AskTOM Office Hours: Graph Database and Analytics

• Welcome (back) to our AskTOM Graph Office Hours series! We’re back with new product updates, use cases, demos and technical tips.
  https://asktom.oracle.com/pls/apex/asktom.search?oh=3084

• Sessions will be held about once a month

• **Subscribe** at the page above for updates on upcoming session topics & dates.
  And submit feedback, questions, topic requests, and view past session recordings

• **Note**: **Spatial** now has a new Office Hours series for location analysis & mapping features in Oracle Database:
  https://asktom.oracle.com/pls/apex/asktom.search?oh=7761
Financial Industry Use Cases for Graph Analytics

Ryota Yamanaka and Melli Annamalai, Graph Product Management
Gautam Pisharam, Oracle Solutions Engineer Hub
Oracle
May 28, 2020
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.
Program Agenda

1. Graph Applications Overview
   Focus on Financial Services
   Ryota

2. Customer Use Case
   Fraud Detection at a Bank
   Gautam

3. Demo App - How to Detect Cycles
   Ryota

4. How to Combine with Machine Learning
   Customer Reference: Paysafe
   Ryota

5. Customer 360 Analysis Using Graphs
   Customer Reference: Banco Galicia
   Melli
Graph Applications Overview
Application 1 - Query to Follow Money

- The graph representation of the transaction data makes queries **intuitive and performant**.
  - E.g. the queries to *find the circular money transfer over multiple intermediate steps*, can be easily written and executed.

- Traversal queries make users possible to detect **hidden relationships** between accounts and their owners (e.g. paradise papers).

- Complex rules to score **suspicious accounts** (e.g. closeness to known fraud accounts) can be also expressed in graph queries and executed in very short response time.
Application 2 - 360 Degrees Visualization

- The graph visualization component can show the original graph data on web browsers. This is **useful for manual inspection** against "suspicious" accounts (which are often detected by rules or machine learning models).

- Users can "expand" the graph by clicking particular nodes to see the related information.

- Users can also run **PGQL queries** to find the accounts in certain conditions. This helps users examine the existing rules, as well as try and create new rules.
Application 3 - Enhancing ML Model

- New scores for accounts are calculated based on relationships and by rules and graph algorithms.
- The scores are used as input features to enhance the prediction models.

<table>
<thead>
<tr>
<th>ID</th>
<th>Feature1</th>
<th>Feature 2</th>
<th>Feature 3</th>
<th>Feature 4</th>
<th>Feature 5</th>
<th>Feature 6</th>
<th>Feature 7</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>1002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No</td>
</tr>
<tr>
<td>1003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>
Use Case in Practice
Evolution of Fraud: Examining the Cases
Evolution of Fraud: Connecting the Dots

CUSTOMER
- Mobile
- Empowered
- Socially Connected

THE BANK
- Regulations
- Digital Transformation
- ATMs
- Branches

Fintechs
- Payment Processing
- Peer – Peer Lending

Blockchain
Crowdsourced Lending
Identity Verification
E-Wallets
Crowdsourced Lending

Wallet Share
Loyalty
Evolution of Fraud: Connecting the Dots

Property Graph:

- **Name of graph**: financial_transactions
- **Vertex**:
  - Account 1 (number: 10039)
  - Account 2 (number: 2090)
  - Person 1 (name: Camille, worksFor: Company)
  - Person 2 (name: Nikita, ownerOf: Account 3 (number: 8021))
  - Person 3 (name: Liam, ownerOf: Account 4 (number: 1001))

- **Edge**:
  - Transaction amount: $9900.00 (between Account 1 and Company)
  - Transaction amount: $1500.30 (between Person 1 and Account 3)
  - Transaction amount: $3000.70 (between Account 3 and Account 4)
  - Transaction amount: $9999.50 (between Account 4 and Person 3)

- **Label of vertex**:
  - Account: ownerOf, amount
  - Person: name
  - Company: name

- **Label of edge**:
  - Transaction amount

- **Property of vertex**:
  - Account: number
  - Person: name

- **Property of edge**:
  - Transaction amount
Why Oracle

1. Proven scalability and performance
2. Built-in algorithms and visualizations
3. Easy detection of cycles
Connect the Organization

1. IT & Operations
   - Raw Data

2. Database Management

3. Business Analytics
   - Graph
   - Algorithms
   - Statistics

4. Data Science
   - Vertex tables:
     - Person
       - id | name | dob
       - 1  | Riya | 1995-03-20
       - 2  | Kathrine | 1994-01-15
       - 3  | Lee | 1996-01-29
   - Edge tables:
     - knows
     - person1_id | person2_id
       - 1 | 1
       - 2 | 2
       - 3 | 3

5. Risk Analytics
   - student_network
     - Person
       - name: Riya | dob: 1995-03-20
       - name: Kathrine | dob: 1994-01-15
       - name: Lee | dob: 1996-01-29
     - University
       - name: UC Berkeley
     - knows
     - studentOf
       - person_id | university_id
         - 1 | 1
         - 2 | 1
         - 3 | 1

Copyright © 2020, Oracle and/or its affiliates. All rights reserved.
Our Approach
Approach 1 vs Approach 2

In-memory engine

Graph Viz

Graph

Relational

Database 19c

Raw Data

In-memory engine

Graph Viz

Graph

Relational

Database 19c

Raw Data
Graph Analytics - 50+ Built-in Algorithms

Detecting Components and Communities
- Strongly Connected Components
- Weakly Connected Components
- Label Propagation
- Conductance Minimization
- Infomap

Ranking and Walking
- PageRank
- Personalized PageRank
- Degree Centrality
- Closeness Centrality
- Vertex Betweenness Centrality
- Eigenvector Centrality
- HITS
- SALSA
- Random Walk with Restart

Evaluating Structures
- Adamic-Adar Index
- Conductance
- Cycle Detection
- Degree Distribution
- Eccentricity
- K-Core
- LCC
- Modularity
- Reachability Topological Ordering
- Triangle Counting

Path-Finding
- Shortest Path (Bellman-Ford, Dijkstra, Bidirectional Dijkstra)
- Fattest Path
- Compute Distance Index
- Enumerate Simple Paths
- Fast Path Finding
- Hop Distance

Link Prediction
- WTF (Who to follow)

Others
- Minimum Spanning-Tree
- Matrix Factorization
Basic Graph Pattern Matching

```
SELECT v3.name, v3.age
FROM socialNetworkGraph
MATCH (v1:Person) -[:friendOf]-> (v2:Person) -[:knows]-> (v3:Person)
WHERE v1.name = 'Amber'
```

**Query:** Find all people who are known by friends of ‘Amber’.

... and parallel graph mutation operations

- **Create Bipartite Graph**
- **Create Undirected Graph**
- **Sort-By-Degree (Renumbering)**
- **Simplify Graph**
- **Filtered Subgraph**

The original graph

Left Set: “a,b,e”

Filter-Expression
Example: Determining Betweenness Centrality

Code snippet

```java
analyst.vertexBetweennessCentrality(pg).getTopKValues(15)
```

Identify influencers
Graph Visualization
Performance Results

- 73 Million Customers
- 137 Million Transactions
- 0.3 Million Cycles
The Oracle Graph Edge

Enable data scientists to create graph models, faster

Comprehensive Platform to Consume all Data quickly

Digital Ecosystem for Complete Graph Cycle

Ensure Business Change
Demo App - How to Detect Cycles
Demo - How to Detect Cycles

How we can find (the shortest) cycle in graphs, starting from one specific node:

Mock data
• 1,000 nodes (bank accounts) + 5,000 edges (money transfers)
• Each node has 5 outgoing edges to randomly selected nodes
• Not consider: transaction date, transaction amount, ...

Scenario
• Select one specific account by its ID
• Show how many nodes are connected in k hops from this account
• Search for a cycle starting from and ending at this account
Demo - How to Detect Cycles

Count the number of the accounts connected in 6 hops from account ID=1

```
SELECT DISTINCT COUNT(*)
MATCH (n)-[:transfer{1,6}]->(m)
WHERE ID(n) = 1
```

Search for the shortest cyclic path starting from and ending at the account ID=1
  • The 1st shortest path has no edge (= 0 hop)
  • The 2nd shortest path is the shortest cyclic path

```
SELECT n, ARRAY_AGG(ID(m)), ARRAY_AGG(ID(e))
MATCH TOP 2 SHORTEST ((n) -[e:transfer]->(m))* (n)
WHERE ID(n) = 1
```
Demo - How to Detect Cycles

This demo content is available here:

The Docker containers include
• Mock data (in files)
• Graph Visualization
• Zeppelin notebook
Using Built-in Tools

The cycles will be queried and visualized using the following built-in tools:

- Zeppelin notebook
- Graph Visualization app
Using Built-in Tools

![Image of Zeppelin notebook showing detected cyclic paths and a query for finding shortest paths.](image)
Using Built-in Tools
Implementing Custom App

Custom applications can be implemented with **Graph Client Java API**.

In this demo, the components are:

- **REST server** written in Java (using Javalin)
- Custom visualization in JavaScript (using D3.js)
Implementing Custom App

This sample REST API returns subgraph and cycle in JSON format.

- Graph in 6 hops: `http://<host:port>/travarsal?node_ids=1&iteration=6`
- Cycle: `http://<host:port>/cycle?node_ids=1`

```json
```
Implementing Custom App
How to Combine with Machine Learning
Machine Learning

**Information coverage** of the training dataset is important to make good predictive models.
Graph database potentially provides more information because of its:
  • **flexible** model
  • **algorithm** capability
Example - Mule Account Detection

- **Mule accounts** are often stolen accounts and transfers money illegally
- Suspicious accounts are flagged by human curation
  - The system should be able to predict the human decision (= **objective variable**)
- If machine learning can make the predictive model?
  - But the accounts themselves has limited information (= **explanation variables**)
- Is it possible to generate **more features** based on the relationships between accounts (e.g. transaction patterns, family relationships, ...)

![Diagram showing relationships between accounts and flagged accounts with is_fraud variable](image)
Example - Mule Account Detection

Feature 1

• If the owner of this account is sharing personal information with others

```
SELECT a1, COUNT(s)
MATCH (a1)<-[[:owns]]-(c1)-(s)-(c2)-[:owns]->(a2)
WHERE a1 != a2 AND c1 != c2
GROUP BY a1
```
Example - Mule Account Detection

Feature 2

- How many fraud accounts exist in the same money transfer community

```
alyst.communitiesLabelPropagation(G, 100)
```

```
SELECT a.community, COUNT(a)
MATCH (a)
WHERE a.type = 'Account'
    AND a.is_fraud = 'true'
GROUP BY a.community
```
Example - Mule Account Detection

Feature 3
- Closeness to know fraud accounts

```python
analyst.personalizedPagerank(G, fraud_accounts)
```

```sql
SELECT a.account_no, a.pagerank, a.is_fraud
MATCH (a) WHERE a.type = 'Account'
ORDER BY a.pagerank DESC
```
Example - Mule Account Detection

Database

Account

- name
- age
- occupation
- branch
- balance
- ... 
- share_info
- fraud_community
- fraud_closeness
- ...
- is_fraud
Example - Mule Account Detection

<table>
<thead>
<tr>
<th>ID</th>
<th>age</th>
<th>occupation</th>
<th>balance</th>
<th>share_info</th>
<th>fraud_com</th>
<th>fraud_clo</th>
<th>is_fraud</th>
<th>prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes 85%</td>
</tr>
<tr>
<td>1002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No</td>
<td>Yes 78%</td>
</tr>
<tr>
<td>1003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No</td>
<td>No 62%</td>
</tr>
</tbody>
</table>
More Algorithms

**RandomWalk**
- Closeness to other nodes

**DeepWalk**
- Collect sequences by RandomWalk
- Use the sequences as input of Skip-gram
- Obtain the tensor for each node
Customer Reference - Paysafe

• Providing **online payment** solutions
  • Real-time payments, e-Wallets
  • 1B revenue/year, 500K payments/day

• Strong demand for **fraud detection**
  • Combination with rule-based approach and machine learning
  • In real-time, upon money movement

• More information
  • AnD Summit: [2020 slides](#), [2019 slides](#)
  • YouTube: [video](#)
Customer Reference - Paysafe (Query)

- The queries to follow transactions **multiple hops** need to join the tables multiple times.
- The existing complex SQL queries were rewritten into PGQL queries.
- Queries become **much simpler**
  - e.g. 32 lines --> 7 lines
- Queries become **much faster**
  - e.g. 50 min --> 0.5 sec
- Possible to run the queries that didn't complete in reasonable time by SQL.

---

### Performance Benchmark

Payments up to the 4th hop on an active customer

- SQL created by Paysafe (32 lines)
  - 1 day: 50 min 20 sec
  - 1 week: (cancelled after 4 hours)
  - 1 month: (did not even try)
- SQL optimized by Oracle (62 lines)
  - 1 day: 20.3 sec
  - 1 week: 8 min 33 sec
  - 1 month: (cancelled after 6 hours)
- 4 PGQL queries (7 lines each)
  - 1 day: 0.547 sec
  - 1 week: 0.588 sec
  - 1 month: 0.597 sec

---

[Analytics and Data (AnD) Summit 2020 Presentation](#)
Customer Reference - Paysafe (Visualization)

- Visualization of money flows is essential to manually check the activities of suspicious accounts, as well as to understand the common fraud patterns.
- Using this custom visualization application, interesting networks such as possible money laundering flows (typically consists of 3 steps: placement, layering, integration) were detected.
- Using device fingerprints, the multiple accounts using the same device, and their money flows are also visualized.
Customer Reference - Paysafe (Machine Learning)

- Introduction of graph visualization and graph analytics helps the investigations of fraud specialists, and saves a lot of time and effort.

- For further automation, the fraud patterns should be also detected without human intervention. For this goal, machine learning will be combining with graphs.

- Graph embedding techniques such as DeepWalk was experimented, and this feature improved the accuracy of the machine learning model significantly.

Analytics and Data (AnD) Summit 2020 Presentation
Customer 360 Analysis Using Graphs
Money Transfer Use Case

Actionable Targets

• **Cash Deposits**: Identify Galicia’s accounts with many cash deposits that then transfer their money to self accounts in another Banks

• **Value Chain**: Identify Galicia’s big accounts who pay their providers through other banks
Graph Design

Account 100

Account 101

Account 100 Bank A

Account 100 Bank B

Account 101 Bank A

Account 101 Bank B

Account 200

Account 300

Account 400

Account 500

Account 600

has

transfers_t

transfers_t

transfers_t

transfers_t

transfers_t

transfers_t

transfers_t

transfers_t

transfers_t

Copyright © 2020, Oracle and/or its affiliates. All rights reserved.
Money Transfer Use Case - Discoveries
Money Transfer Use Case - Discoveries

imp_emitido_usd = 0.0
Comunidad = 1763330
imp_recibido_ars = 4.0215E9
PageRank = 1.2558642E-7
cuenta_origen = 5079001023
imp_emitido_ars = 6.159450035E8
Betweenness = 4094.3164
cant_ingresos_ars = 125
codigo_entidad_destino = 7
entidad_destino = BANCO DE GALICIA Y BUENOS AIRES S.A.
imp_emitido_usd = 0.0
cant_egresos_ars = 838
tipo_nodo = CUENTA ORIGEN - BANCO DESTINO
OutDegree = 42
InDegree = 1
entidad_origen = BBVA BANCO FRANCES S.A.

imp_emitido_ars = 4.0215E9
tipo_arco = CUENTA - CUENTA ORIGEN_BANCO DESTINO
codigo_entidad_destino = 17.0
entidad_destino = BBVA BANCO FRANCES S.A.
tipo_transferencia = Interbanking
imp_emitido_usd = 0.0
cant_egresos_ars = 12.1
entidad_origen = BANCO DE GALICIA Y BUENOS AIRES S.A.
tipo_operacion = CTAS. PROPIAS
Money Transfer Use Case - Discoveries

- **Discoveries**
  - On its Galicia Bank account CUSTOMER receives 73MM ars in cash from more than 10M deposits
  - From its Galicia Bank account CUSTOMER transfers more than 400MM ars to a self account on Banco Frances.
  - Banco Frances account transfers more than 600MM ars to different providers with accounts on Galicia Bank

- **Actionable**: Payment to providers should be done through Banco Galicia and not Banco Frances so that the money stays within Galicia’s circuit. Otherwise a higher fee could be charged given this huge amount of deposits
Credit Entity: NO Banco Galicia
Money Transfer Use Case - Discoveries
Money Transfer Use Case - Discoveries

Debit Entity: Citibank
Money Transfer Use Case - Discoveries

• Discoveries
  • 50% of operations between self accounts
  • Providers payment mainly through Citi Bank
  • CUSTOMER uses Galicia’s Investment Funds to then withdraw money + interests and pay Providers
  • CUSTOMER’s chain of providers is outside Galicia’s circuit
• Actionable: Payment to providers should be done through Galicia accounts instead of Citi’s
Useful Links
AskTOM Office Hours: Graph Database and Analytics

- Welcome (back) to our AskTOM Graph Office Hours series! We’re back with new product updates, use cases, demos and technical tips https://asktom.oracle.com/pls/apex/asktom.search?oh=3084
- Sessions will be held about once a month
- **Subscribe** at the page above for updates on upcoming session topics & dates
  And submit feedback, questions, topic requests, and view past session recordings
- **Note**: **Spatial** now has a new Office Hours series for location analysis & mapping features in Oracle Database:
  https://asktom.oracle.com/pls/apex/asktom.search?oh=7761
Helpful Links

- Graphs at Oracle
  https://www.oracle.com/goto/graph

- Oracle Property Graph
  http://www.oracle.com/goto/propertygraph

- Blog: Examples, Tips and Tricks


- Social Media
  - Twitter: @OracleBigData, @SpatialHannes, @JeanIhm, @ryotaymnk
  - LinkedIn: Oracle Spatial and Graph Group
  - YouTube: youtube.com/c/OracleSpatialandGraph
The Spatial & Graph User Community

• A part of Analytics and Data Oracle User Community (formally BIWA)
• Vibrant community of tech enthusiasts including customers, partners, students
• We share knowledge online, and at conferences and events
• Global – Americas, Europe, Africa, Asia

LinkedIn Oracle Spatial and Graph group
linkedin.com/groups/1848520/

@oraspatsig
oraclespatialsig@gmail.com
AskTOM Office Hours: Graph Database and Analytics

• Today’s session will be repeated for EMEA/Americas on May 28 (17:00 Central Europe | 11:00 New York | 08:00 San Francisco)

• Next session - save the date:
  • **Topic:** Building Recommendation Systems with Graphs
  • **Date:** Likely June 25th -- *Check back at landing page for details*

• Recording of today’s session will be available at the landing page

• **Subscribe** for updates on upcoming session topics & dates

• Submit feedback, questions, topic requests, and view past session recordings [https://asktom.oracle.com/pls/apex/asktom.search?oh=3084](https://asktom.oracle.com/pls/apex/asktom.search?oh=3084)
Thanks for attending! See you next time.

https://asktom.oracle.com/pls/apex/asktom.search?oh=3084