MCKNIGHT CONSULTING GROUP

Understanding the Modern Applications of Graph Databases

Presented by: William McKnight

"#1 Global Influencer in Big Data" Thinkers360

President, McKnight Consulting Group

A 2-Time Inc. 5000 Company

@williammcknight

www.mcknightcg.com

(214) 514-1444

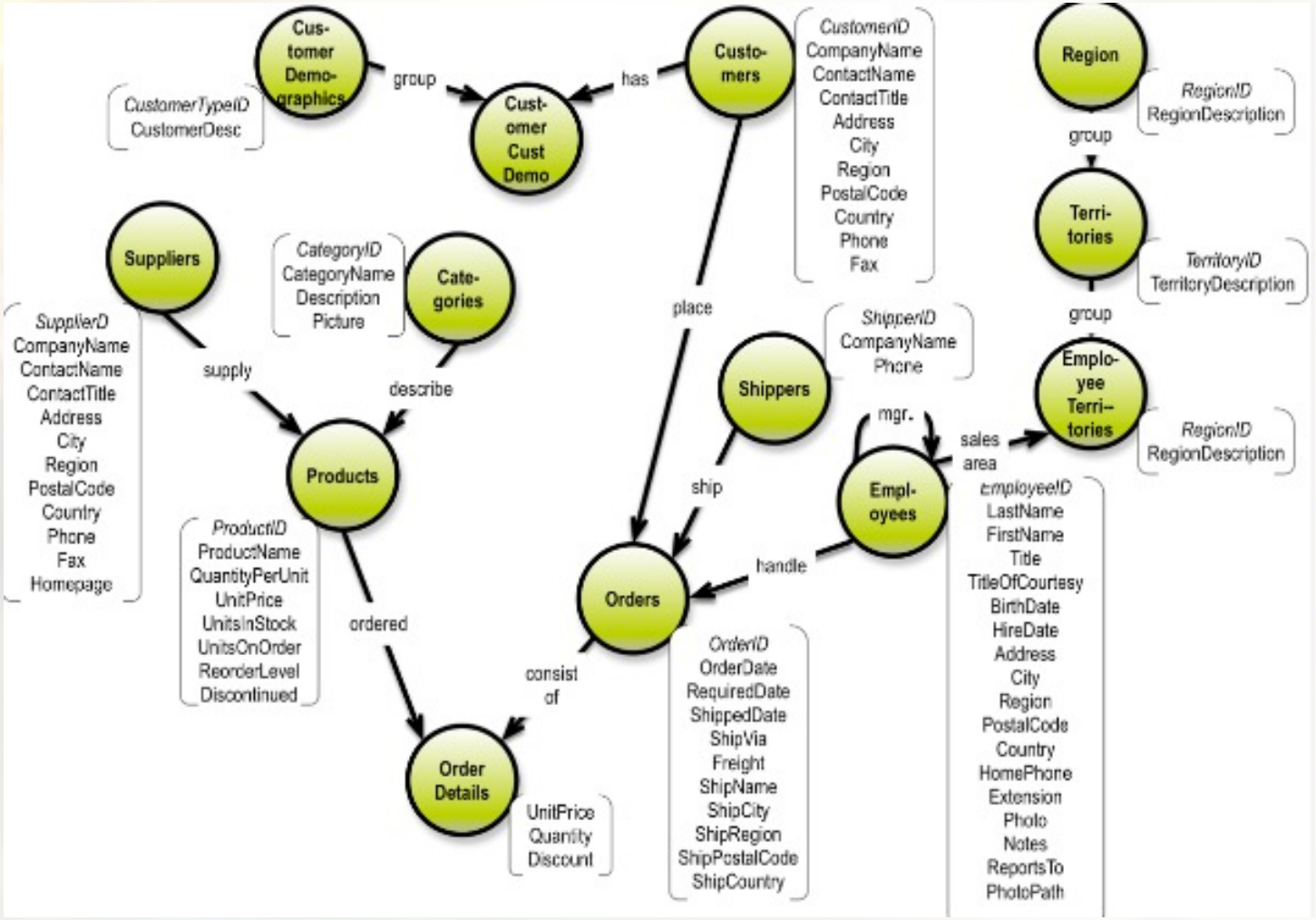
ADVANCED ANALYTICS

Second Thursday of Every Month, at 2:00 ET

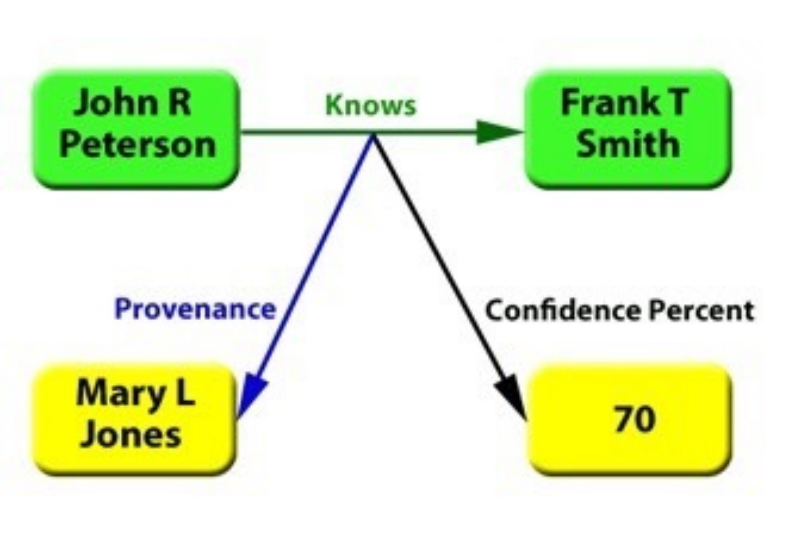
Graph > Relational Databases

- Avoiding Complex Joins
- Relational database performance degrades with number and levels of relationships, and database size
- Adding new types of data and relationships to a relational database requires schema redesign, increasing time to market

Property Graph: The Domain Model

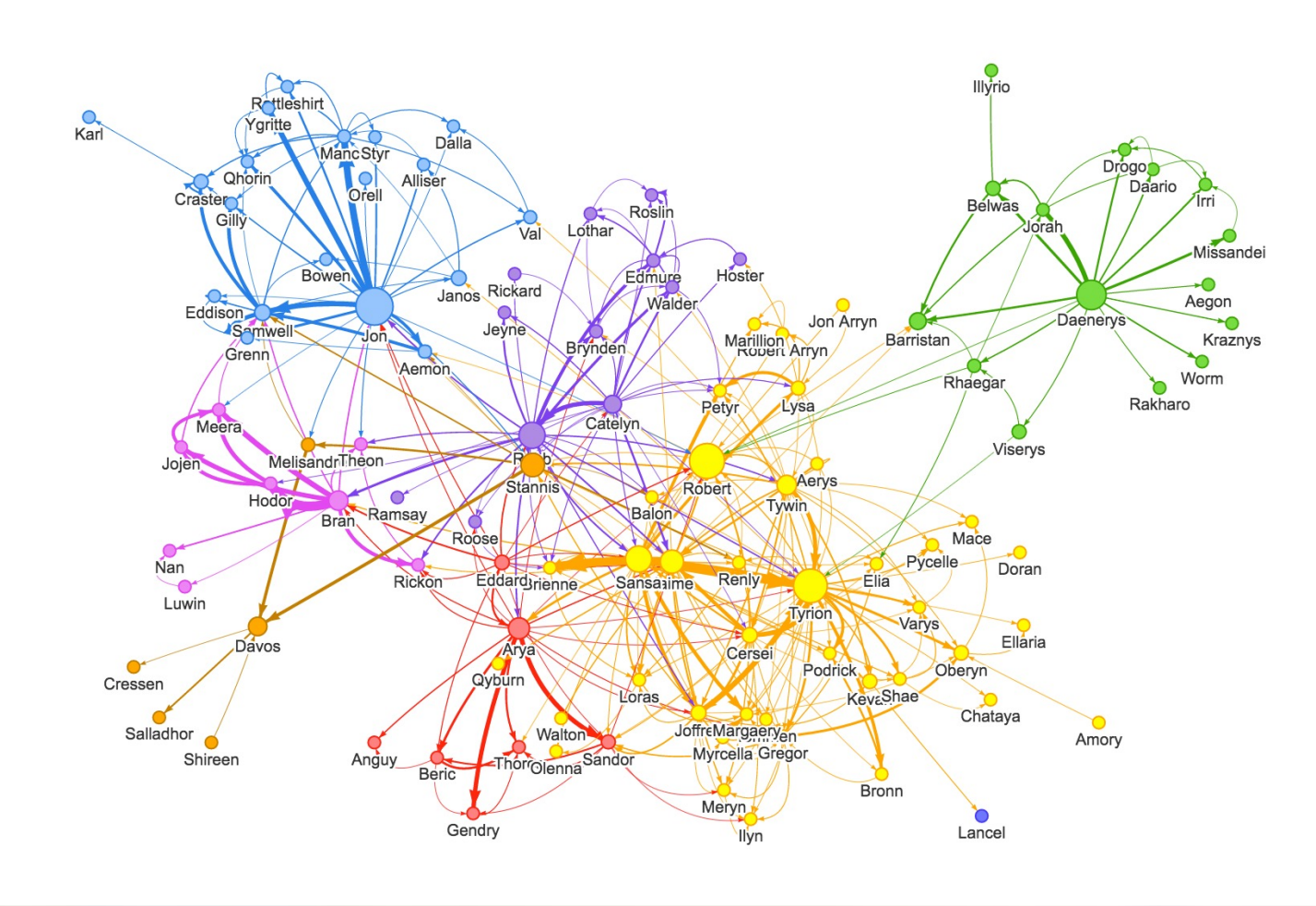


Semantic/RDF Graphs

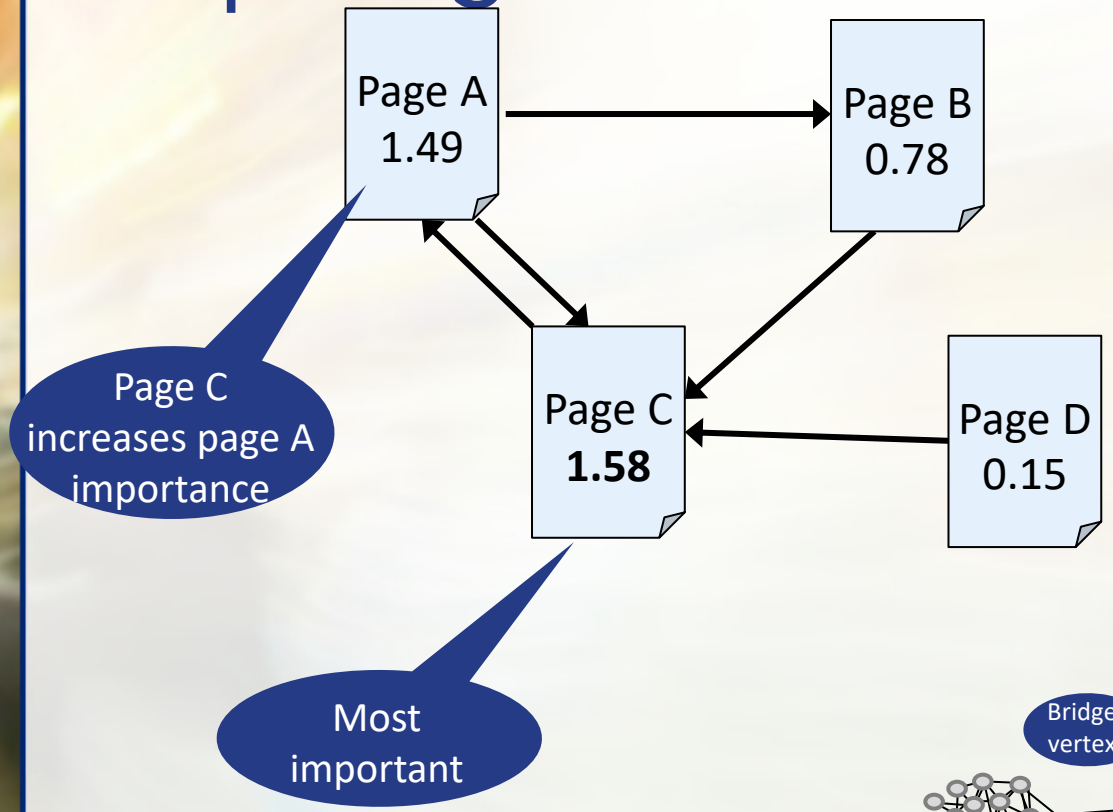


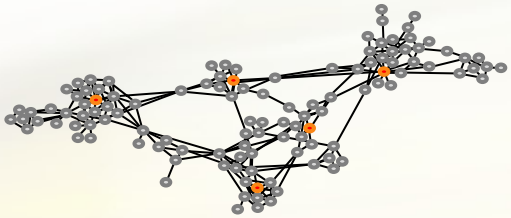
- A triple is a data entity composed of subject-predicate-object
- *In the image:*
  - *Triple 1: Subject*: John R Peterson *Predicate*: Knows *Object*: Frank T Smith
  - *Triple 2: Subject*: Triple #1 *Predicate*: Confidence Percent *Object*: 70
  - *Triple 3: Subject*: Triple #1 *Predicate*: Provenance *Object*: Mary L Jones

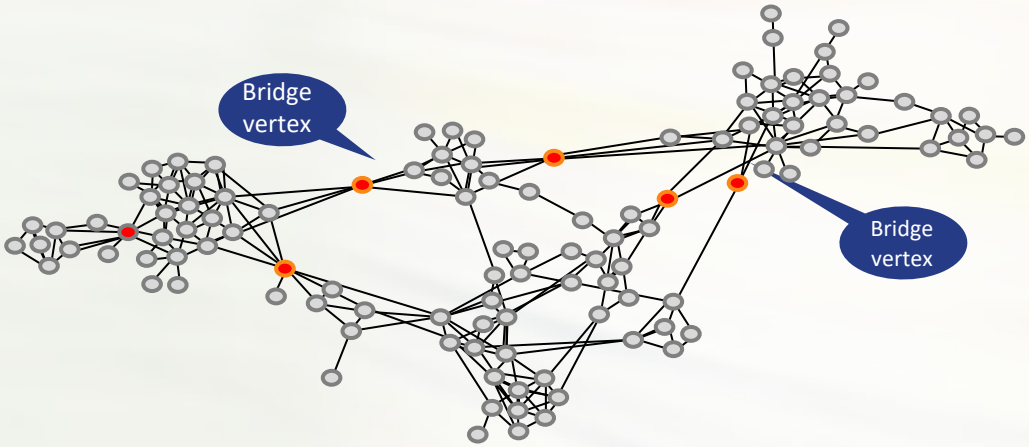
Graph Visualization



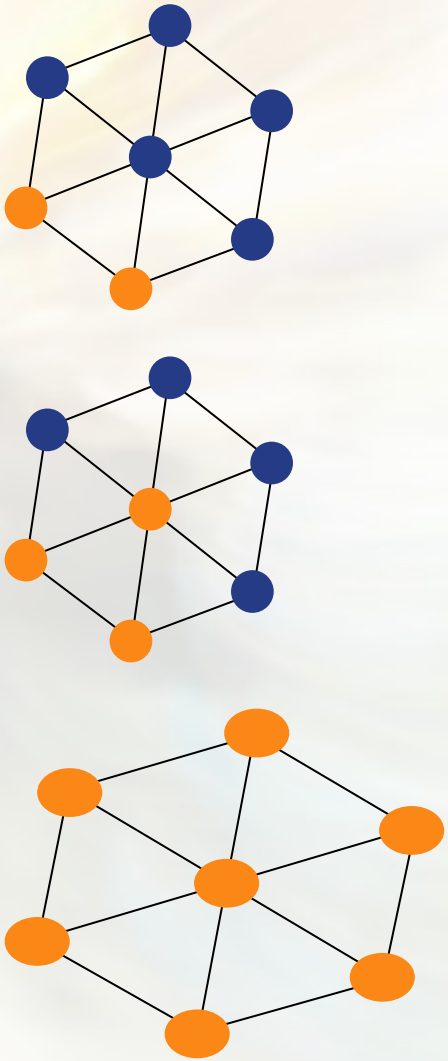
Graph Algorithms

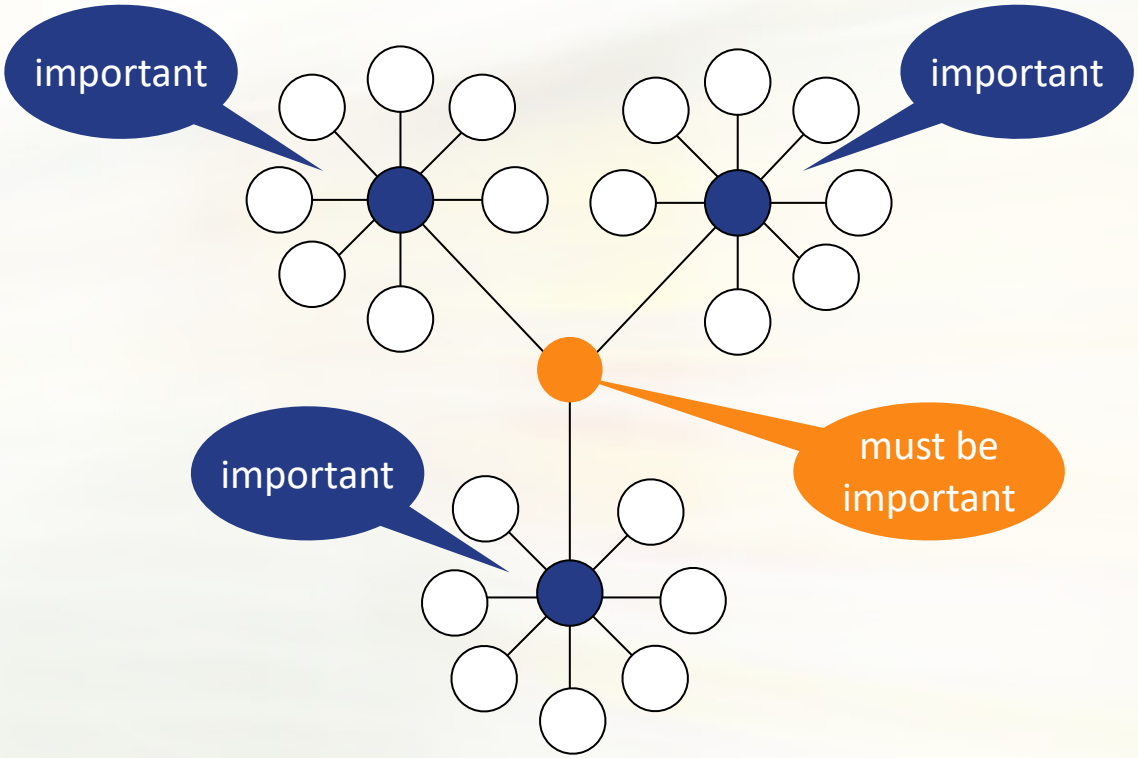






Graph Algorithms





Great Questions for Graph Databases

- In what order did a specific set of related events happen?
- Are there patterns of events in our data that seem to be related by time?
- How far apart in a (social or physical) network are two nodes and how strong is their relationship?
- What are the identifiable social groups and what are the general patterns of such groups?
- How important is any given "actor" in any given network and event?
- What type of messages emanate from a specific area?

How to Identify a Graph Workload

- Workload is identified by "network, hierarchy, tree, ancestry, structure" words
- You are planning to use relational performance tricks
- Your queries will be about pathing
- You are limiting queries by their complexity
- You are looking for "non-obvious" patterns in the data

Graph Databases

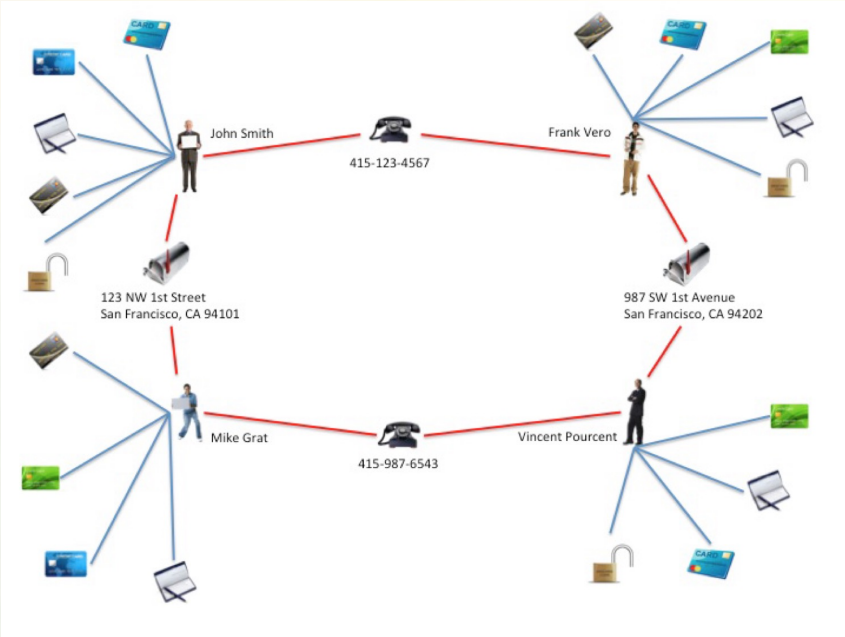
Graph Maturity

- Graph Visualization
- Graph Algorithms – Betweenness, PageRank, Closeness, Eigen Centrality, Clustering Coefficient
- Graph AI/ML – Identifying Subcommunities, Training Probabilistic ML models on graph

Graph Use Cases

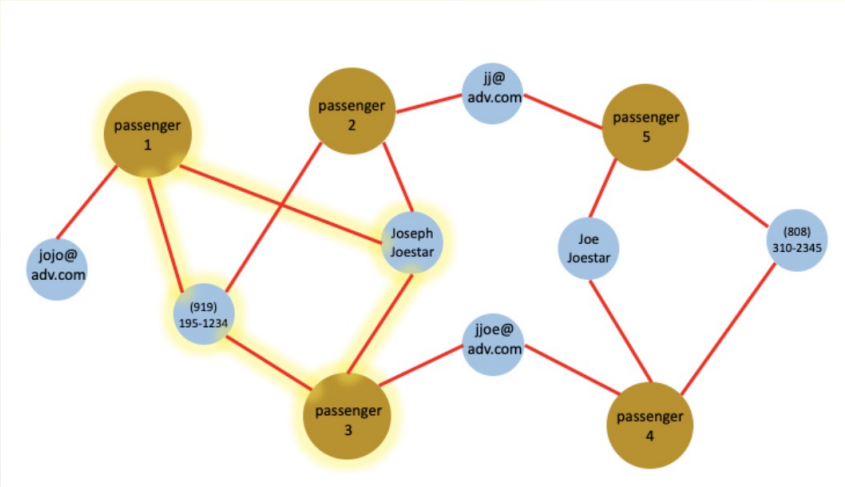
Fraud Detection

- By examining the connections between entities, graph databases can spot fraudulent behavior patterns.
- Graph databases can identify fraudulent behavior patterns and alert users or stop transactions by continuously analyzing data streams.
- Quickly and effectively spot patterns and connections between objects that appear to be unrelated.
- Graph databases can identify behaviors that are indicative of fraud and use those patterns to predict future fraud by training machine learning models on historical data.



Entity Resolution

- The ability to resolve and identify relationships between data
- Using the relationships between different entities, graph databases can be used to determine if they refer to the same real-world entity
- The graph database stores the relationship between the entities, allowing the system to quickly determine if two records correspond to the same entity
- This eliminates the need to manually compare multiple fields for each record
- Can be useful for matching records from different sources

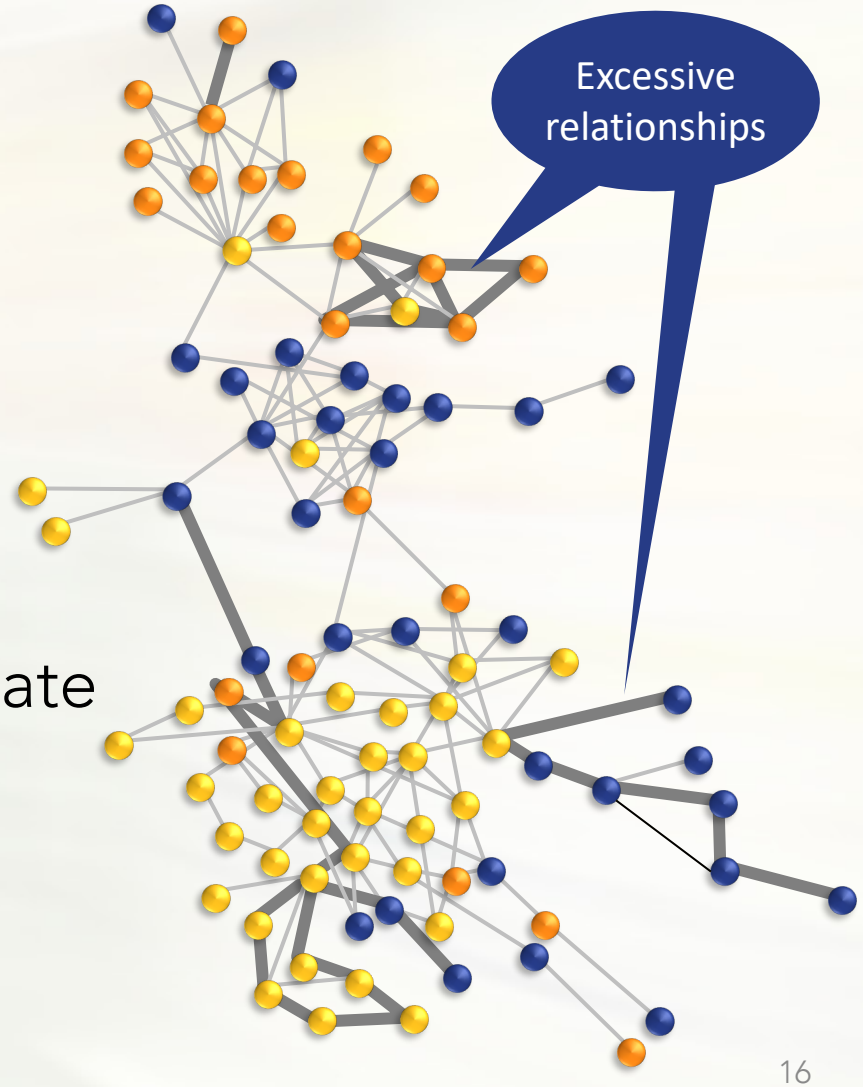


Network Attack

- Design a graph schema to represent the different nodes and relationships among the entities in the network
- Represent network objects (switches, routers, firewalls, servers, workstations, etc.) as nodes and the connections between them as edges
- Store data regarding traffic flow between the objects as graph properties
- Use analytics and machine learning techniques to identify malicious network activity and block it automatically
- Quickly traverse the network graph and process massive volumes of data in real-time

Healthcare Fraud

- Monitor drugs and treatments
  - Excessive prescribers
  - Excessive consumers
- Patients connected to
  - Doctors, pharmacies, medications
- Use Graph Access
  - Find outliers and investigate



Online Shopping

- Bring fast context to a shopping experience
- Need to recall past similar interactions
- Need probabilistic models
  - Product catalog
  - Shopper attributes

Major Insurer

- Insight into risk environment
- Risks such as
  - People appearing in multiple policies and claims
  - Premium leakage i.e., Underestimated mileage, undeclared drivers, false garaging
  - Padded claims
- Policyholder graph with risk indicators
  - Risk indicators spread in graph
- Worker's Compensation Fraud

Television, Magazine and Media

- Analyze content and consumption for personalization
- Most users don't "log in"
- Identified anonymous users through unique cookies
  - Cookies unstable, used third-party to enrich; needed to vet
- Determine valuable (connected) providers, audience segments
- Enabled evaluation of the accuracy of vendor data
  - And cut the cost of using unreliable data

Preventive Maintenance

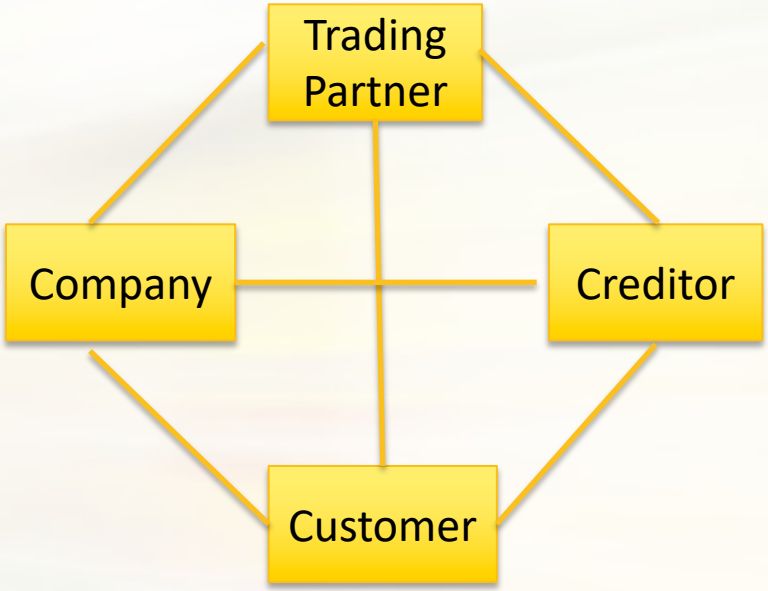
- Identify which robotic parts were about to fail so they could replace the failing parts all at once
- Able to reconcile data to the same piece of the production line machinery
- Able to identify when a part is about to fail so they can pre-plan and avoid unnecessary breaks in the production assembly line

Pharmaceutical/Research

- Need to connect data from disparate parts of the company to increase research and operational efficiency, increase output, and accelerate drug research
  - Allow analysts to quickly and easily access the full body of institutional knowledge
- Graph allowed bioinformaticians to more easily identify useful signals within large sets of noisy data and to answer highly-specific questions
- Link targets, genes, and disease data across different parts of the company

Financial Services

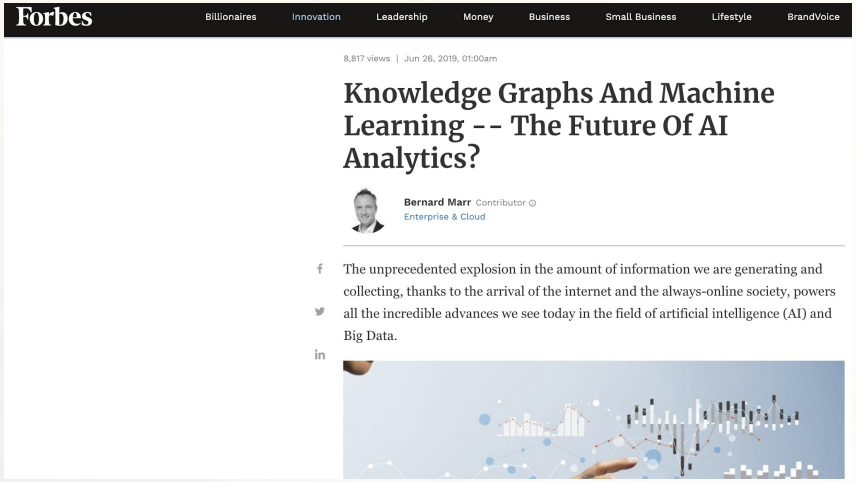
- Anti-Money Laundering
  - Identify connections
  - Display the connections surrounding a specific point
  - Identify which connections and situations of interest lead to productive investigations and inform work



Closing Thoughts

Integration with LLMs and ML

- Graph databases are developing natural language interfaces for knowledge graphs and creating knowledge graphs from unstructured data.
- Combines the domain specific knowledge in a knowledge graph with the general knowledge in an LLM
- Combines the structured knowledge in a knowledge graph with the unstructured knowledge in an LLM



Vector Databases

- Data structures called graph embeddings are used to compare similar data structures quickly
- The attributes are all compressed by graph embedding
- Embeddings are calculated using machine learning algorithms
- The process of creating a new embedding vector is called “encoding”
- Embeddings can be thought of as a low-level representation of an item
- Embeddings are ideal for “fuzzy” match problems
- Embeddings work with graph algorithms

Vector Databases vs Graph Databases

- Graph databases are better suited to processing data with complex relationships, whereas vector databases are better suited to handling high-dimensional data, such as images and video
- Graph databases are made for queries involving relationships, while vector databases excel at similarity searches
- Graph databases may not be as scalable due to the complexity of the data model

Conclusion

- Graph is a Fast Growing data category
- It's all about the Use Case; Good for Graph:
  - Real-time Recommendations
  - Fraud Detection and Risk
  - Network and IT operations
  - Entity Resolution
  - Identifying Relative Importance
- Graph databases are made for queries involving relationships, while vector databases excel at similarity searches

2023 Advanced Analytics Topics

1. ~~2023 Trends in Enterprise Analytics~~
2. ~~Showing ROI for your Analytic Project~~
3. ~~Architecture, Products and Total Cost of Ownership of the Leading Machine Learning Stacks~~
4. ~~Competitive Analytic Architectures: Comparing the Data Mesh, Data Fabric, Data Lakehouse and Data Cloud~~
5. ~~Why Analytics Leaders deploy Master Data Management~~
6. ~~What Does Information Management Maturity Look Like in 2023~~
7. ~~Understanding the Modern Applications of Graph Databases~~
8. Common Misconceptions About Master Data Management
9. Organizational Change Management: Will it Hold Back Artificial Intelligence Deployments?
10. Open-Source vs Commercial Vendor Software in the Enterprise
11. Data Quality: The ROI of Adding Intelligence to Data
12. Strategies for Machine Learning Success

MCKNIGHT CONSULTING GROUP

Understanding the Modern Applications of Graph Databases

Presented by: William McKnight

"#1 Global Influencer in Data Warehousing" OnAlytica

President, McKnight Consulting Group

A 2-Time Inc. 5000 Company

@williammcknight

www.mcknightcg.com

(214) 514-1444

ADVANCED ANALYTICS

Second Thursday of Every Month, at 2:00 ET