

# Graph Data Science

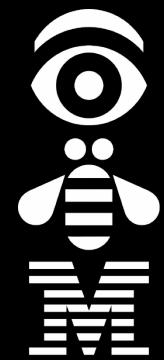
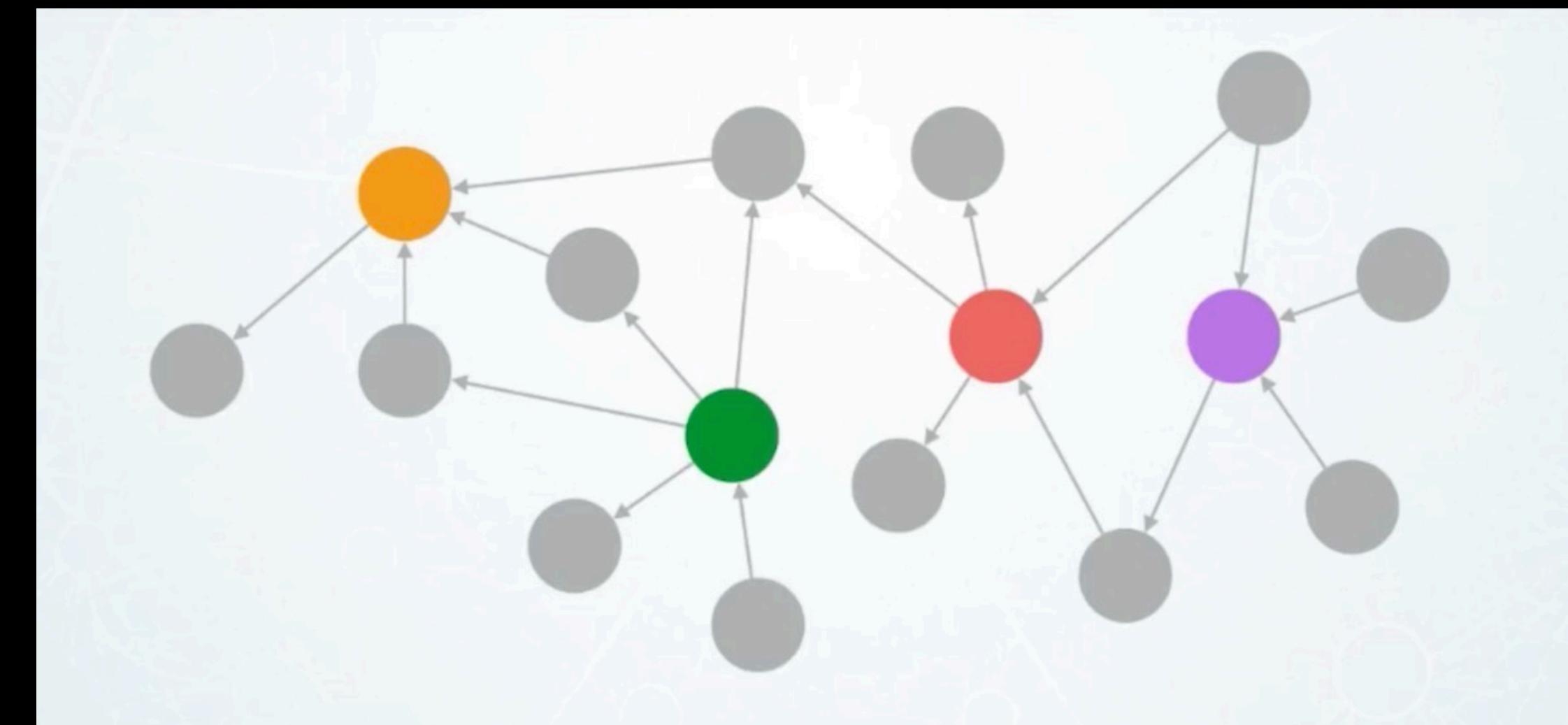
Ivan Portilla  
AI Leader  
Systems, TSS



ivanp@us.ibm.com

February 24 2022

# Graph Data Science with Neo4j Graph Algorithms



<https://github.com/jiportilla/gds-intro>

# Let me tell you a story...



Introducing Connie:  
A Collaboration Between Hilton and IBM Watson

<https://www.youtube.com/watch?v=ifgf6bZhxiE>

<https://medium.com/@sweetmantech/hilton-ibm-watson-7c5f5f1a611>

# Today's presentation

Graph Databases

Graph Data Science

Demonstration

# Graph Data Science Featured Services

<https://neo4j.com>

## Neo4j Graph Data Science Library

Used by data scientists and developers.

Enterprise graph analytics and graph-native machine learning at scale.

Graph algorithms are used to compute metrics for graphs, nodes, or relationships.

Neo4j Graph Data Science library uses a specialized in-memory graph format to represent the graph data.

<https://neo4j.com/product/graph-data-science/>

## Watson Studio

Build, run and manage AI models.

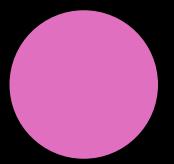
Prepare data and build models anywhere using open-source code or visual modeling.

## Watson NLU

Powerful advanced text analytics for your data

<https://www.ibm.com/cloud/watson-studio>

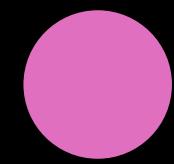
<https://www.ibm.com/cloud/watson-natural-language-understanding>



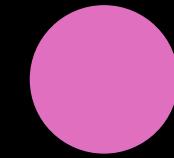
.

..

Graph Databases



Graph Data Science

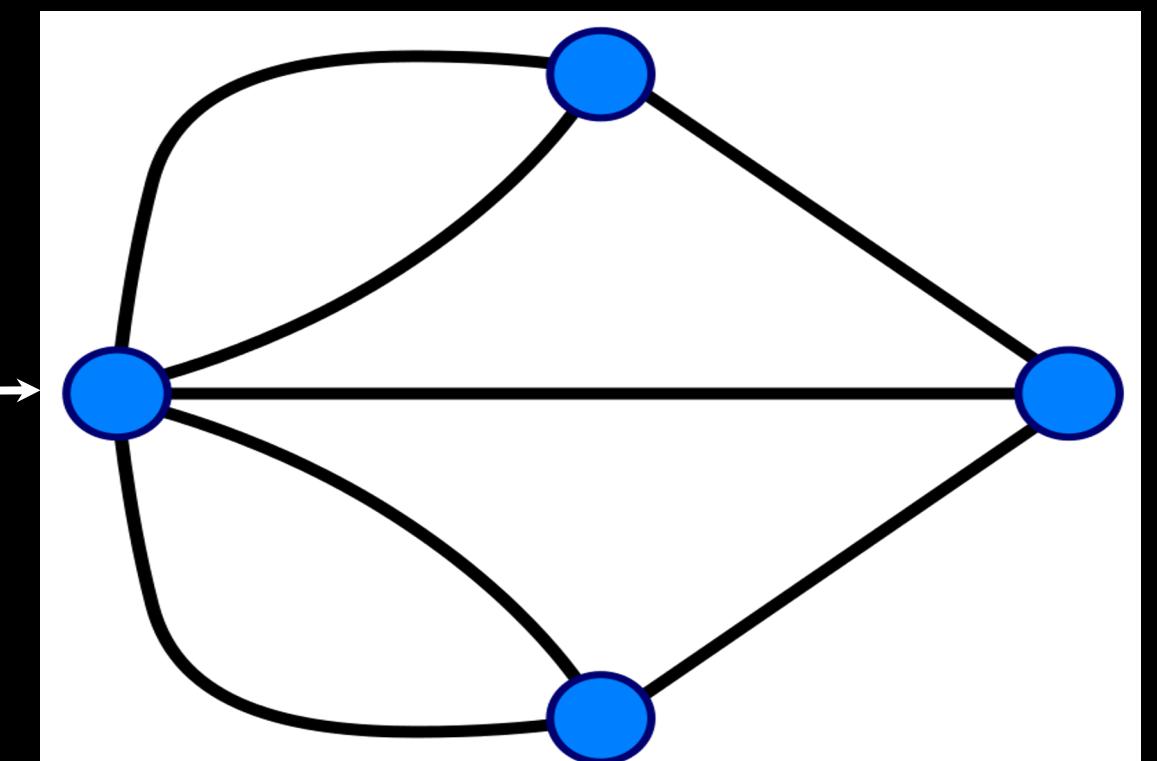
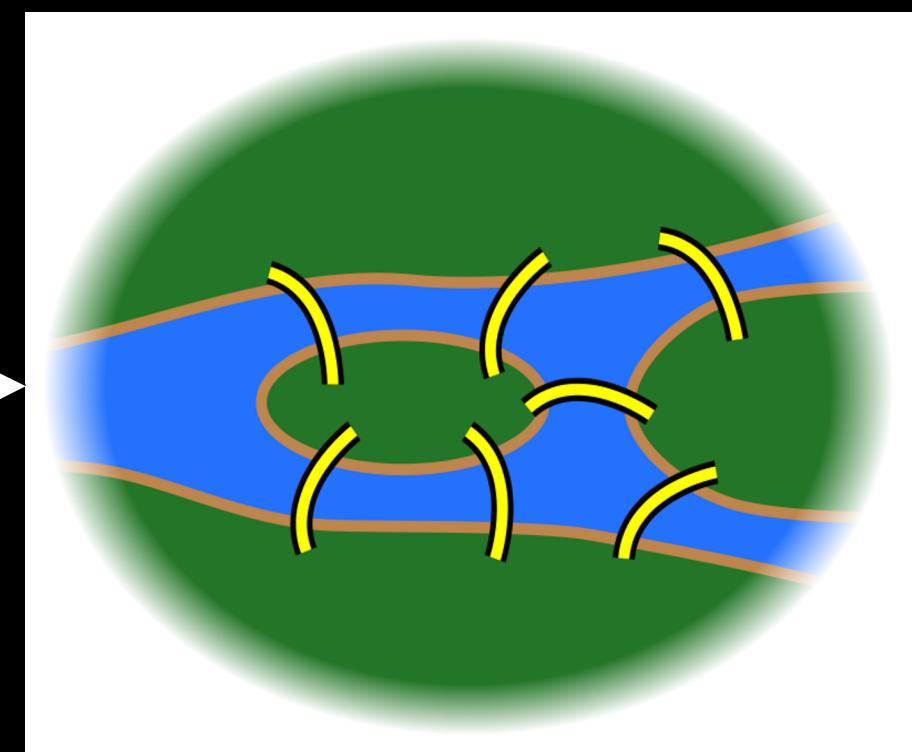
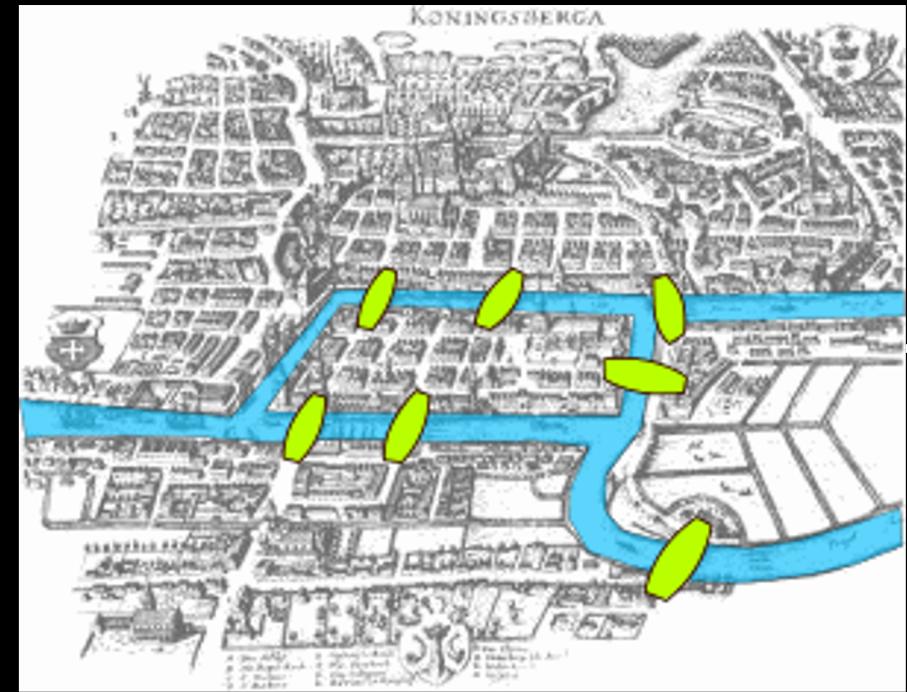


Demonstration

# What is a graph?

# A graph is ...

...a set of discrete objects, each of which has some set of relationships with the other objects

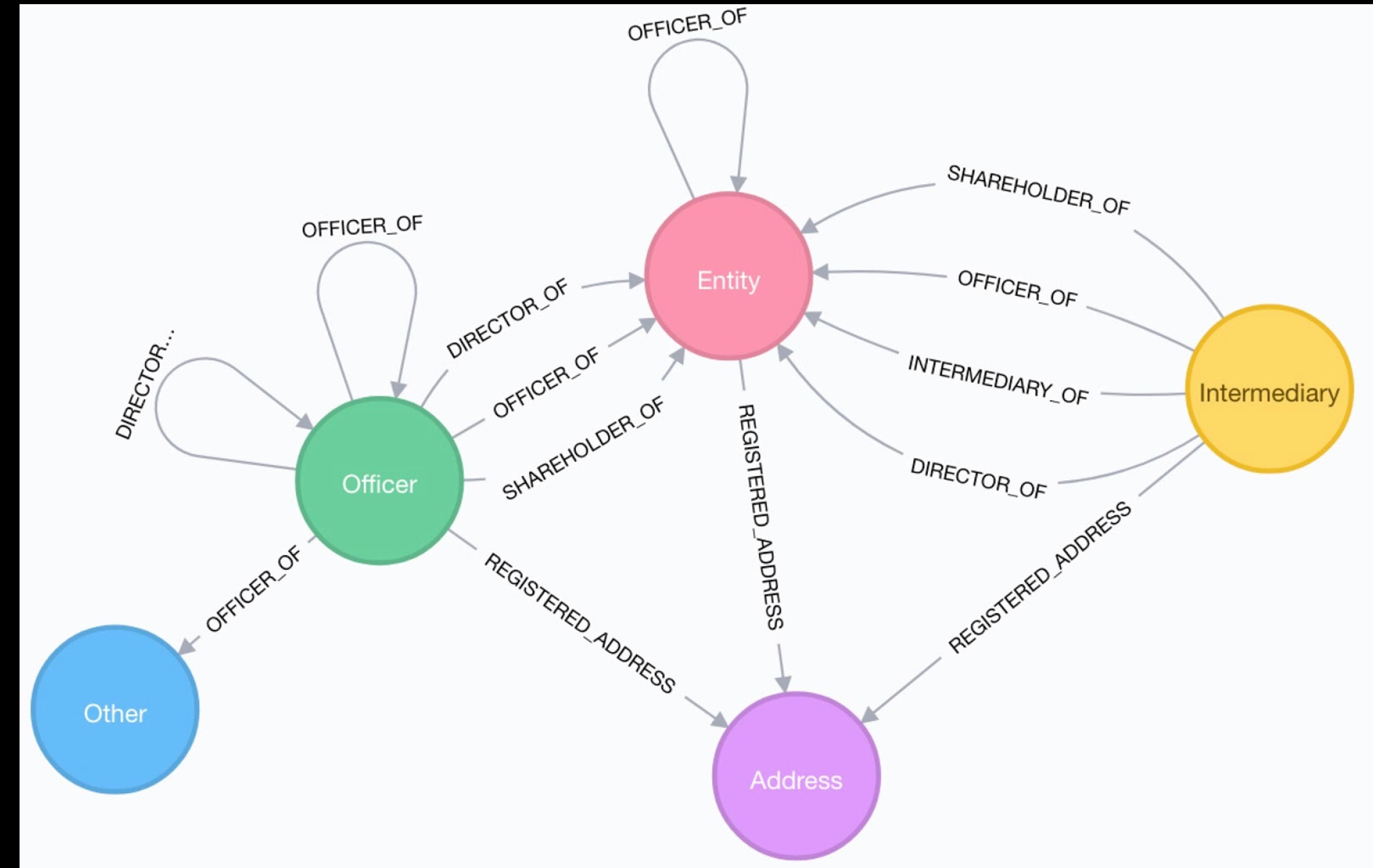


*Seven Bridges of Königsberg problem. Leonhard Euler, 1735*

[https://en.wikipedia.org/wiki/Seven\\_Bridges\\_of\\_Königsberg](https://en.wikipedia.org/wiki/Seven_Bridges_of_Königsberg)

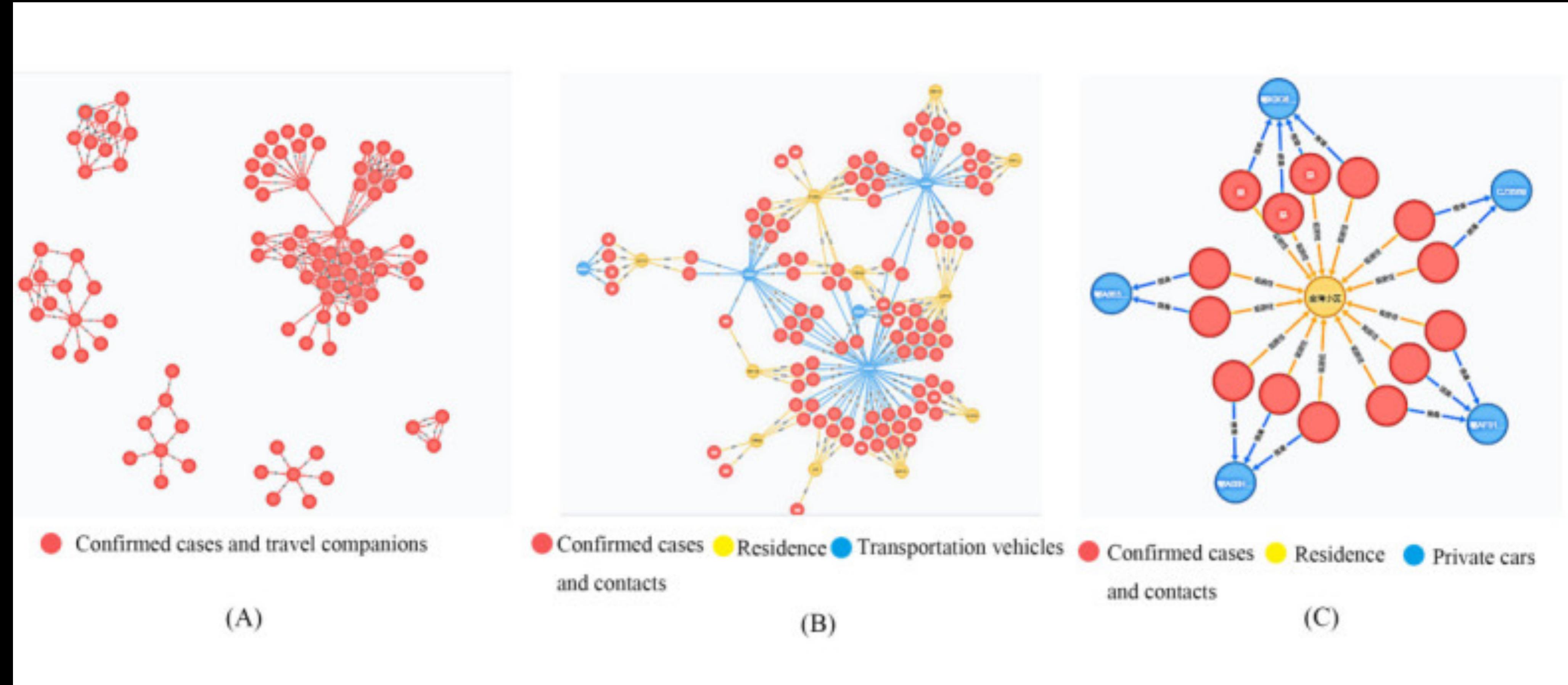
# Why do we want graphs?

The Pandora,  
Paradise,  
Panama papers  
model ...



# Why do we want graphs?

Covid-19  
contact  
tracing...



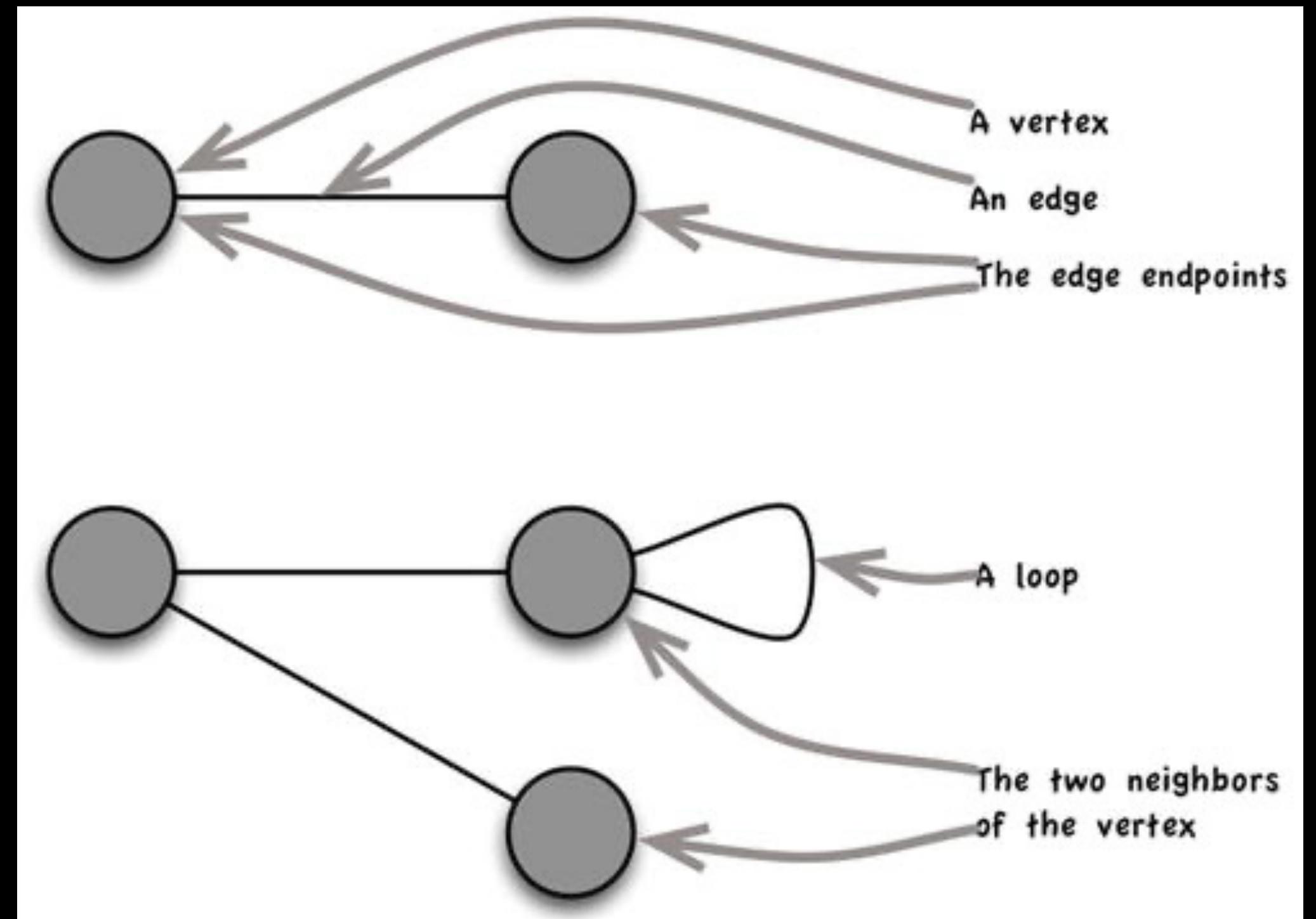
# Graph Components

## Node (Vertex)

- The main data element from which graphs are constructed

## Relationship (Edge)

- A link between two nodes. Has:
  - Direction
  - Type
- *A node without relationships is permitted.*
- *A relationship without nodes is not*





Graph Databases

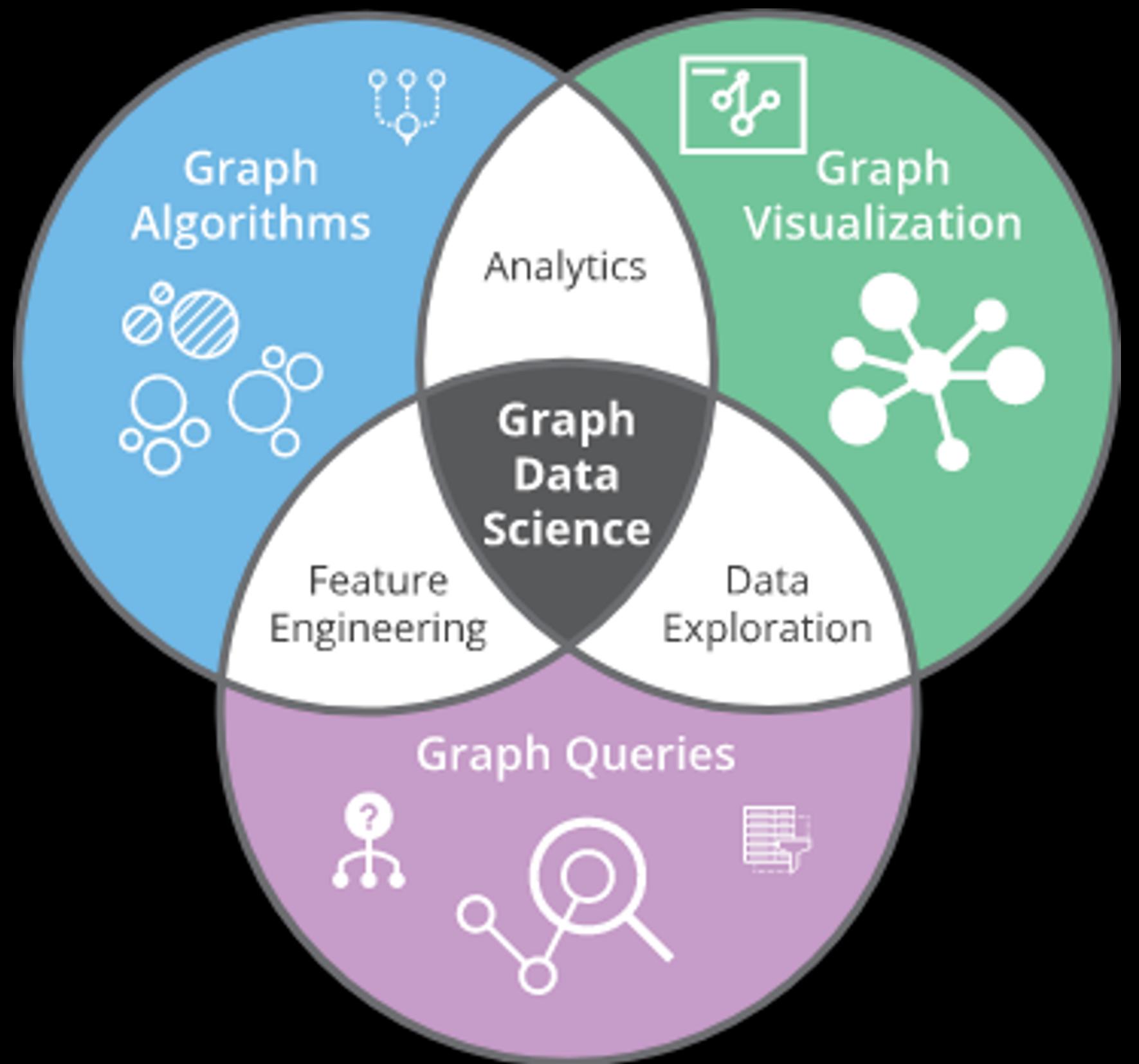
Graph Data Science

Demonstration

# Graph Data Science

- ✓ What is Graph Data Science
- ✓ What are Graph Algorithms
- ✓ The Neo4j Graph Data Science (GDS) Library
- ✓ GDS Library in Neo4j Browser
- ✓ Bloom with GDS for data visualization
- ✓ No code GDS with Graph Algorithms  
Playground (Neuler)

# 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.**

# What is *GRAPH* data science?

Data science when *relationships matter*

## Queries

(e.g. Cypher/Python)

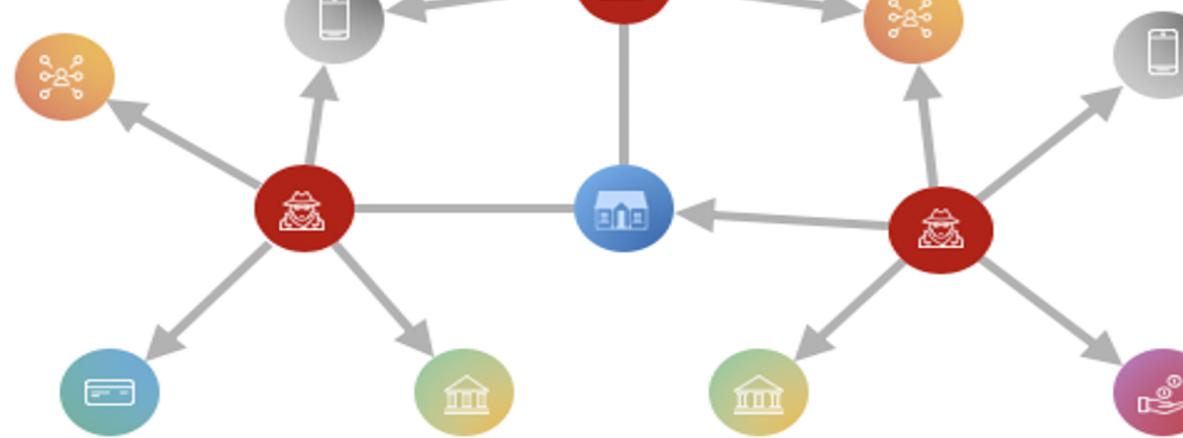
Fast, local decisioning  
and pattern matching

## Graph Algorithms

(e.g. Neo4j library, GraphX)

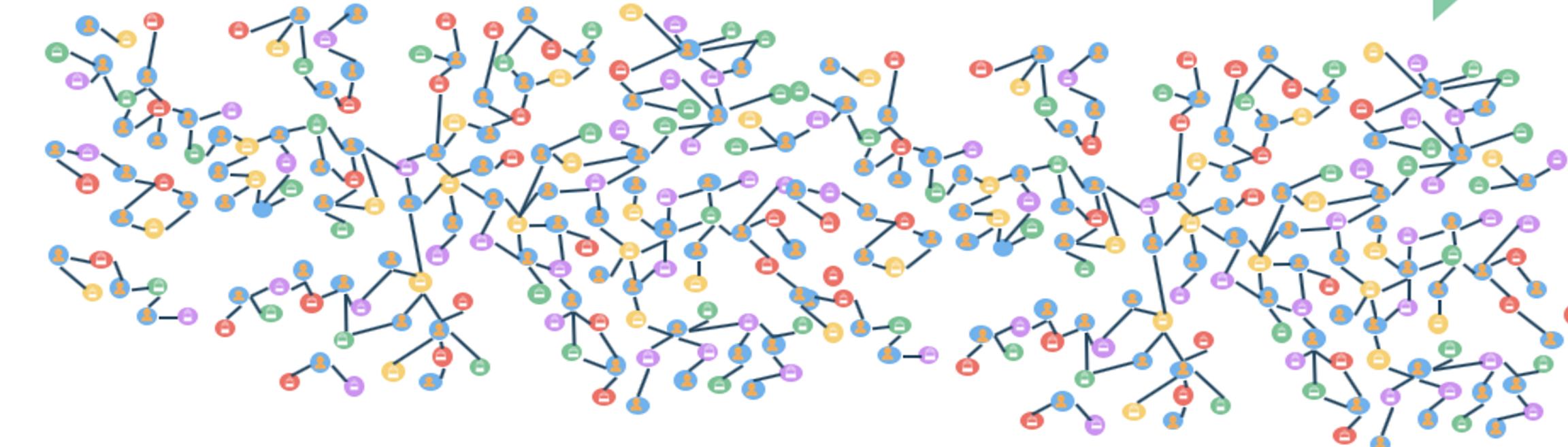
Global analysis  
and iterations

## Local Patterns



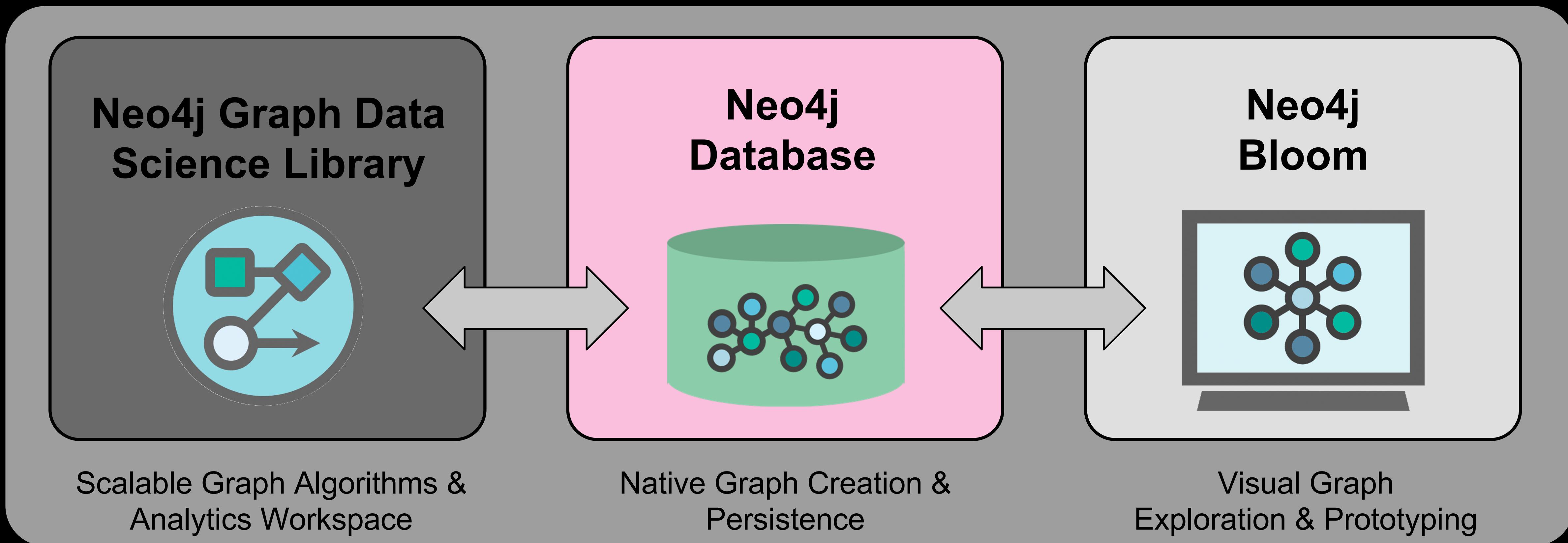
You know what you're  
looking for and  
making a decision

## Global Computation



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

# Neo4j's Graph Data Science Framework

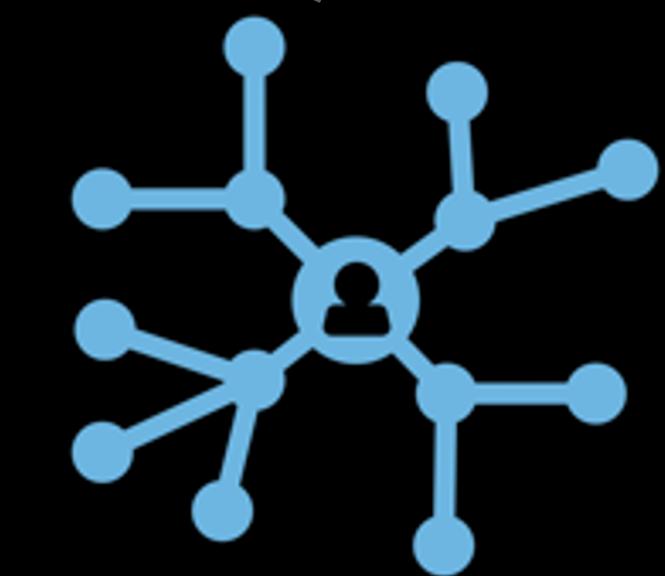


# Graph Algorithm types



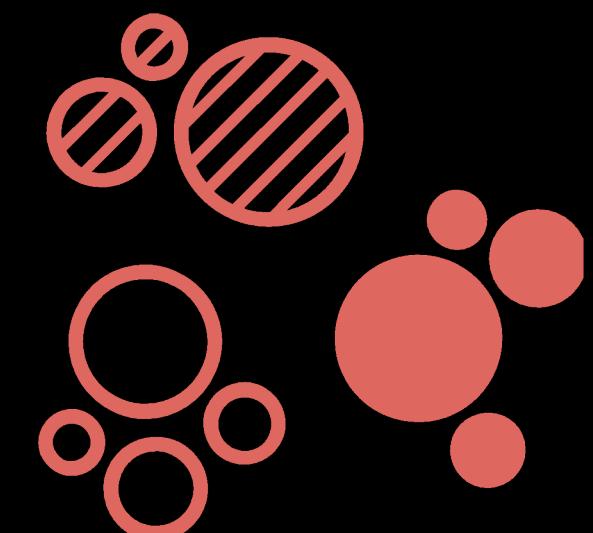
**Pathfinding  
and Search**

E.g. shortest  
path between  
two points



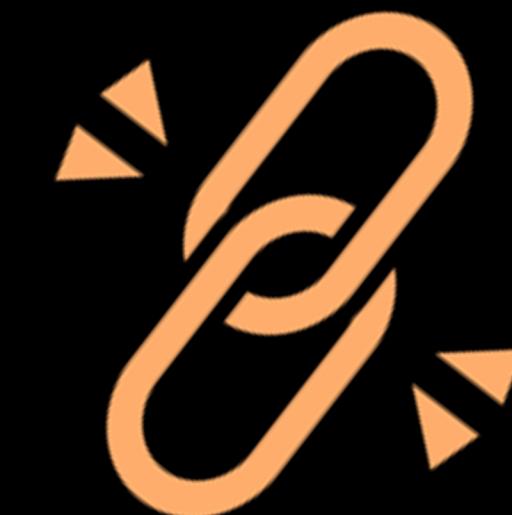
**Centrality  
(Importance)**

E.g. influencers  
in the network



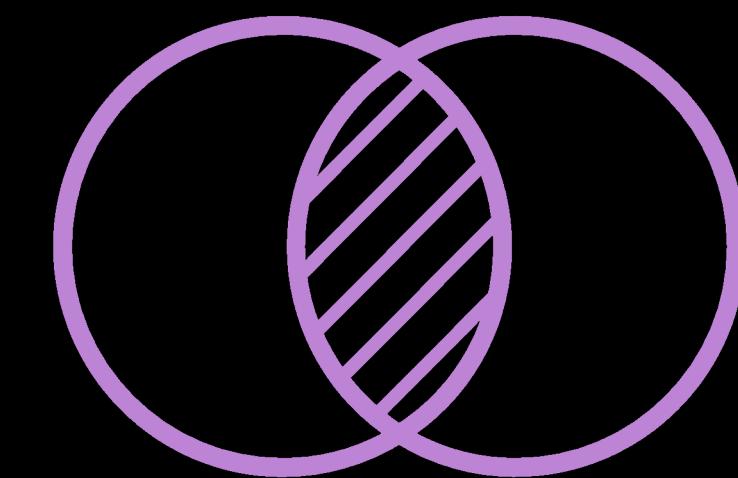
**Community  
Detection**

E.g. grouping  
similar entities  
together



**Heuristic  
Link Prediction**

E.g. hidden  
links between  
two points



**Similarity**

E.g. finding  
common  
entities based  
on graph  
properties

# GDSL - Business Use Cases

Customer classification  
based on purchases

**Node Similarity**

Influencers in an  
organization

**Centrality**

Fraud detection,  
Cybersecurity

**Community (Louvain)**

Root cause analysis

**Centrality (Pagerank)**

Minimum delay in  
telecommunication network

**Path finding**

Most Influential Health Care  
Providers in a Community?

**Community +  
Centrality**

Important routers in a  
telecom network

**Centrality**

# Where do I find the GDS tools?

The screenshot displays the Neo4j Graph Algorithms Playground interface, specifically the NEuler section. The main area shows a network graph visualization titled "Game of Thrones". The graph consists of numerous nodes (represented by colored circles) and edges (represented by lines). A large purple node is at the center, with many red and blue nodes branching out. To the left of the graph, there is a detailed node properties panel for "Daenerys Targaryen", listing properties like age (21), birth\_year (284), and community (332). Above the graph, a Louvain algorithm configuration panel is visible, showing options for "Caption for Officer", "Node Size", and "Node Color". The top navigation bar includes tabs for Louvain, Label Propagation, Connected Components, Strongly Connected Components, Triangles, and Triangle Count. On the far left, a sidebar provides access to the Graph Algorithms interface, showing the active database (Neo4j 3.5.15) and other sections like Files and Plugins (APOC).



**Graph Databases**

**Graph Data Science**

**Demonstration**

# Demonstration

Step 1. Create the ontology data

Step 2. Run Jupyter Notebook, create Neo4j db

Step 3. Visualize the results with NEuler & Bloom

# Demonstration

Knowledge Graphs: Ontology

Centrality: Pagerank

Community: Louvain

# Get started

1. Install Neo4j desktop  
(or cloud sandbox, docker container, OSS community edition)

<https://neo4j.com/product/developer-tools/>  
<https://neo4j.com/sandbox/>

2. Install GDSL plugin (& Bloom)  
<https://neo4j.com/download-center/>

3. Watch the tutorials  
<https://neo4j.com/videos/>  
<https://neo4j.com/video/bite-sized-neo4j-for-data-scientists/>

Q&A

Ivan Portilla

[ivanp@us.ibm.com](mailto:ivanp@us.ibm.com)

@iportilla

