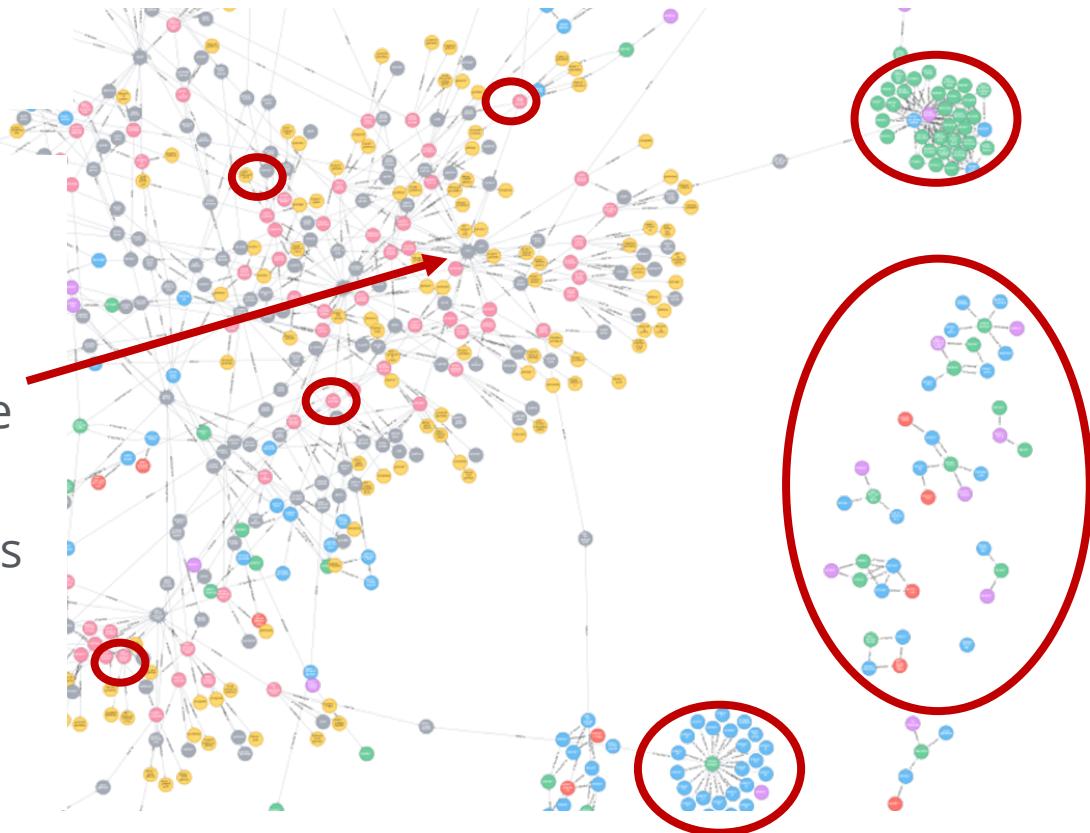


Connected components identify disjointed group sharing identifiers

PageRank to measure influence and transaction volumes

Louvain to identify communities that frequently interact

Jaccard to measure account similarity



Graph Algorithms for Drug Discovery

Identify drug mechanisms and new targets based on **network structure**

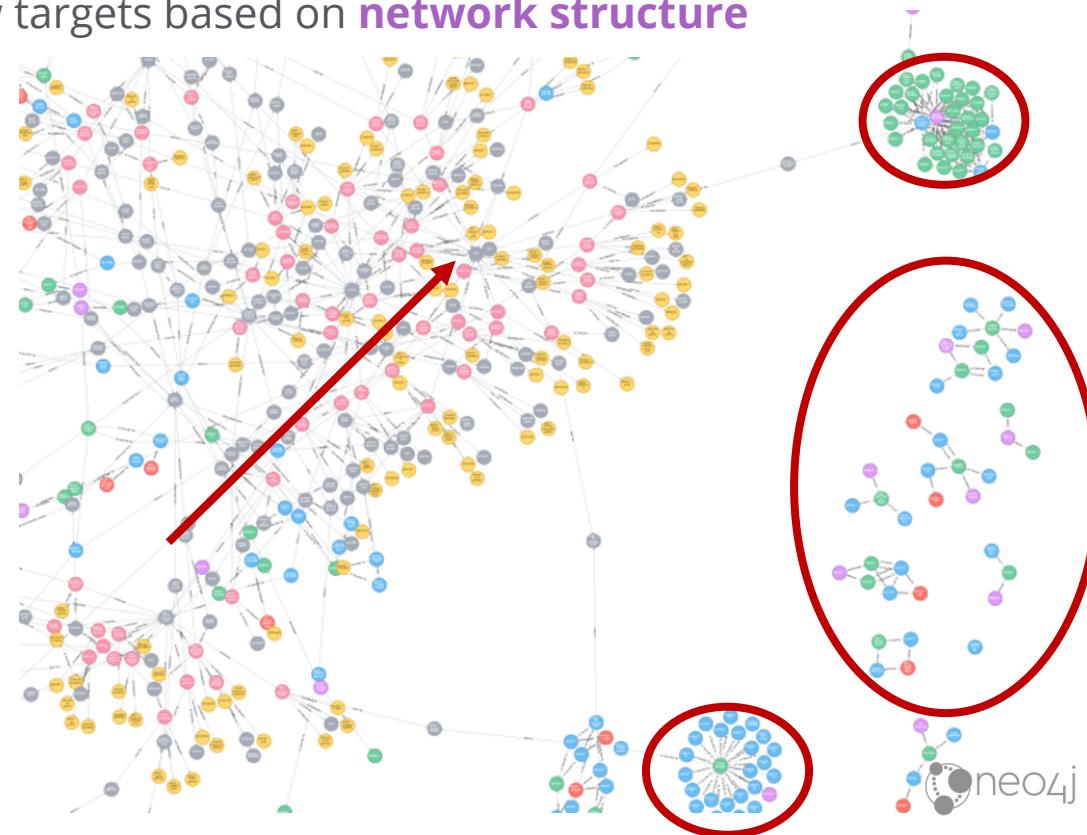
PageRank & Betweenness to identify essential regulatory genes or drug targets

Louvain to identify protein regulatory networks

Shortest path to link drug targets to possible outcomes or side effects

Node Similarity to find structurally similar chemicals

Link Prediction to estimate likelihood of interactions



Graph Algorithms in Retail

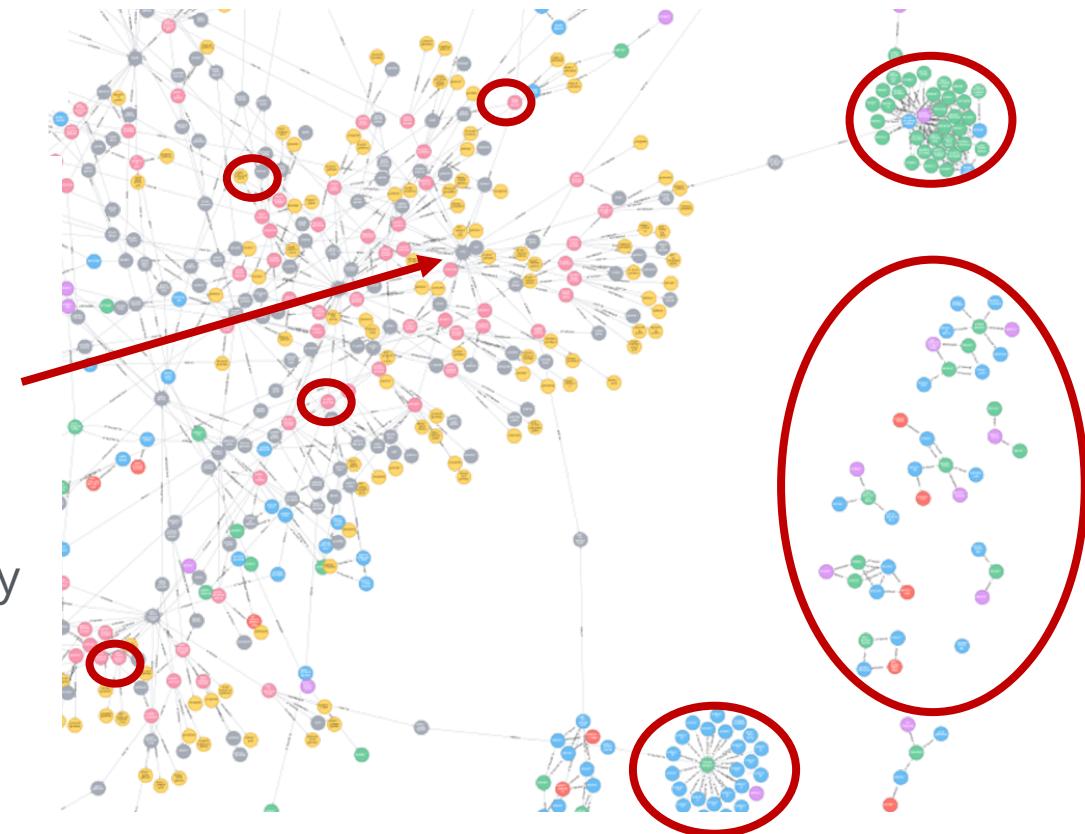
Graph algorithms enable reasoning about **network structure**

Connected components identify unique users

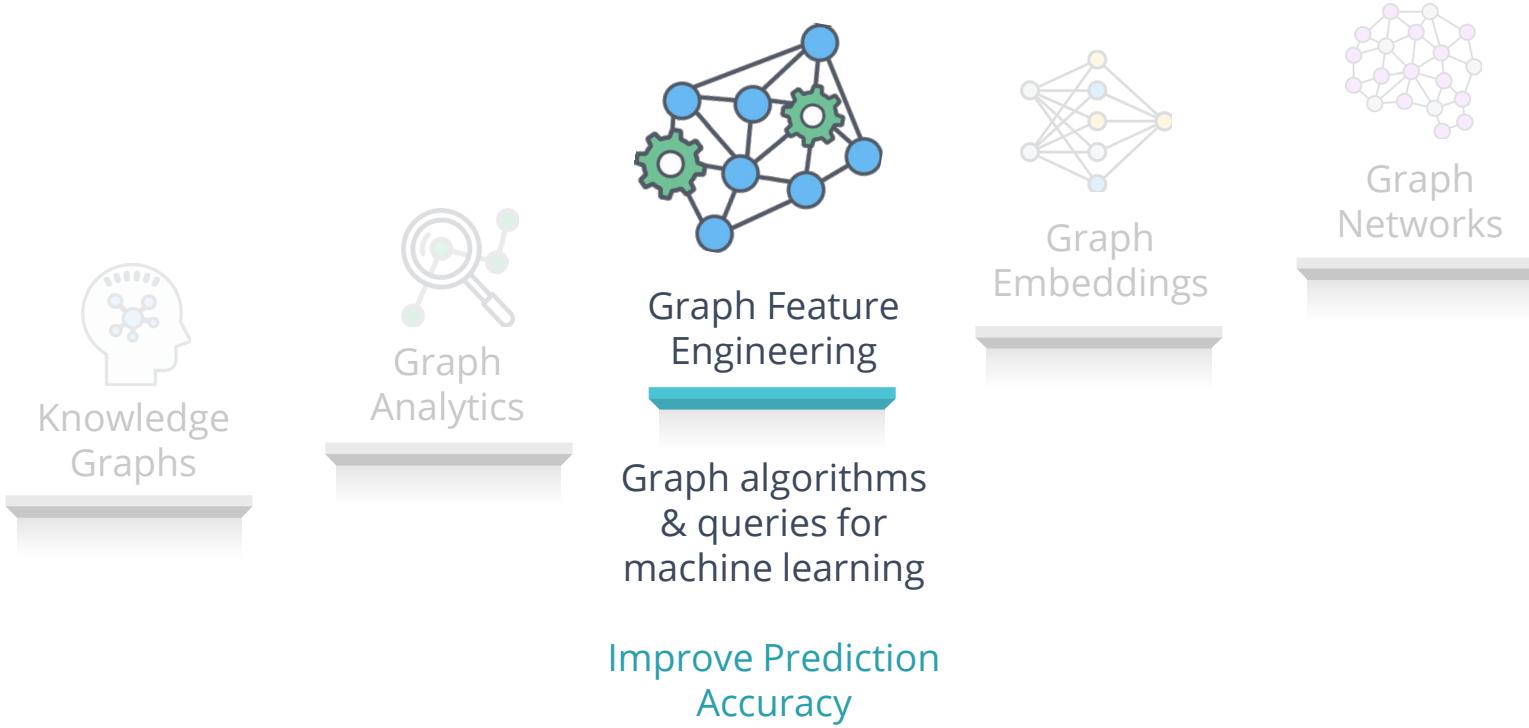
PageRank to measure transaction volumes

Louvain to identify customer segmentation based on topology

Jaccard to measure purchasing similarity

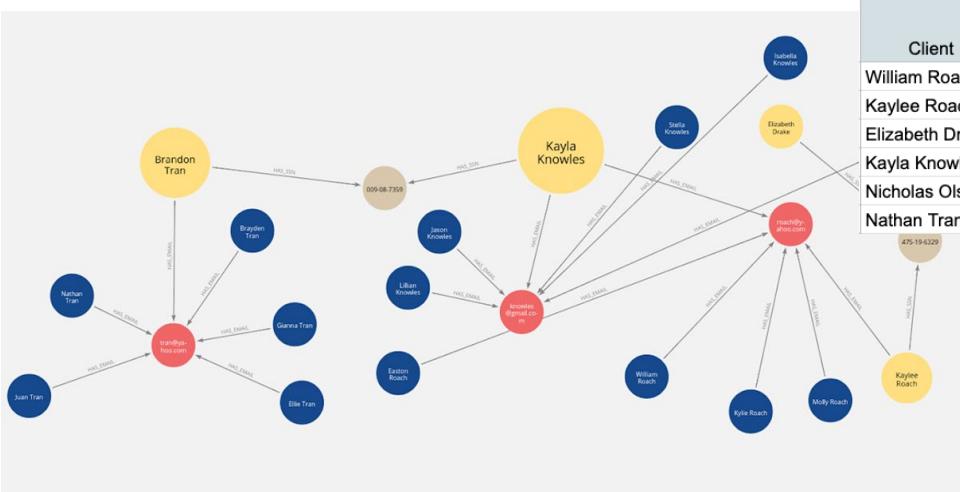


Evolution of Graph Data Science



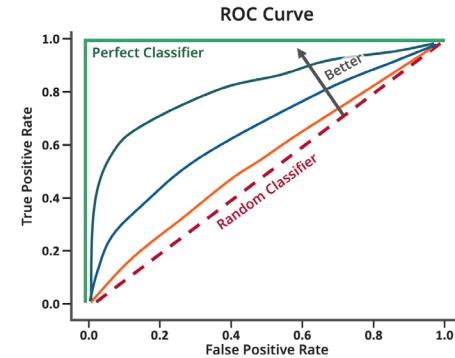
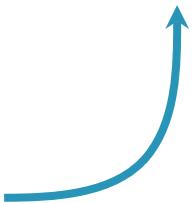
Graph Feature Engineering in Fraud

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.



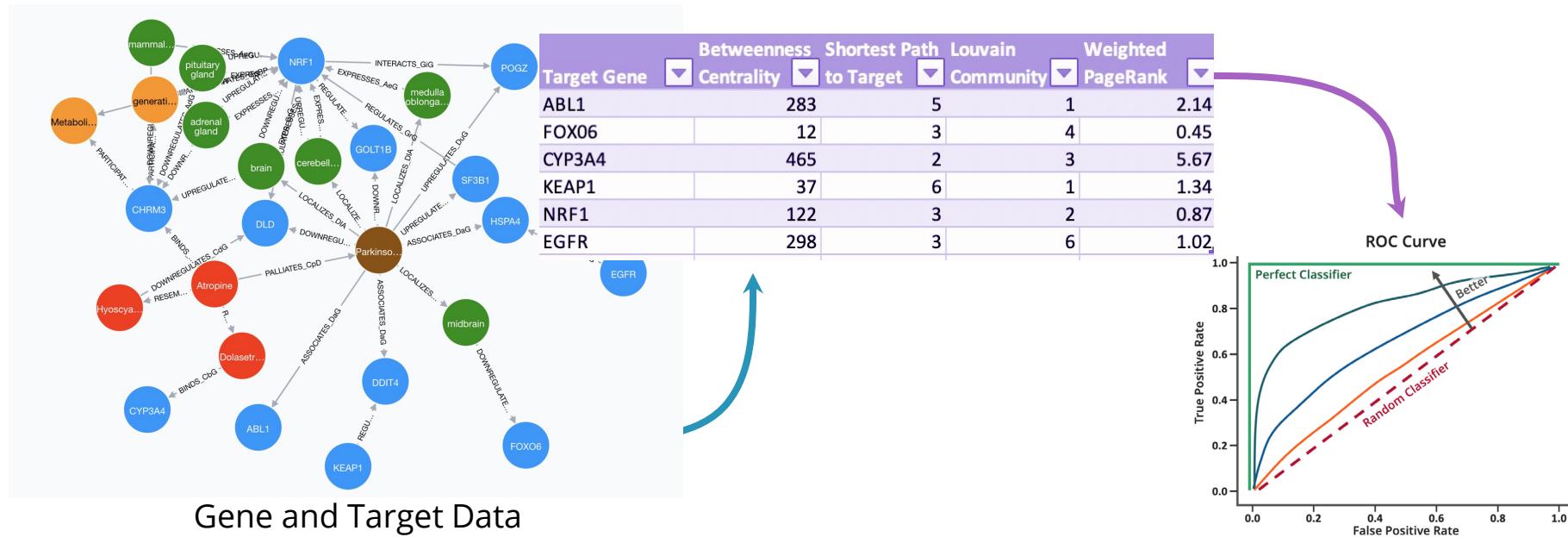
Financial Transaction Data

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



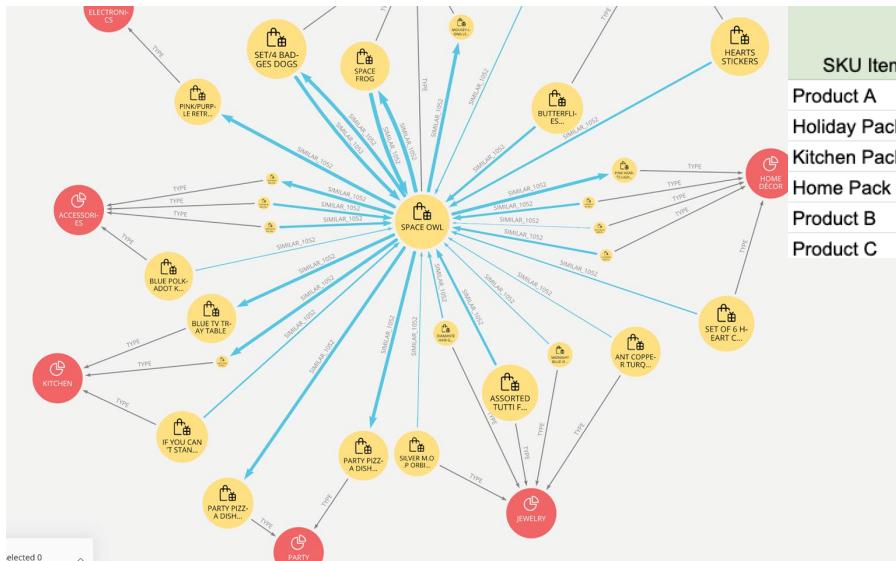
Graph Feature Engineering in Drug Discovery

Feature Engineering is how we combine and process the data to create *new, more meaningful features*. Using graphs we can base ML on genes with similar network topology.

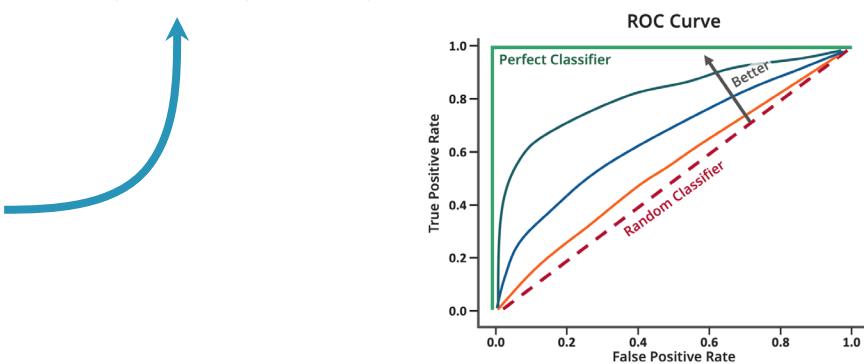


Graph Feature Engineering in Marketing

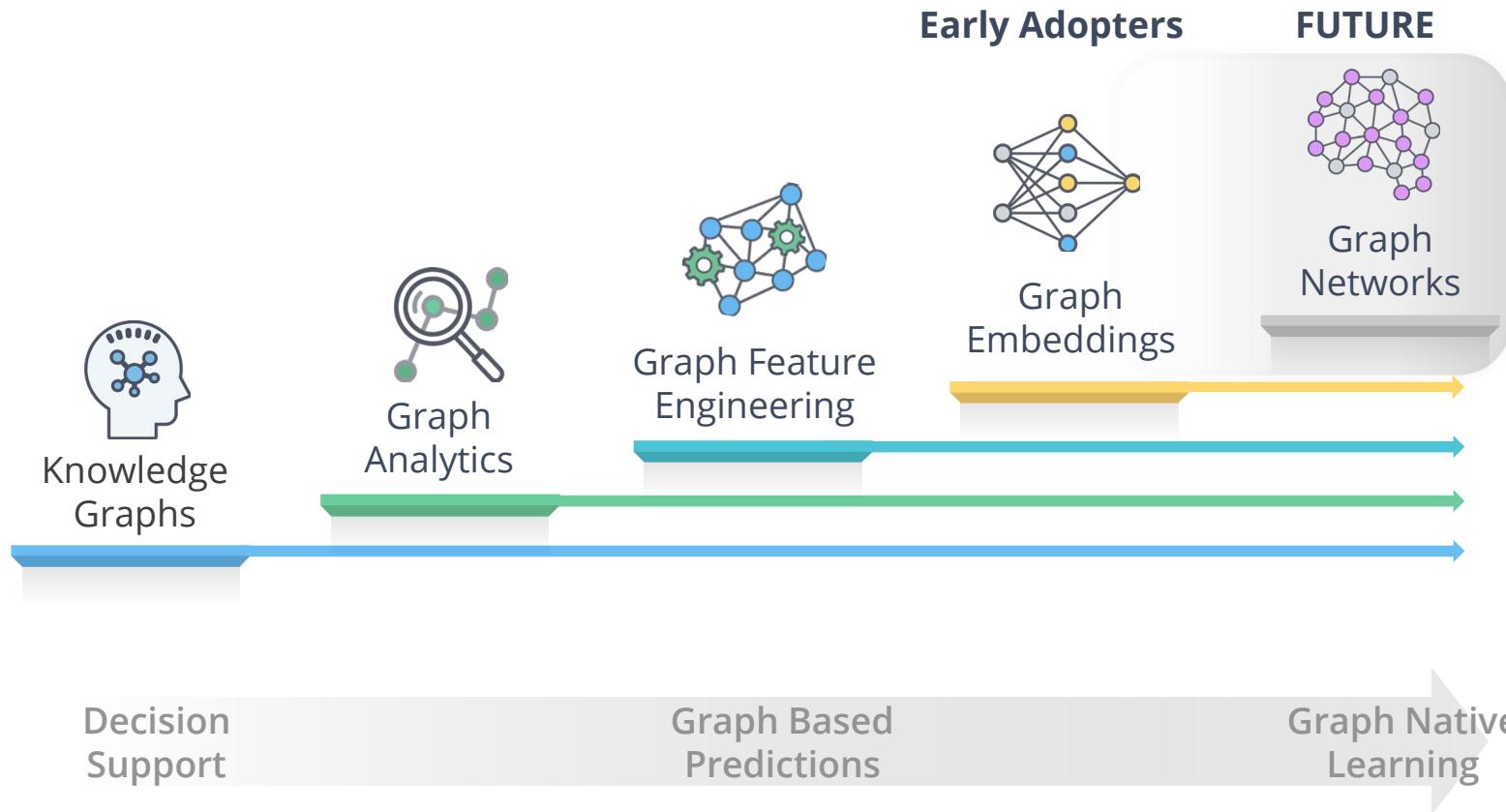
Feature Engineering is how we combine and process the data to create *new, more meaningful features*. Using graphs we can base ML on influential products by different groups.



SKU Item	Betweenness	Unique Shared Components	Weighted PageRank	Abandoned Cart?
Product A	0	1	1	No
Holiday Pack	32	2	4	Yes
Kitchen Pack	0	1	20	Yes
Home Pack	192	3	3	Yes
Product B	0	1	2	No
Product C	0	1	0	No



Evolution of Graph Data Science



The background of the slide features a complex, abstract network graph composed of numerous small, semi-transparent grey dots connected by thin grey lines. There are several larger, solid dark grey circles of varying sizes scattered throughout the graph, which appear to be hubs or specific nodes of interest. The overall effect is a sense of interconnectedness and data flow.

Neo4j Makes this Evolution Possible

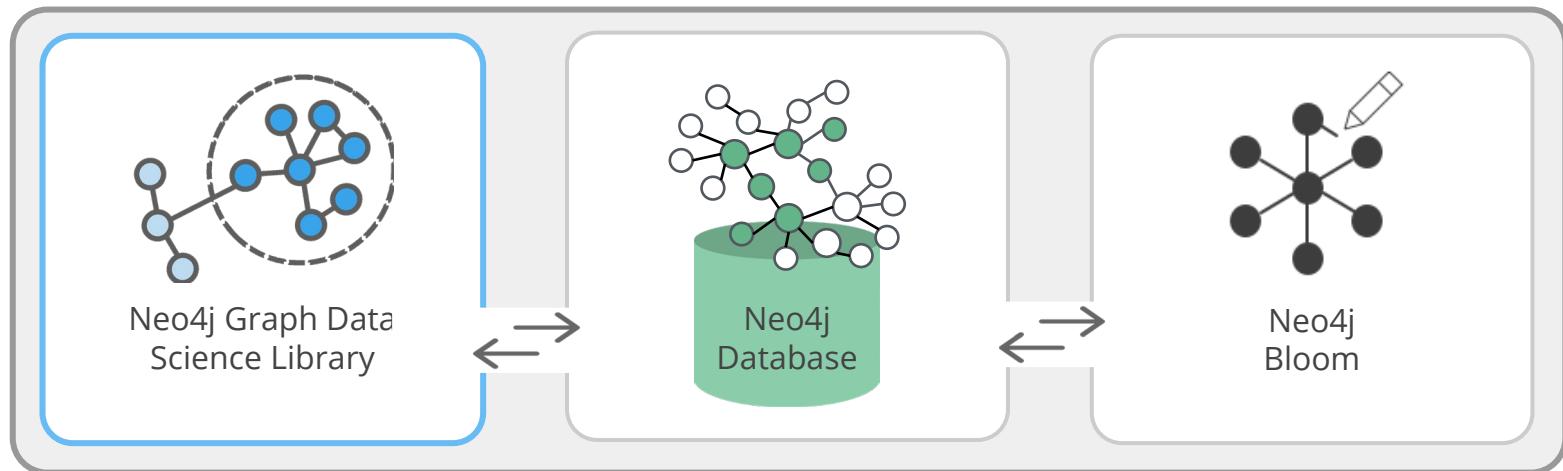


neo4j for Graph Data Science™

Scalable Graph
Algorithms & Analytics
Workspace

Native Graph
Creation & Persistence

Visual Graph
Exploration
& Prototyping

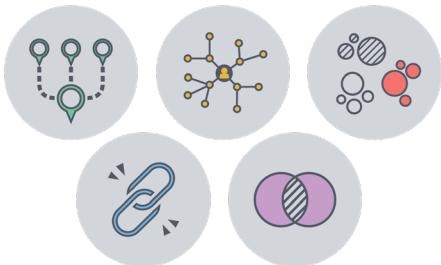


Practical

Integrated

Intuitive

The Neo4j GDS Library

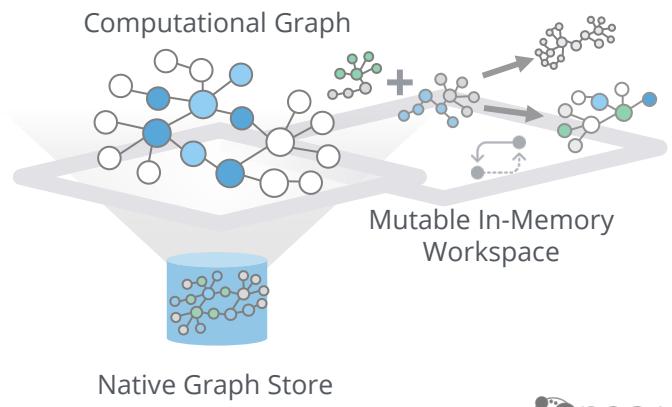


Robust Graph Algorithms

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

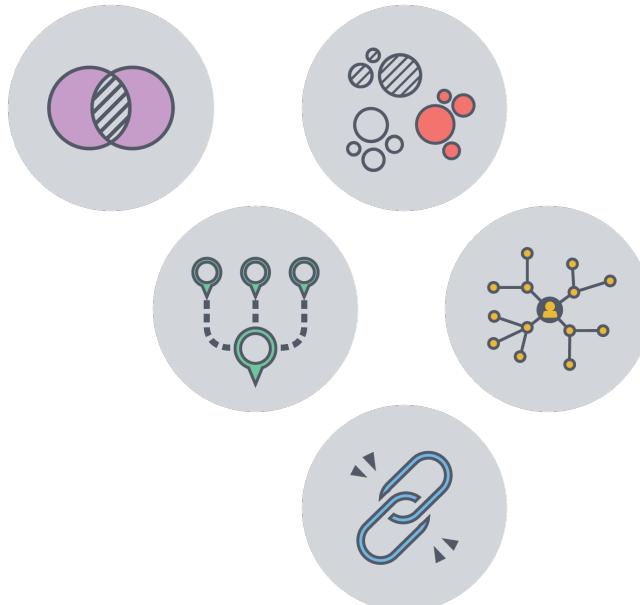
Efficient & Flexible Analytics Workspace

- 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**.

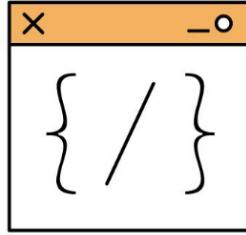


Answer previously intractable questions
with the data you already have

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

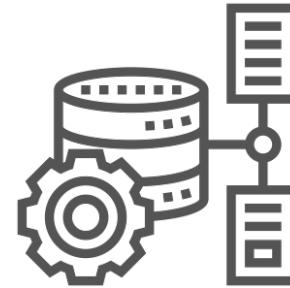
Massive scale to 10's billions of nodes with optimized
algorithms

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**

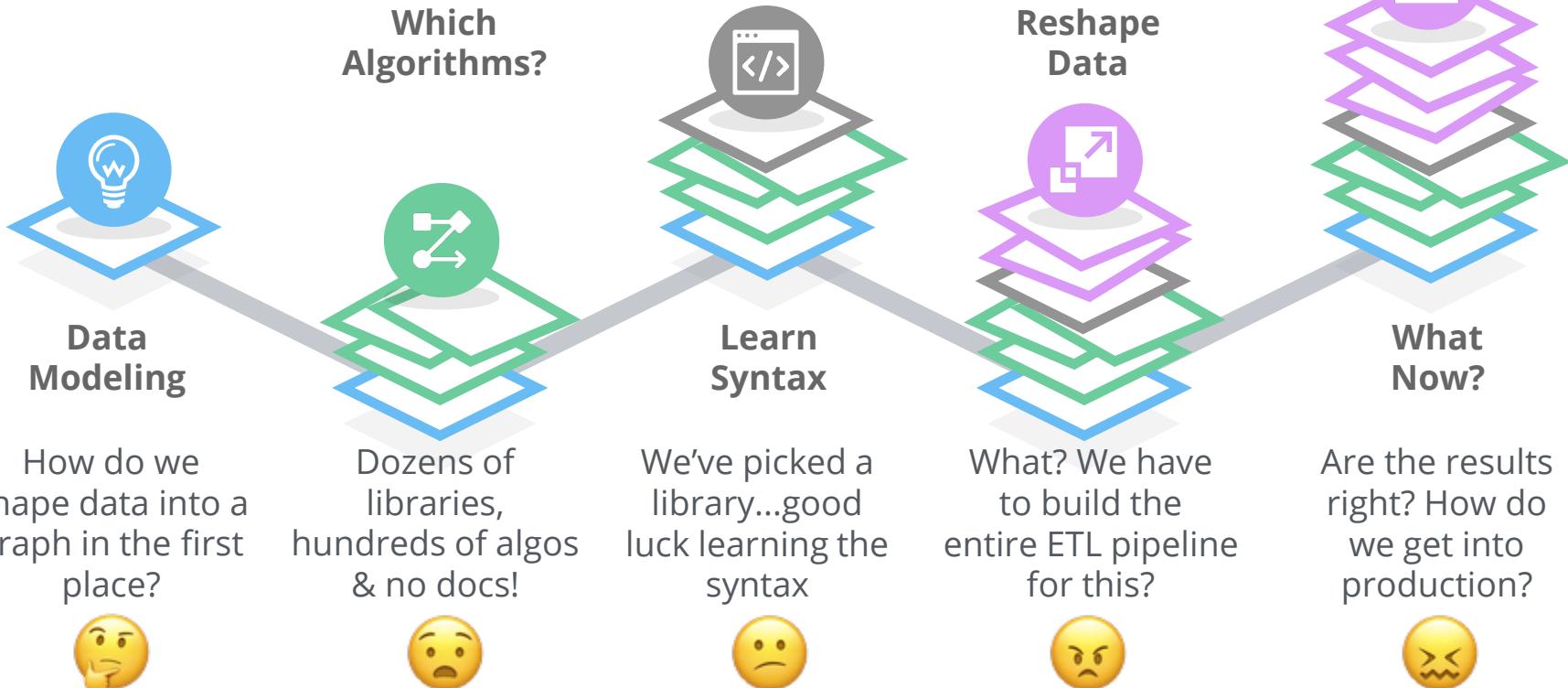
Reshaping, node & relationship aggregation / deduplication and multipartite algos



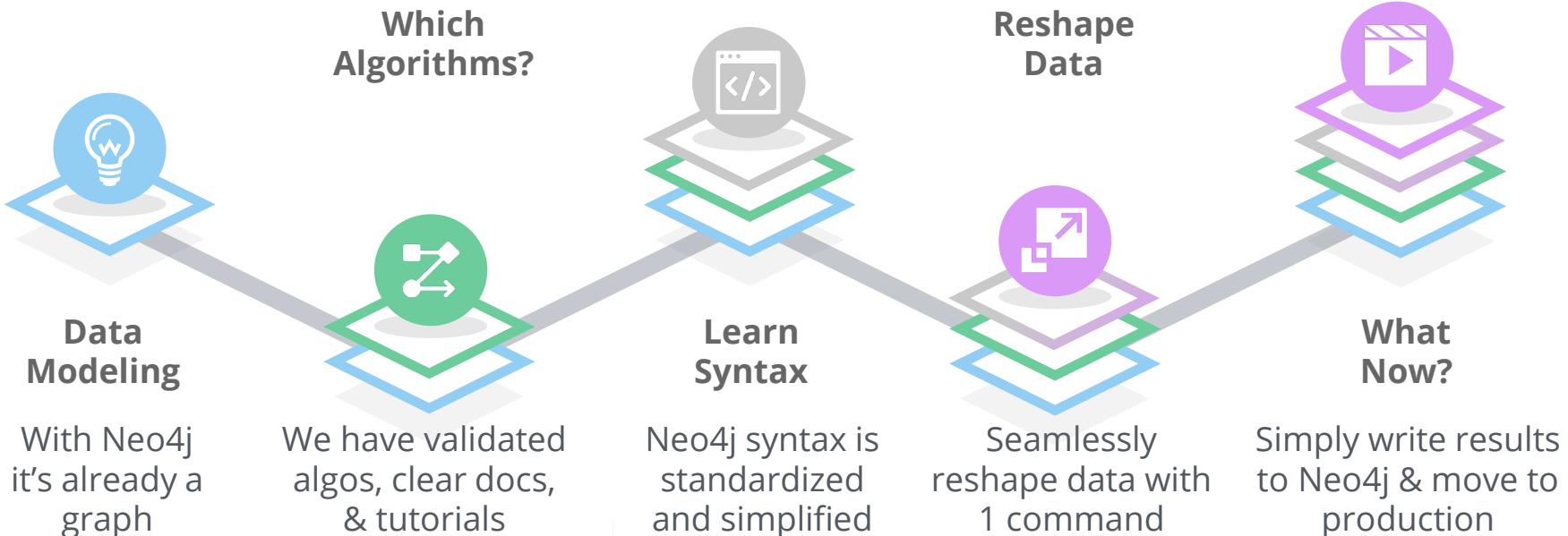
Explore graphs and algorithm results visually with **Bloom**

Share insights across teams for **better collaboration**

Data Science - It's Complicated



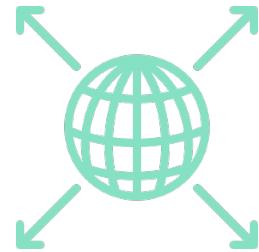
Simplify Your Experience



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