

A person wearing a yellow shirt is crawling on a large map of the United States that is spread out on a wooden floor. The map is oriented with the top of the country towards the top left of the frame. The person is positioned in the upper right quadrant of the map, appearing to be looking at or touching the map. The background is a warm, light brown color, matching the floor.

Grafting Grifters: Identify & Display Patterns of Corruption With Oracle Graph

TechCast Demo Days 2022:

Graph Day

May 25, 2022

Jim Czuprynski

@JimTheWhyGuy

Zero Defect Computing, Inc.

Who Am I, and What Am I Doing Here?



ORACLE®
ACE Director

ORACLE®
Certified Professional



- E-mail me at jim@jimthewhyguy.com
- Follow me on Twitter (@JimTheWhyGuy)
- Connect with me on LinkedIn (Jim Czumprynski)

BEYOND TECH SKILLS

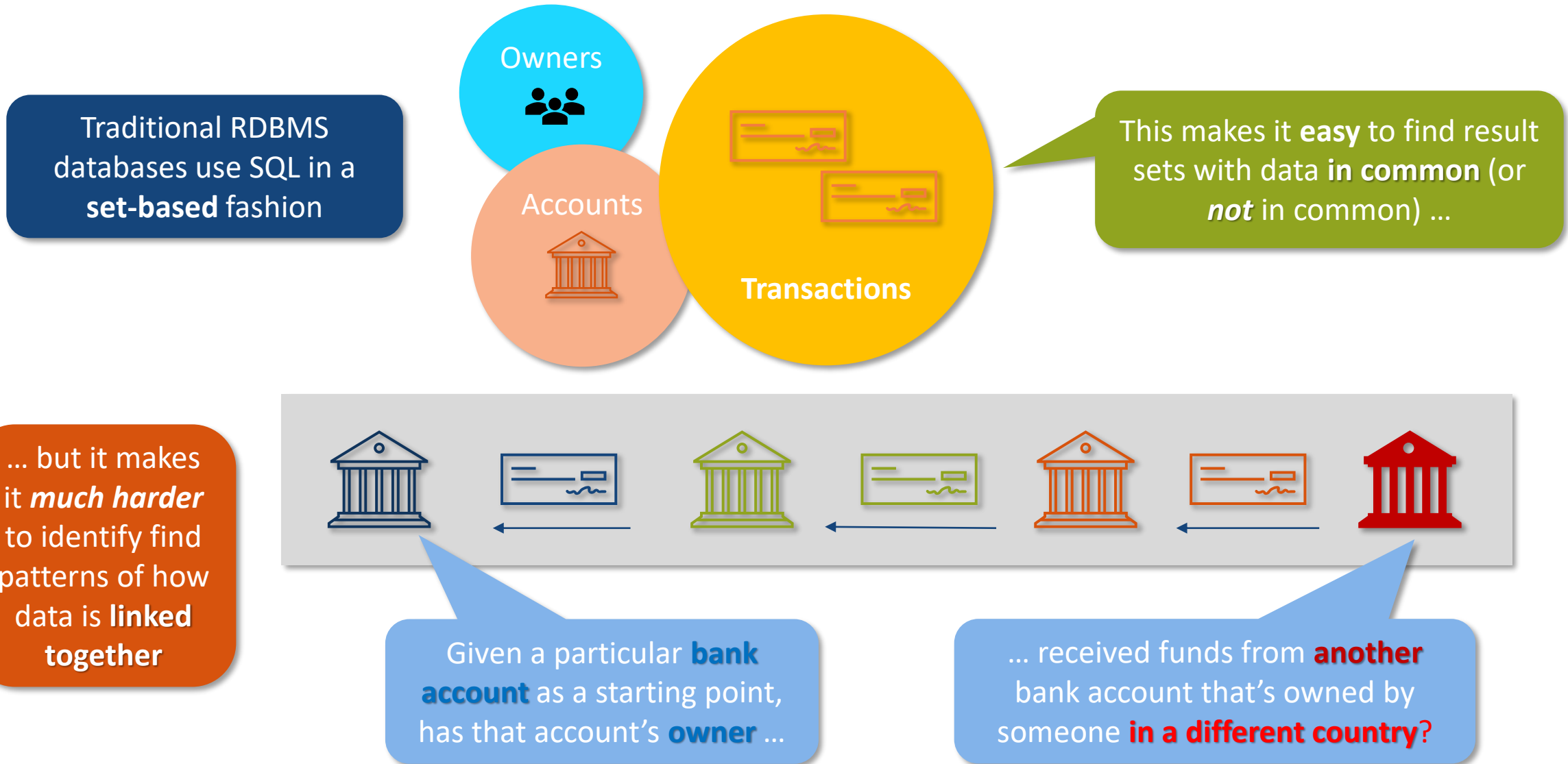
Liron Amitzi

Jim Czaprynski

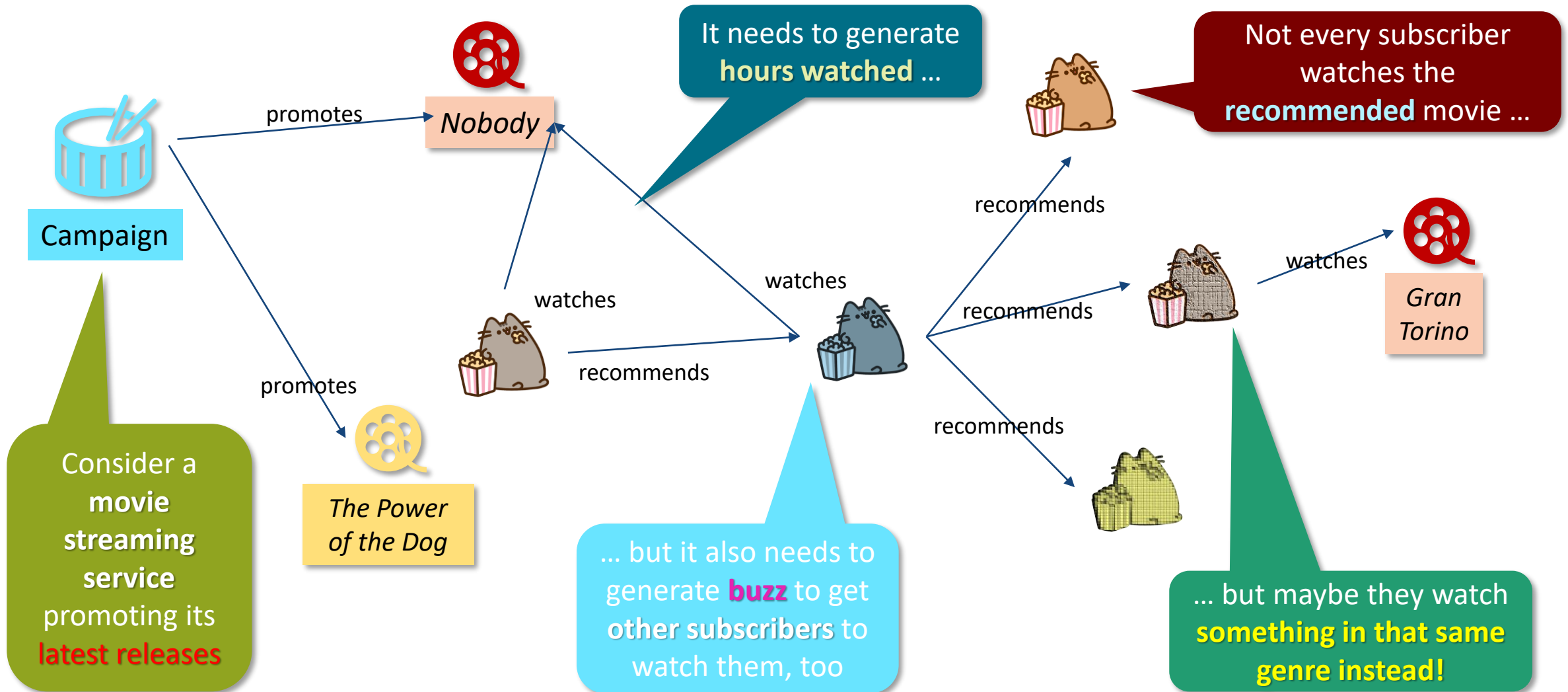
The podcast that talks about everything tech – except tech.™

<https://www.beyondtechskills.com>

Property Graphs: Finding Patterns *Between* Data Elements



Property Graphs: Not About Data *Itself*, But How They're *Connected*



Identifying Corruption Via MOE (Mark One Eyeball)

	A	B	C	D	E	F
1	owner_id	name	address	city	state	country
2	166	Vladimir Ivan Ivanovich		Moscow		
3	266	Jakub Wojciechowski	579 Allen Rd	Krakow		
4	366	Fraud Assurance	37 West 57th Street	New York	NY	
5	466	Bordeaux Oenoophilia GmbH	1000 Pennsylvania Ave	Washington	DC	
6				Harvest		

	A	B	C	D	E	F
1	src_acct_id	tgt_acct_id	tx_date	tx_amount	tx_type	tx_memo
2	50566	10166	2021-02-01	8500000	RFW	Investment withdrawal
3	50566	10166	2021-01-29	6500000	RFW	Investment withdrawal
4	98466	50566	2021-01-31	3250000	RFC	Investment funding
5	98466	50566	2021-01-30	2750000	RFC	Investment funding
6	98466	50566	2021-01-29	2500000	RFC	Investment funding
7	40466	50566	2021-01-28	2717068	RFC	Investment funding
8	40466	50566	2021-01-23	3782932	RFC	Investment funding
9	10166	98166	2020-11-12	15000000	RFC	Investment funding
10	98366	40466	2020-12-27	4500000	RES	Property sales
11	98366	40466	2020-12-23	1500000	RES	Apartment sale
12	30366	98366	2021-01-22	8000000	REP	Parking garage purchase
13	30366	98366	2020-11-18	7000000	REP	Apartment purchase
14	98166	30366	2021-01-21	2500000	CON	Consulting services
15	98166	30366	2020-12-29	7500000	CON	Consulting services
16	98166	30366	2020-11-14	5000000	CON	Consulting services
17	98366	98466	2021-01-28	4500000	CHR	Donation
18	98366	98466	2021-01-24	500000	CHR	Donation
19	98366	98466	2021-01-23	1250000	CHR	Donation
20	98366	98466	2021-01-22	1250000	CHR	Donation
21	98366	40466	2020-12-28	500000	CHR	Donation
22	98366	98466	2020-12-25	1000000	CHR	Donation

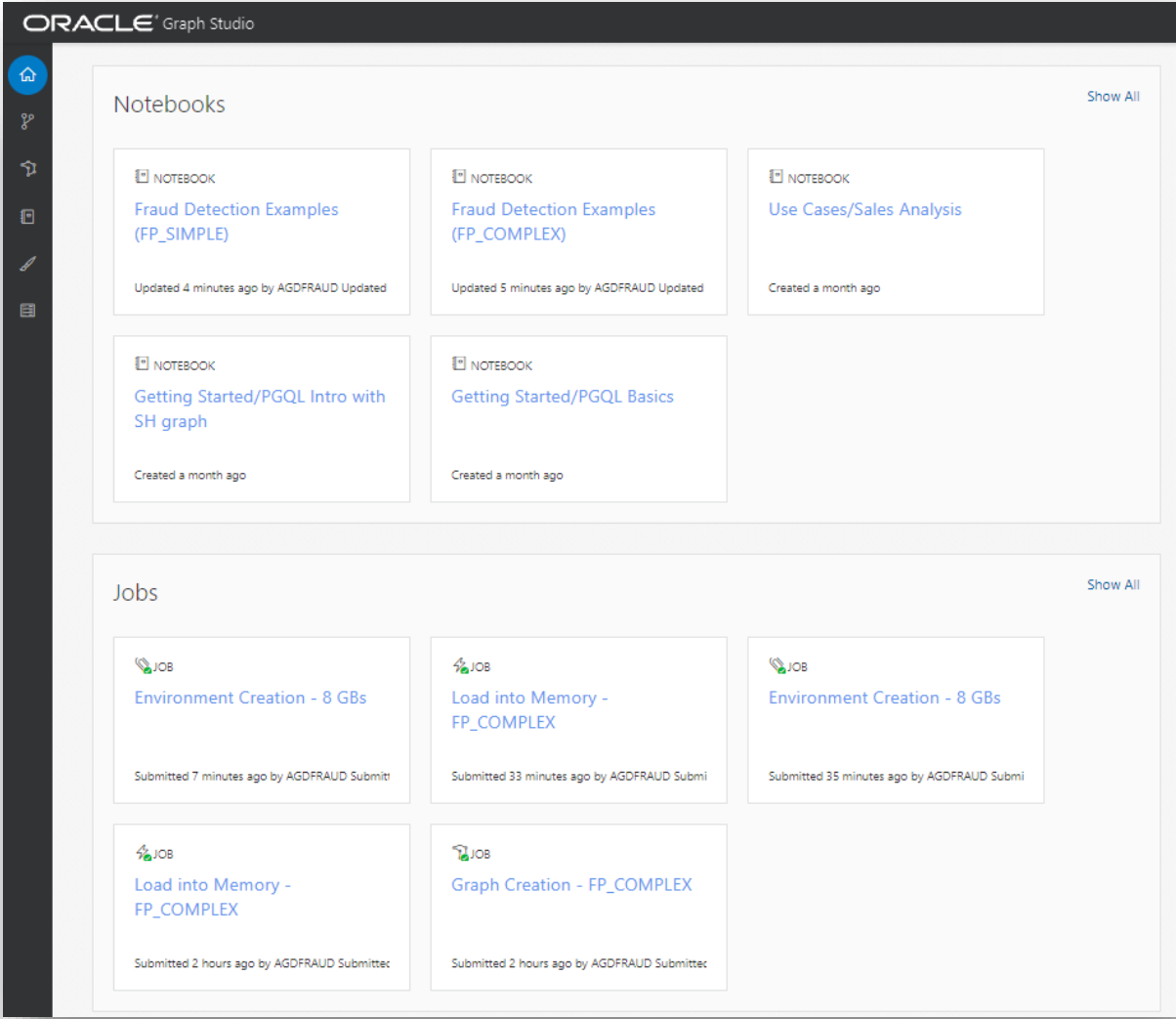
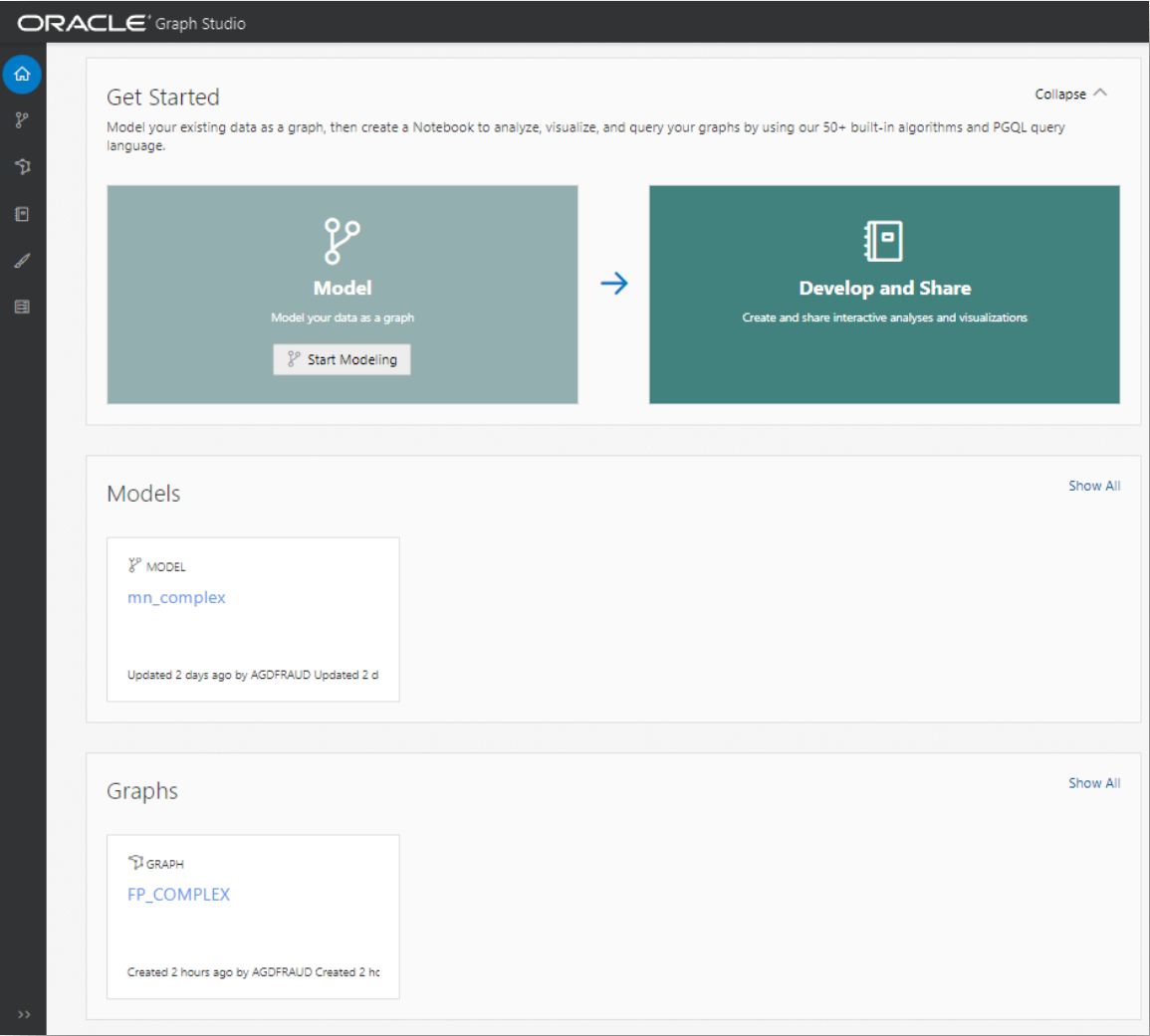
	A	B	C
1	acct_id	acct_owner_id	acct_name
2	10166	166	Ilyich Investment Holdings
3	98166	266	Dacha Retirement Planning
4	30366	366	Anti-Corruption Specialists Inc.
5	98366	366	Fraud Assurance LLC
6	40466	466	Bordeaux Oenoophilia GmbH
7	98466	466	JLR Legal Defense Fund
8	50566	566	Cotroceni Civic Holdings

Quickly, now:
Prove that these
transactions
show distinct
evidence of
money
laundering

	A	B
1	tt_type	tt_desc
2	CHR	Charitable Donation
3	CON	Consulting
4	CSX	Cash Transfer
5	LDF	Legal Defense Fund
6	REP	Real Estate Purchase
7	RES	Real Estate Sale
8	RFC	Retirement Fund Contribution
9	RFW	Retirement Fund Withdrawal
10	TRV	Travel Expenses

Accessing Graph Studio Tools

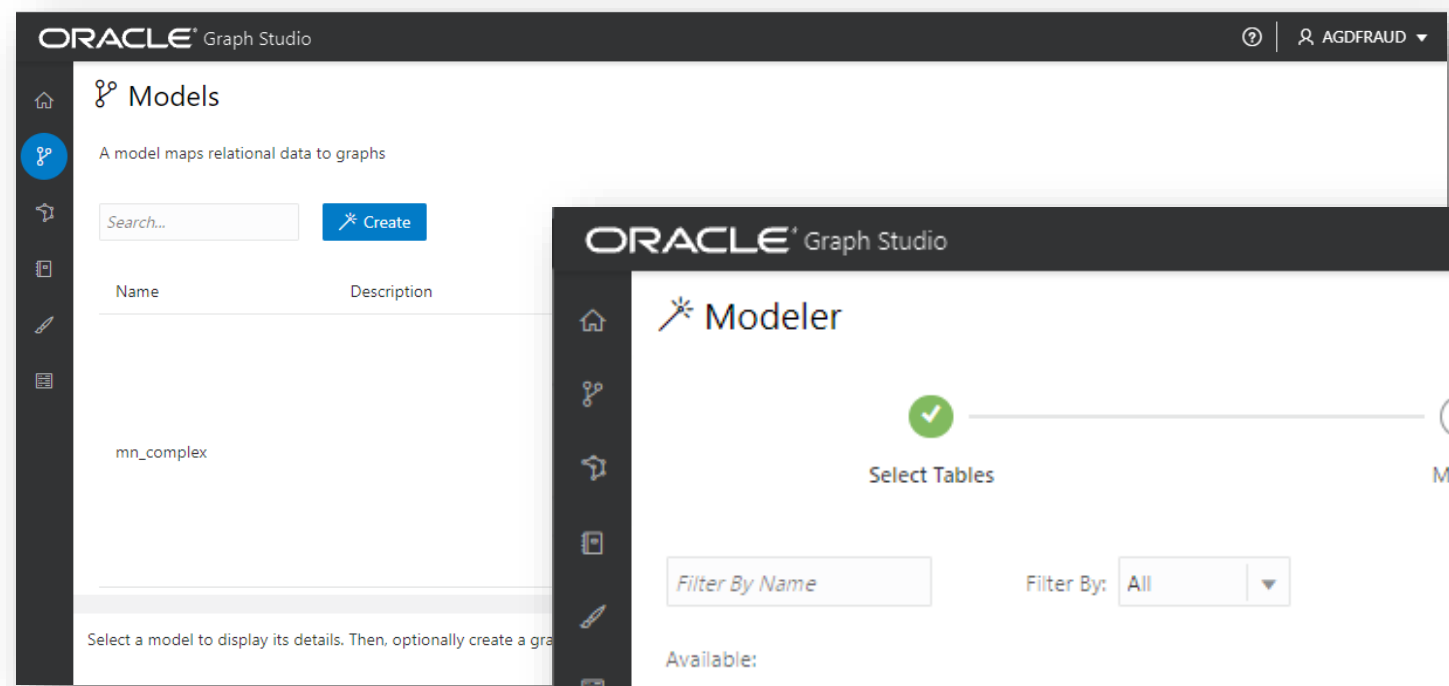
- 1 You can **model** new graphs from tables and views, as well as **visualize** the materialized graphs ...



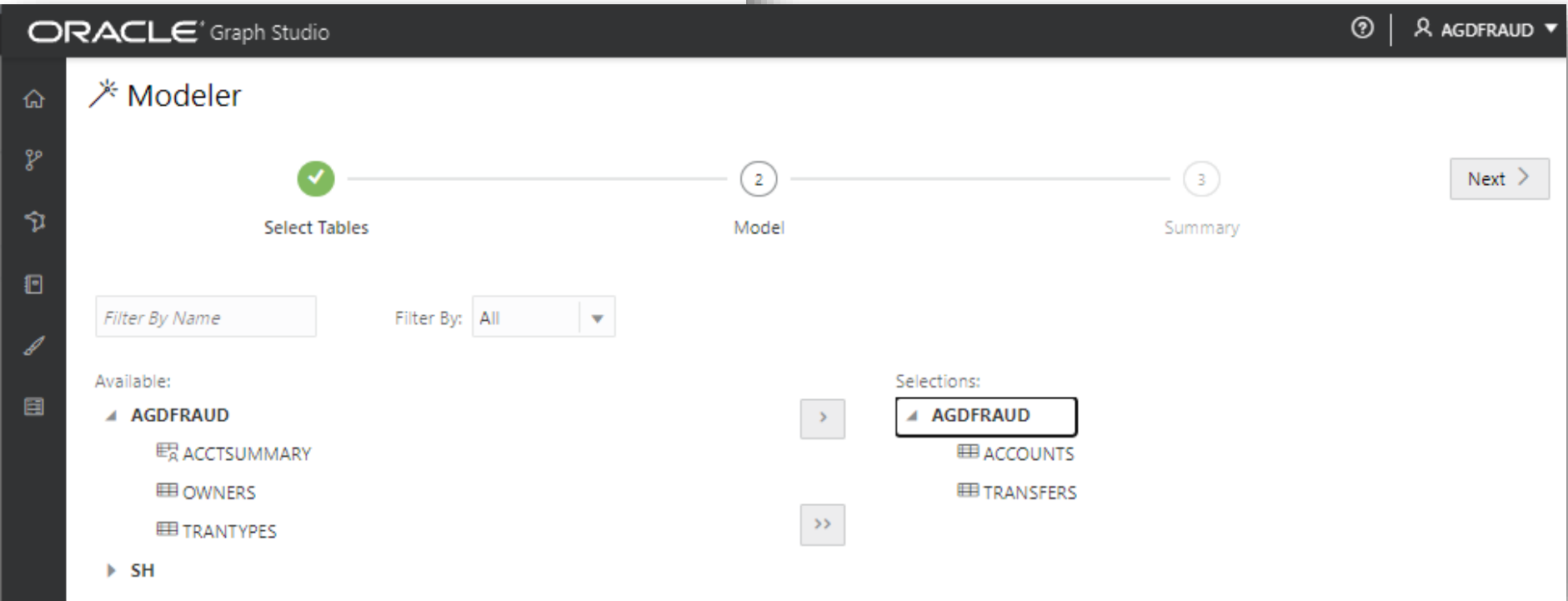
- 2 ... access existing Zeppelin **notebooks**, and review any **tasks** that have recently executed

Creating and Accessing Property Graphs (1)

1 It's easy to model a new Property Graph from existing database tables and views



2 Once entities are chosen, Oracle automatically builds the **CREATE PROPERTY GRAPH** PGQL commands to create the graph using existing PK and FK constraints



Creating and Accessing Property Graphs (2)

3 You can display properties from each **Vertex** ...

ORACLE Graph Studio

Modeler

Back

Select Tables Model Summary

Designer Source Preview

Input Tables / Views (2)

AGDFRAUD.ACCOUNTS AGDFRAUD.TRANSFERS

Vertex Tables (1)

ACCOUNTS

Edge Tables (1)

TRANSFERS

ACCOUNTS

Source Table

AGDFRAUD.ACCOUNTS

Vertex Key

ACCT_ID

Vertex Label

ACCOUNTS

Vertex Properties (3)

Name	Source Column
ACCT_ID	# ACCT_ID
ACCT_NAME	A ACCT_NAME
ACCT_OWNER_ID	# ACCT_OWNER_ID

4 ... as well as each **Edge** in the Property Graph, and even **exclude** specific columns from the final Graph

ORACLE Graph Studio

Modeler

Back

Select Tables Model Summary

Designer Source Preview

Input Tables / Views (2)

AGDFRAUD.ACCOUNTS AGDFRAUD.TRANSFERS

Vertex Tables (1)

ACCOUNTS

Edge Tables (1)

TRANSFERS

TRANSFERS

Source Table

AGDFRAUD.TRANSFERS

Source Vertex

ACCOUNTS

Destination Vertex

ACCOUNTS

Edge Label

TRANSFERS

Edge Properties (6)

Name	Source Column	Options
SRC_ACCT_ID	# SRC_ACCT_ID	✎ ✕
TGT_ACCT_ID	# TGT_ACCT_ID	✎ ✕
TX_AMOUNT	# TX_AMOUNT	✎ ✕

Creating and Accessing Property Graphs (3)

5 Glimpse the **data** within each **Vertex** and **Edge** ...

The screenshot shows the Oracle Graph Studio Modeler interface. On the left, a breadcrumb trail indicates the path: Select Tables > Model. Below this, the draft ID 'DRAFT_1632940706569' is shown. The main area displays a table titled 'Sample - AGDFRAUD.ACCOUNTS' with two columns: '# ACCT_ID' and '# ACCT_OWNER_ID'. The table contains five rows of data. At the bottom, there is a pagination control showing 'Page 1 of 2 (1-5 of 10 items)' and a 'Sample Size' dropdown set to 10.

# ACCT_ID	# ACCT_OWNER_ID
10166	166
98266	266
30366	366
98366	366
40466	466

6 ... supply **names** and **descriptors** for the new Model and Graph ...

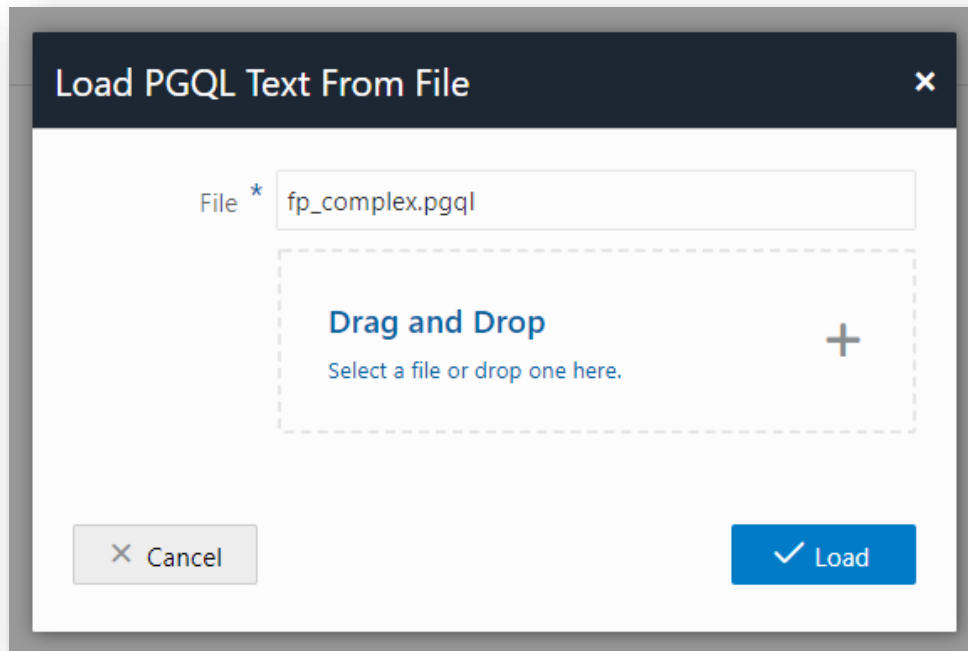
The screenshot shows the Oracle Graph Studio Jobs panel. It lists asynchronous operations executed by Graph Studio. The table has columns for Type, Name, Description, Created By, Status, and Time Created. The first job, 'Graph Creation', is currently 'In Progress'. The other jobs, 'Environment Creation' and 'Load Into Memory', are 'Succeeded'.

Type	Name	Description	Created By	Status	Time Created
Graph Creation	Graph Creation	Create a graph from an existing model - FP_SIMPLE	AGDFRAUD	In Progress	a few seconds ago
Environment Creation	Environment Creation	Create environment with 8 GBs	AGDFRAUD	Succeeded	17 minutes ago
Load Into Memory	Load Into Memory	Load graph FP_COMPLEX from PG schema objects into an in-memory representation	AGDFRAUD	Succeeded	44 minutes ago
Environment Creation	Environment Creation	Create environment with 8 GBs	AGDFRAUD	Succeeded	an hour ago
Load Into Memory	Load Into Memory	Load graph FP_COMPLEX from PG schema objects into an in-memory representation	AGDFRAUD	Succeeded	3 hours ago

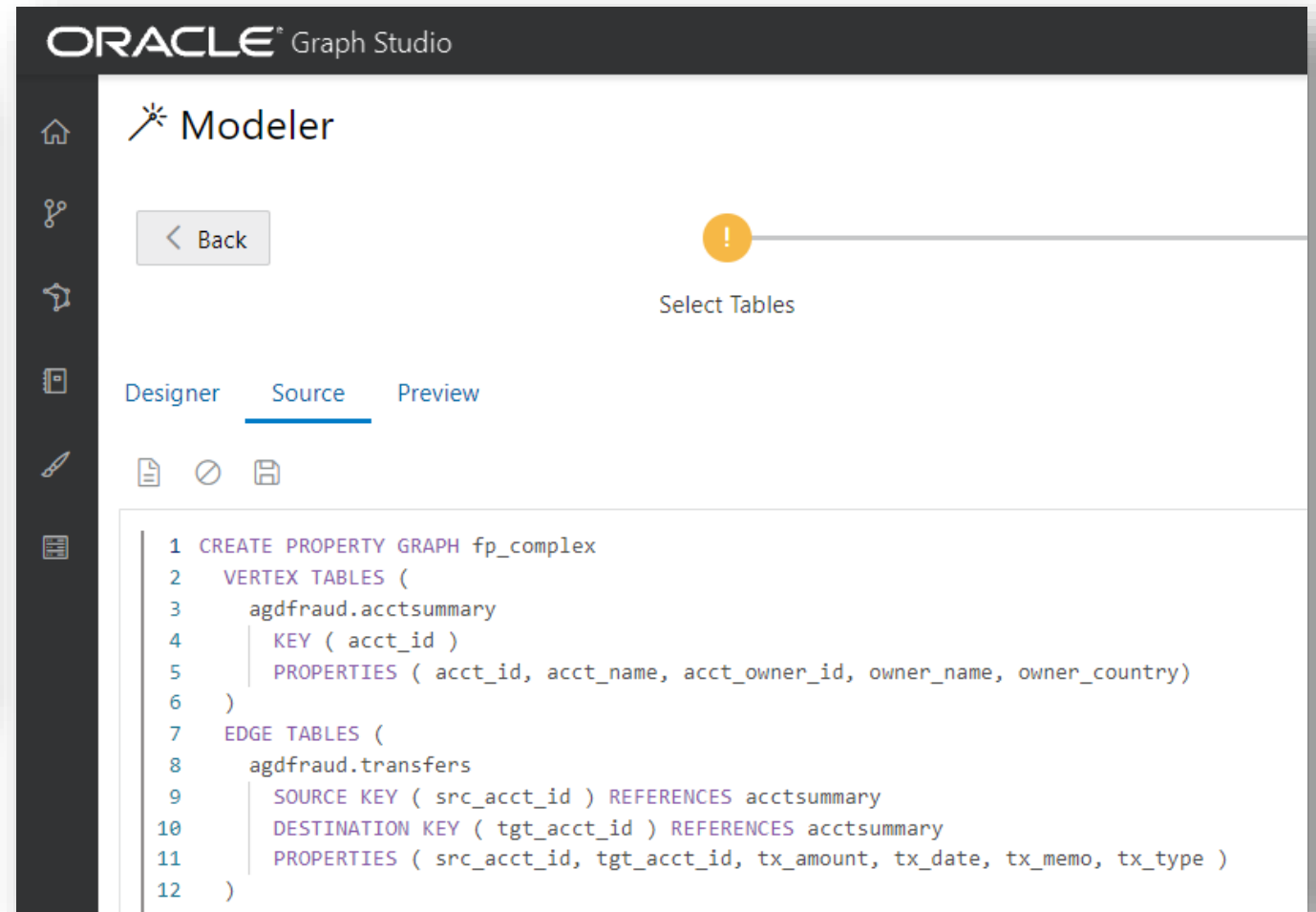
7 ... and **monitor** the successful creation of the new Graph and its corresponding Model

Creating and Accessing Property Graphs (4)

- 8 If you have a PGQL file containing a **CREATE PROPERTY GRAPH** statement ...



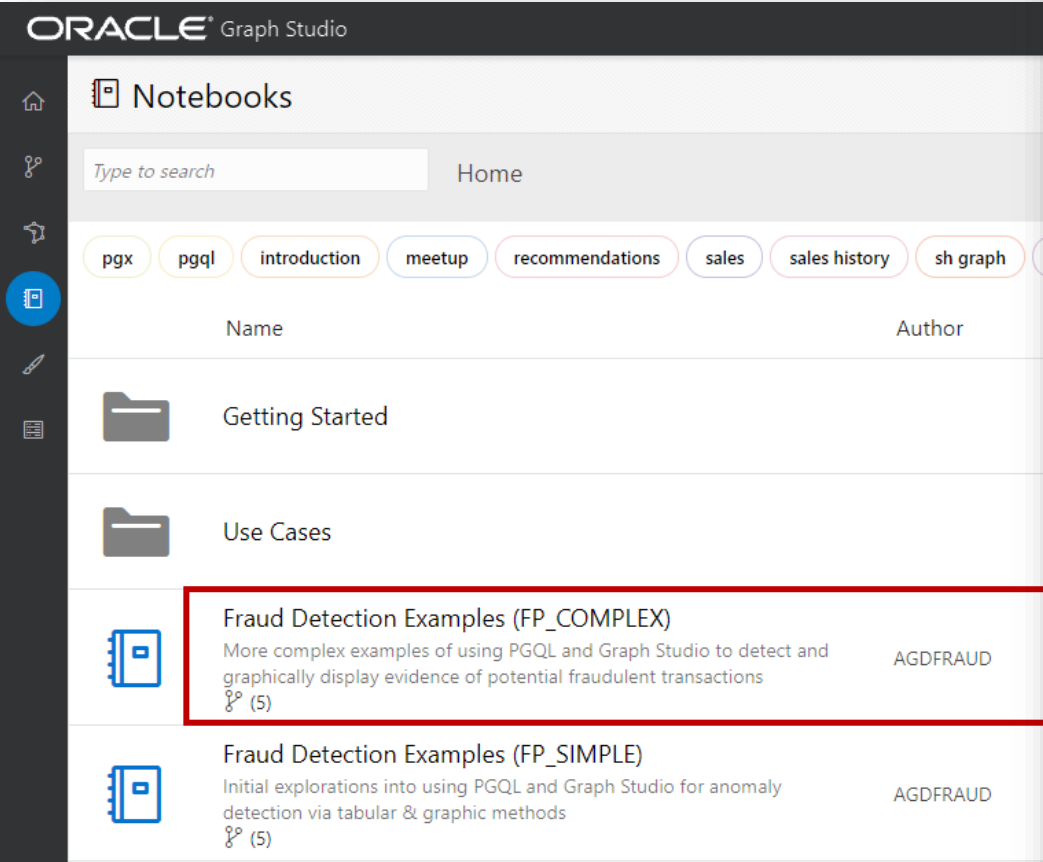
- 9 ... you can supply it within the Modeler instead of building the Graph and Model graphically



Leveraging Zeppelin Notebooks To Probe & Display Property Graphs (1)

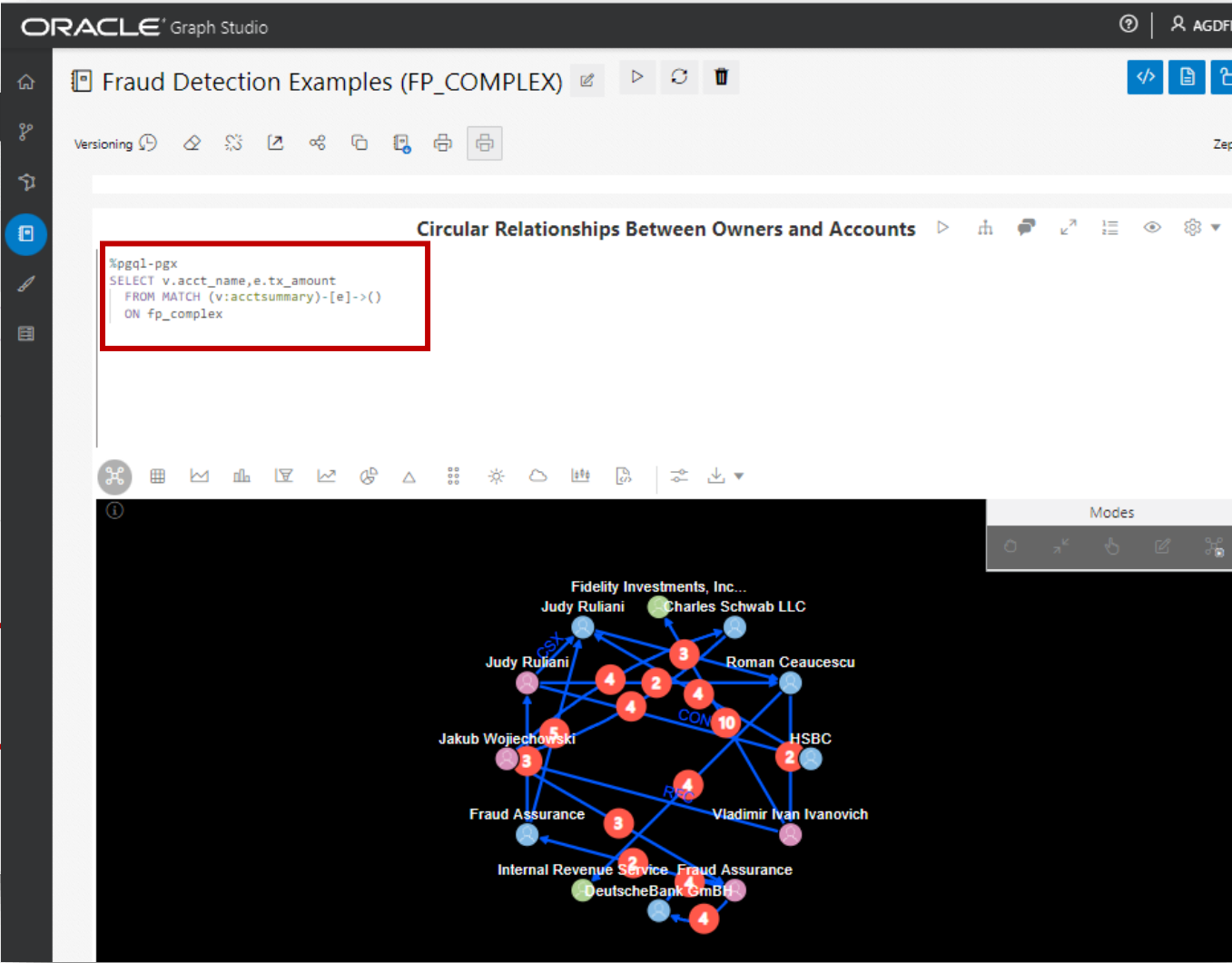
1

Zeppelin notebook technology allows even inexperienced PGQL users to immediately dive into property graph content ...



2

... and with a simple PGQL statement and some mouse clicks, suspicious patterns are immediately evident!



Leveraging Zeppelin Notebooks To Probe & Display Property Graphs (2)

4 ... and, with just a little polish the apparently fraudulent activity becomes even more evident..

ORACLE Graph Studio

Fraud Detect

Versioning

```
%pgql-pgx
SELECT
  a.acct_owner,
  a.owner_name,
  a.acct_id,
  a.acct_name,
  out_degree(a),
  in_degree(a),
  (in_degree(a))
FROM MATCH (a)
ON fp_complex
ORDER BY acct_ow...
```

Type to search

ACCT_OWNER_ID
101
166
201
266
301
366
366

ORACLE Graph Studio

Fraud Detection Examples (FP_COM...

Versioning

```
%pgql-pgx
/* Isolate potential fraudulent transactions by filtering out severak apparently "legitimate" acco
SELECT v.acct_name,e.tx_amount
FROM MATCH (v:acctsummary)-[e]->()
ON fp_complex
WHERE e.src_acct_id NOT IN (10101, 20201, 30301, 40401, 50501)
AND e.tgt_acct_id NOT IN (10101, 20201, 30301, 40401, 50501)
```

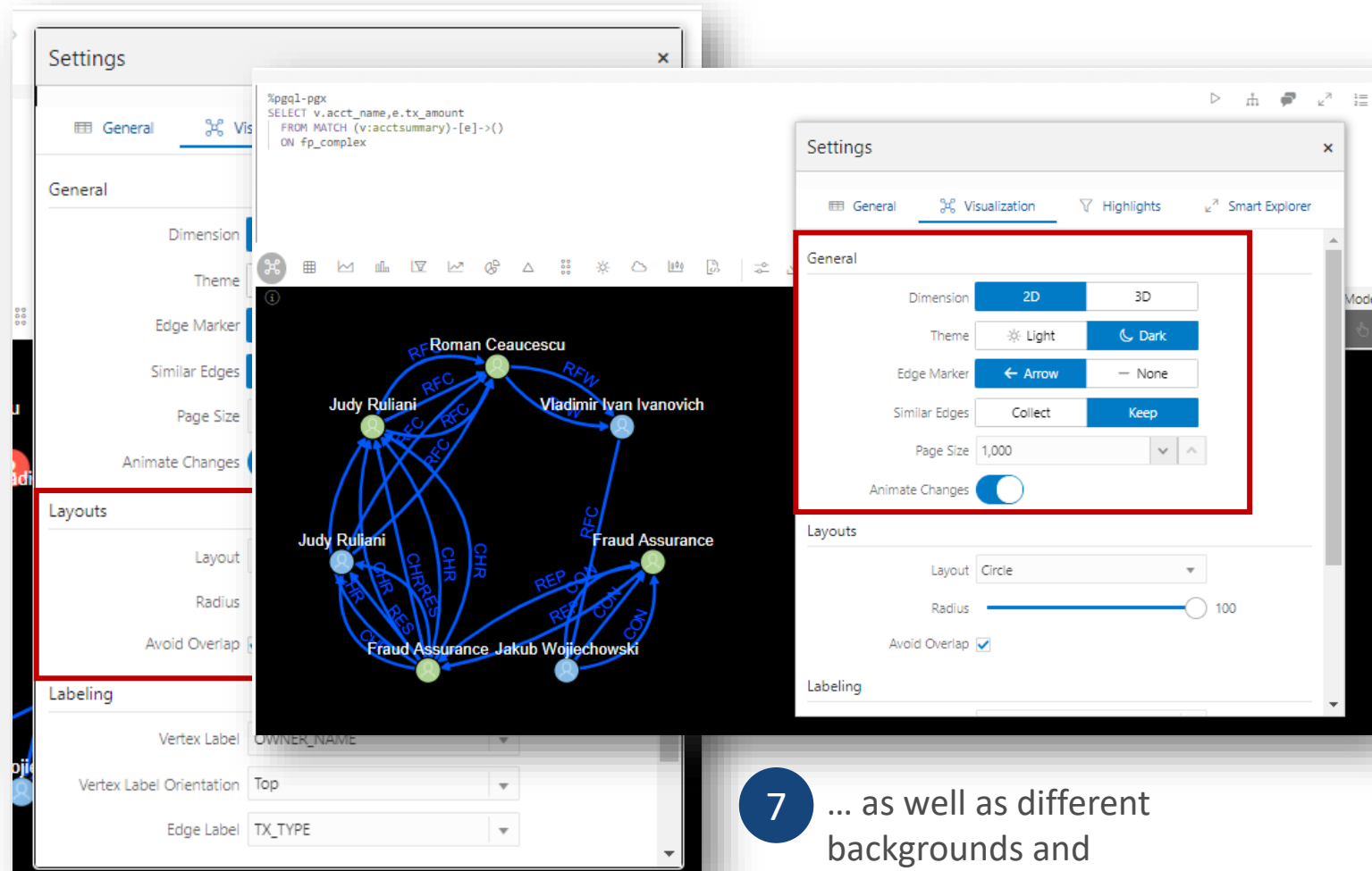
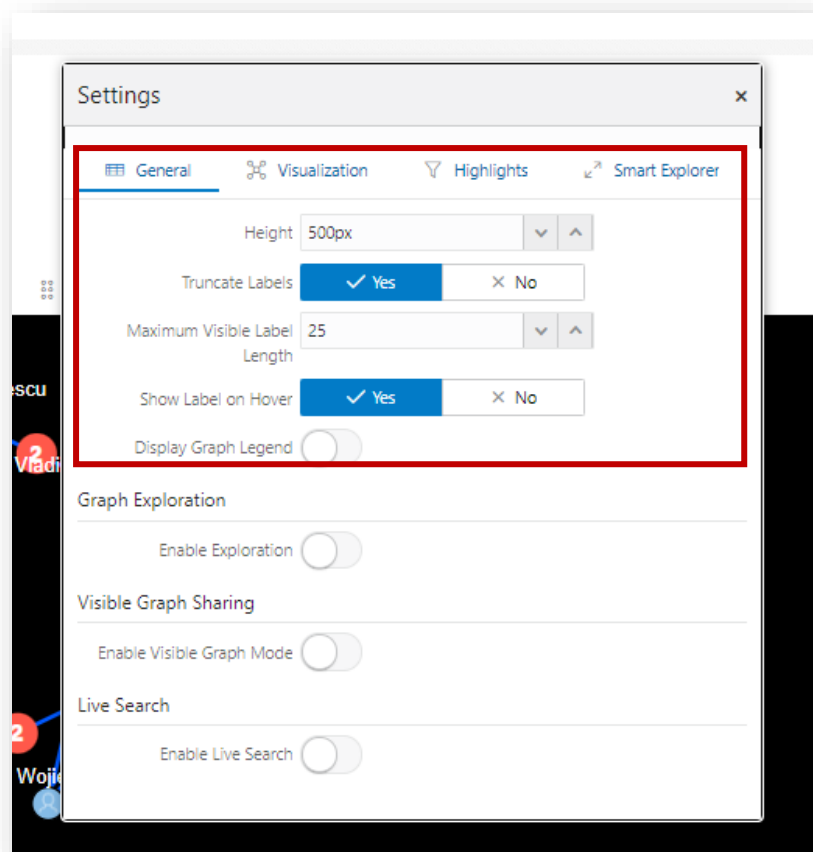
AGDFRAUD

Zeppelin

TRX_TOTAL
10
13
8
12
8
10
13

Leveraging Zeppelin Notebooks To Probe & Display Property Graphs (3)

5 Property graph display options are flexible and precise ...



7 ... as well as different backgrounds and representations of vertices

Live Demonstration: *Like, Wow, That's Amazing, Miss Information!*



Just 12 People Are Behind Most Vaccine Hoaxes On Social Media, Research Shows

Updated May 14, 2021 · 11:48 AM ET ⓘ

Heard on All Things Considered



SHANNON BOND



To illustrate how easy it is to leverage **Graph Studio**, here's an example of how to **detect patterns** in (fictitious!) social media postings and possibly identify which Twitter accounts **are being spread by bots and "sock puppets"** versus posts by **actual human beings**



What could
possibly go wrong?

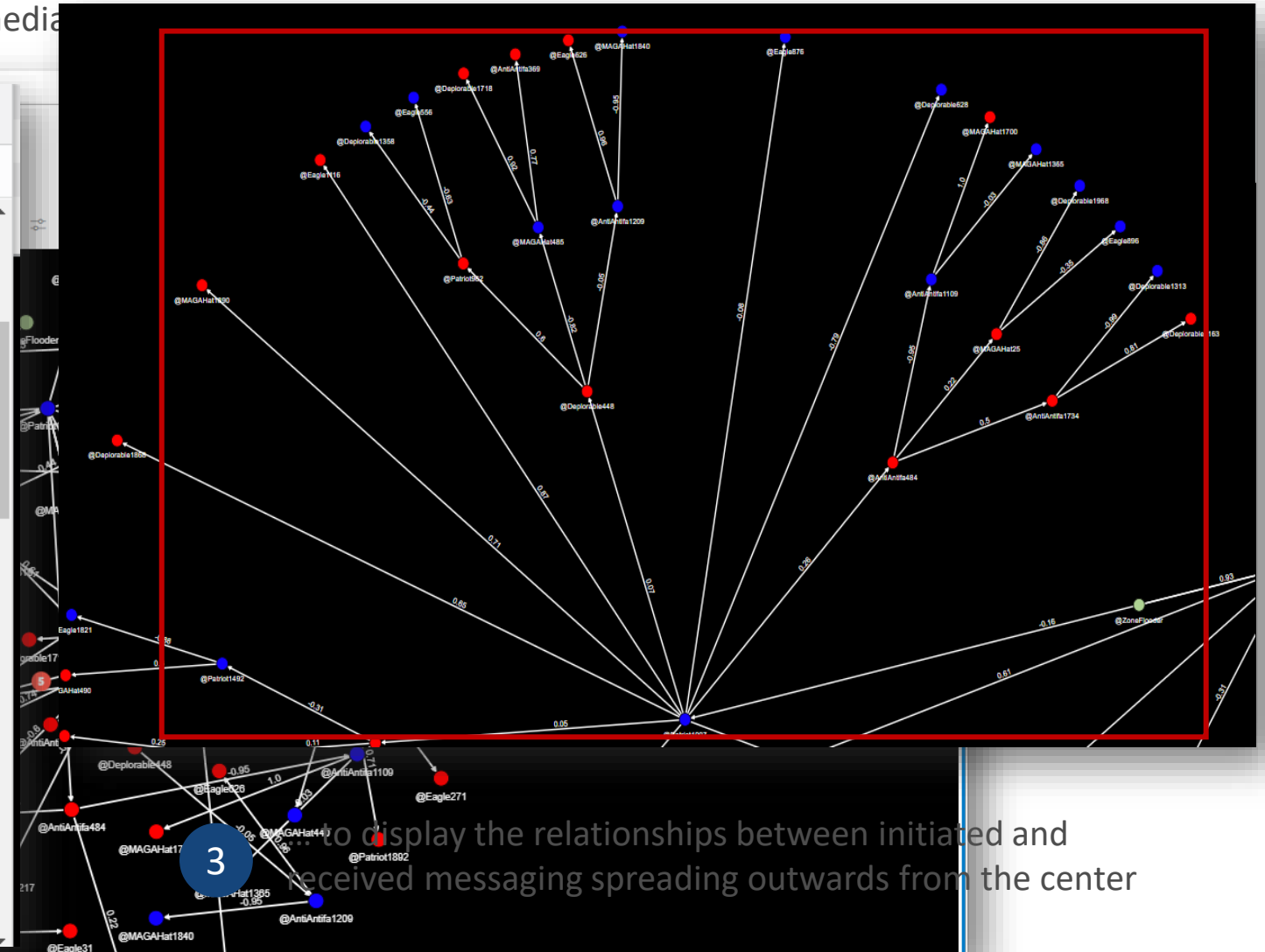
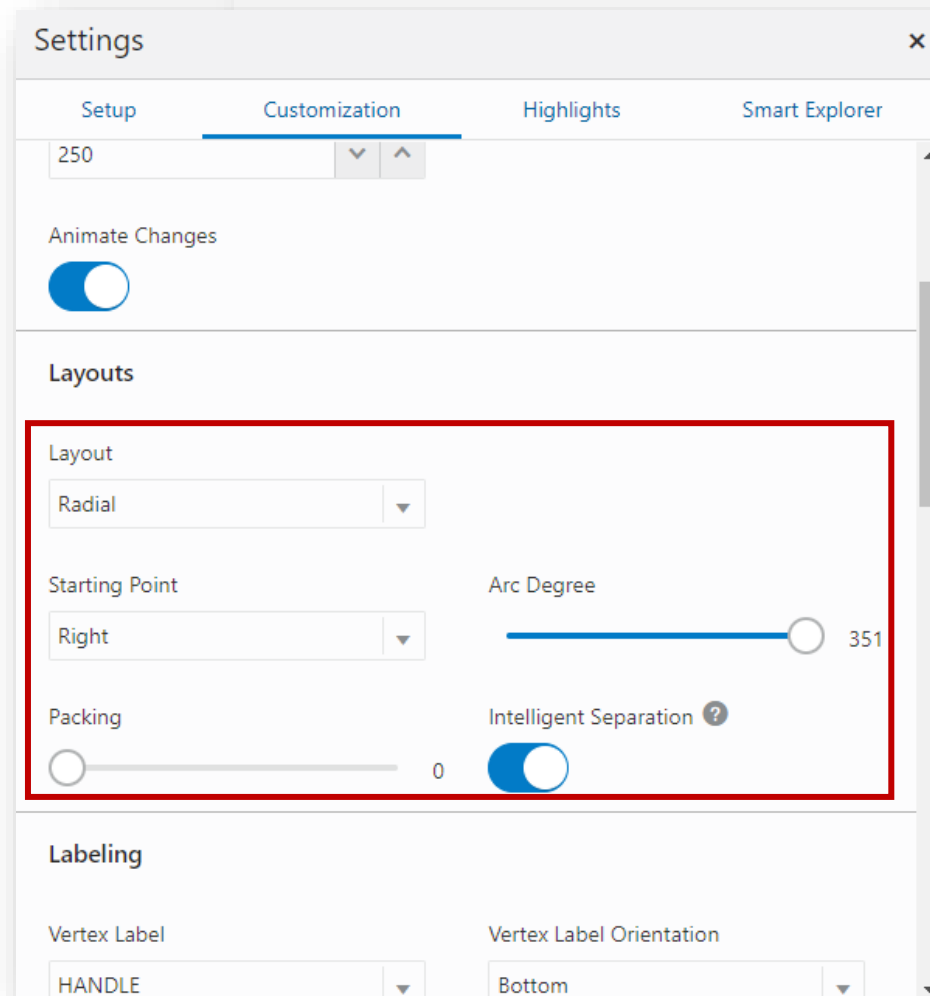


Identifying Social Media Information Spreaders vs. Consumers

2

For a different graphic representation, we can change

from the default **Force** layout to a **Radial** one



3 to display the relationships between initiated and received messaging spreading outwards from the center

Different Viewpoints, With a Few Mouse Clicks

4

One other small change, and here's now **yet another** different viewpoint ...

Settings

Setup

Customization

Highlights

Smart Explorer

Animate Changes

Layouts

Layout

Concentric

Minimum Vertex Spacing

45

Labeling

Vertex Label

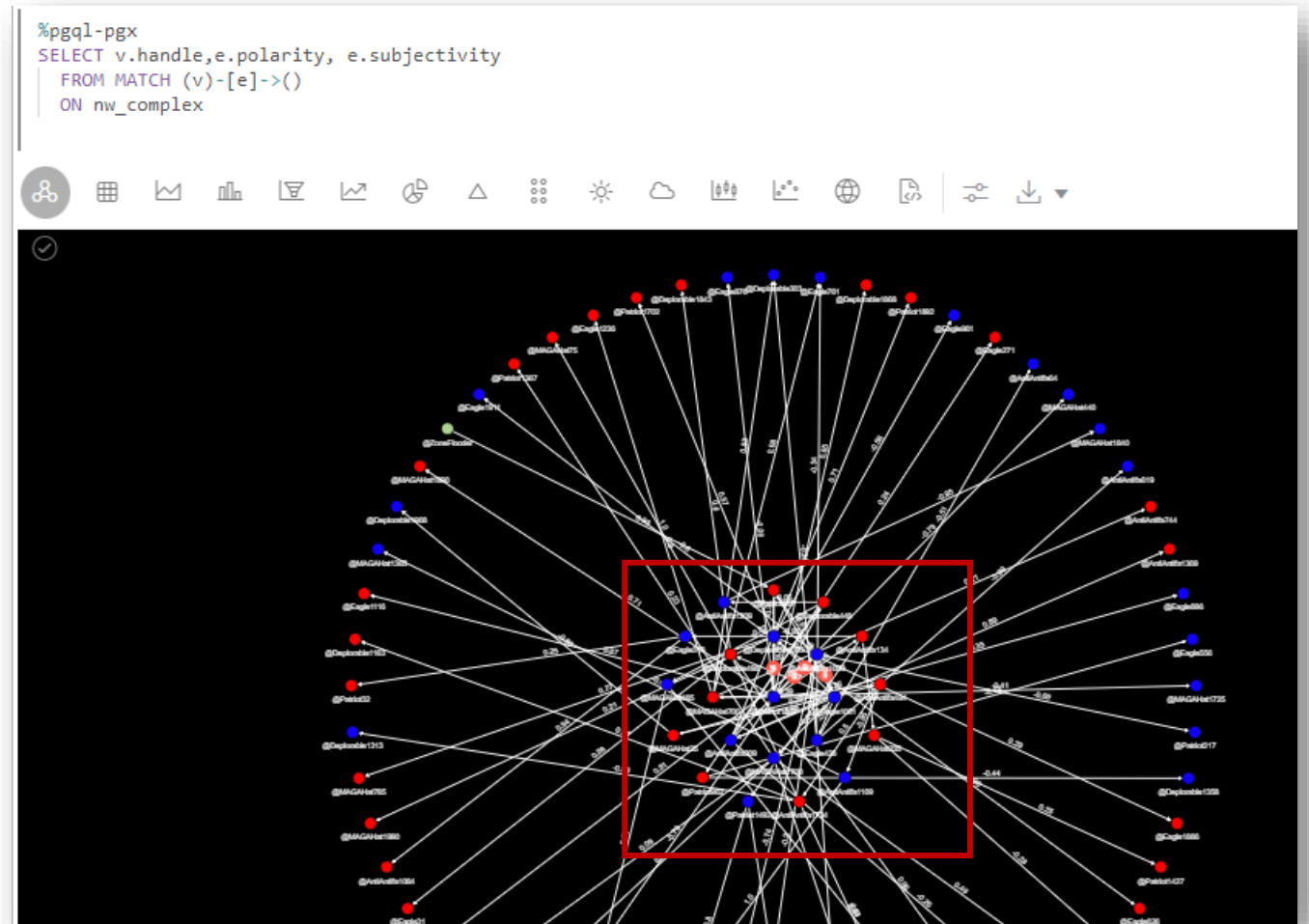
HANDLE

Vertex Label Orientation

Bottom

Edge Label

POLARITY



5

... focused more on the **centrality** of initial SM **senders** and those who **relayed** the posts

PGQL Analytic Functions: A Closer Look

6 can also leverage PGQL analytic functions like **PAGERANK()** to identify the relative importance of nodes in a network

7 can also leverage PGQL analytic functions like **PAGERANK()** to identify the relative importance of nodes in a network

%pgql-pgx
SELECT x.p
FROM MAT
ON nw
ORDER BY
LIMIT 12

Type to search

PAGERANK
0.00421687
0.00344187
0.00301630
0.00275407
0.00253132
0.00253132

%pgql-pgx
SELECT
| x.handle as "Handle"
| ,x.followers as "# of Followers"
| ,x.following as "# of Following"
| ,x.pagerank as "PageRank"
| ,x.betweenness as "DistBtwn"
| ,x.authority as "Hits - Authority"
| ,x.hubs as "Hits - Hubs"
FROM MATCH (x)
ON nw_complex
WHERE x.betweenness > 0
ORDER BY x.betweenness DESC

Type to search

Handle	# of Followers	# of Following	PageRank	DistBtwn	Hits - Authority	Hits - Hubs
@Eagle1001	20	5	0.0034418799889605984	118.0	0.027406761249333042	0.4854152337546281
@Patriot987	50	35	0.0030163043478260874	79.0	5.4171665766392065E-188	0.1545814580986151
@Patriot1007	21	18	0.0022713994565217396	74.0	0.01774315361502717	0.860344755549369
@AntiAntifa1009	52	12	0.0027540718410326093	44.000000000000001	0.5115031862658065	0.01683820803240701
@Deplorable448	6	80	0.00172696925951087	27.0	0.09875200653015585	3.204579685542998E-141
@AntiAntifa484	9	22	0.00172696925951087	27.0	0.09875200653015585	3.204579685542998E-141

Beyond PGQL: Other Property Graph Tools

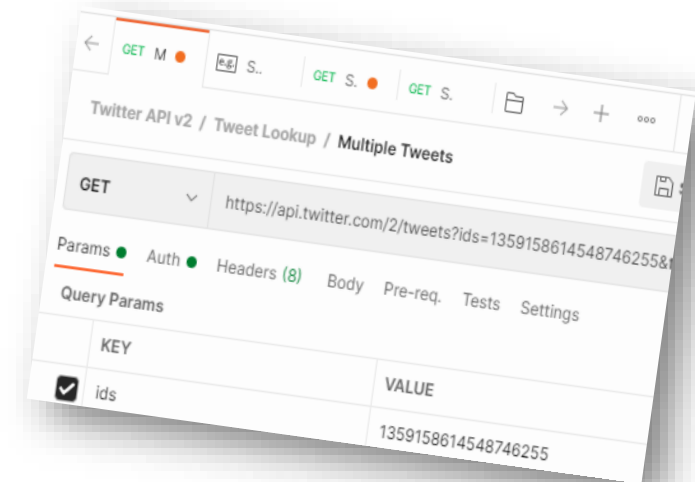
Property Graph toolsets can be accessed through just about any programming language, including **Java**, **Python**, and even **SQLcl** and **PL/SQL**



60+ powerful **graph algorithms** offer the ability to:

- Detect graph **components** and **communities**
- Evaluate graph **structures** for patterns
- **Rank** and “**walk**” graphs
- **Identify paths** through graph nodes
- Build **machine learning models**

Leverage the **Graph Client API** to build custom applications



PGX Analytic Functions Provide a Deeper Look *Within* Graphs

Analytic Function	Provides Useful Intelligence For:
PageRank	Determines <i>which node in a graph is most important</i> based on its number of incoming edges
Closeness Centrality	Calculates <i>how “close”</i> a node is to other nodes within a graph
Betweenness Centrality	Detects <i>how much a node may influence</i> the flow of information within a graph
HITS	Points to which <i>web page</i> is <i>likely to contain the most meaningful information</i> based on its <i>HyperLink-Induced Topic Search</i> score
WTF	Projects <i>Whom To Follow</i> within a social network for <i>maximum positive impact to social standing</i>

Using PGX To Enhance Graph Content

1 You can apply any one of **60+** additional graph **analytic functions** to an **existing** in-memory graph

2 ... and you can use these new attributes for further **analyses** or even different **visualizations**

Handle	# of Followers	# of Following	PageRank	DistBtwn	Hits - Authority	Hits - Hubs
@Eagle1001	20	5	0.0034418799889605984	118.0	0.027406761249333042	0.4854152337546281
@Patriot987	50	35	0.0030163043478260874	79.0	5.4171665766392065E-188	0.1545814580986151
@Patriot1007	21	18	0.0022713994565217396	74.0	0.01774315361502717	0.860344755549369
@AntiAntifa1009	52	12	0.0027540718410326093	44.000000000000001	0.5115031862658065	0.01683820803240701
@Deplorable448	6	80	0.00172696925951087	27.0	0.09875200653015585	3.204579685542998E-141
@AntiAntifa484	9	22	0.00172696925951087	27.0	0.09875200653015585	3.204579685542998E-141

PGX ML Toolset (And You Thought “Normal” ML Was Tough To Grok!)

Model	Description	Examples of Real-World Use Cases
<u>DeepWalk</u>	Computes random walks for every vertex, then generates new vector representations	How likely is it that a <i>new post</i> will spread quickly through a network of connected friends via a mobile social media application?
<u>Supervised GraphWise</u>	Based on GraphSage , it's an inductive vertex representation learning algorithm against vertex feature information	Based on a customer's <i>prior ordering habits</i> , what new products or offerings can we suggest that they're actually interested in ?
<u>Unsupervised GraphWise</u>	Based on Deep Graph Infomax , it applies an inductive vertex representation learning algorithm against vertex information	Can we quickly identify brain abnormalities to detect autism spectrum disorder (ASD) by comparing 4-D MRI brain scans of new patients against those of patients <i>already diagnosed with ASD</i> ?
<u>Pg2Vec</u>	Generates graphlets that can be compared for matching patterns	Based on <i>prior known patterns</i> , is a new set of financial transactions a warning sign that money laundering may be occurring ?

Plans for Future Experimentation



Expand beyond Graph Studio for Autonomous Database to use powerful PGX tools **in native mode**



Use publicly-available **social media data** to refine methods **identifying spread of misinformation**



Leverage data **captured from Twitter in real time** to explore available **Machine Learning** algorithms

Sample Use Cases For Property Graphs

- **Social Media Sentiment Analysis**

<https://towardsdatascience.com/sentiment-analysis-74624b075842>

- **Graphs Analytics for Fraud Detection**

<https://towardsdatascience.com/graphs-analytics-for-fraud-detection-83ee3af81ec7>

- **Detecting Fake Users on Social Media with a Graph Database**

<https://journals.uvic.ca/index.php/arbutus/article/view/20027>

- **Just 12 People Are Behind Most Vaccine Hoaxes On Social Media, Research Shows**

<https://www.npr.org/2021/05/13/996570855/disinformation-dozen-test-facebooks-twitters-ability-to-curb-vaccine-hoaxes>

Useful References

- **Graph Databases and Analytics: How to Use Them**

<https://www.oracle.com/it/a/tech/docs/sg-oow2019-graph-databases-and-analytics.pdf>

- **Property Graph Developer's Guide**

<https://docs.oracle.com/en/database/oracle/property-graph/20.4/spgdg/oracle-graph-property-graph-developers-guide.pdf>

- **PGQL: Vertex and Edge Functions**

<https://pgql-lang.org/spec/1.4/#vertex-and-edge-functions>

- **Using the Machine Learning Library (Pgxml) for Graphs**

<https://docs.oracle.com/en/database/oracle/property-graph/22.1/spgdg/using-machine-learning-library-pgxml-graphs.html>