
Neo4j

— 07: Intro Neo4j - IPSA - Hiver 2025 —
10/03/2025 - 2h

Last time on NoSQL

- Projets
- Embeddings
- Vector search
- Weaviate
- RAG

Cette semaine

Graph databases, Graphs networks, Graph knowledge

- Usages et avantages des graphs databases
- Cypher le langage de query de Neo4j
- Graph networks et graph data science
 - mesures de centralités
 - algorithmes de shortest path
 - détection de communauté, d'anomalies
- Knowledge graphs - LLMs

Cette semaine : Neo4j

Aujourd'hui Lundi

- Intro graph database
- Pratique sur Aura

Et Jeudi

- Intro graph database
- Pratique sur Aura

Et Vendredi

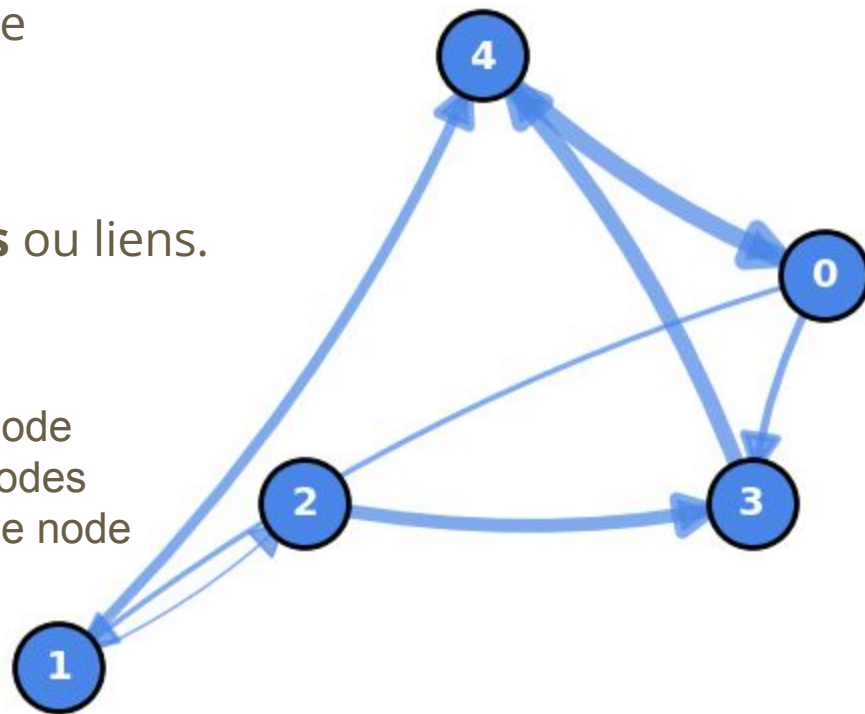
- Graph data science

graph networks

Graphes

Un graphe est une structure composée de

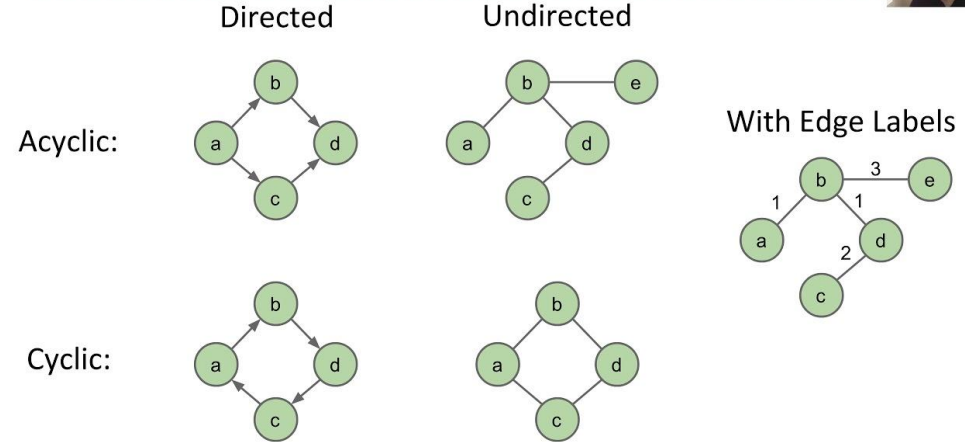
- **nœuds** ou sommets.
 - Nodes
- les noeuds sont reliés par des **arêtes** ou liens.
 - Edges, vertices
- **Degree**: Number of edges connected to a node
- **Path**: Sequence of edges connecting two nodes
- **Cycle**: Path that starts and ends at the same node



acyclic, cyclic, directed graphs

- **Orienté / Directed** graph :
direction explicite entre les noeuds :
 - parent -> enfant,
- **acyclique** : impossible de revenir en arrière (DAG)
 - $A \rightarrow B \rightarrow C$
- **cyclique** : il existe au moins un cycle qui permette de revenir en arrière:
 - $A \rightarrow B \rightarrow C \rightarrow A$

Graph Types



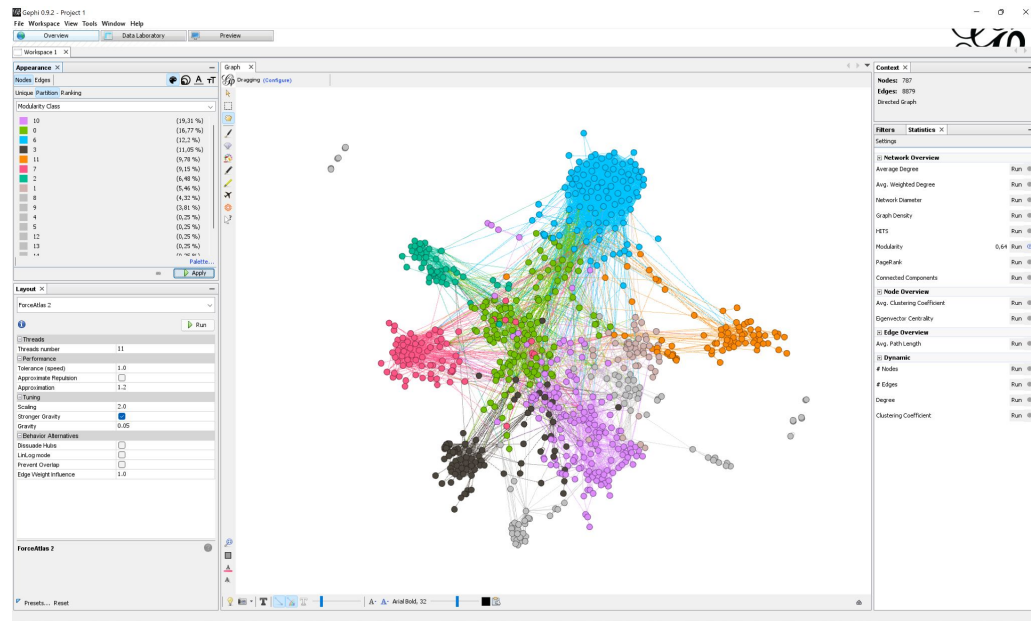
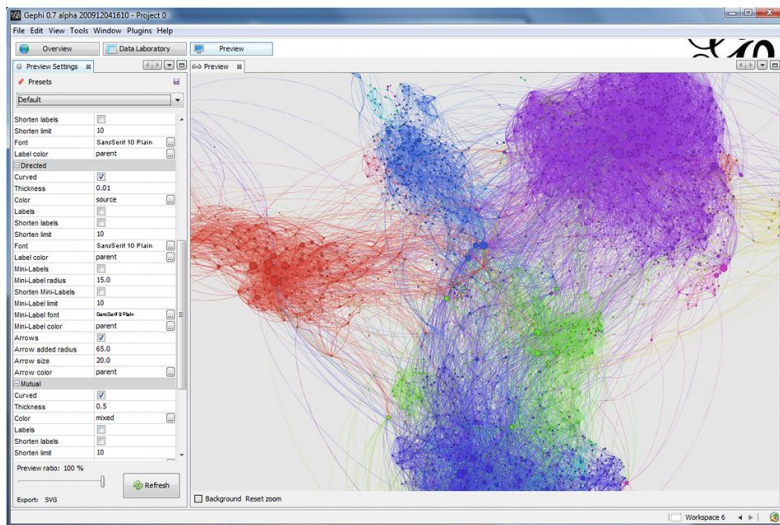
<https://youtu.be/anRVomtGXFc?t=204>

Cas d'usage

- Analyse de réseaux sociaux : Étudier les relations entre individus.
- Réseaux de transport : Optimiser les itinéraires et les flux de trafic.
- Biologie : Modéliser les interactions entre protéines.
- Informatique : Représenter les dépendances entre composants logiciels.
- Sociologie numérique: Comportements, modélisation
 - Scientométrie : evolution de la science, références croisées

Outils d'analyse : Gephi

- Gephi : Logiciel interactif pour visualiser et explorer des graphes.
- Mathieu Jacomy - <https://gephi.org/> - <https://gephi.wordpress.com/>



Outils d'analyse : networkx

- NetworkX : Bibliothèque Python pour la création, la manipulation et l'étude de graphes.

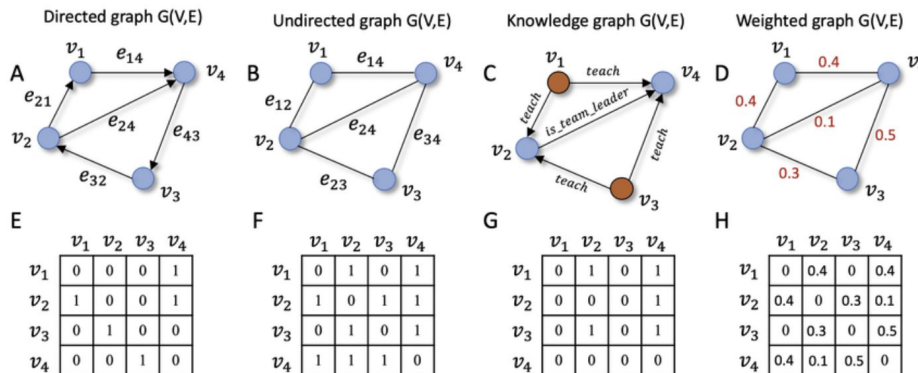


<https://www.toptal.com/data-science/graph-data-science-python-networkx>

Représentations des données

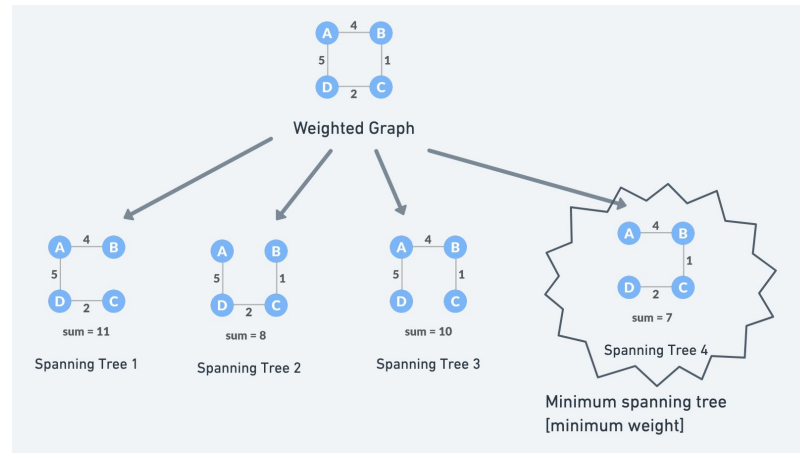
- **Adjacency Matrix:** $N \times N$ matrix where element (i,j) indicates connection
- **Adjacency List:** Each node stores a list of its connected neighbors
- **Edge List:** Simple list of all node pairs that form edges
- **Incidence Matrix:** Rows represent nodes, columns represent edges

Adjacency Matrix



Algorithmes d'analyse

- **Traversal**: BFS (Breadth-First Search), DFS (Depth-First Search),
- **Shortest Path / Plus Court Chemin** : Dijkstra, Bellman-Ford, A*
- **Minimum Spanning Tree** : Kruskal, Prim, [lien](#), [wikipedia](#)
- **Mesures de Centralité** : Degré, Intermédierité, Proximité, Vecteur Propre
- **Détection de Communautés** : Recherche de clusters ou groupes de nœuds



Neo4j

Take a deep breath!



SQL vs MongoDB vs Neo4j

Neo4j est avant tout une **Graph Database**.

- **PostgreSQL**: performance, optimisation et normalisation des données.
- **MongoDB**: flexibilité du schéma, pipelines d'agrégation et la performance pour des gros volumes.

Les **Graph databases** comme Neo4j apportent une dimension d'analyse des données modélisées sous forme de graphes

- graph datascience & knowledge graphs

Graph data science

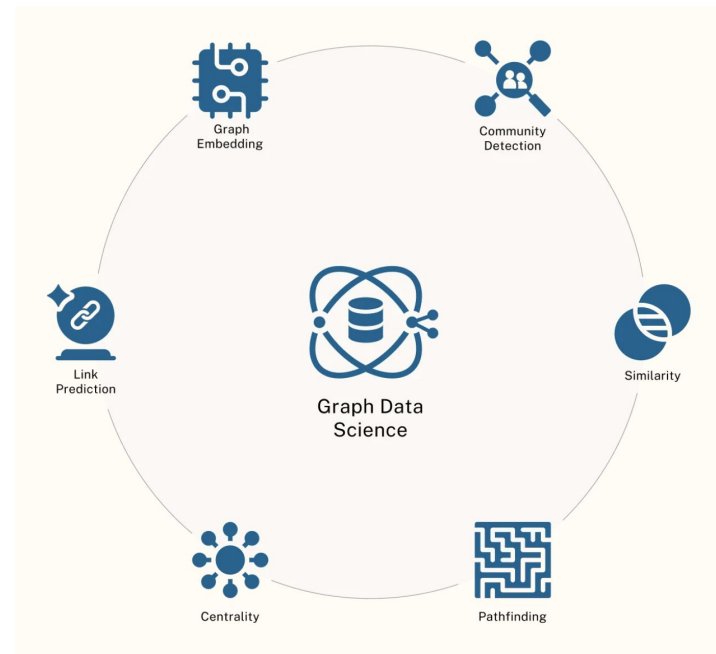
Analyser des données complexes modélisées sous formes de graphes.

- **algorithmes** d'identification de patterns, de communautés ou d'anomalies
- **Étude** des relations entre les entités (nœuds) et leurs connexions (arêtes) pour extraire des informations.
 - analyse de réseaux,
 - détection de communautés,
 - prédiction de liens
- **Applications** : réseaux sociaux, bioinformatique, recommandation de produits....

Graph data science

- Centrality Algorithms: Find important/influential nodes in network
- Community Detection: Identify groups of densely connected nodes
- Path Finding: Discover optimal routes between nodes
- Similarity Algorithms: Calculate how similar nodes/communities are
- Machine Learning: Node embeddings and predictions
- Node Embedding: Convert graph structure to vector space
- Link Prediction: Predict future/missing relationships

<https://neo4j.com/use-cases/knowledge-graph/>



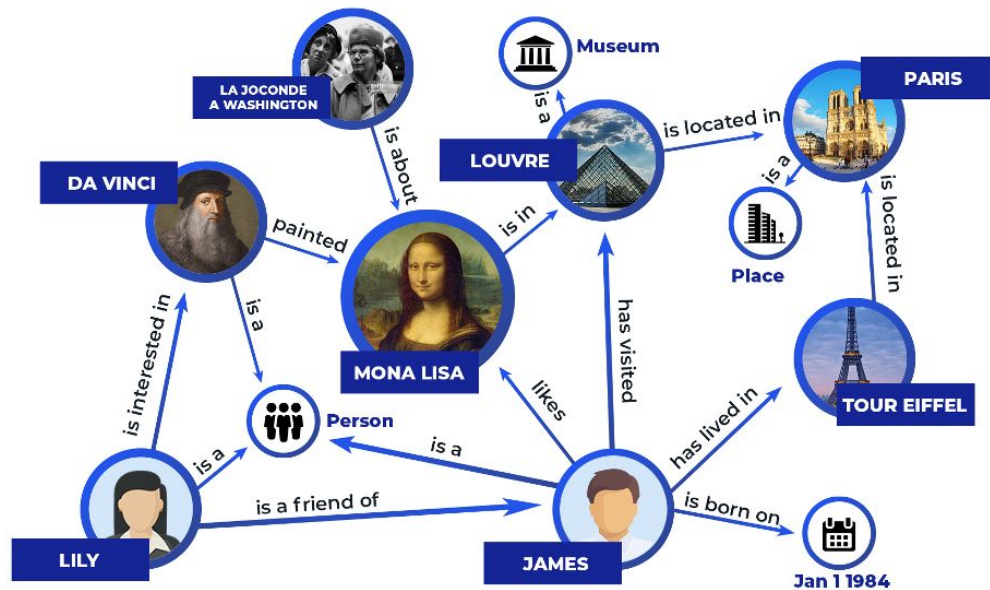
Knowledge Graphs

Représentation structurée de faits et de relations entre des entités.

- Nœuds (Entités) : Personnes, lieux, objets, concepts.
- Arêtes (Relations) : Comment les entités sont connectées
 - Paris est la capitale de la France

Pourquoi utiliser des graphes de connaissances ?

- recherche
- raisonnement
- inférence



Google Knowledge Graphs

Il y a une API!

<https://developers.google.com/knowledge-graph>

AllImagesNewsVideosShort videosForumsWebMore

Tools

Sponsored

Neo4j

<https://www.neo4j.com>

Neo4j Graph Database & Analytics

Learn how graph databases are powering fraud detection, personalization and more. Gain holistic visibility, accurate predictions, and data-driven insights.

Neo4j Pricing

Compare pricing and features for Neo4j Graph Database products.

Start Free with AuraDB

The world's leading graph database. Start free, no credit card needed.

Get Started Free

Sign up for AuraDB Free with no credit card required.

Graph Database Demos

Register now for weekly live demos with Neo4j experts. Live Q&A.

Case Studies

View customer success stories and use cases from leading companies.

Neo4j

<https://neo4j.com>

Neo4j Graph Database & Analytics | Graph Database Management ...

Uncover Hidden Patterns With Graph. In a connected world, your data's true value lies in its relationships. Neo4j stores your data as a network of connected ...

Download Neo4j Desktop

Download Neo4j Desktop · Get Started Now · Experience Neo4j ...





What is a graph database

What is a graph database. A Neo4j graph database stores data as ...

Documentation

Neo4j

System software







Neo4j is a graph database management system developed by Neo4j Inc. The data elements Neo4j stores are nodes, edges connecting them, and attributes of nodes and edges. [Wikipedia](#)

Programming language: [Java](#)

Initial release date: 2007

License: Source code: [GPLv3](#) and [AGPLv3](#) with Commons Clause; Binaries: [Freemium](#) registerware

People also search for



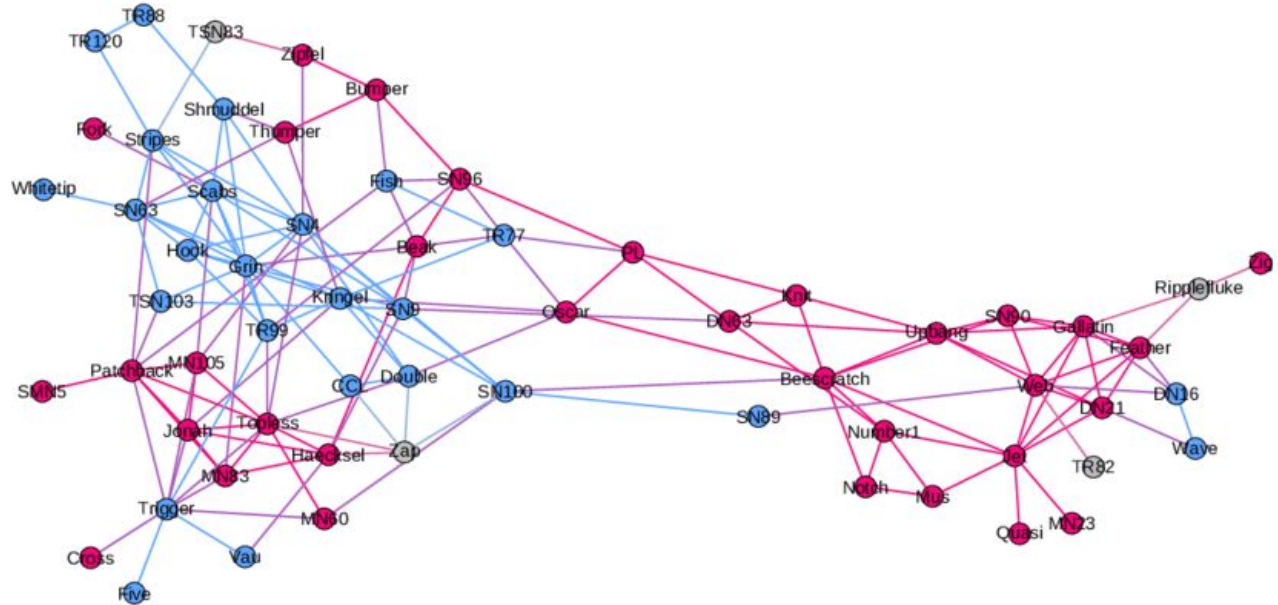
MongoDBArangoDBRedisPostgreSQL

<https://www.google.com/search?q=Neo4j>

19

Graph

Nodes & Edges



A node-edge diagram of the **dolphin social network**.
Node labels are the names researchers assigned to the individual dolphins.
Node color indicates male (pink), female (blue) or unknown (gray).

from [Algorithms and Software for the Analysis of Large Complex Networks](#)

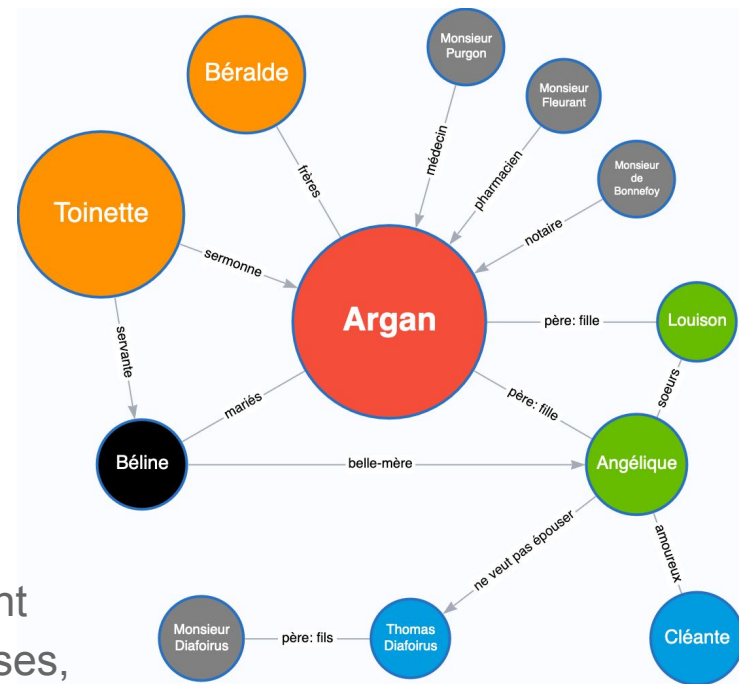
Graphs dans Neo4J: weights, properties, labels

Les noeuds et les relations (Aretes) peuvent avoir

- des poids en fonction de leur importance relative
- des propriétés
- des étiquettes / label

Les relations sont des **verbes** ou des **locutions verbales** qui décrivent l'action, la connexion entre les nœuds.

Elles définissent la signification sémantique de la façon dont les nœuds interagissent. Comme les verbes dans les phrases, ils doivent être clairs et spécifiques sur la nature de la relation.



[Moliere.love](https://moliere.love/)

Research Field: Co-citation of Cited
Translational Research Triangle (CFE)

- **laboratory research** -
vegetal model organism
- **clinical research** - central
role played by application
to solid cancer
- **instrumental research** -
bioinformatics and
biostatistics as necessary
tools

Research Field: Co-citation of Cited Journal
Translational Research Triangle (CER corpus)

Exemple métro parisien

Noeuds

- étiquettes / labels: Metro / RER
- propriétés:
 - nom
 - accessibilité,
 - aérien ou sous terre
 - propriétés événementielles

Relations

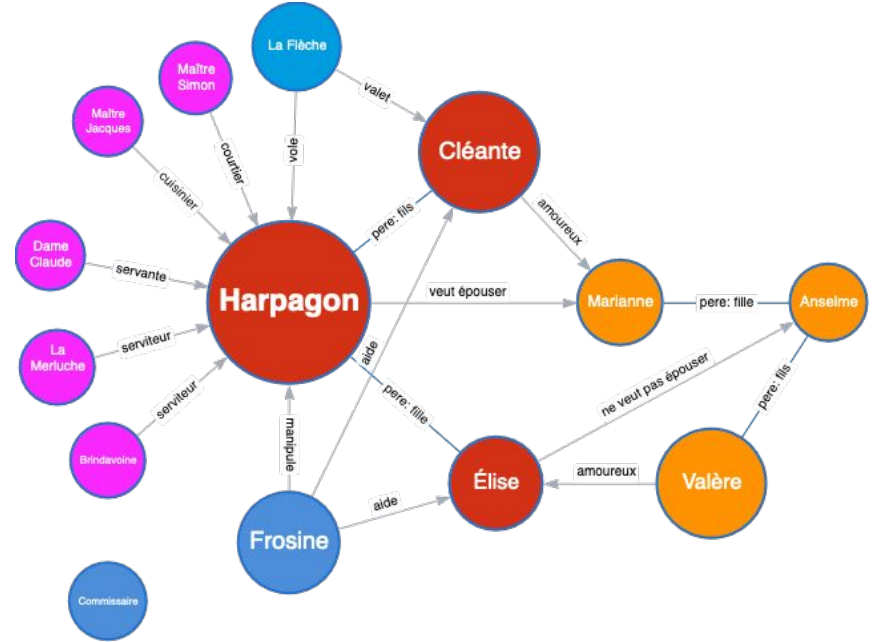
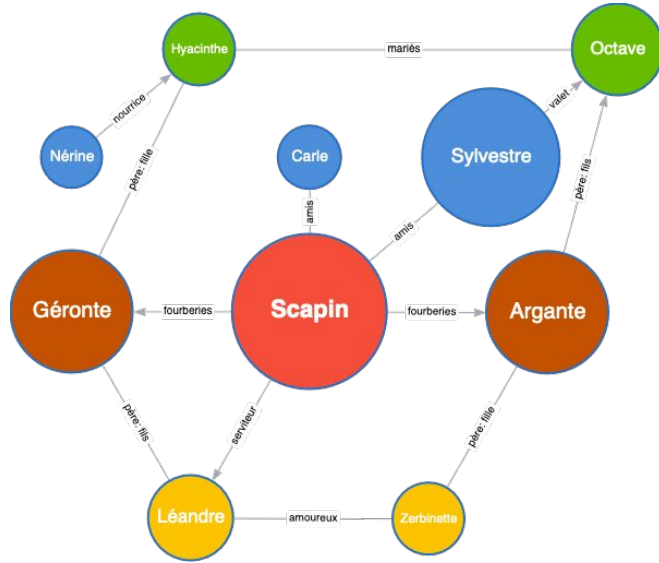
- BELONGS_TO_LINE
 - properties: line number, maintenance time, distance
- PASSENGER_CONNECTS_TO
 - can be unidirectional or both ways
 - properties: distance, stairs, ...
- CONNECTS_TO_LINE
 - properties: line number, ...

Relationship **types** (BELONGS_TO_LINE, PASSENGER_CONNECTS_TO, CONNECTS_TO_LINE) are **verbs** or **verb phrases** that describe the action, the connection between nodes.

They define the semantic meaning of how nodes interact.

Like verbs in sentences, they should be clear and specific about the nature of the relationship.

Théâtre



from [Moliere.love](https://moliere.love)

Neo4J : Property Graph Model or KG

What Makes a Graph in Neo4j a "Knowledge Graph" vs. Just a Graph?

1. Semantics & Ontology

- If your Neo4j graph follows a **formal ontology** (e.g., "Person" → "won" → "Award"), then it is a proper **knowledge graph**.
- If it's just a graph without clear entity relationships, it's a **general graph**.

2. Querying for Facts & Reasoning

- If you use Neo4j to **infer new facts** (e.g., "If Einstein won the Nobel Prize, he must have been a scientist"), it's closer to a **knowledge graph**.
- If you use Neo4j for network analysis (e.g., shortest path between nodes), it's just a **graph database application**.

So, Is Every Graph in Neo4j a Knowledge Graph?

👉 Yes, if it represents structured knowledge and relationships.

👉 No, if it's just a general graph for things like social networks, logistics, or recommendation systems.

In short:

- Neo4j CAN store and query knowledge graphs.
- But it can also be used for non-KG applications.

Use cases

<https://neo4j.com/graphgists/>
<https://neo4j.com/use-cases/>

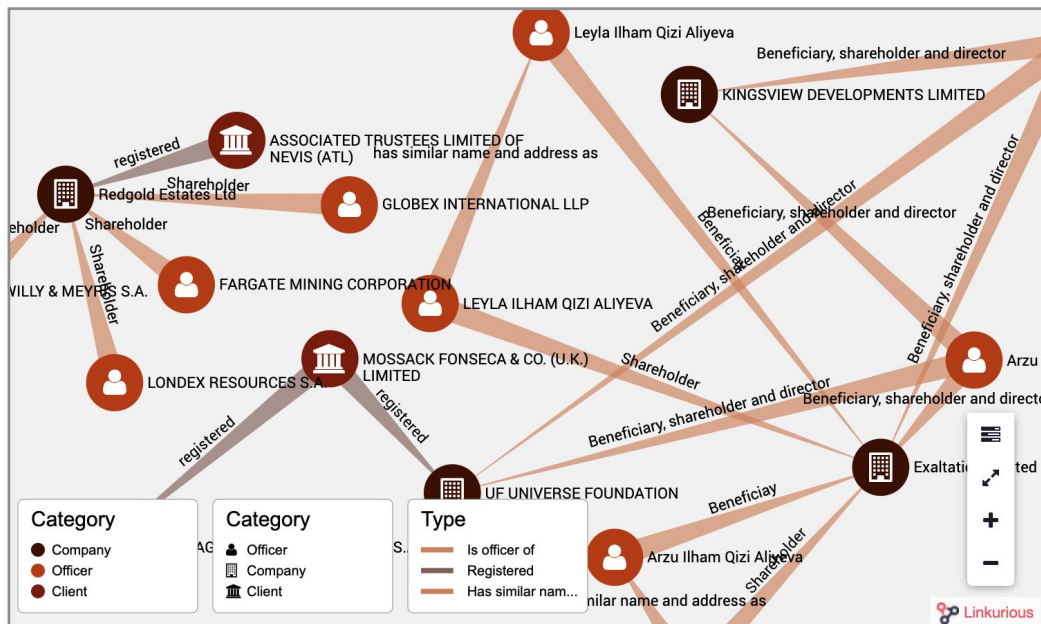
Investigative journalism - The Panama Papers

What Happened?

- **Leak:** 11.5 million documents (2.6 TB) from Mossack Fonseca.
- **Revealed:** Offshore companies used for **tax evasion, money laundering, and corruption.**
- **Investigated by:** ICIJ + 400+ **journalists** across **80+ countries.**

Why Use Neo4j?

- **Traditional databases failed** due to complex relationships.
- **Graph database = better for uncovering hidden networks.**



Panama Papers

Graph-Based Investigation

- **Nodes:** People, companies, addresses, banks.
- **Edges:** Relationships like *"owns"*, *"directs"*, *"registered at"*.

Key Techniques

- **Pattern discovery:** Found secret owners of offshore firms.
- **Cypher queries:** Traced hidden financial flows.
- **Visualization:** Used [Linkurious](#) for easy exploration.

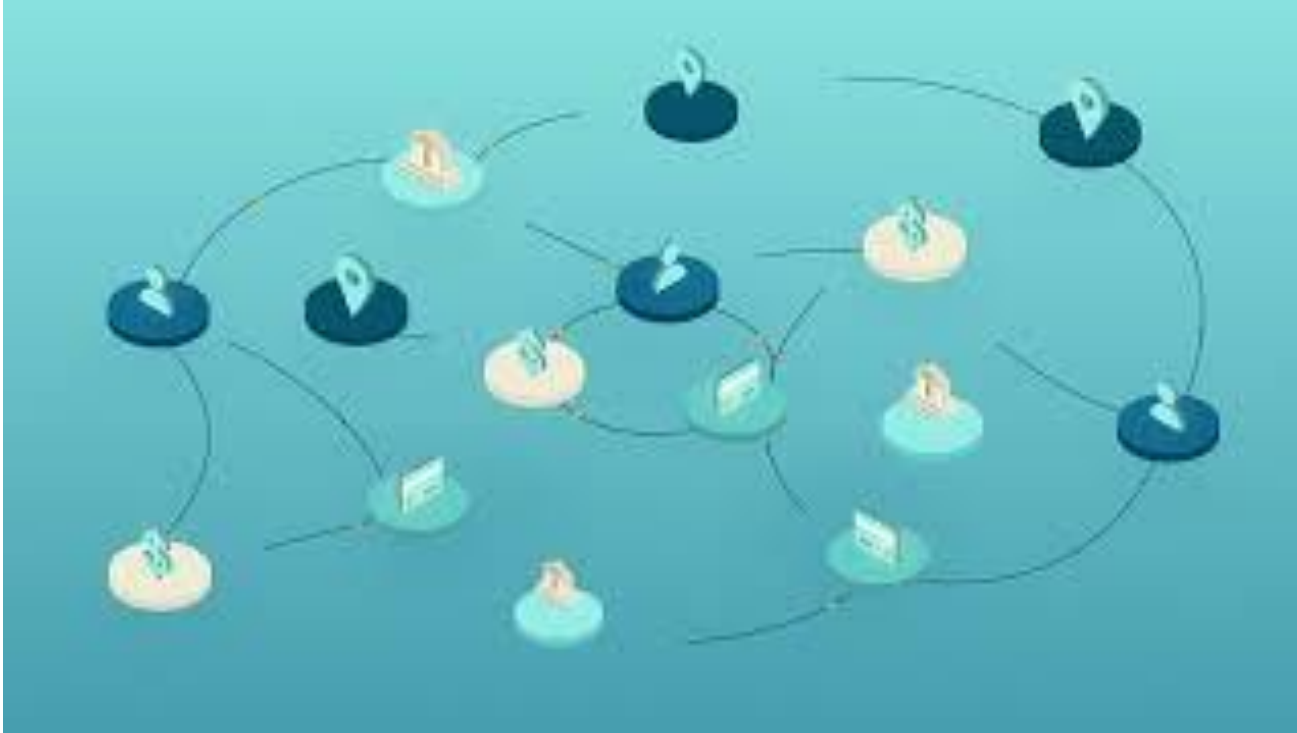
Major Discoveries

- ✓ **143 politicians** & world leaders involved.
- ✓ **Banks & law firms enabled** tax havens.
- ✓ **Complex money trails exposed.**

Global Impact

- 🔥 **\$1.36B recovered** in unpaid taxes.
- 🔥 **Iceland's PM resigned**, global investigations launched.
- 🔥 Led to **Paradise Papers & Pandora Papers.**

Fraud detection



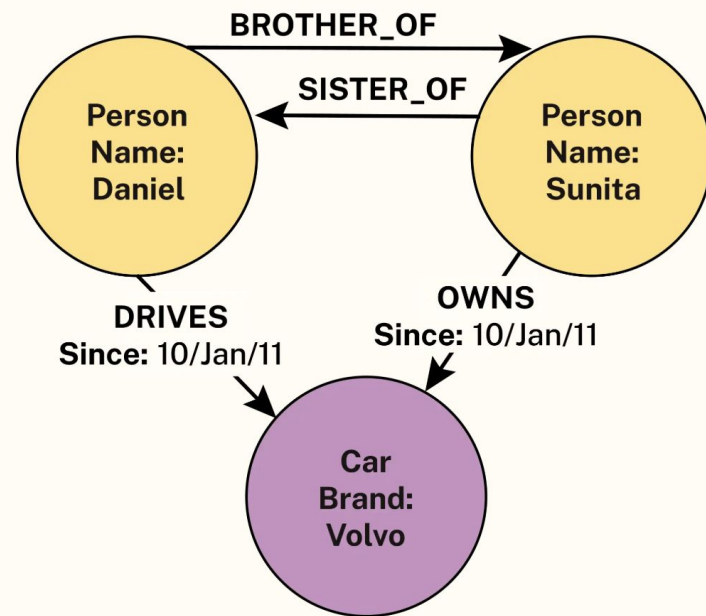
Characteristics

Property Graph Model

pas besoin de transformer les objets *naturels* en tables, foreign keys, ...

Les relations s'expriment naturellement

<https://neo4j.com/use-cases/knowledge-graph/>



Property graph model of siblings with a car matches the conceptual model

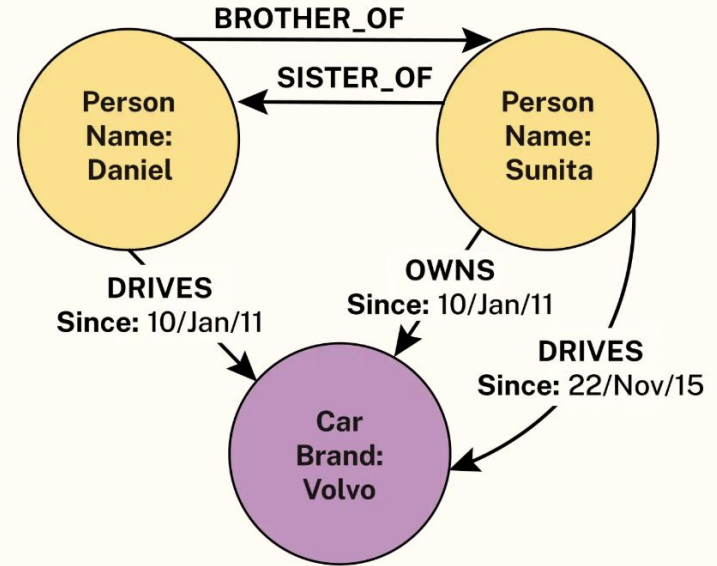
Schema flexible

On retrouve le **schéma flexible** de MongoDB

avec en plus une facilité accrue pour modifier les relations entre les objets.

Dans cet exemple on transforme une relation **one to many** en une relation **many to many** en ne créant qu'une relation supplémentaire

Les attributs de noeuds et relations peuvent aussi évoluer sans contraintes



Changing from one -> many to many->many relationships
doesn't require redesign or recoding

Index Free Adjacency

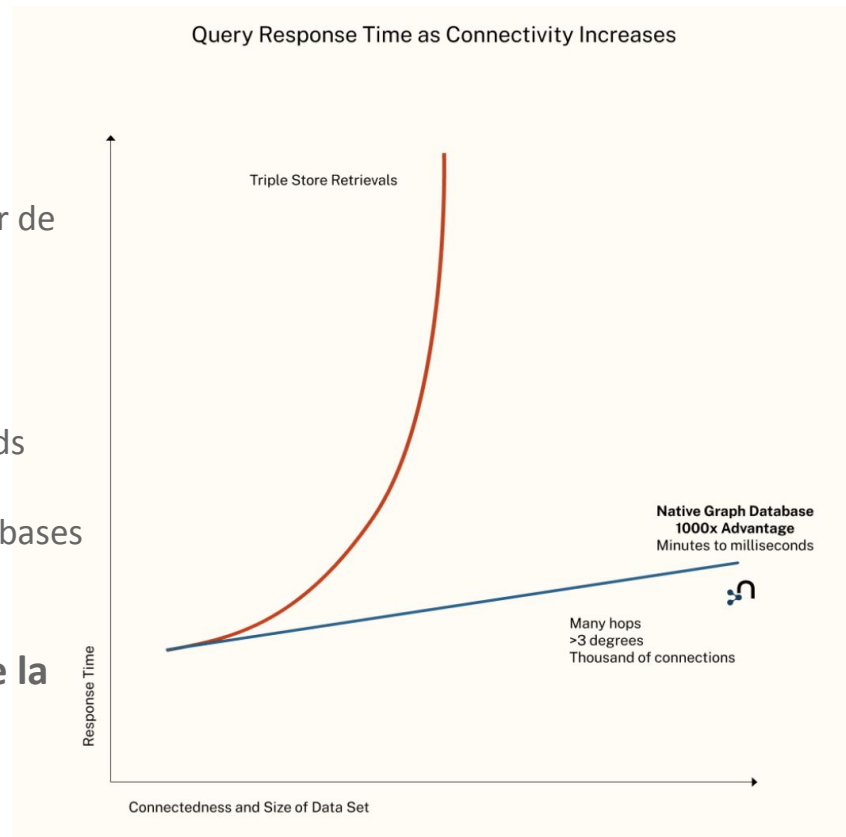
Les noeuds référencent directement leurs relations sans utiliser de recherche d'index.

Quand un noeud doit trouver ses voisins :

- Il suit directement les pointeurs physiques vers les noeuds connectés
- Évite les recherches d'index coûteuses requises dans les bases relationnelles

Temps de parcours constant quelle que soit la taille de la base

La relation est une sorte de micro index



Ecosysteme

core platforms

- [Neo4j](#) Graph Database (Community & Enterprise): a native graph database that stores and processes connected data
- [GraphAcademy](#): Free official learning platform with structured courses and certifications
- [Neo4j AuraDB](#): Fully managed cloud service that handles infrastructure, updates, and scaling automatically
- [Cypher](#): Declarative graph query language designed to be visual and intuitive, like ASCII art for queries

Outils

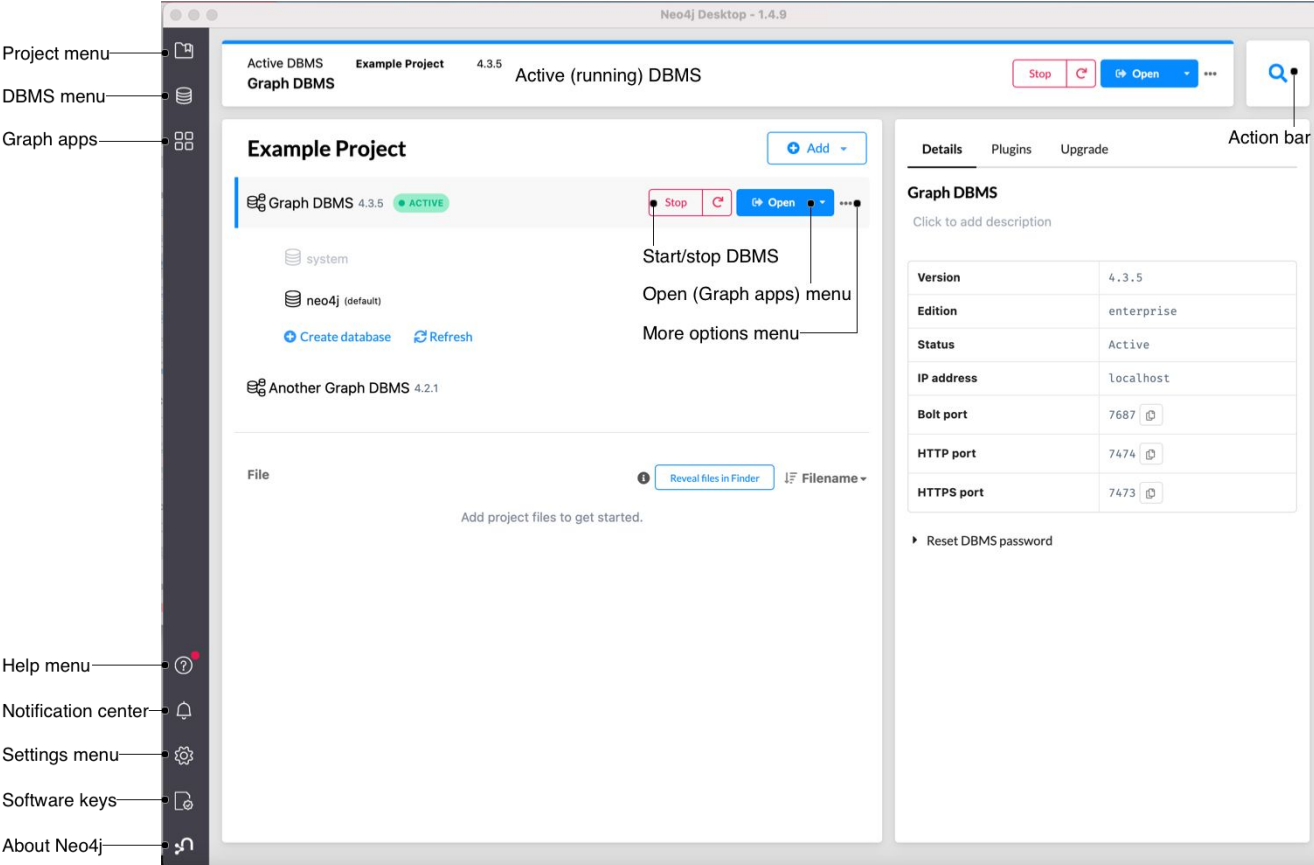
- [Neo4j Desktop](#) : All-in-one development environment for local database management and project organization
- [Neo4j Browser](#): Web-based interface for writing queries, visualizing results, and exploring your graph
- Neo4j Bloom: Business-friendly tool for visual graph exploration without writing queries
- [APOC library](#): Swiss army knife of useful functions that extend Neo4j's capabilities
- [GDS \(Graph Data Science\)](#) library: Machine learning and graph algorithms for advanced analytics
- Official drivers: Native connectivity for major programming languages like Python, Java, JavaScript
- [Arrows.app](#) : Free web-based tool for creating and sharing graph data models visually

Pratique

Neo4j desktop

<https://neo4j.com/docs/desktop-manual/current/installation/download-installation/>

Neo4j desktop



Network Dependency Graph

Network Dependency Graph

As a [Colab notebook](#) using the icypher extension

