TigerGraph

Smarter AI with Analytical Graph Databases: Best Practices and Case Studies

Gaurav Deshpande, VP Marketing, TigerGraph

July 14, 2021

It's Great to Meet You all Virtually

Gaurav Deshpande

Vice President of Marketing

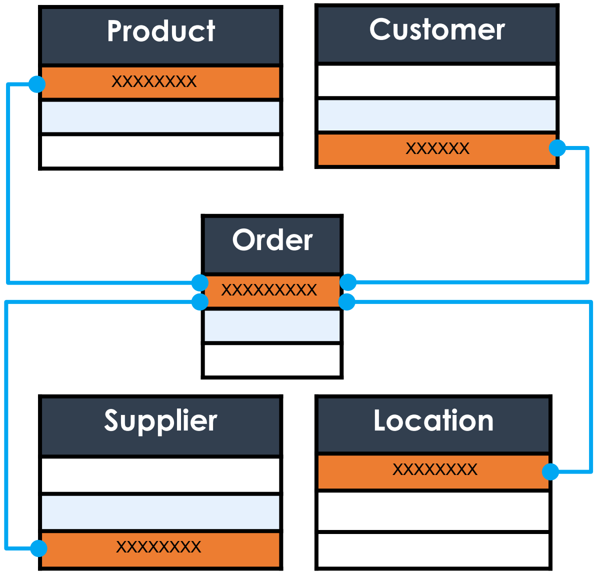
- All things marketing at TigerGraph
- Led 2 startups through explosive growth - i2 Technologies (IPO) & Trigo Technologies (largest MDM acquisition by IBM)
- Patents in supply chain management and big data analytics
- **HUGE GRAPH HEAD**
- Email: gaurav@tigergraph.com

AGENDA

- Who is TigerGraph?
- What's a Graph Database?
- Why Graph + AI?
- Three Basic Approaches for Graph + AI, with Use-Case Examples
    - Unsupervised Learning
    - Feature Enrichment from Graph Features
    - In-Database Learning

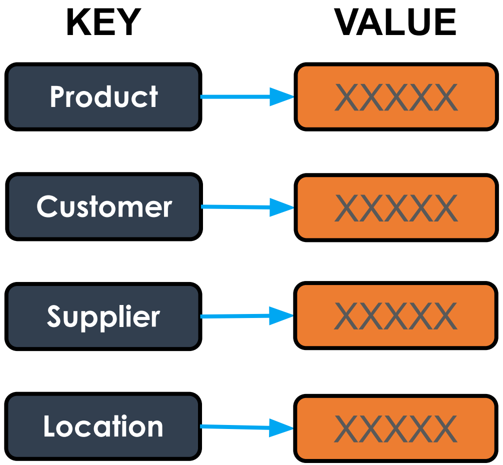
Evolution of The Database Landscape and The Rise of Graph DB

Relational Database



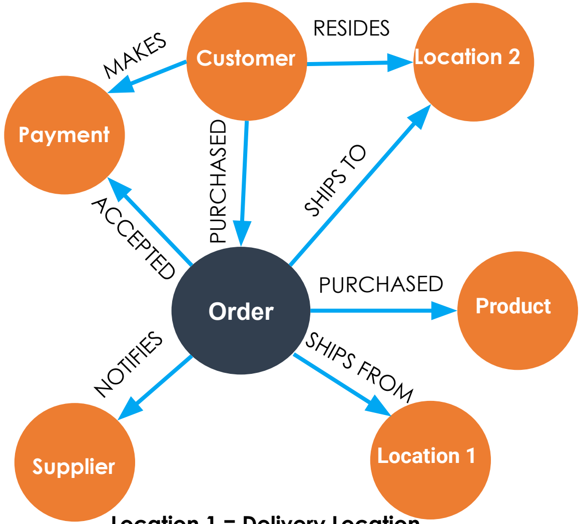
- Rigid schema
- High performance for transactions
- Poor performance for deep analytics

Key-Value Database



- Highly fluid schema/no schema
- High performance for simple transactions
- Poor performance deep analytics

Graph Database



Location 1 = Delivery Location
Location 2 = Warehouse

- Flexible schema
- High performance for complex transactions
- High performance for deep analytics

By 2025, graph technologies will be used in 80% of data and analytics innovations, up from 10% in 2021, facilitating rapid decision making across the enterprise.

Source: Gartner, “Top Trends in Data and Analytics for 2021”, Rita Sallam et al, 2021

Who is TigerGraph?

FORRESTER® WAVE LEADER 2020 Graph Data Platforms

Gartner COOL VENDOR 2020™

We provide **advanced analytics and machine learning on connected data**

- The only scalable graph database for the enterprise: 40-300x faster than competition
- Foundational for AI and ML solutions
- Designed for efficient concurrent OLTP and OLAP workloads
- SQL-like query language (GSQL) accelerates time to solution
- Available on-premise & on: Google GCP, Microsoft Azure, aws

Our customers include:

- The largest companies in financial, healthcare, telecom, media, utilities and innovative startups in cybersecurity, ecommerce and retail

Founded in 2012, HQ in Redwood City, California

Corporate Overview Video

Advanced Analytics and Machine Learning on Connected Data

CONNECT ALL DATASETS AND PIPELINES

Friction-free scale up from GB to TB to Petabyte with lowest cost of ownership

UNITEDHEALTH GROUP®

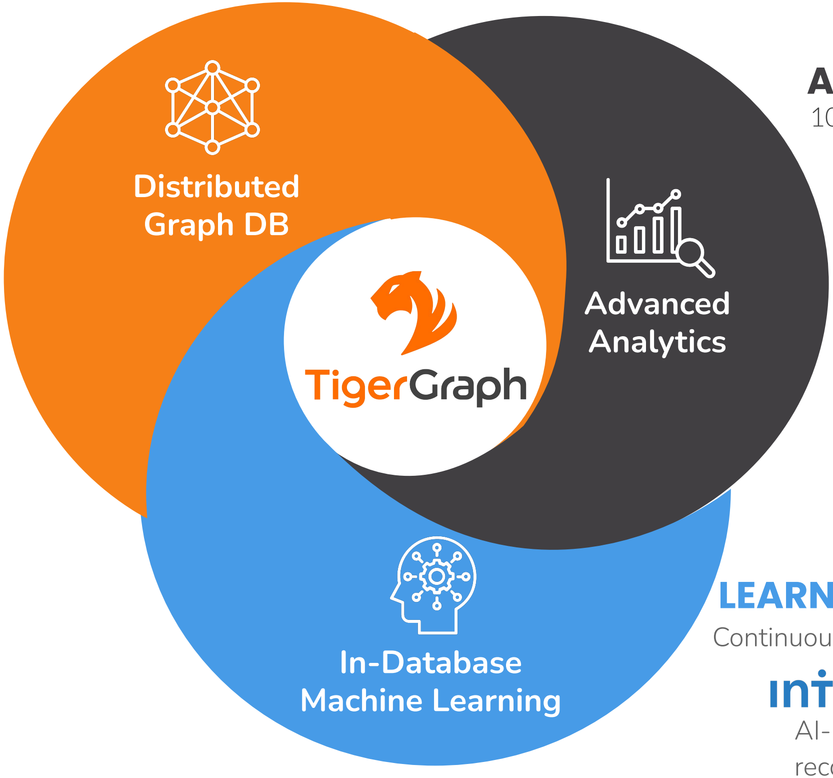
Customer 360 connecting 200+ datasets and pipelines

Fortune 50 Retailer

Item 360 for eCommerce across 100+ datasets

xandr **AT&T**

Identity graph connecting multiple data pipelines



ANALYZE CONNECTED DATA

10-100X faster than current solutions

Jaguar Land Rover

Supply chain planning accelerated from 3 weeks to 45 minutes

China Mobile International

Fraud Detection - batch to real-time for 750 million calls/day

LEARN FROM CONNECTED DATA

Continuous graph-based feature generation and training

intuit.

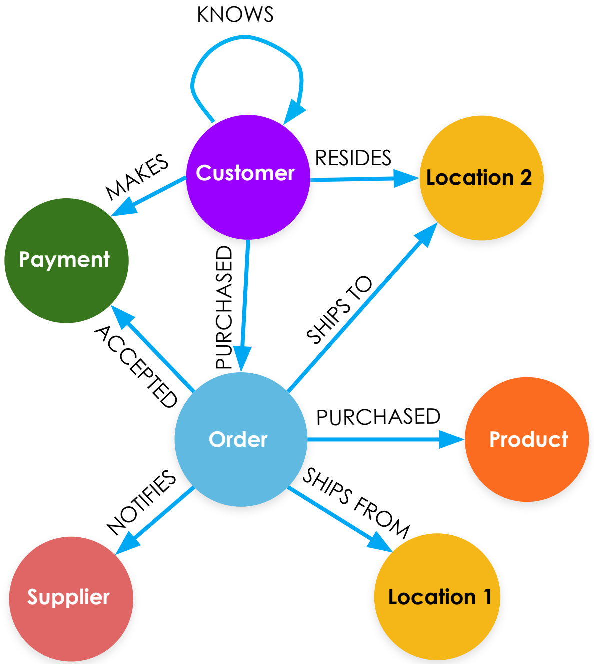
AI-based Customer 360 for entity resolution, recommendation engine, fraud detection

7 out of top 10 global banks

Real-time fraud detection and credit risk assessment

What is TigerGraph Video

Why Graph? Why Graph + AI?



Richer, Smarter Data

- Connections-as-data
- Connects different datasets, breaks down silos

Deeper, Smarter Questions

- Look for semantic patterns of relationship
- Search far & wide more easily & faster than other DBs

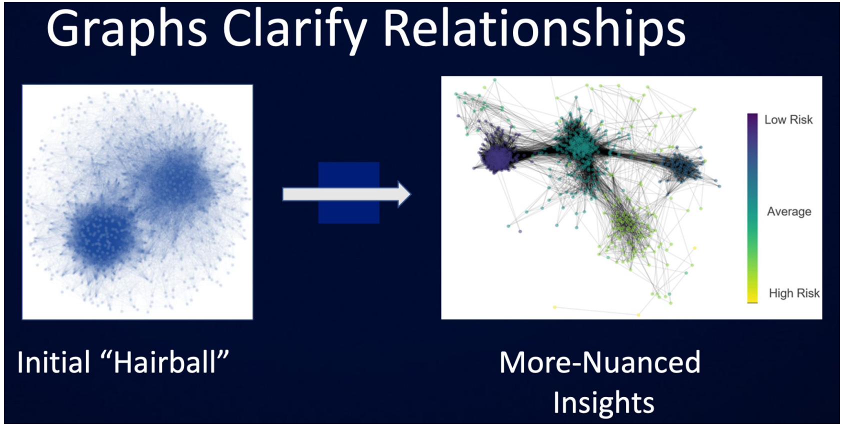
More Computational Options

- Graph algorithms
- Graph-enhanced machine learning

Explainable Results

- Semantic data model, queries, and answers
- Visual exploration and results

Graphs Clarify Relationships and Feed Machine Learning with Nuanced Inputs for Better Models



Brad Spiers
JPMorgan Chase

- “Graphs clarify relationships. Put simply, relying upon entire paths or sets of relationships in a Graph can provide deeper insights than just looking at nearest neighbors, which is what a typical database will afford you.”
- “You can turn an initial hairball *(of data and connections)* into nuanced insights like which portions of my graph whether it’s your supply chain, customers etc. are low risk or high risk. Then machine learning can leverage those nuanced insights, turn those more nuanced inputs into better models.”
- “Graph algorithms scale exponentially. Graph requires scalable software, more so than any of the other situations or challenges you have considered.”

Graph + AI Summit 2021 Keynote

Real-World Better Outcomes from Graph+AI

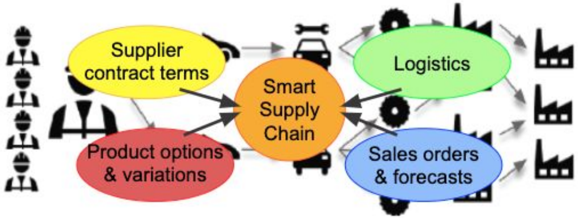
Healthcare: UnitedHealth Group

Real-time recommendations

- 1.3TB graph brain
- Real-time care recommendations
- **Improve customer care for 50 million customers**

Industrial Supply Chain: Jaguar Land Rover

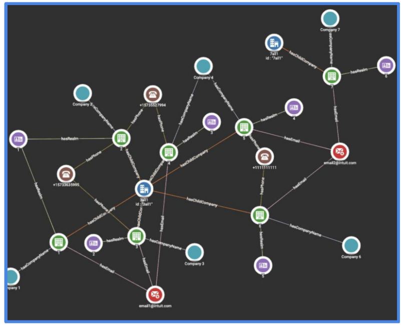
Analytics for decisions



- Analytics: 3 weeks→ 45 minutes
- Reveal opportunities, optimize tactical & strategic decisions
- **Saving £100M+/yr**

Financial Services: Intuit

Real-time fraud detection



- Integrate multiple tools
- "Magical" real-time visual results for investigators
- **Cut the Fraud in Half**

Case 1: Analytical Queries & Graph Algorithms

Types of Graph Algorithms

- Path Finding
- Clustering / Community Detection
    - Lenient clustering - connected component: one connection
    - Strict clustering - clique detection: every possible connection
    - **Relative density** - more connections in-group than between-group
- **Ranking and Centrality**
    - PageRank, HITS
    - SimRank, RoleSim
    - Closeness, Betweenness
- **Similarity**
- **Frequent Pattern Discovery**

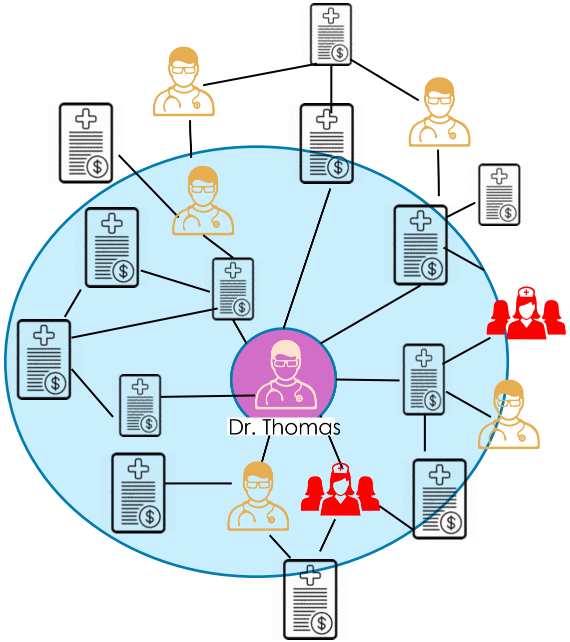
BOLD indicates more complex tasks, with iterative algorithms, which can be considered **unsupervised learning**

Finding the Most Influential Health Care Providers in a Community

- Who is the **most influential provider** in each region for a particular medical condition?
  ⇒ Use **PageRank** to rank each provider based on the relative importance of their referrals

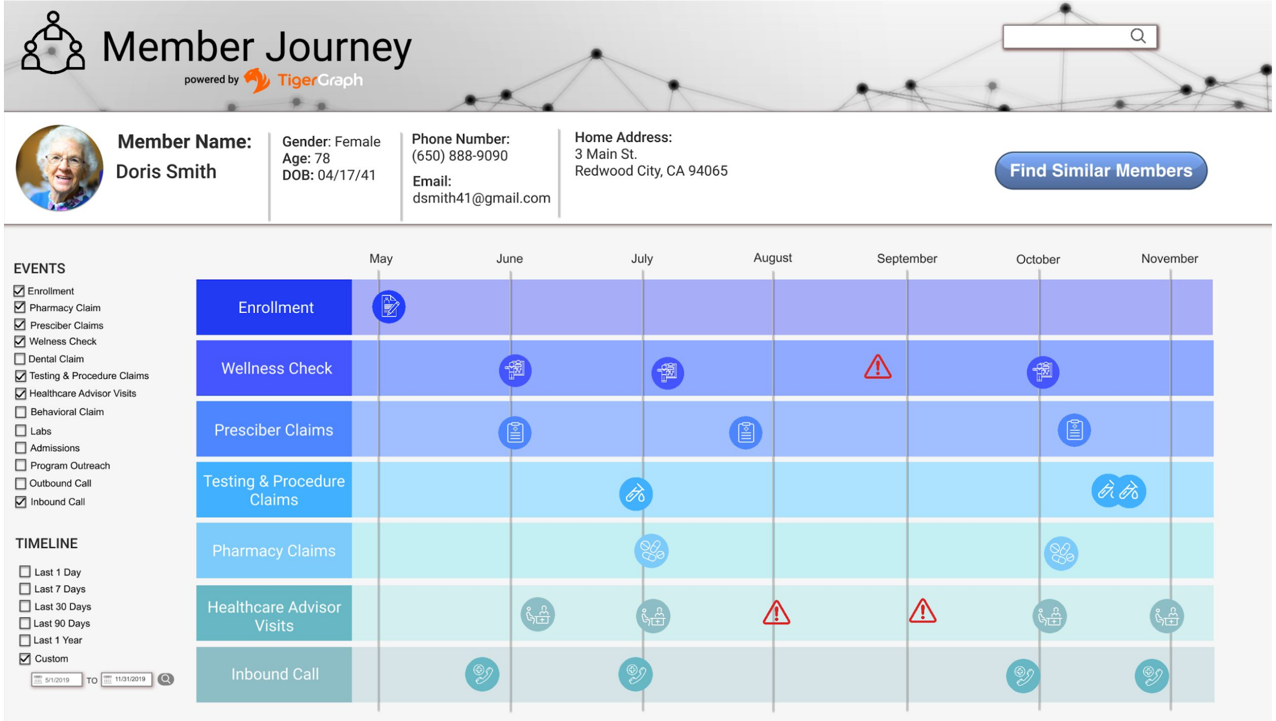
- **Who is influenced** by these leaders (e.g. other doctors, chiropractors, physical therapists, facilities)?

  ⇒ Use **Community Detection** to find the groups surrounding Influencers



Graph with Patients, Providers, and Service Claims

Find Similar Cases to deliver better healthcare

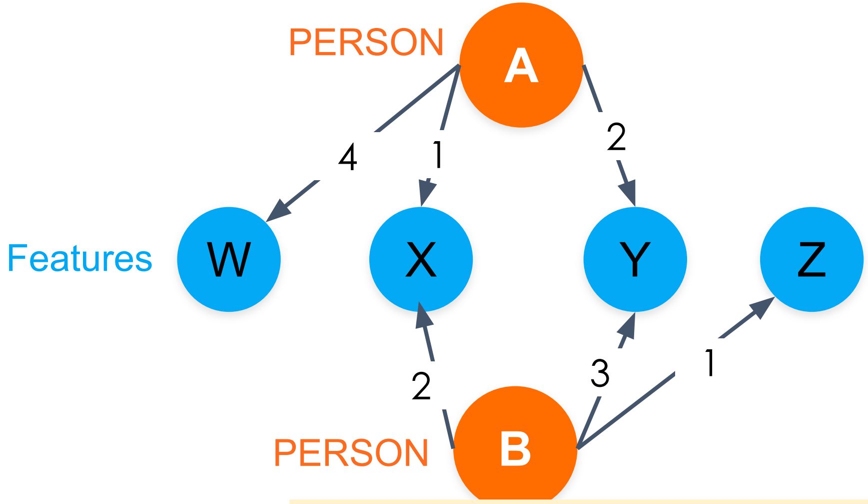


- Seamlessly integrate multiple sources of data to provide unified and comprehensive view for each journey among 50M members
- Find similar members with a click of a button in real-time
- Deliver care path recommendations for similar members

Graph-Based Structural Similarity: Cosine Similarity

Use a vertex's neighbors as its feature set

- **Cosine**: Use edge weights to each neighboring vertex



A's weighted neighbors = {4,1,2,0}
B's weighted neighbors = {0,2,3,1}

Cos(A,B) = 8 / [$\sqrt{21}\sqrt{14}$] = **0.4666**

W,X,Y,Z represent feature vertices, different vertex types than A,B

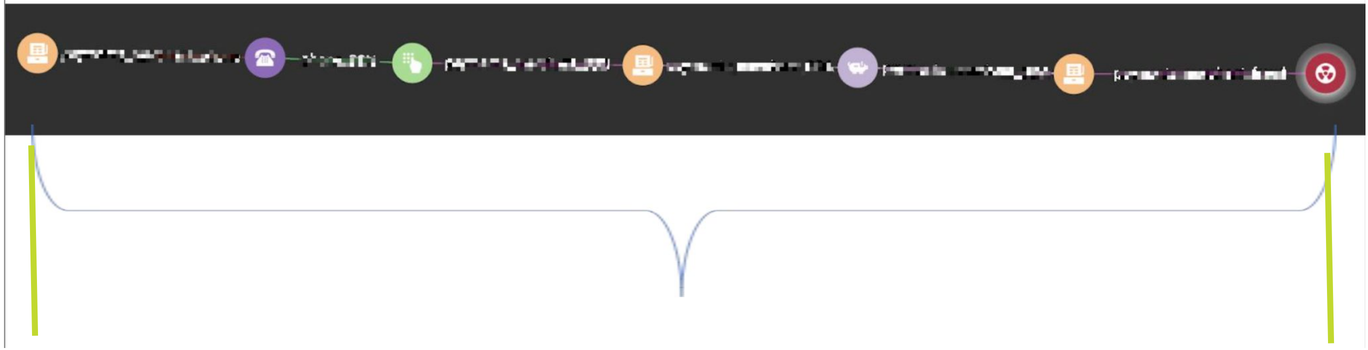
Case 2 - Graph-Based ML Feature Extraction

Traditional ML Features

- **Aggregation-based features:** e.g. count of payments from the same device in past 30 days
- **Ratio-based features:** e.g. number of fraudulent payments to number of legitimate payments for a given device
- **Row-based features:** e.g. Geo-location mismatch between IP address and zip code

Graph Based Features

- Graph based features brings a new perspective to the prediction game
- By adding graph features to fraud detection models we can detect delicate fraud patterns

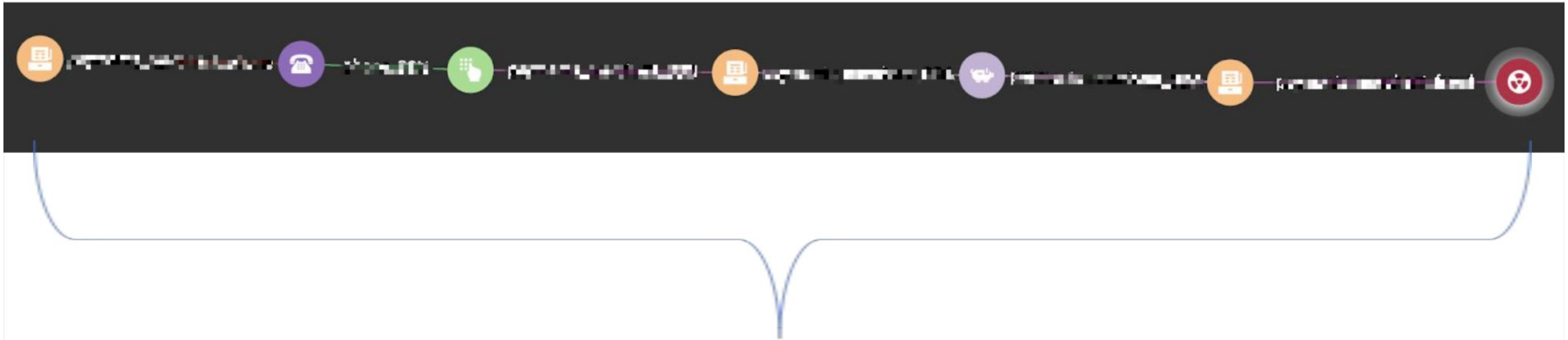


Account **6 Hops** **Confirmed fraud**

Source: [“Using Graph to Boost AI”](), Uri Lapidot, Sr. Product Manager, Intuit at Graph + AI Summit Spring 2021

Graph Based Features

- Graph based features brings a new perspective to the prediction game
- By adding graph features to fraud detection models we can detect delicate fraud patterns

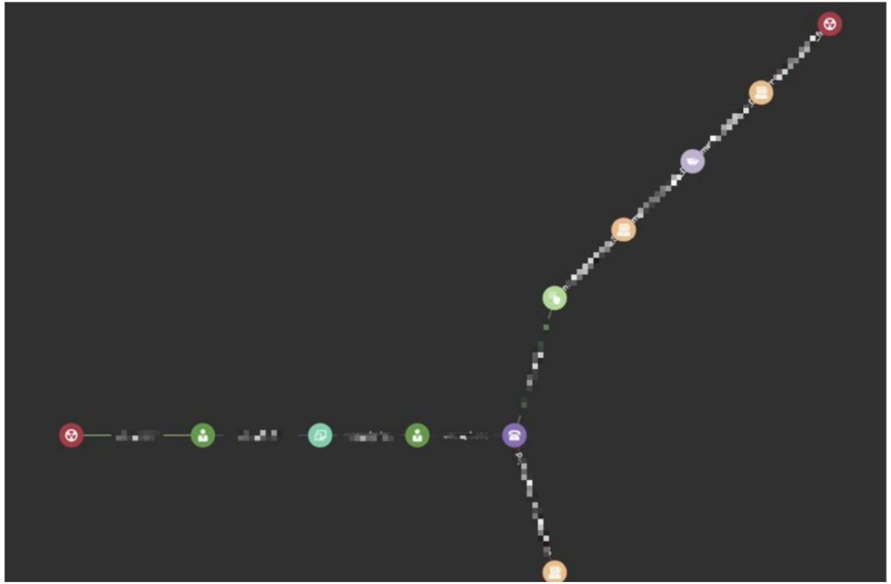


6 hops to fraud

Case 2 - Graph-Based Feature Examples

Here's a few graph basic features:

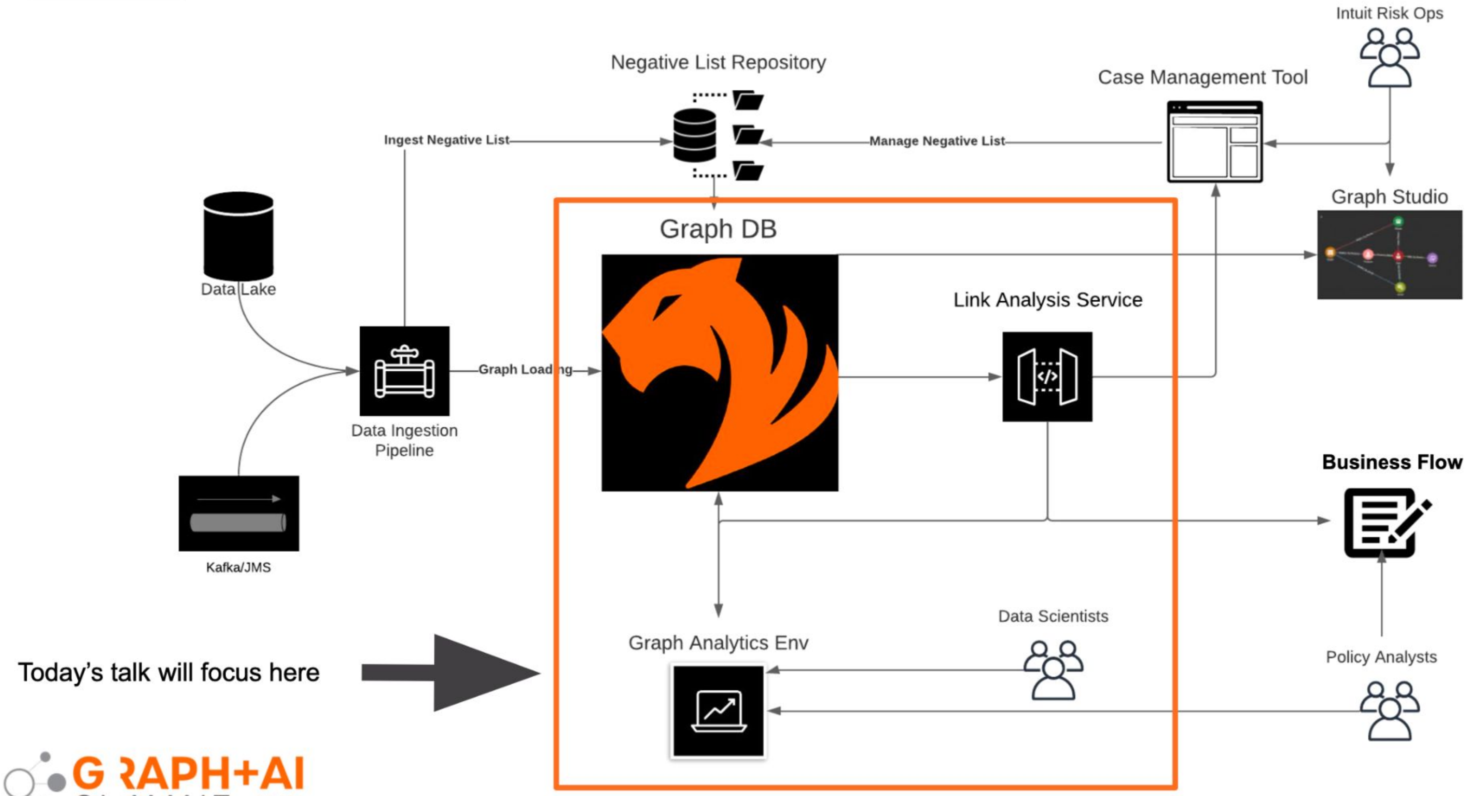
1. Min linked distance
2. Average linked distance
3. Number of links to target
4. Link strength
5. Path to target



Source: "Using Graph to Boost AI", Uri Lapidot, Sr. Product Manager, Intuit at Graph + AI Summit Spring 2021

Intuit's Architecture Using Graph-Based ML Features

Link Analysis Using TigerGraph



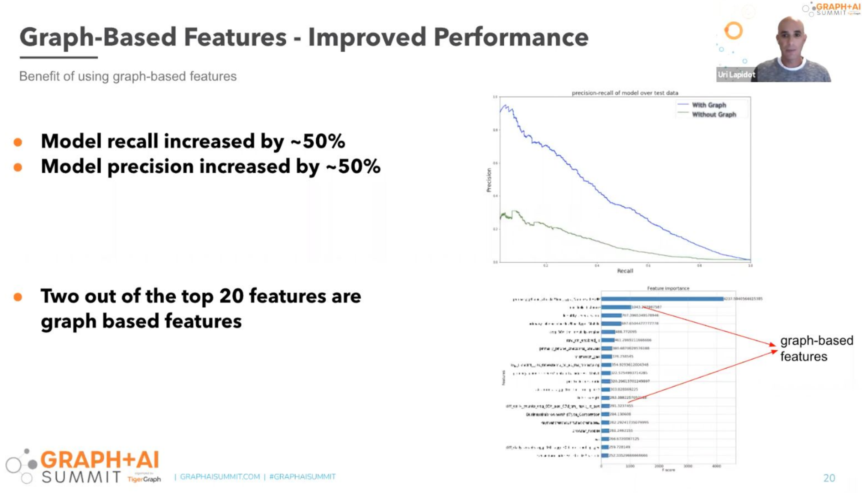
Source - https://info.tigergraph.com/graph-ai-summit-spring-2021-using-graph-to-boost-ai

Intuit Cuts the Fraud in Half with Graph-based Machine Learning

Uri Lapidot
Intuit

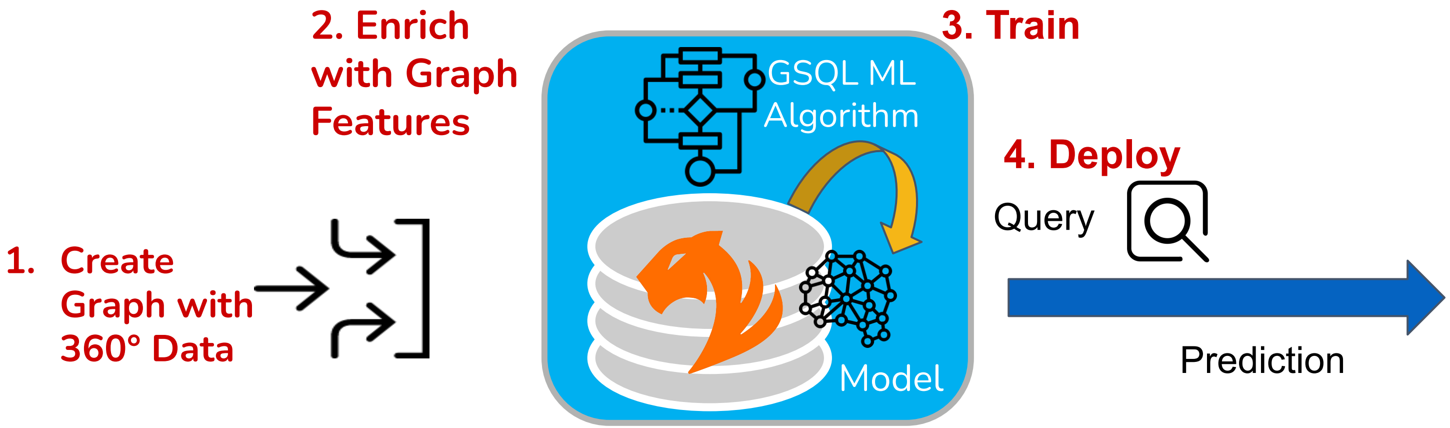
- "Using graph-based *(machine learning)* features, our model shows an amazing improvement in detecting 50% more risk *(fraud)* events and also improving model precision *(reduce false positives)* by 50% at the same time."

- "For us this is a game changing technology that we intend to leverage more and more in the future."



Session slides and recording - https://info.tigergraph.com/graph-ai-summit-spring-2021-using-graph-to-boost-ai

Case 3: In-Database Machine Learning with TigerGraph



- Native graph storage
- Coded once, auto scale-out & scale-up
- Real-time updates
- GSQL Turing-complete language
    - Preprocess data
    - Training: flow-control, accumulator, pattern match
    - Model validation

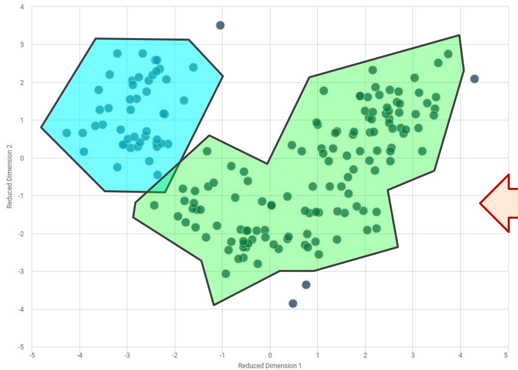
Applications:

- Entity resolution
- Recommendation
- Fraud detection
- ...

Community Detection with Louvain

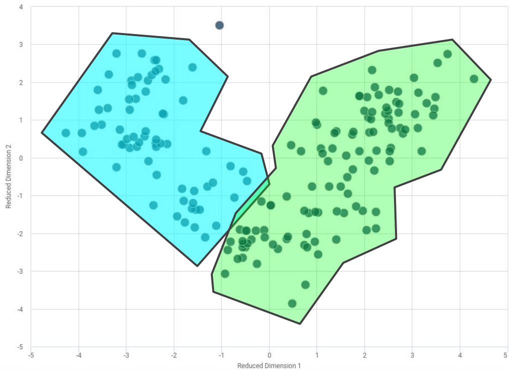
Modularity is a measure of how "good" is the partitioning of a graph

= (fraction of the edges that fall ***within*** the given groups)
minus (the expected fraction ***if edges were distributed at random***)



$$Q = \frac{1}{2m}\sum_{vw}\left[A_{vw} - \frac{k_v k_w}{2m}\right]\frac{s_v s_w + 1}{2}$$

Which has a higher modularity?



- Researchers at the University of **Louvain** developed an especially efficient method for finding the partitioning with the best Q score.

Detecting Fraud Rings with TigerGraph Tier 1 U.S. Bank

Business Challenge
A leading U.S. bank wanted a better way to detect and remove fraudsters from their credit-card network. Prototypes had shown that a combination of advanced graph algorithms gave significant gains – big-data tools and other graph technologies either couldn't scale to the full customer base or gave inconsistent results.

Solution

- Implementing PageRank and Louvain [fraud] community detection in an MPP native-parallel database.
- Leveraging deep analytics to find hidden connections across 20TB+ of data.

Business Benefits

- Able to expose fraud rings, shut down connected cards, and combat fraudulent activity on a massive scale –**35% uplift** and **$50M incremental fraud avoidance**. >**$1.5 million through cost savings** on false positives, infrastructure and TCO

10TB
Card applications data

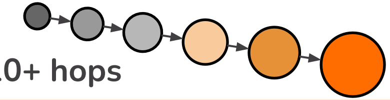
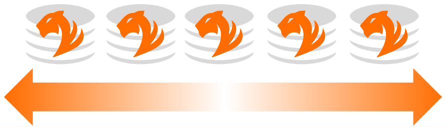
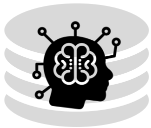
6 weeks
PoC elapsed time

3 months
Time to build and fully deploy platform to production

+$50M
1st year ROI with 35% uplift in fraud detection

CLV Impact > $200M

The TigerGraph Difference

Feature	Design Difference	Benefit
Real-Time Deep-Link Querying  5 to 10+ hops	• Native Graph design • C++ engine for high performance • Storage Architecture	• Uncovers hard-to-find patterns • Operational, real-time • HTAP: Transactions+Analytics
Handling Massive Scale 	• Distributed DB architecture • Massively parallel processing • Compressed storage reduces footprint and messaging	• Integrates all your data • Automatic partitioning • Elastic scaling of resource usage
In-Database Analytics & Machine Learning 	• GSQL: High-level yet Turing-complete language • User-extensible graph algorithm library, runs in-DB • ACID (OLTP) & Accumulators (OLAP)	• Avoids transferring data • Richer graph context • Graph-based feature extraction for supervised machine learning • In-DB machine learning training
NO CODE	• No-code migration from RDBMS • No-code Visual Query Builder	• Democratize self-service analytics to derive new-insights from legacy/external data stores

START FREE

Start in minutes, build in hours and deploy in days with the industry's first and only distributed graph database-as-a-service.

https://www.tigergraph.com/cloud/

TigerGraph

Thank You