



Graph Databases and Analytics: How to Use Them

Melli Annamalai and Jean Ihm

Graph Technologies

Oracle

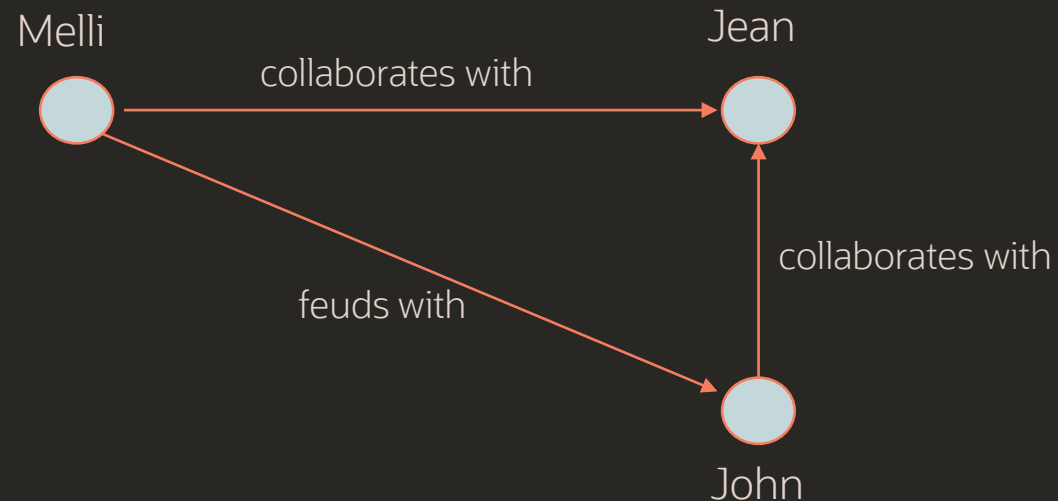
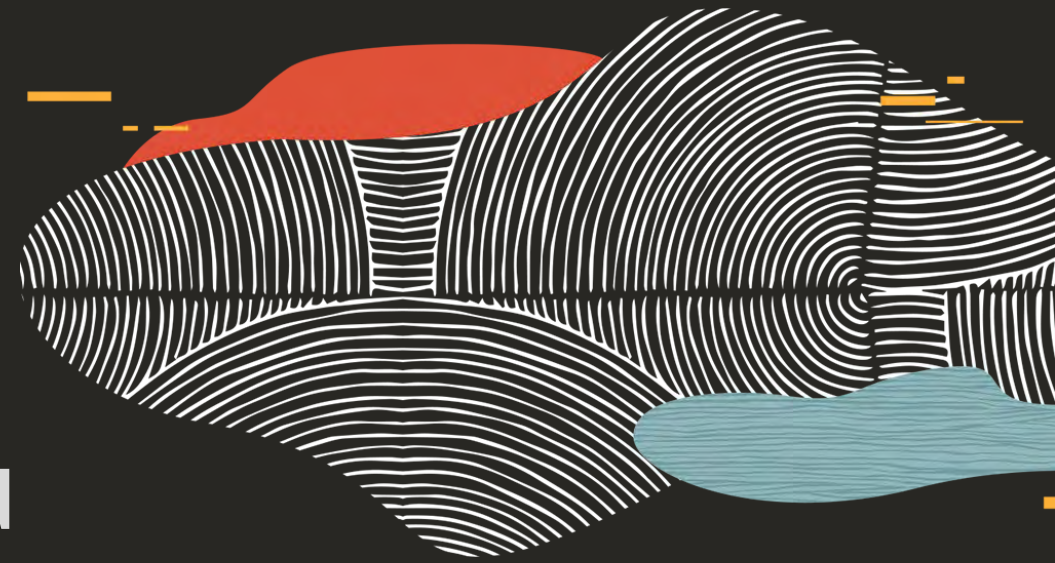
Safe Harbor

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, timing, and pricing of any features or functionality described for Oracle's products may change and remains at the sole discretion of Oracle Corporation.

Statements in this presentation relating to Oracle's future plans, expectations, beliefs, intentions and prospects are "forward-looking statements" and are subject to material risks and uncertainties. A detailed discussion of these factors and other risks that affect our business is contained in Oracle's Securities and Exchange Commission (SEC) filings, including our most recent reports on Form 10-K and Form 10-Q under the heading "Risk Factors." These filings are available on the SEC's website or on Oracle's website at <http://www.oracle.com/investor>. All information in this presentation is current as of September 2019 and Oracle undertakes no duty to update any statement in light of new information or future events.

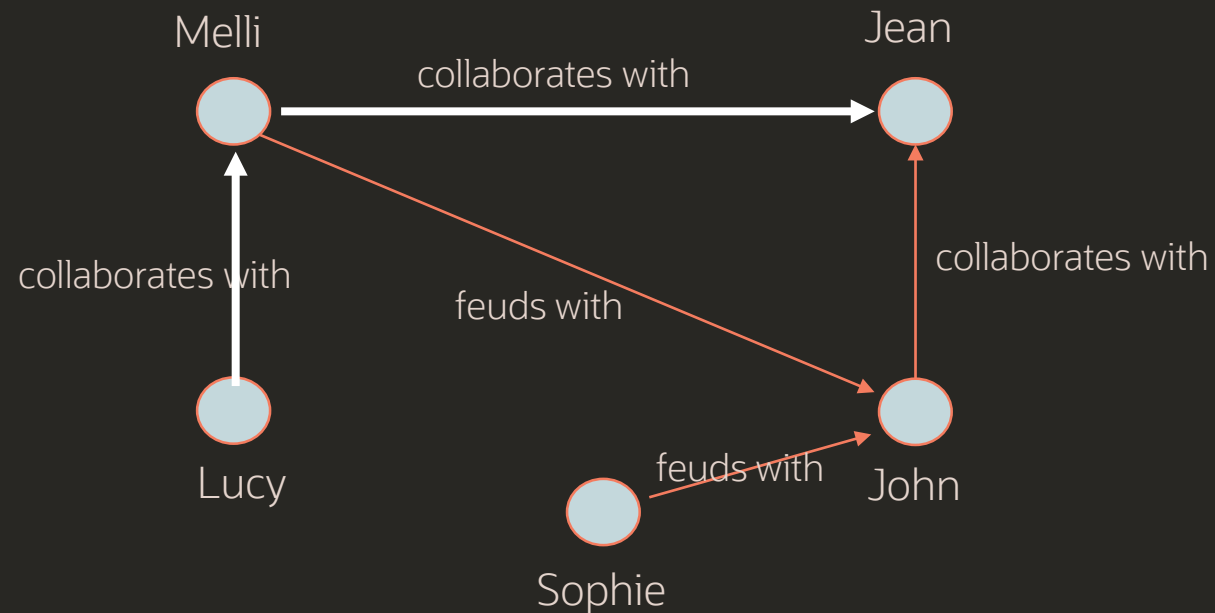
What is a Graph?

A collection of points (vertices) and lines between those points (edges)

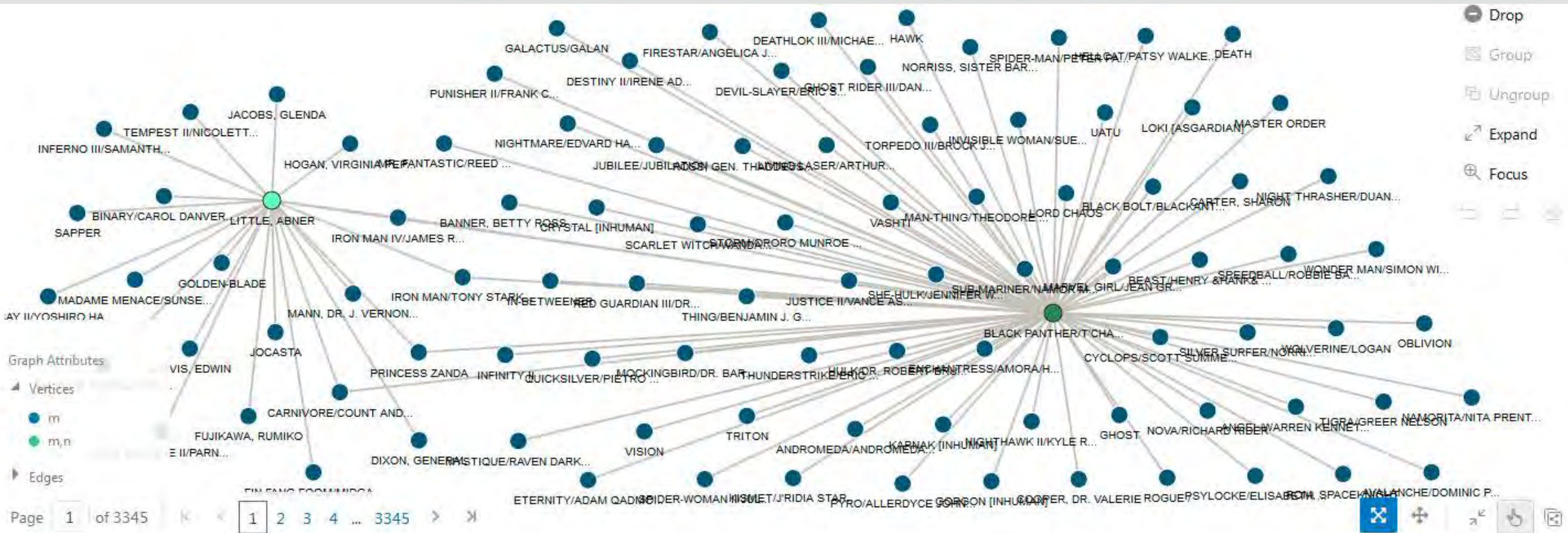


What is a Graph?

A collection of points (vertices) and lines between those points (edges)



A More Complex Graph: Find popular nodes in a social network

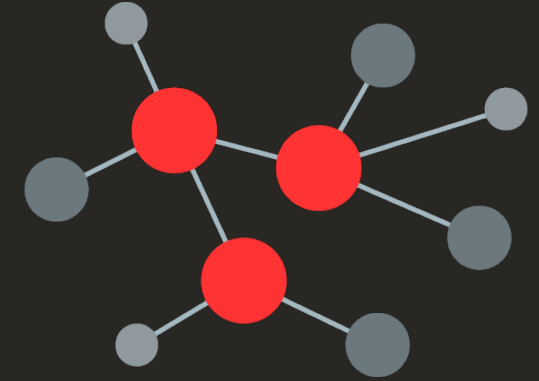


Identify influencers



Graph Analytics

- Community detection and influencer analysis
 - Churn risk analysis/targeted marketing, HR Turnover analysis
- Clustering
 - Product recommendation
- Anomaly detection
 - Identify fraud
- Path analysis and reachability
 - Manage Bill of Materials, Outage analysis in utilities networks, vulnerability analysis in IP networks
- Pattern matching
 - Tax fraud detection, data extraction



Topics

Data Analytics &
Data Science



Andrew Ross

18 February 2019

Gartner: top 10 data and analytics technology trends for 2019

Trend #5: Graph

According to Gartner, graph analytics is a set of analytic techniques that help enterprises explore the relationships between entities of interest such as transactions, processes and staff.

The application of graph processing and graph database management systems will grow at 100% annually through 2022.

Gartner MQ - Data Management Solutions for Analytics



Graph Data Models

Property Graph Model

- Path Analytics
- Graph Analytics
- Detect patterns and anomalies

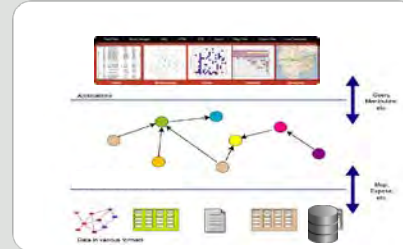


- Financial
- Retail, Marketing
- Social Media
- Smart Manufacturing

Shipping for 3+ years

RDF Graph Model

- Data federation
- Knowledge representation
- Semantic Web



- Life Sciences
- Health Care
- Publishing
- Finance

Shipping for 12+ years

Graph Model

Use Cases

Industry Domain

Customer Success Stories


Graph Technology Applications

- Oracle's graph technology has been applied to solve many real problems for internal/external customers

Detecting Malicious Network Traffic via Graph Analysis

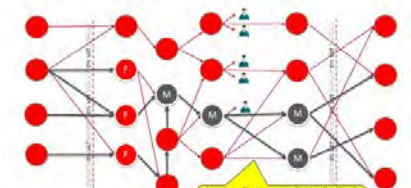
- Approach (from DSN'17)
 - Rather than analyzing payload of each packet, collect up packet traces as graphs
 - Extract characteristics of those graphs via graph algorithms
 - Train model → Differentiate malware traces from normal ones

Observation: Malwares have different trace patterns than normal activities



Oracle
DSN'17
Angle: Examining Cryptomall.com on 12/11/2013

VAT Fraud Detection: Find Intermediaries




Key is to find intermediaries (hubs) that sit in the middle of transactions through them

Oracle

Anti-Money Laundering with Graph ML

- Task #1
 - Given a new case, can we find existing cases in history that look similar to this one? (as a reference for investigation)
 - Use Pg2Vec to train and find
- Task #2
 - Train from existing cases, learn a classifier
 - i.e. system recommends that "this case looks serious". Recommended for official investigator



Oracle

Product Recommendation


- Taobao from Alibaba
 - Base Graph Embedding (BGE)
 - DeepWalk to learn the embeddings of each item
 - Edge-weights used in the transition-probability
 - Graph Embedding with Side Info (GSI)
 - Use side info (e.g., price, stock, category, ...) to mitigate cold-start item issue
 - Enhanced GSI (EGSI)
 - Includes a weighted average layer
- Pinsage from Pinterest
 - Pinterest is used to save and organize Pins
 - Pinsage
 - Random-walk based Graph Convolutional Net
 - Learn embeddings for billions of Pins



Oracle

Diagnosis Prediction from Patient Records (Health-Science)

- Dataset: MIMIC-III, v1.4
 - MIMIC is a relational database containing data of patients who stayed within the intensive care units of Beth Israel Deaconess Medical Center. Commonly found in the scientific research & literature
- Approach
 - Captures sparse information as graph and train the model from embedding
 - Combines multiple patients' history
 - Combine knowledge graph

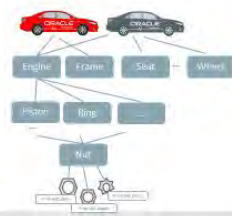


Oracle

Example: Bill of Materials for a Vehicle Manufacturer

- Represent bill of materials as a graph
 - Parts and bits and their suppliers
- Ask critical questions via graph query
 - How many suppliers are associated with model RX-78 (blue-and-white color)?
 - Which frame parts are shared between model RX-78 and RX-79?
 - Which models get impact, if the supplier named "ABC parts" become unavailable?

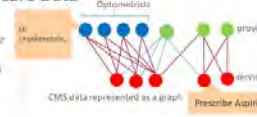
As BOM gets deeper, graph-based query gets much faster than Relational DB (Join-based query process becomes too expensive)



Oracle

Anomaly Detection in Medicare Data

- Anomaly in the Data Set
 - Doctors of the same-specialty provide similar services
 - What if a doctor perform a lot of treatments that typically belong to other specialties?
 - E.g. a cardiologist doing plastic surgery?
 - How do we find such cases?
 - By applying graph algorithm




Oracle

Asset Risk Analysis (Electricity Network)

- Electric Power Distribution Network as a graph
 - Analyze importance of nodes in the network
 - What nodes to pay attention during emergency (e.g. Storm season)

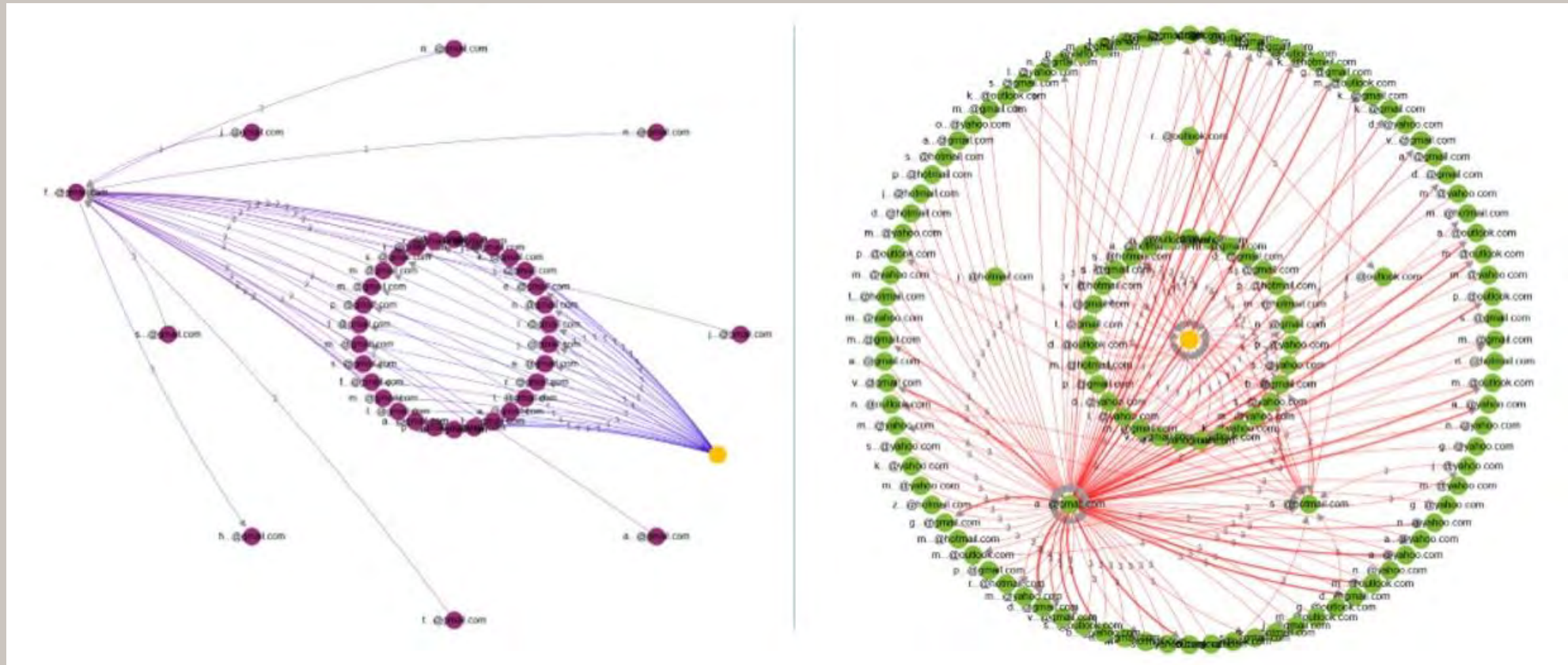
Failure of this node has higher impact than others



Oracle



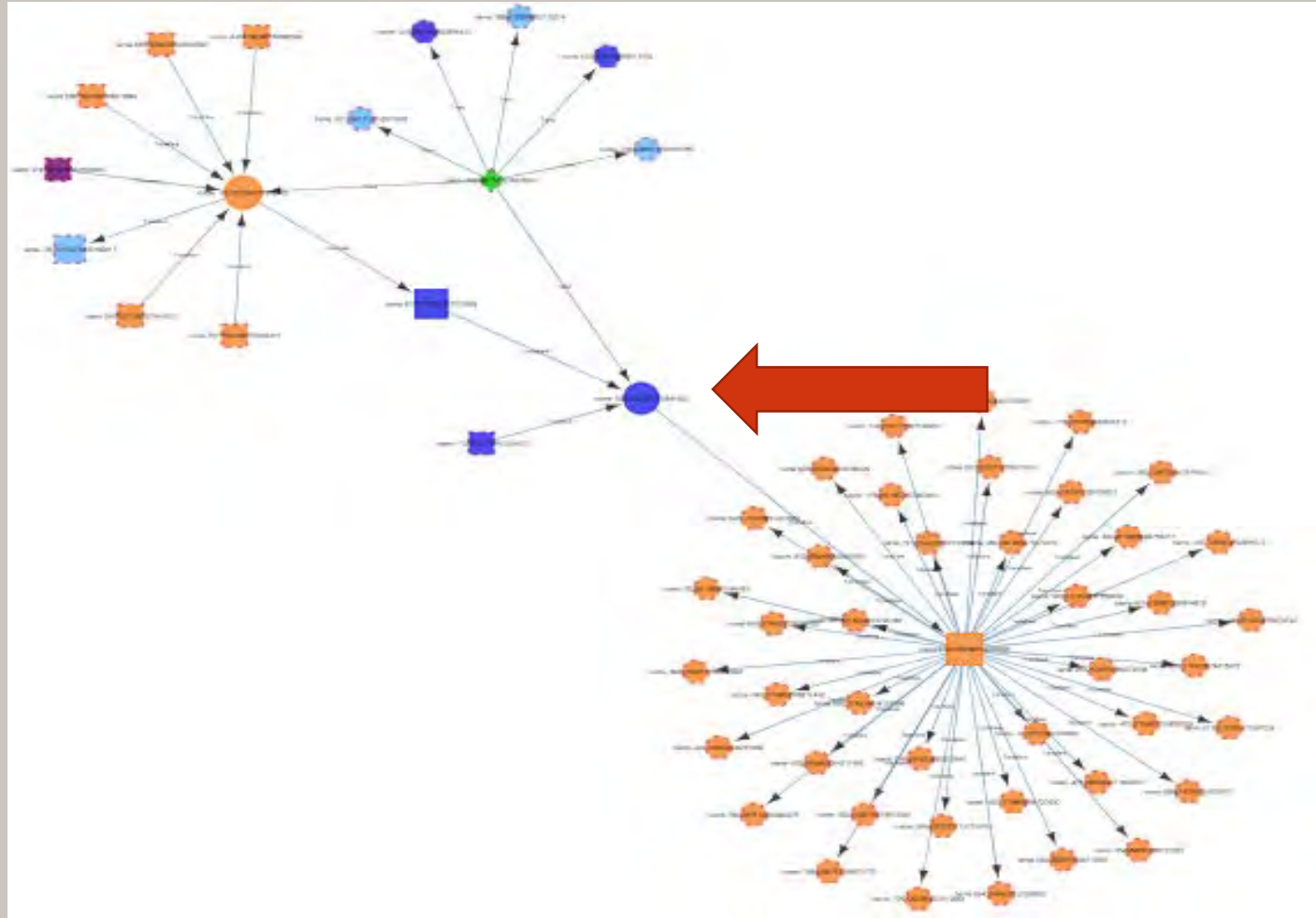
Paysafe: Money Transfer Fraud Detection



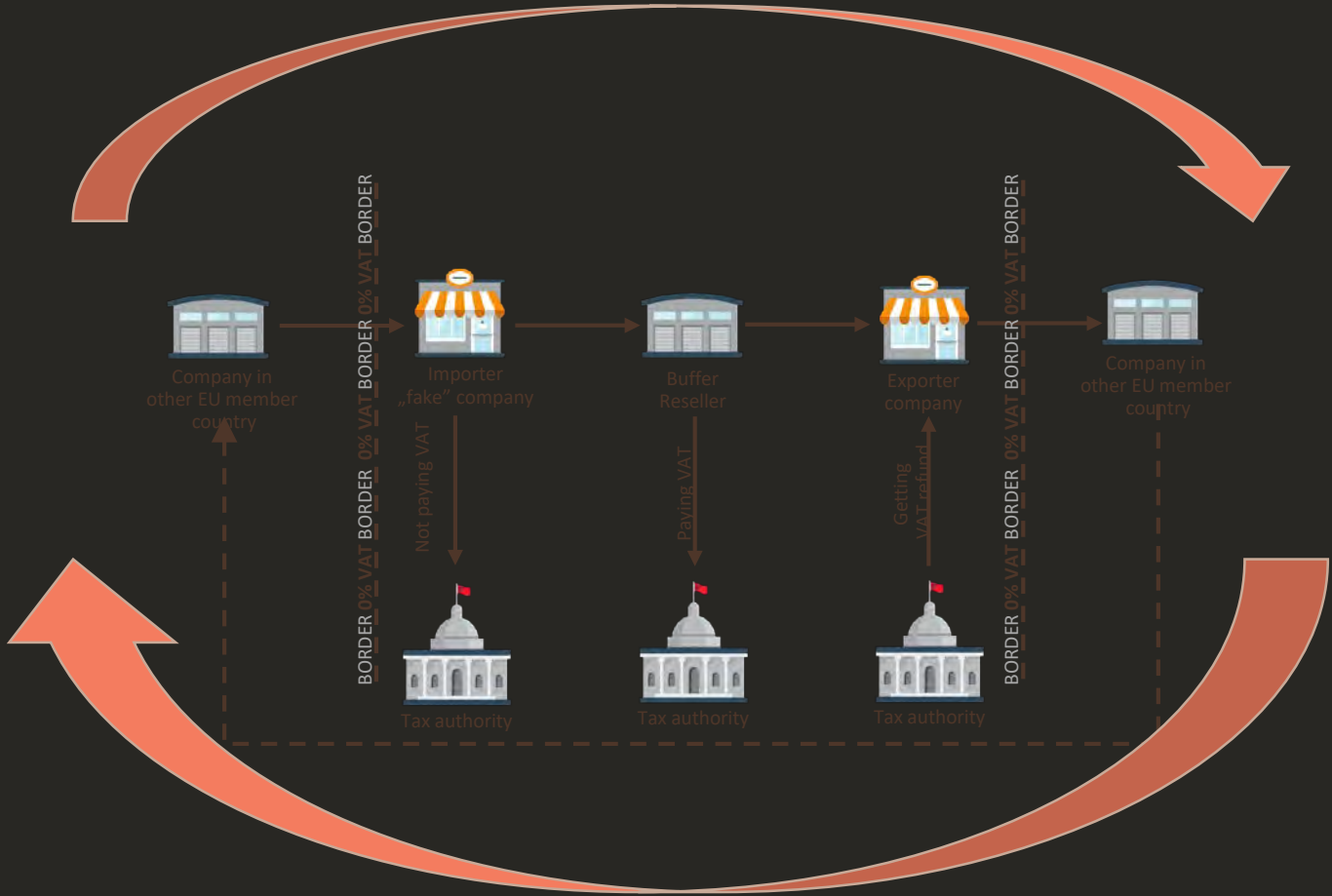
Multiple paths going to the same destination

Limited number of source and destination vertices

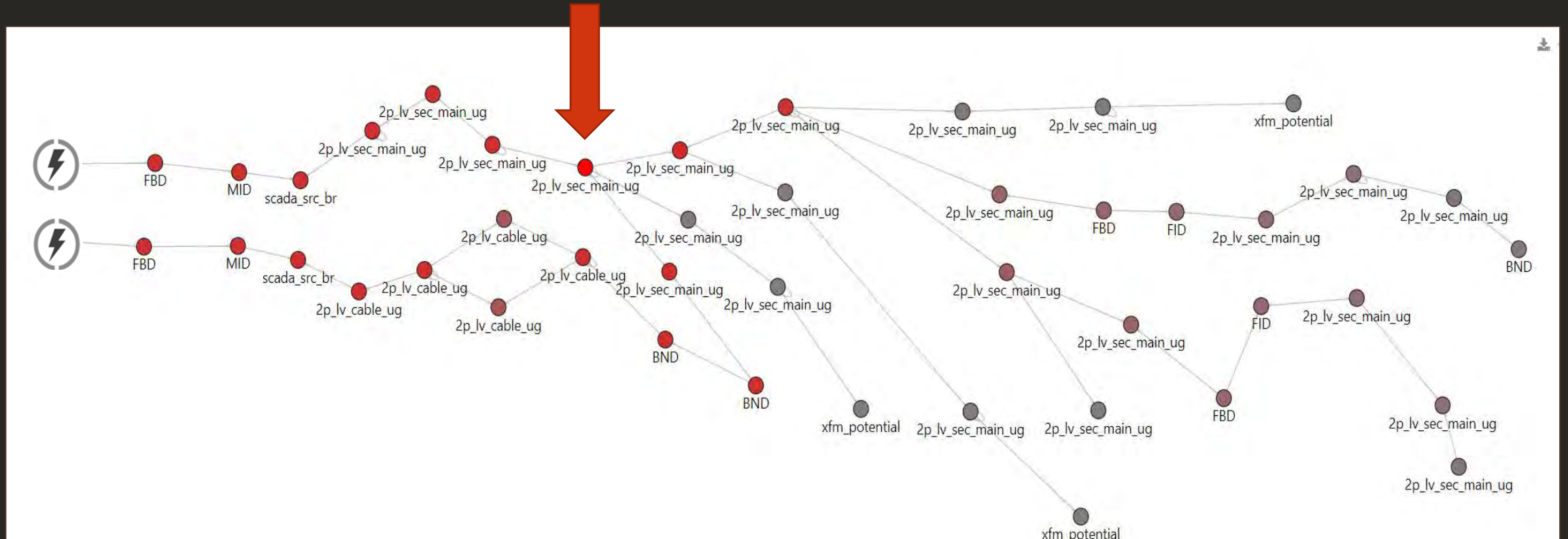
Banco De Galicia: Which Nodes are Transferring money to other Banks?



Circular Payments and Tax Fraud



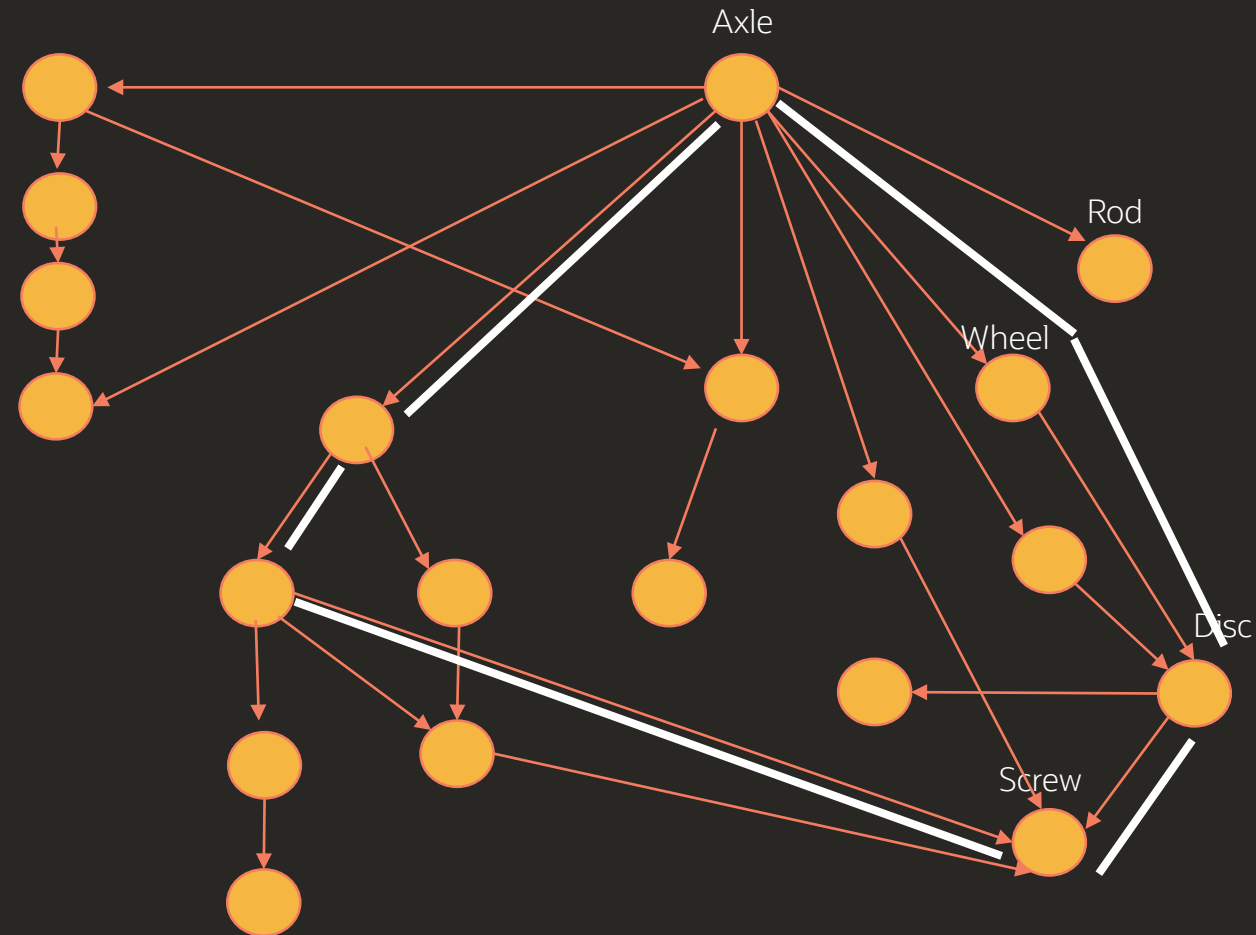
Find nodes that are in most shortest paths between other nodes



Identify nodes that can cause vulnerabilities in a grid

Manufacturing: What is the Impact of Changing this Part?

- A car has 30,000 parts



Southern European Police Force

Analyzing suspicious travel patterns
Detecting potential threats early

Data sources

Integrated Operations Management System (SIGO, built by Accenture)

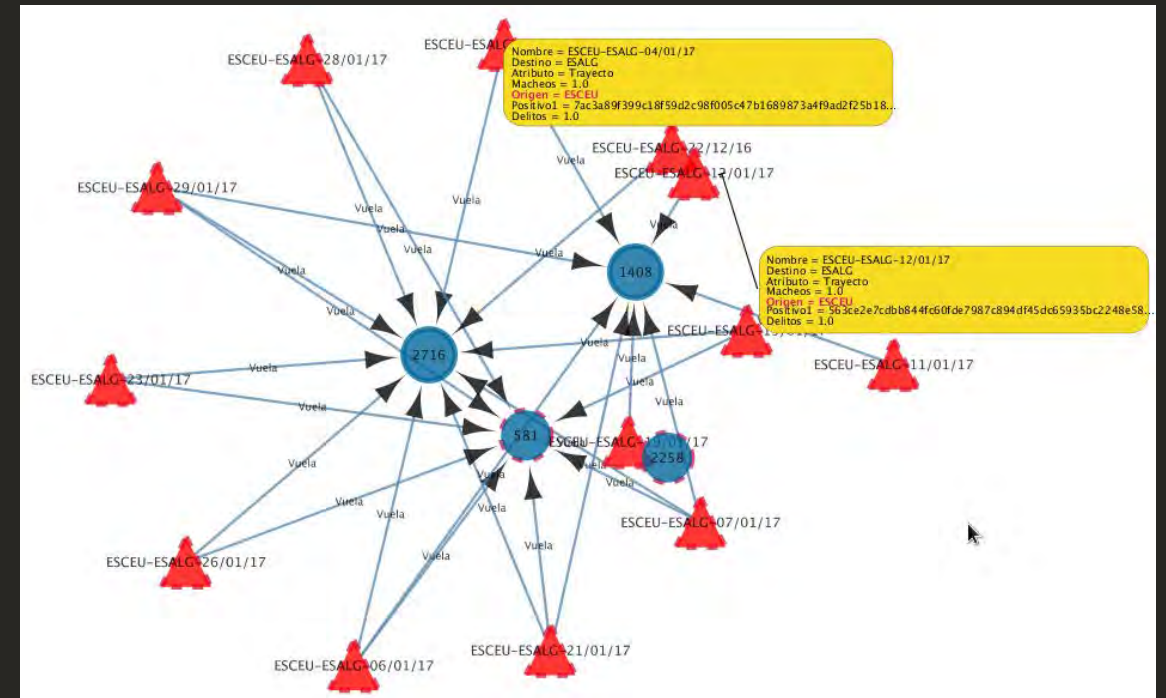
Advance Passenger Information System (APIS, data from Non-Schengen Airport and Ports)

Combining legal requests and co-travelers

eg. determining „hot flights“ with at least one known criminal on board

Looking at passenger relationships

eg. Betweenness centrality to determine recurring travelers



Property Graphs

Property Graph Product Overview

- Store, manage, query and analyze graphs
- Highly scalable in-memory analytics
 - 10s of billions of edges and vertices
- 50+ pre-built graph analysis algorithms
 - Detecting components and communities
 - Tarjan's, Kosaraju's, Weakly connected components, label propagation, etc.
 - Ranking and walking
 - Pagerank, personalized pagerank, betweenness centrality, etc.
 - Evaluating community structures
 - Conductance, modularity, triangle counting, Adamic-Adar, etc.
 - Path-finding
 - Path distance, Dijkstra's, Bellman-Ford's, etc.

Property Graph Product Overview

- PGQL: Powerful graph query language

SQL-like language for specifying graph patterns
Working on graph additions to the SQL standard

- Java API for analytics
- Rich user interface

Notebook
Shell UI
Graph Visualization

- Enterprise capabilities – built on Oracle infrastructure

Manageability, fine-grained security, high availability, integration, and more

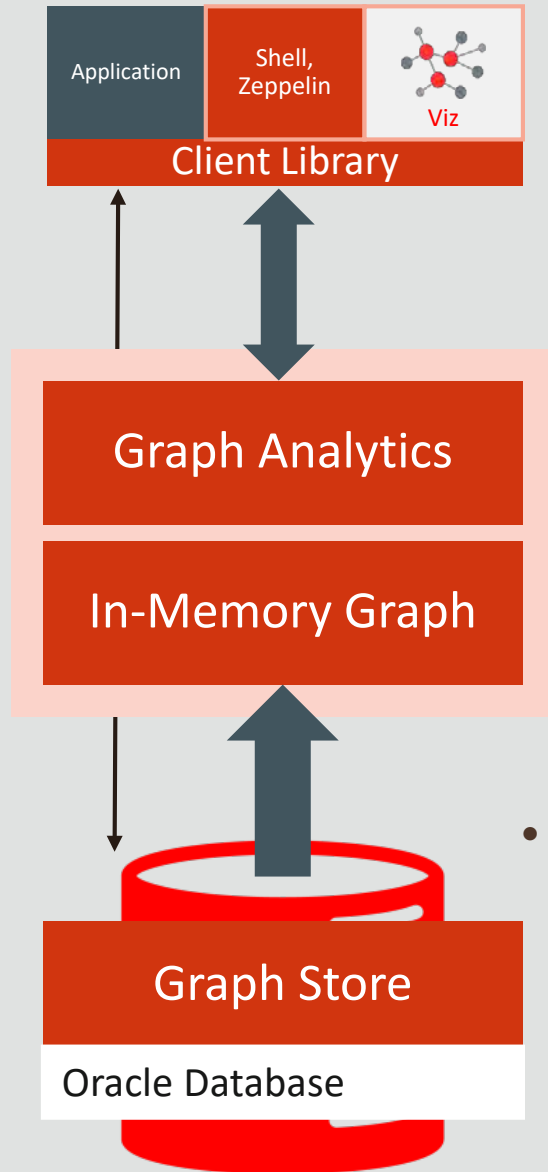
PGQL example:

```
PATH any_edge as ()-[]-()  
SELECT n, m MATCH(n) -/: any_edge/ -> m  
WHERE n.name='Lucy' and m.name='Jean'
```

Architecture

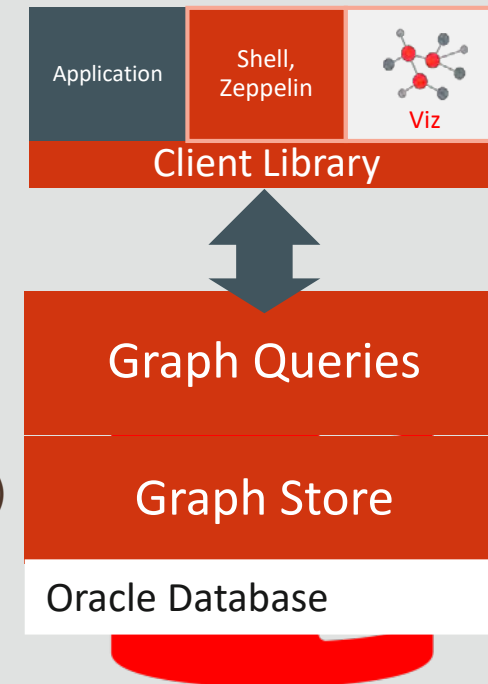
Product:
Oracle Database
Spatial and Graph

- In-memory parallel graph analytics server (**PGX**)
 - Load graph into memory for analysis
 - Automate graph refresh



- Client libraries
 - Java API to develop applications
 - Command-line submission of graph queries
 - Graph visualization tool
 - APIs to update graph store

- In-database parallel graph traversal
 - Run PGQL queries (converted to SQL) in the database

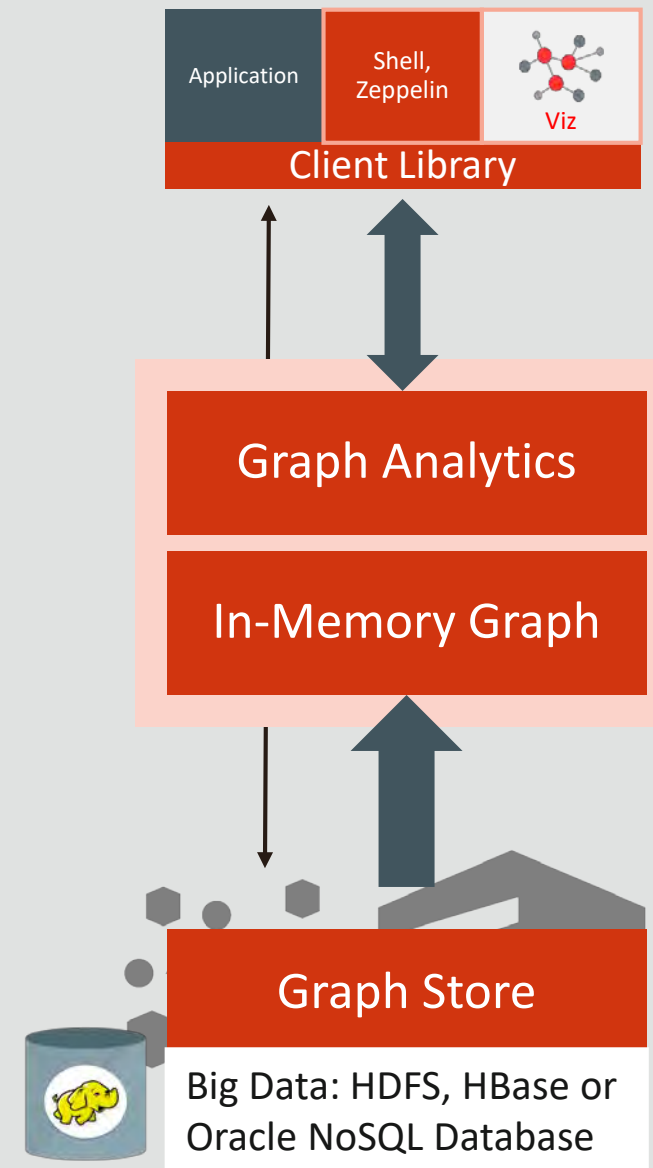


Using Graphs on a Big Data Platform

Product: Oracle Big Data Spatial and Graph

Runs on the Hadoop platform

- In-memory parallel graph analytics server (**PGX**)
 - Load graph into memory for analysis
 - Automate graph refresh
- Client libraries
 - Java API to develop applications
 - Command-line submission of graph queries
 - Graph visualization tool
 - APIs to update graph store



From Tables to a Graph

Organizations

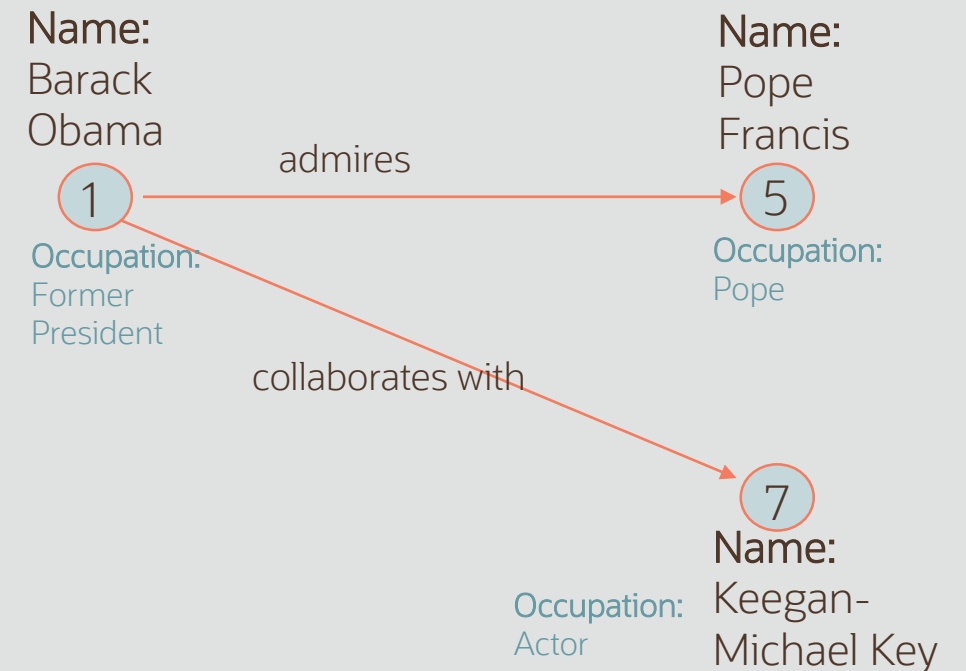
org_id	name	type	religion	genre	country
	CBS	TV Netw			
	Amazon	Company			

People

person_id	name	company	occupation	country
5	Pope Francis		Pope		Vatican
1	Barack Obama		44 th Pres.		USA
7	Keegan-Michael Key		Actor		USA

Relationships

relation_id	from_id	from_type	to_id	to_type	relation_type
	1	person	5	person	admires
	1	person	7	person	collaborates



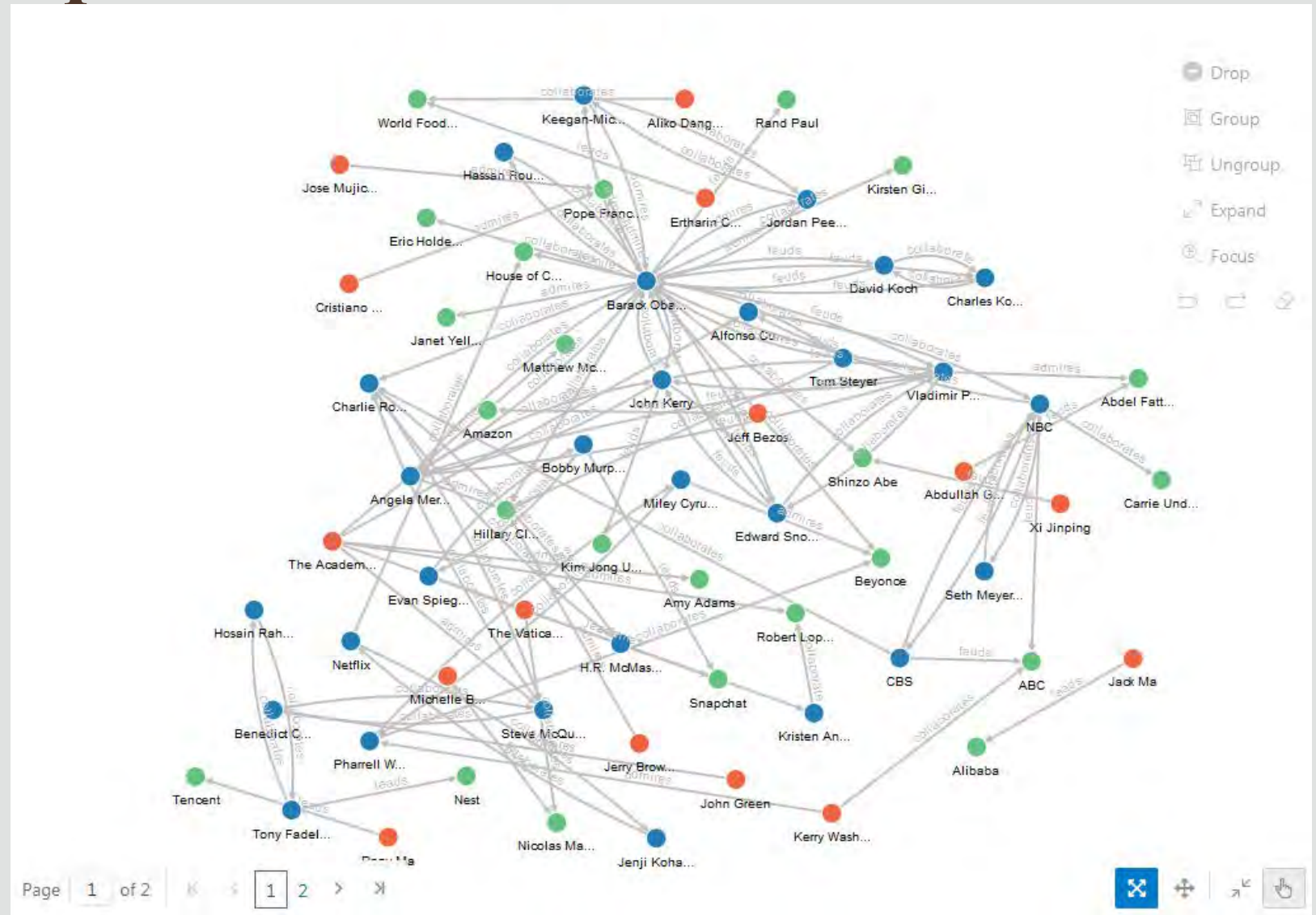
Query Graphs

PGQL Graph Query Language

- Graph pattern matching
 - (person) –[:collaborates] -> (person)
- Basic patterns and reachability patterns
 - Can we reach from A to B with an arbitrary number of hops?
- Familiarity for SQL users
 - Similar language constructs and syntax
 - SELECT ... WHEREGROUP BY ... ORDER BY
 - “Result set” (table) as output

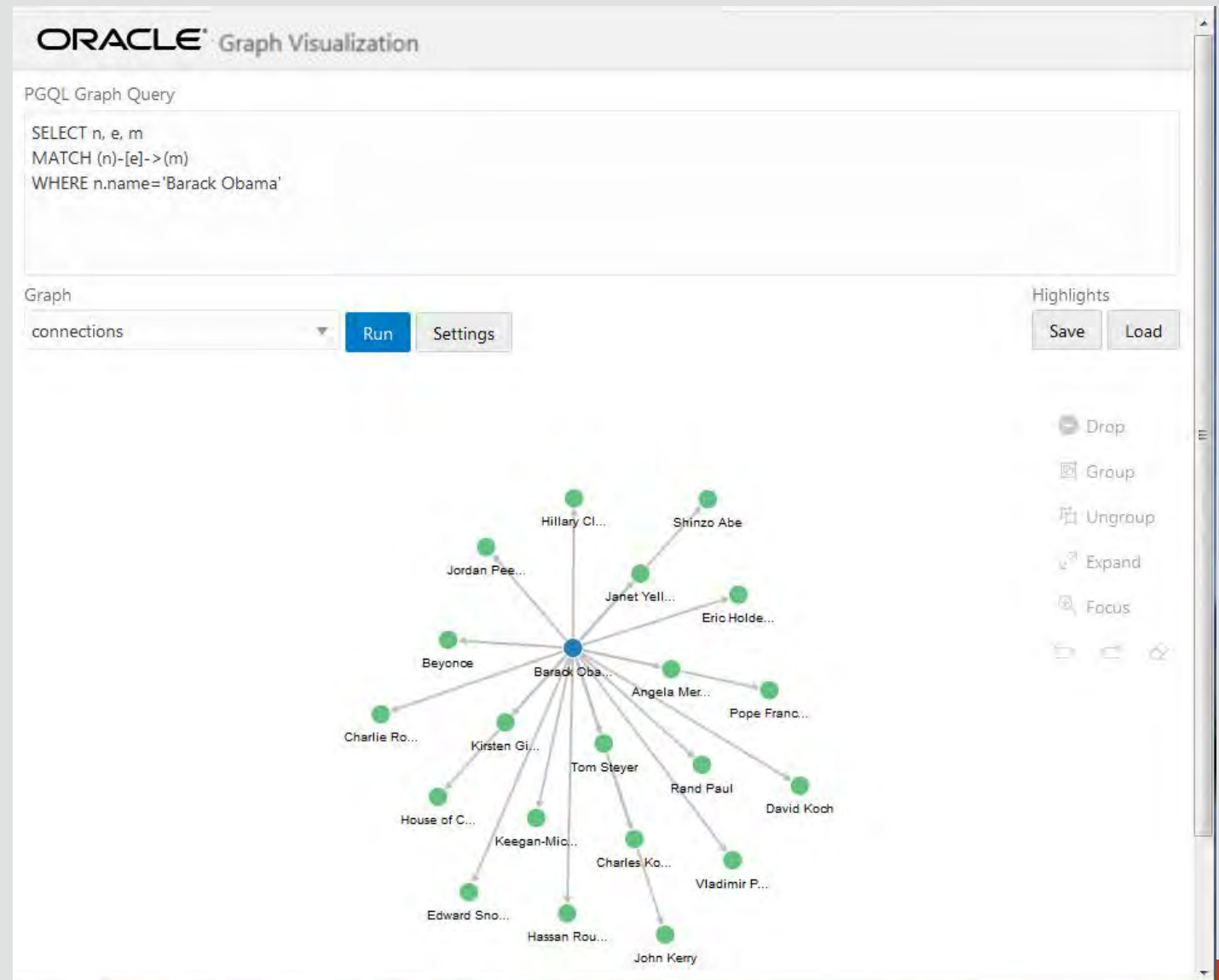
PGQL Examples

```
SELECT e  
MATCH ()-[e]->()
```



PGQL

```
SELECT e
MATCH (n)-[e]->(m)
WHERE n.name='Barack Obama'
```



PGQL

```
PATH any_edge as ()-[]-()  
SELECT n, m MATCH(n) -/: any_edge*/ ->(m)  
WHERE n.name = 'Barack Obama'
```

```
PATH any_edge as ()-[]-()  
SELECT n, m MATCH(n) -/: any_edge*/ ->(m)  
WHERE n.name = 'Barack Obama'
```

Graph

connections

Highlights

- Drop
- Group
- Ungroup
- Expand
- Focus

Page 1 of 1



PGQL

```
SELECT label(e), count(*)  
MATCH () -[e]-> ()  
GROUP BY label(e)
```

label(e)	count(*)
admires	28
leads	9
feuds	45
collaborates	82

PGQL

```
SELECT n, e, m  
MATCH (n)-[e:collaborates]->(m)
```

ORACLE Graph Visualization

PGQL Graph Query

```
SELECT n, e, m  
MATCH (n) - [e:collaborates]->(m)
```

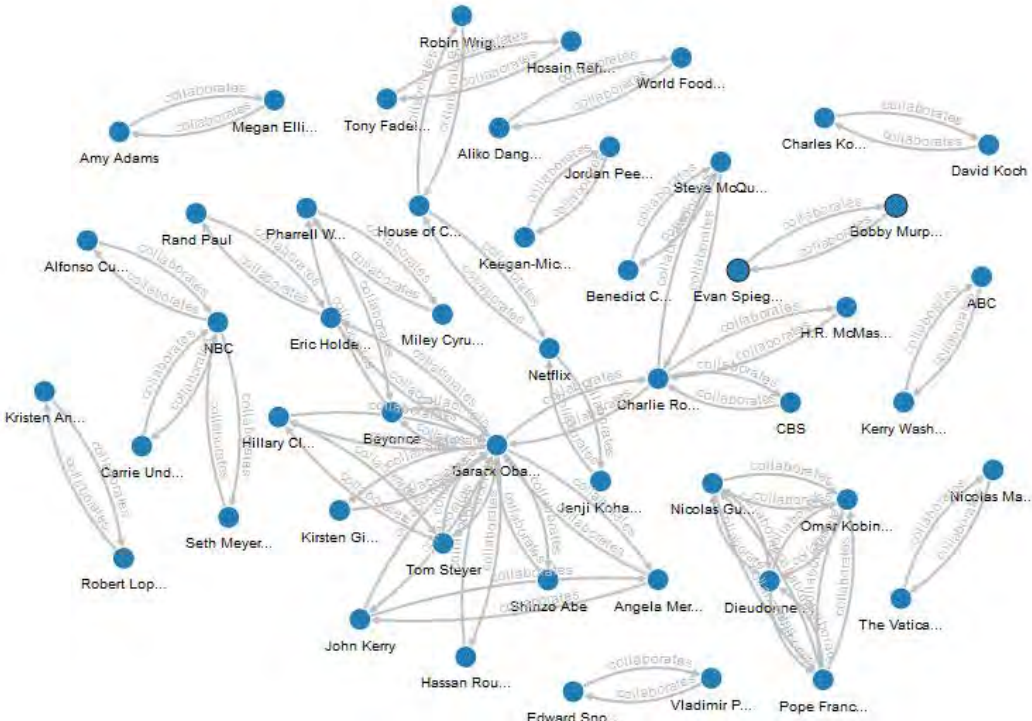
Graph

connections

Run Settings

Highlights

Save Load



- Drop
- Group
- Ungroup
- Expand
- Focus



Analyze Graphs

Analyzing the Marvel Graph

```
g = session.readGraphWithProperties("config.json")  
  
analyst.pagerank(g)  
  
analyst.vertexBetweennessCentrality(g)  
  
g.publish(VertexProperty.ALL, EdgeProperty.ALL)
```


Displaying Query Results with Pagerank

ORACLE Graph Visualization

PGQL Graph Query

```
1 SELECT n, n0, n1, e0, e1, e2, n.pagerank, n0.pagerank, n1.pagerank
2 MATCH (n)-[e0]-(n0)-[e1]-(n1), (n)-[e2]-(n1)
3 WHERE ID(n0) = 'IRON MAN/TONY STARK '
4 ORDER BY n.pagerank DESC, n0.pagerank DESC, n1.pagerank DESC LIMIT 30
```

Graph

sub-graph_5 ↶ ↷

Highlights

-
-
-
-
-

Graph Attributes

- Vertices
 - n1
 - n
 - n0
- Edges

Page 1 of 1 | 1

Copyright © 2014, 2019 Oracle and/or its affiliates All rights reserved.

PGQL Graph Query

```

1 SELECT n, n0, n1, e0, e1, n.betweenness, n0.betweenness, n1.betweenness
2 MATCH (n)-[e0]-(n0)-[e1]-(n1)
3 WHERE ID(n) = 'SPIDER-MAN/PETER PAR' AND n1.betweenness > 0 AND n0.betweenness > 0
4 ORDER BY n.betweenness DESC, n0.betweenness ASC, n1.betweenness ASC LIMIT 500
    
```

Graph

sub-graph_5

Run

Settings



Highlights

Save

Load

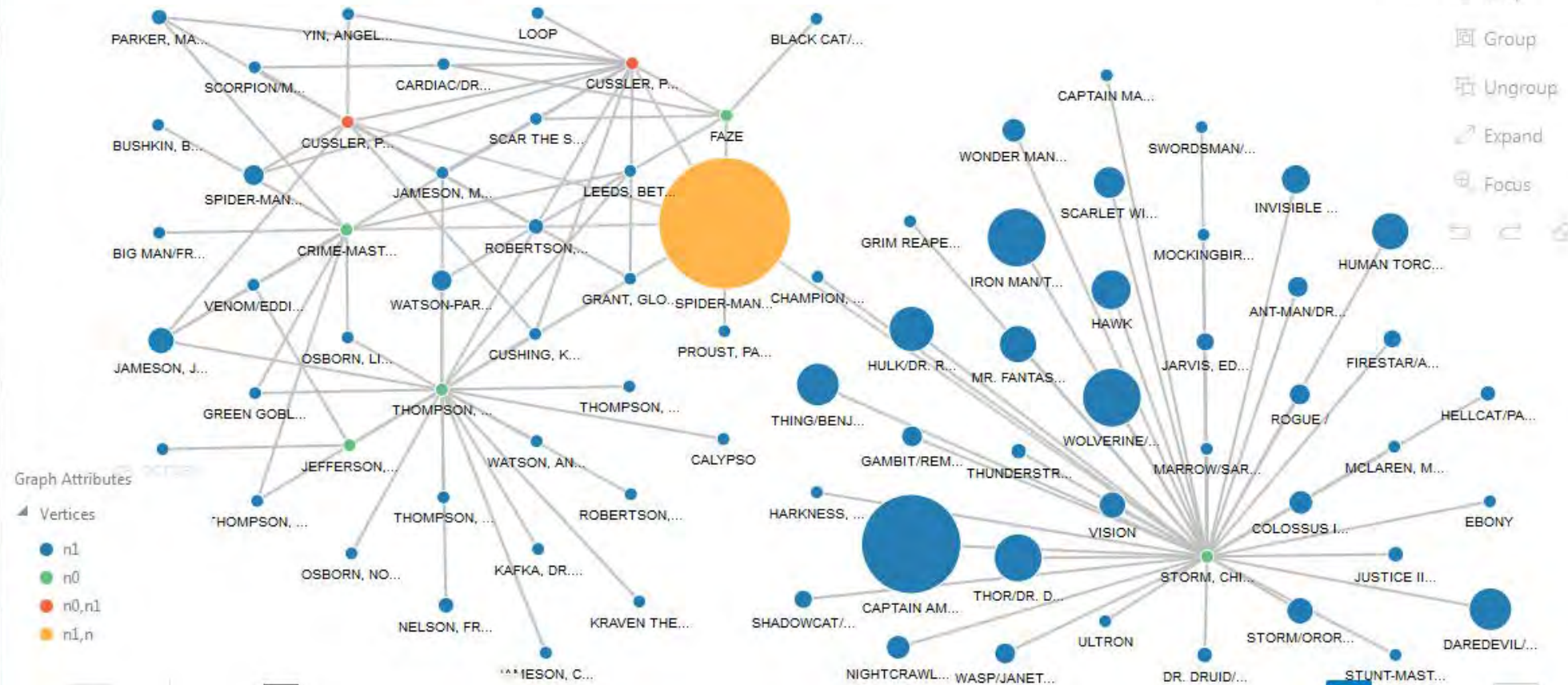
Drop

Group

Ungroup

Expand

Focus



Graph Attributes

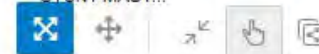
Vertices

- n1
- n0
- n0,n1
- n1,n

Page

1 of 5

1 2 3 4 5 >



Compute Shortest Path

Find shortest path between “WARR” and “MERREE”

```
w = g.createEdgeProperty(PropertyType.DOUBLE, "weight")
w.fill((Double) 1.0)
src = g.getVertex("WARR")
dst = g.getVertex("MERREE")
path = analyst.shortestPathDijkstra(g,src,dst,w)

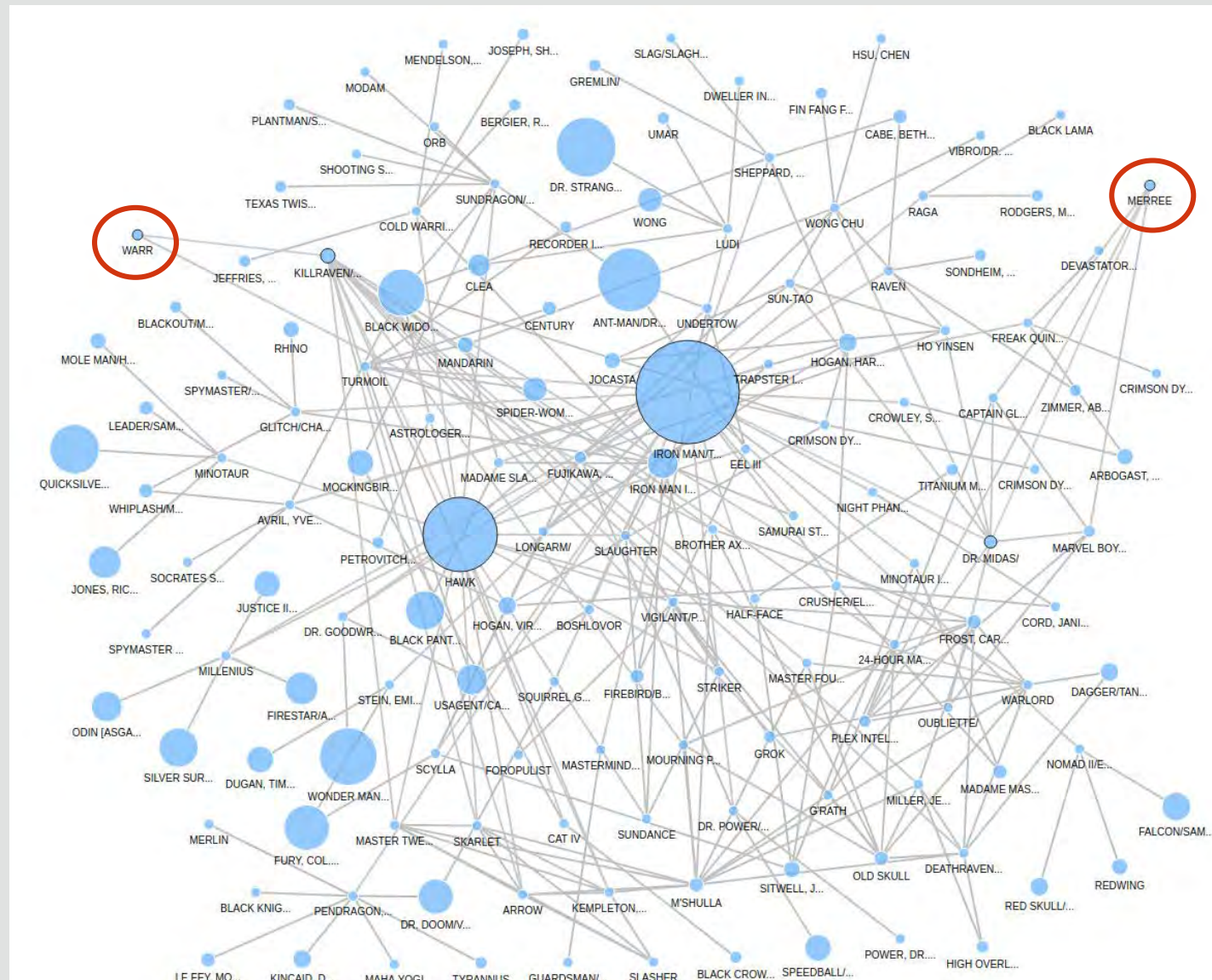
// Creating boolean properties to highlight vertices and edges in the path easier
in_path = g.createEdgeProperty(PropertyType.BOOLEAN, "in_path")
for (PgxE edge: path.edges) {
    in_path.set(e, true);
}

v_in_path = g.createVertexProperty(PropertyType.BOOLEAN, "v_in_path")
for (PgxVertex v: path.vertices) {
    v_in_path.set(v, true);
}

// Shortest Path
SELECT n, e, m, e2, m2, e3, m3, n.pagerank, m.pagerank, m2.pagerank, m3.pagerank, e.in_path, e2.in_path, e3.in_path,
n.v_in_path, m.v_in_path, m2.v_in_path, m3.v_in_path
MATCH (n)-[e]-(m), (n)-[e2]-(m2), (m2)-[e3]-(m3)
WHERE e.in_path AND e2 != e AND e3 != e AND e2 != e3 AND NOT m2.v_in_path AND OUT_DEGREE(m2) < 10 AND NOT m3.v_in_path
ORDER BY n.pagerank ASC, m2.pagerank ASC LIMIT 500
```

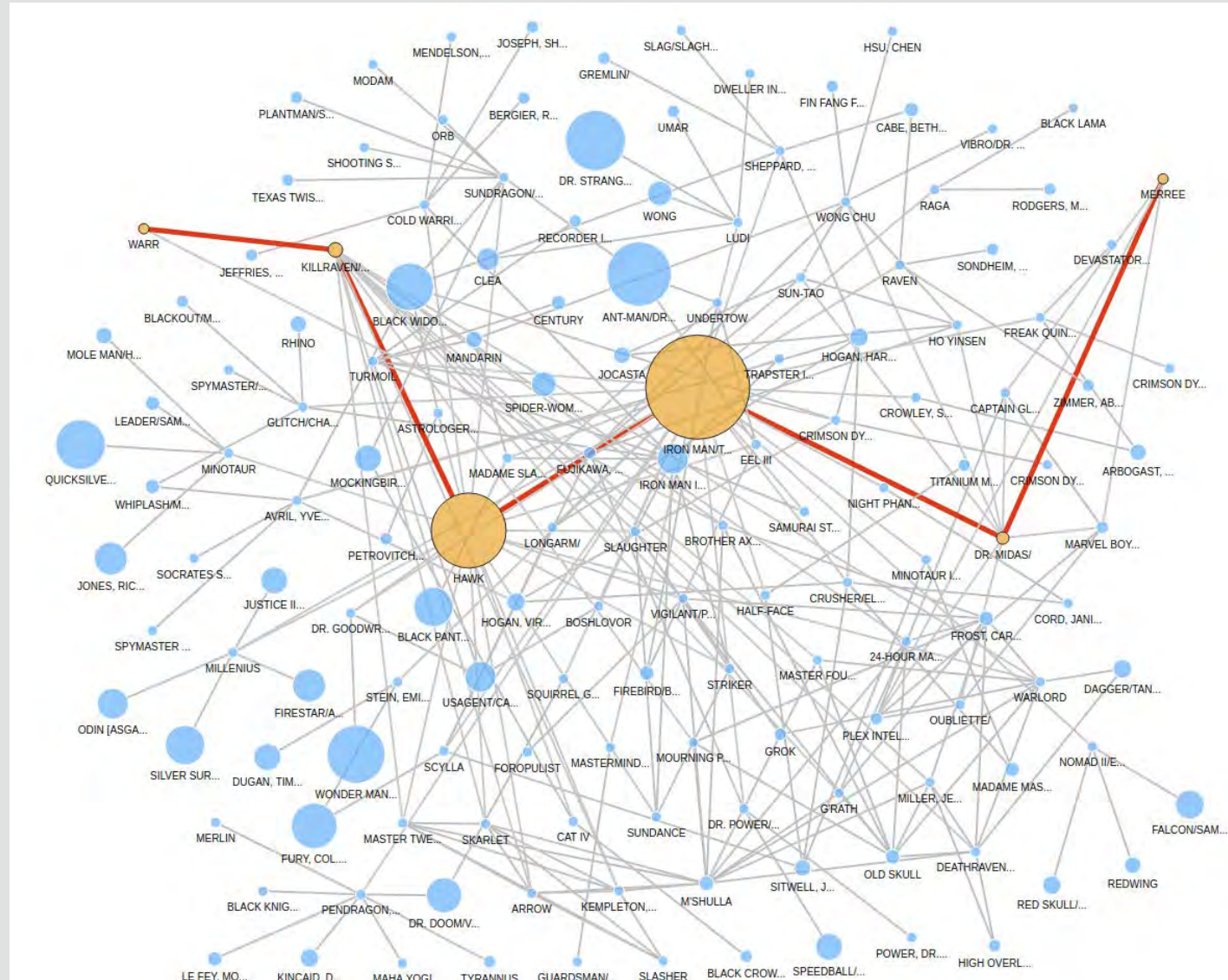
Compute Shortest Path

Find shortest path
between "WARR"
and "MERREE"



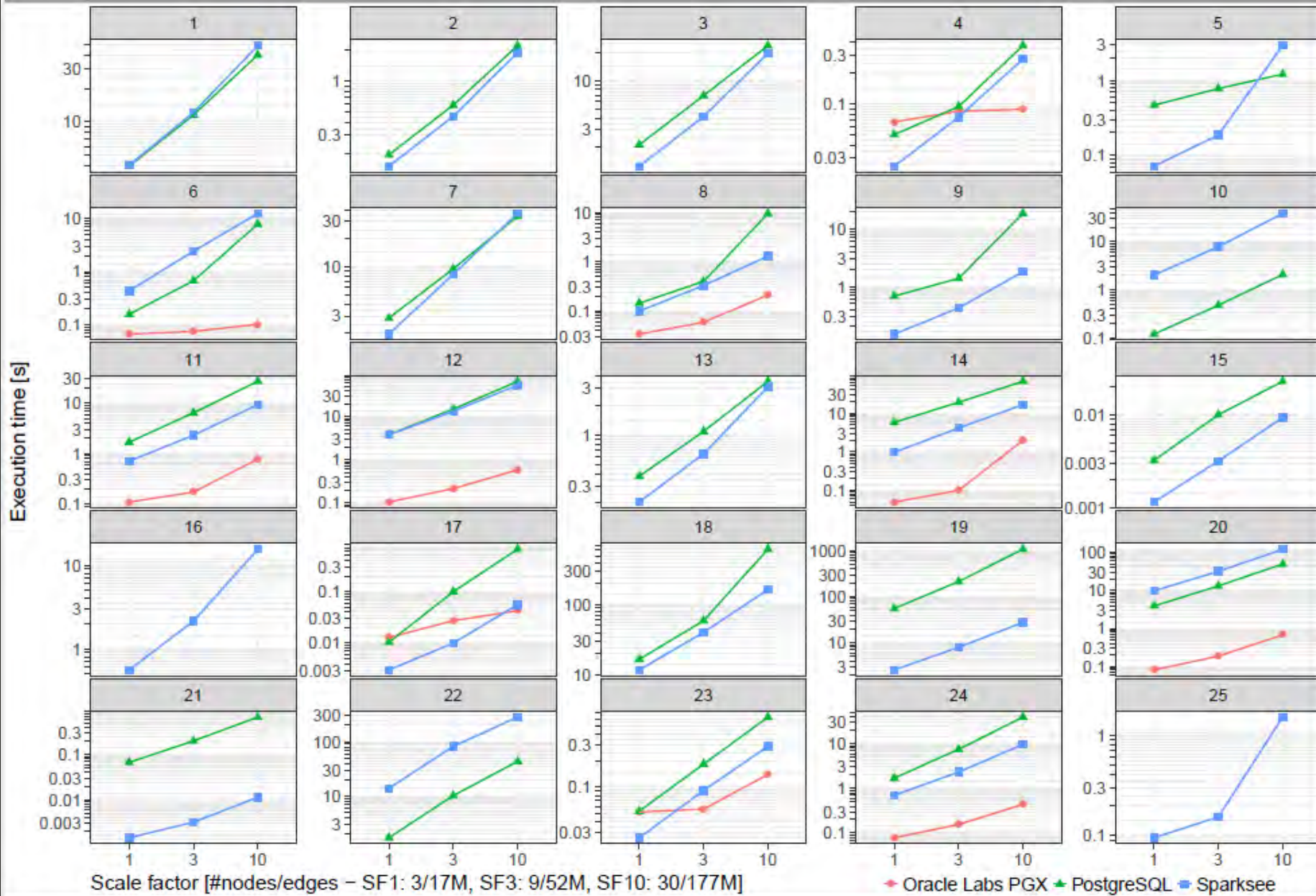
Compute Shortest Path

Find shortest path
between “WARR”
and “MEREE”



Performance and Scale





LDBC benchmark

Scale factor 1
 Number of vertices: 3,181,724
 Number of edges: 17,256,038
 Memory footprint: 1.3GB

Scale factor 3
 Number of vertices: 9,281,922
 Number of edges: 52,695,735
 Memory footprint: 4.0GB

Scale factor 10
 Number of vertices: 29,987,835
 Number of edges: 176,623,445
 Memory footprint: 13.3 GB

- Oracle
- ▲— Postgres SQL
- Sparksee



Property Graph Sizing Recommendations

Table 1-1 Property Graph Sizing Recommendations

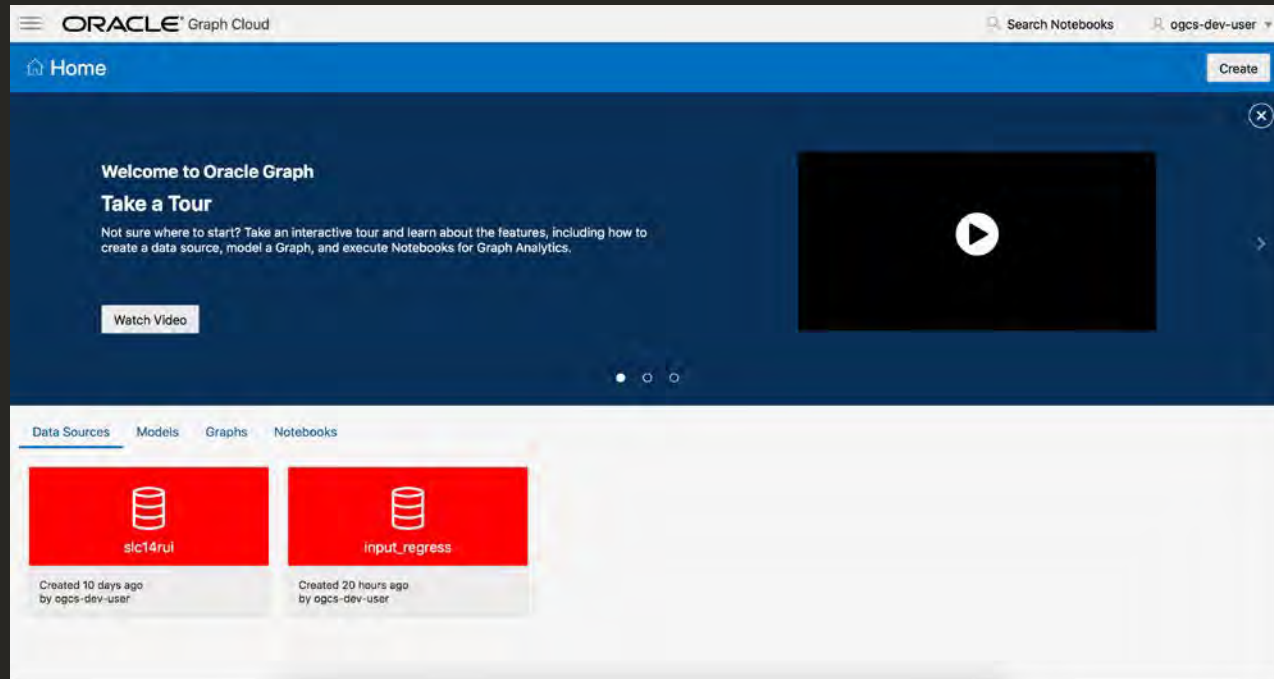
Graph Size	Recommended Physical Memory to be Dedicated	Recommended Number of CPU Processors
10 to 100M edges	Up to 14 GB RAM	2 to 4 processors, and up to 16 processors for more compute-intensive workloads
100M to 1B edges	14 GB to 100 GB RAM	4 to 12 processors, and up to 16 to 32 processors for more compute-intensive workloads
Over 1B edges	Over 100 GB RAM	12 to 32 processors, or more for especially compute-intensive workloads

Future Plans

Graph Cloud Service



Graph Cloud Service



Automated, end-to-end analytic service

Fully managed

- Create graphs
 - Automated modeler creates graphs from database tables and file formats
- Explore connections
 - Using visualization tools, notebooks, and query languages
- Analyze relationships
 - With pre-built analytics, visualization tools, and query languages

Why use Graph Cloud?

Quick to Deploy

Graph Studio, a new paradigm for data scientists and analysts to build graph applications

Quick to Get Started

Industry-specific analytic workflows

Simplified Graph Application Development

Automatic graph model creation from relational tables

Easy to Share Results

Publish and share analysis in notebooks or REST endpoints

Graph Cloud



Home

Graph Modeler

Models

Graphs

Import

Notebooks

Developer and Data Scientist Friendly

The image displays two screenshots of the Oracle Graph Cloud interface, demonstrating its capabilities for developers and data scientists.

Top Screenshot: Sales Analysis

The interface shows a workspace titled "Sales Analysis". The main content area contains the question "Which products did a customer buy?" and a sub-heading "Expressing a query to find out what products a specific customer bought becomes natural using PGQL pattern matching:". Below this, there is a query editor with a PGQL query:

```
1 %pgql
2
3 // Let's take a look at (part of) the graph
4
5 SELECT n,e,m FROM fdGraph MATCH (n) -[e]-> (m) LIMIT 5
```

The graph visualization below the query shows a network of nodes and edges. The nodes are labeled with alphanumeric strings, and the edges represent relationships between them.

Bottom Screenshot: Fraud Detection and Money Launderin...

The interface shows a workspace titled "Fraud Detection and Money Launderin...". The main content area contains a query editor with a PGQL query:

```
1 %pgql
2
3 // Let's take a look at (part of) the graph
4
5 SELECT n,e,m FROM fdGraph MATCH (n) -[e]-> (m) LIMIT 5
```

The graph visualization below the query shows a network of nodes and edges. The nodes are labeled with alphanumeric strings, and the edges represent relationships between them. The interface also includes a "Graph Attributes" panel on the left, showing "Vertices" and "Edges" with corresponding icons.

Graph Studio

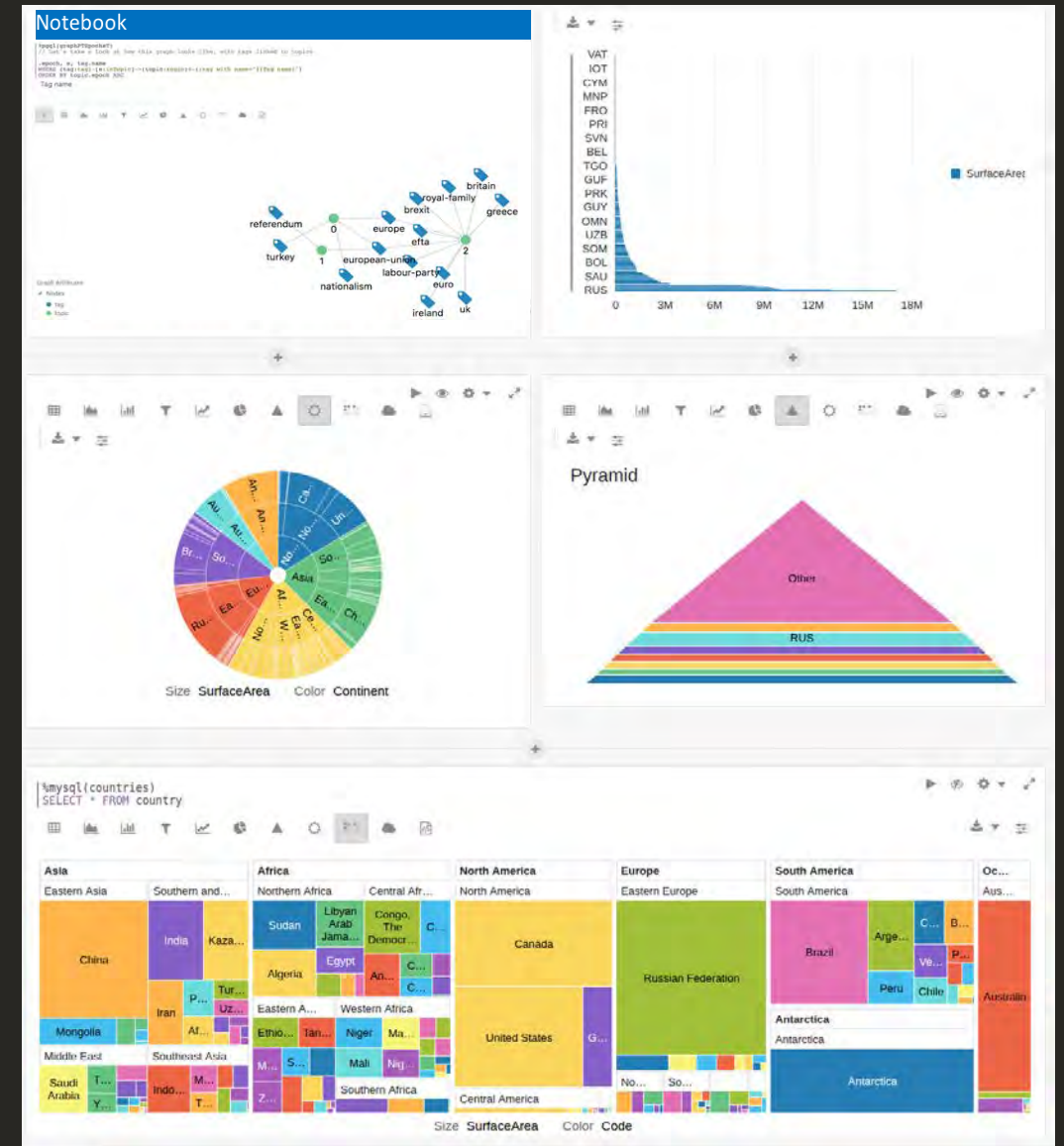
Graph Notebook

Apache Zeppelin-compatible interpreters

User Interface built with Oracle JET
Consistent look-and-feel

Powerful and interactive visualization

Custom graph visualization



Extensions to the SQL Standard



SQL/PGQ

- SQL extensions to query property graphs
 - Our team is working with ISO and ANSI committees
 - Target: Next version of SQL (~2020/21)

Create a Property Graph using SQL Data Definition

```
CREATE PROPERTY GRAPH myGraph
  VERTEX TABLES (Person, Message)
  EDGE TABLES (
    Created SOURCE Person DESTINATION Message,
    Commented SOURCE Person DESTINATION Message )
```

Query a Property Graph in SQL

```
SELECT GT.creationDate, GT.content
FROM myGraph GRAPH_TABLE (
  MATCH
    (Creator IS Person WHERE Creator.email = :email1)
    -[ IS Created ]->
    (M IS Message)
    <-[ IS Commented ]-
    (Commenter IS Person WHERE Commenter.email = :email2)
    WHERE ALL_DIFFERENT (Creator, Commenter)
  ONE ROW PER MATCH
  COLUMNS (
    M.creationDate,
    M.content )
) AS GT
```


SQL/PGQ: Implementation Plan

Transition graph support to align with proposed SQL/PGQ standard

External graph analytics engine
not required

Query and analysis within database

Performance trade-off

Large data: use existing SQL engine

In-memory columnar graph index
for performance

External engine still useful

For alternative deployment models

Big Data audience: Spark, Hadoop,
HBase ...

Scalable distributed in-memory
execution

What's Ahead

Tuesday

- 11:15-12:00** New Tools to Fight Against Financial Crime [CON6222]
Moscone South – Room 204
- 12:30-1:15** Using Graph Analysis and Fraud Detection in the Fintech Industry (Paysafe customer session) Moscone South – Room 152C
- 12:30-1:15** Blazing-Fast Distributed Graph Query Processing: 100x as Fast as Spark [DEV3712] Moscone South – Room 307
- 3:15-4:00** Introducing Oracle Graph Cloud: Automating Graph Analysis [TRN4754]
Moscone South – Room 159B

Wednesday

- 10:00-10:45** Graph Database and Analytics: How To Use Them [TRN 4755]
Moscone South – Room 152C
- 1:30-2:15** Traversing and Querying Graphs with PGQL and Gremlin with Oracle Spatial and Graph [DEV4084] Moscone South – Room 202
- Meet the Experts**
- 1:30-2:20** Graph Database and Analysis
- 2:30-3:20** Graph Cloud Service: Automating Graph Analysis
Lounge C, Code One Groundbreakers Hub, Moscone South Level 1

Demogrounds

Date/Time	Title	Location
Monday 10:00 am – 4:00 pm Tuesday 10:30 am – 5:30 pm Wednesday 10:00 am– 4:30 pm	Spatial and Graph: Database, Analytics and Cloud	Moscone South Exhibit Hall (“The Exchange”) • Oracle Demogrounds > Data Management area > Kiosk # ODB-017



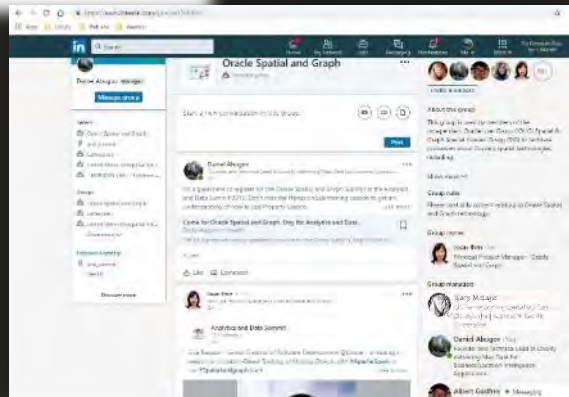
Spatial & Graph
Demos



The Spatial & Graph SIG User Community

Now part of BIWA User Group

We are a vibrant community of customers and partners that connects and exchanges knowledge online, and at conferences and events.



Meet us at OpenWorld! Monday-Wednesday
Moscone West, Level 3, User Group area
at the *BIWA/Analytics Community* table

Join us online

tinyurl.com/oraclespatialcommunity



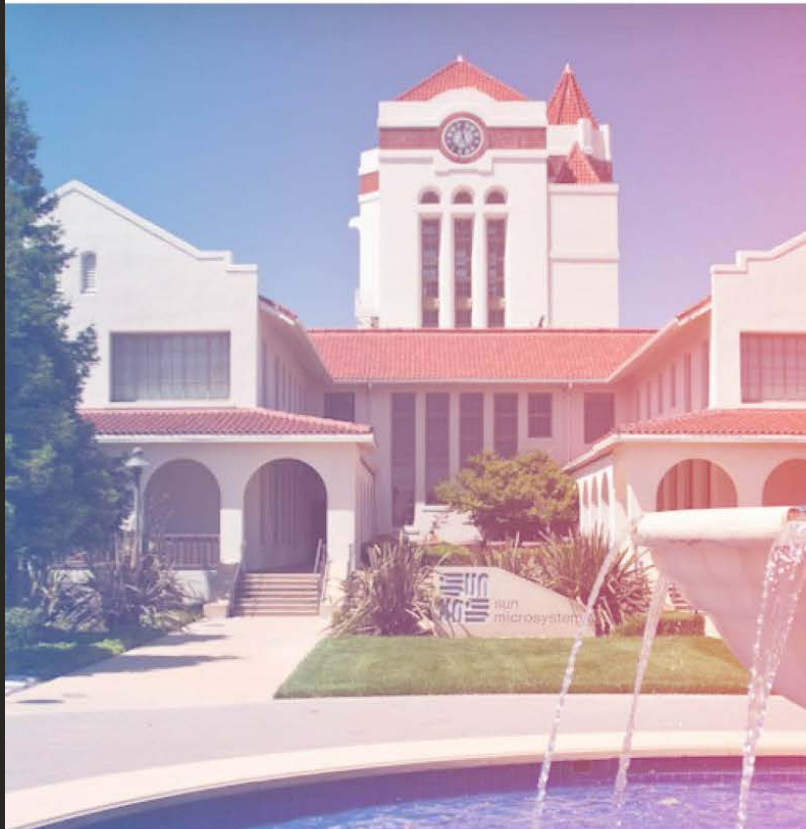
LinkedIn



@oraspatialsig



oraclespatialsig@gmail.com



SAVE THE DATE

ANALYTICS AND DATA SUMMIT 2020

All Analytics. All Data.
No Nonsense.

February 25-27, 2020

Call for Speakers Now Open!

SIGN UP FOR OUR NEWSLETTER

Formerly the BIWA Summit with the Spatial and Graph Summit.

@AnalyticAndData



analyticsanddatasummit.org

Seeking customer use cases and technology sessions
Dedicated Spatial & Graph track with 20+ sessions

Thank You

Melli Annamalai
(Melliyal.Annamalai@oracle.com)

Session Survey

Help us make the content even better. Please complete the session survey in the Mobile App.

