Incorporating Graph Databases into the Survey Lifecycle

Tech Demo using RTI Merge and Neo4j

Nestor Alexis Ramirez, PhD
Center for Education Survey Research and Analysis

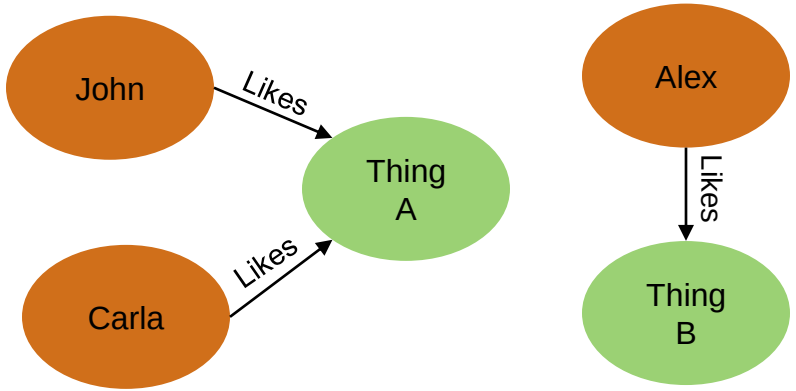
What is a graph database?

- A NoSQL database that is built to store data *relationships*
- Graph databases excel at leveraging complex data relationships and unstructured data
- In a graph model, ***nodes*** or ***vertices*** are data objects that are connected to each other through ***edges***

Relational Database

ID	Name	ID	Likes_thing
10029	John	10029	A
21047	Carla	39102	B
39102	Alex	21047	A

Graph Database



Why use graph databases?

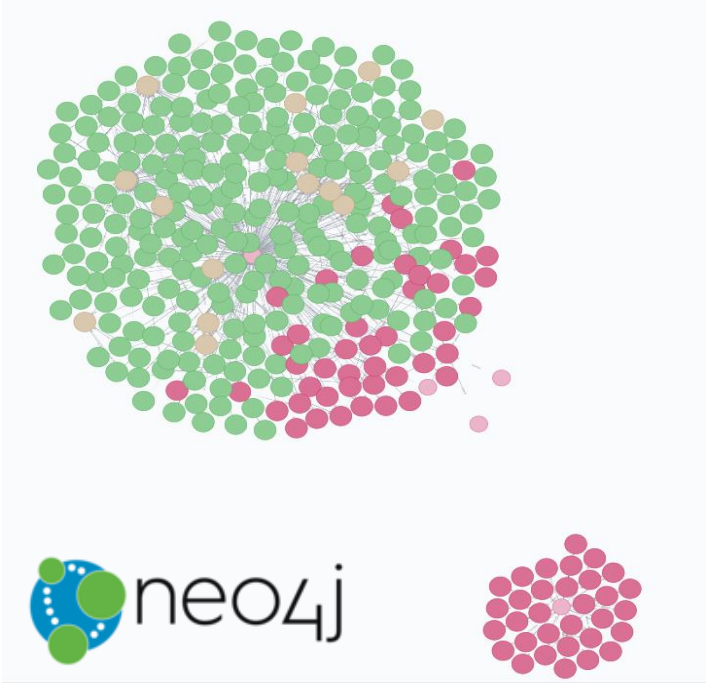
- Graph databases are ideal for:
  - Master data management
  - Real-time recommendation platforms
  - Fraud detection
  - Social media data
  - Advanced data science analytics and AI

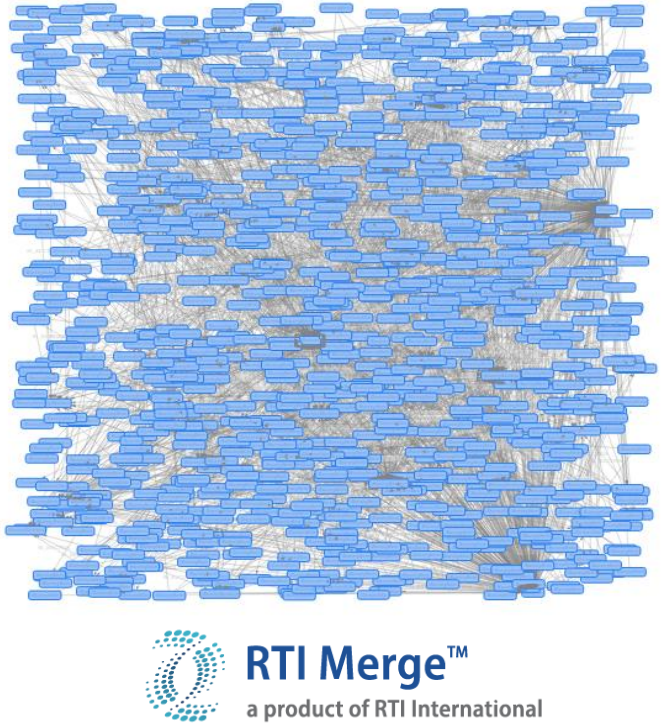
- Graph databases can also be useful for survey work:
  - Consolidate pre- and post-data collection information that is traditionally siloed
  - **Graph survey instrument metadata to aid in future decision-making**
  - **Map derived variable relationships to aid in data QC**

Graph Database Systems:

- **Neo4j**
- JanusGraph (used by **RTI Merge**)
- Amazon Neptune
- OrientDB
- ArangoDB
- …and many more!

2016–2020 Baccalaureate and Beyond (B&B:16/20)
Graph query demonstration using Neo4j and RTI Merge





More Information

Nestor Alexis Ramirez

Research Education Analyst

nramirez@rti.org