GRAPH+AI SUMMIT

organized by TigerGraph

Graph Data Science Algorithms for Analytics and Machine Learning

Dr. Victor Lee, VP of ML/AI, TigerGraph

Outline

- The Power of Graph Analytics and Algorithms
- In-Database Graph Data Science
- Three ways to leverage Graph Algorithms for ML
    - Unsupervised Learning
    - Feature Enrichment from Graph Features
    - In-Database Learning

Smarter Analytics and AI

Good business decisions require data with **context**, **recency**, and **relationship**

- Context - it is situational?
- Recency - is it fresh?
- Relationship - how do different facts relate to one another?

Good AI requires training data with the right **content** and right **structure**

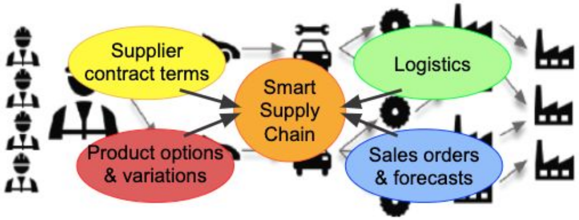
Graphs provide the relationships, rich content, and flexible structure for better analytics and decision making

Real-World Better Outcomes from Graph+AI

Healthcare:
Real-time recommendations

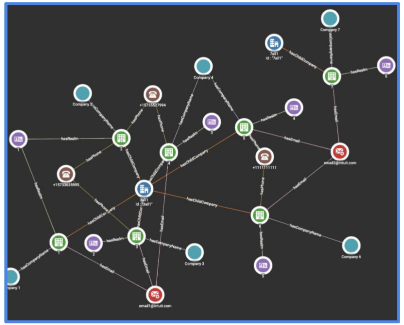
- 1.3TB graph brain
- Real-time care recommendations
- Improving healthcare, lowering cost

Industrial Supply Chain:
Analytics for decisions



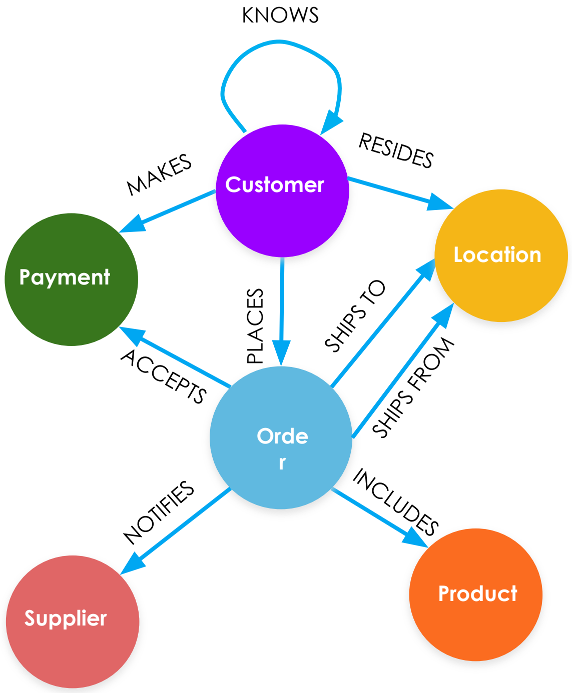
- Analytics: weeks→ minutes
- Reveal opportunities, optimize tactical & strategic decisions
- Saving $100M+/yr

Financial Services:
Real-time fraud detection



- Integrates multiple tools
- "Magical" real-time visual results for investigators
- Scalable for growth

Power of Graph



Richer, Smarter Data with Relationship

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

Deeper, Smarter Questions

- Look for semantic patterns of relationship
- Search farther than other DBs

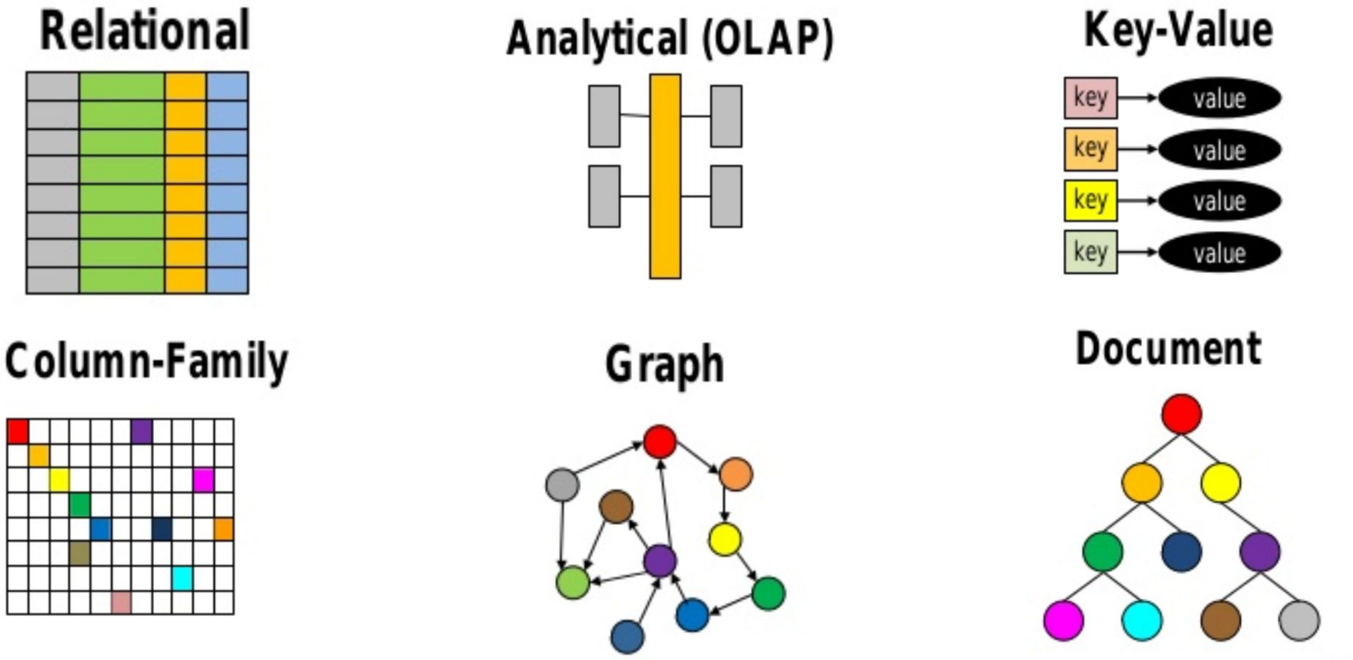
More Computational Options

- Graph algorithms
- Graph-enhanced machine learning

Explainable Results

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

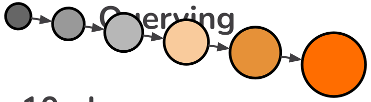
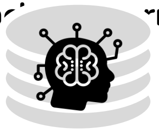
How is Graph Model different than other Data Models ?



Graph Model can represent all other Data Models **Naturally** !

Author: Dan McCreary: https://www.slideshare.net/Dataversity/nosql-now-nosql-architecture-patterns-23589170

Analytical Graph Database Characteristics

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
	• 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 & Ma[illegible]rning 	• 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

Three ways to leverage Graph Algorithms for ML

1. Unsupervised Learning
2. Feature Enrichment from Graph Features
3. In-Database Learning

(1) Graph Algorithms for Unsupervised Learning

Types of Graph Algorithms

- Path Finding
- **Ranking and Centrality**
  - PageRank, HITS
  - Closeness, Betweenness
- **Similarity**
  - Cosine, Jaccard
- **Clustering / Community Detection**
  - Louvain modularity
- **Frequent Pattern Discovery**

BOLD indicates more complex tasks, which produced a value for each vertex and can be considered **unsupervised learning**

Graph Algorithm Use Cases

Algorithm	Key Use Cases
Shortest Path (X to Y)	● Most efficient **route/process** from X to Y, physical or conceptual ● Likelihood of two parties X & Y being aware of one another
PageRank of X	● Predicting X's **influence** on others ● Likelihood of an arbitrary person/entity being aware of Entity X
Closeness/ Betweenness	● Most efficient **location** for a hub
Community Detection	● Identifying **social groups** or groups of mutual awareness/influence
Similarity between (X,Y), Classification	● **Recommendations** and **Substitutions** (If you liked X, you'll like Y) ● **Prediction** (If something was true for X, it may be true for Y) ● Classification (X and Y belong to the same category)

TigerGraph In-Database *Graph Data Science Library*

Signals our commitment to serving the needs of data scientists

- More algorithms (15 released this week)
  - Graph Embedding
  - Link prediction
  - Similarity
  - Centrality

- More than just algorithms

In-Database

- No export needed
- Live, updatable data
- Scaleable, Ultra-fast engine
- GSQL query language

For Data Scientists

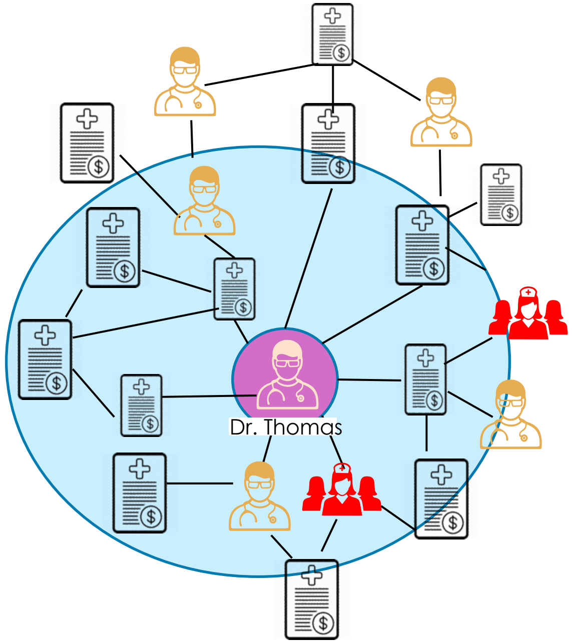
- easier and faster to run
- include ML, such as graph embeddings
- will integrate with feature & model management
- will integrate with Graph+ML Workbench

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

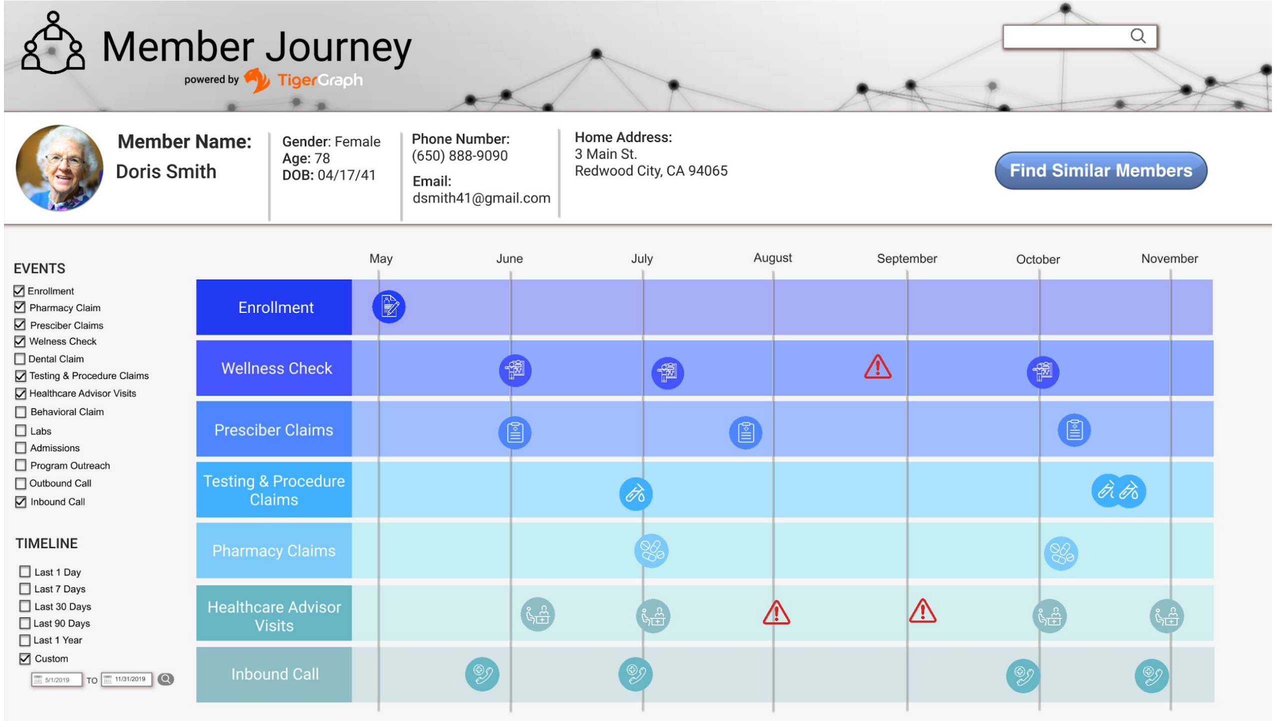
- **Who is influenced** by these leaders (e.g. other doctors, chiropractors, physical therapists, facilities)?

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



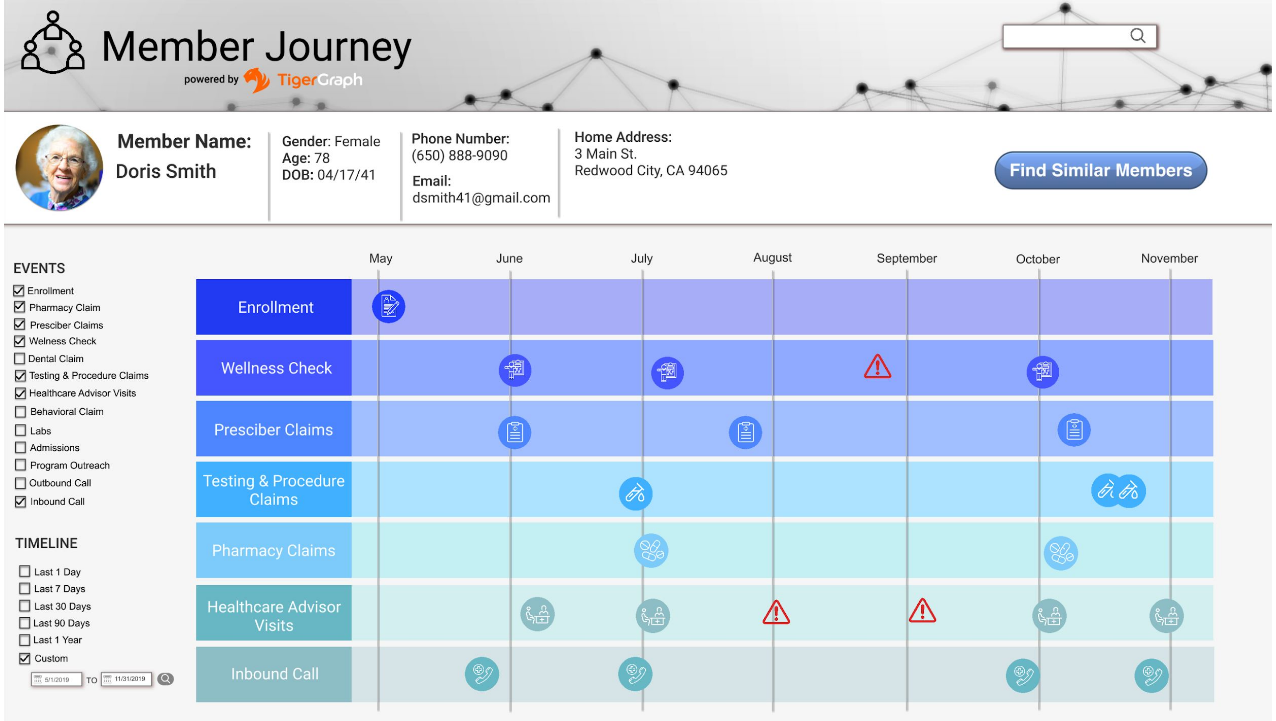
Graph with Patients, Providers, and Service Claims

Finding Similar Cases to deliver better healthcare



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

Finding Similar Cases to deliver better healthcare

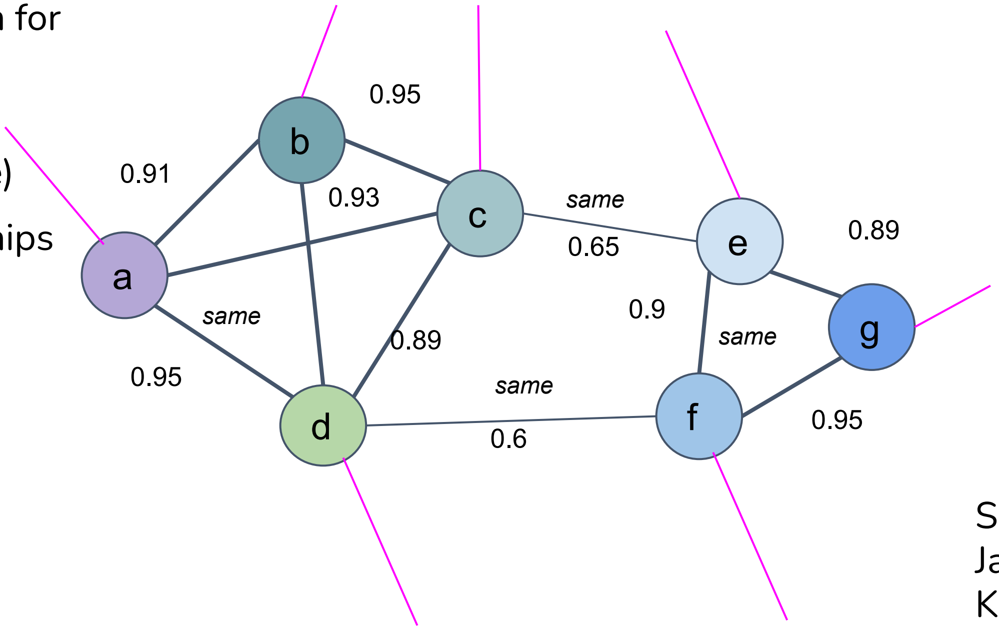


- Seamlessly integrate multiple sources of data to provide unified and comprehensive view for each journey among 50M Medicare members
- Find similar members with a click of a button in real-time
- Deliver care path recommendations for similar members

Entity Resolution using Similarity Scores

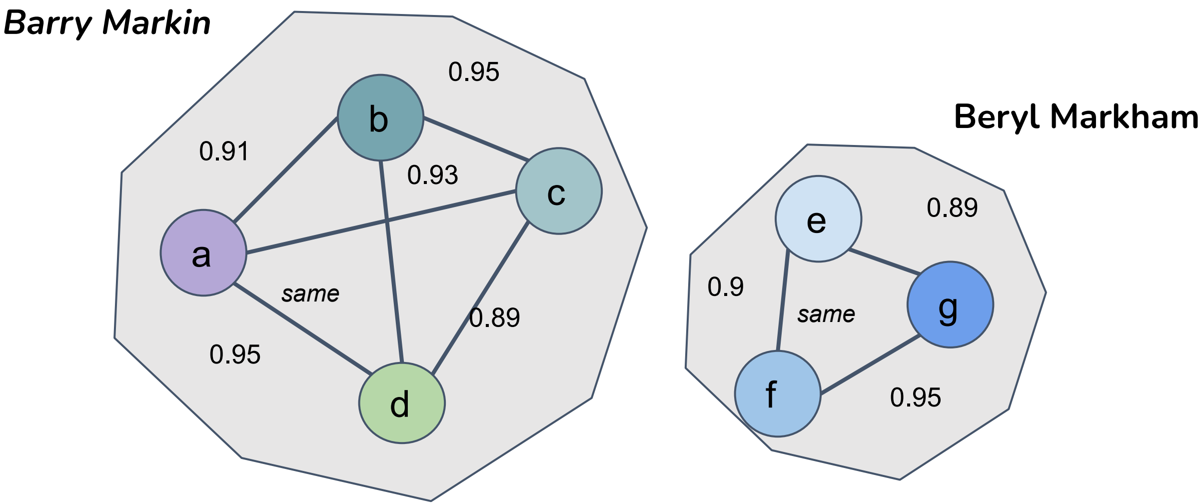
Apply a scoring system for comparing entities:

- Similar attribute values (e.g. name)
- Similar relationships (school, work, activities, ...)

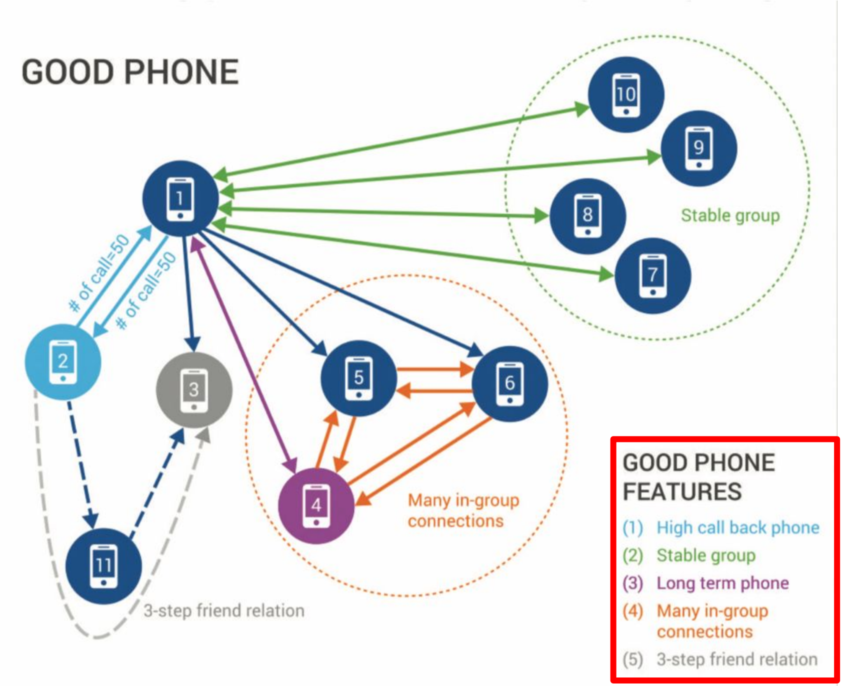


Several scoring systems: Jaccard, Cosine similarity, Kolmogorov distance, etc.

Entity Resolution using Similarity Scores



(2) Graph Feature Extraction - Better Training Data



Challenge

Find and report fraudsters among billions of calls per week.

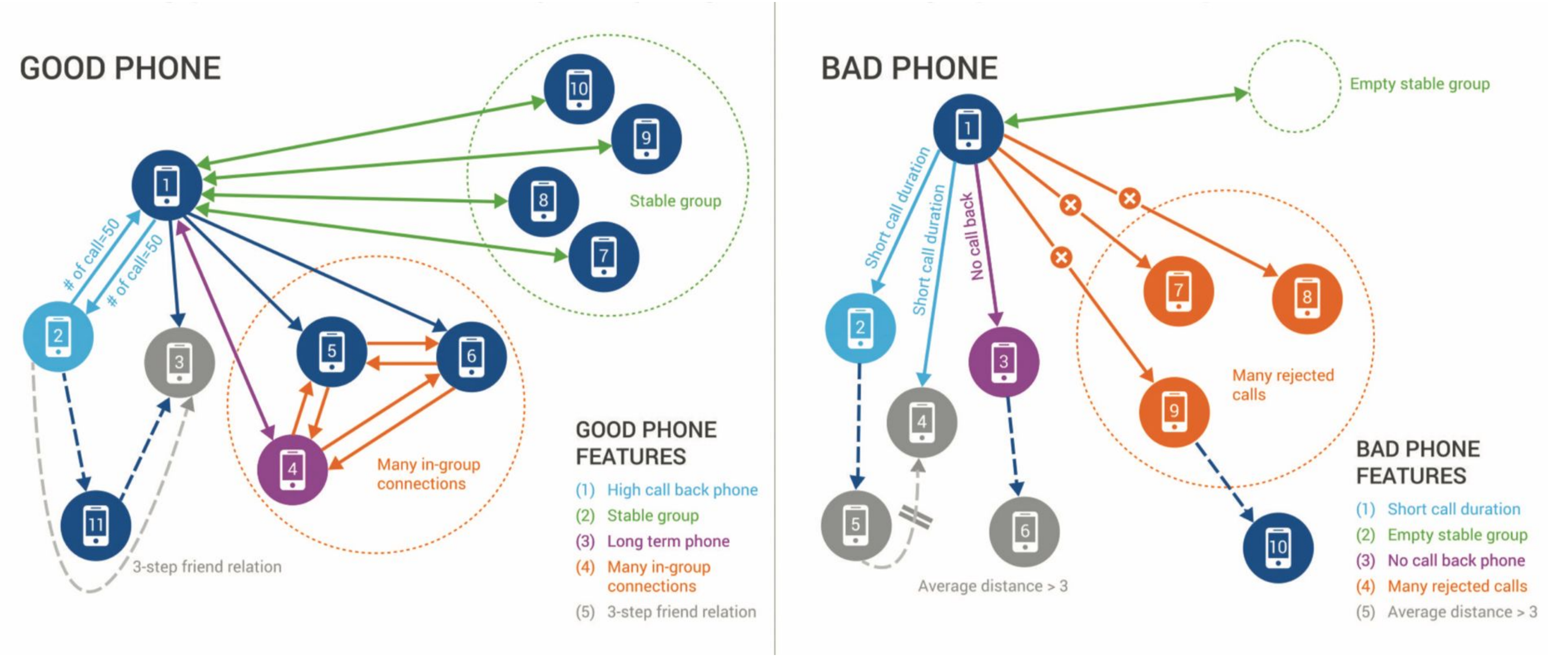
Solution

- **Build graph**: Real-time operational graph with 600M phone nodes & 15B call detail records.
- **Get features and labels**: Domain experts write GSQL queries to extract 118 features/phone. Some past calls are labeled for 3 types of unwanted calls.
- **Train**: Feed machine learning with training data for fraud detection with 118 features/phone for 30M calls.
- **Deploy**: For each incoming call, extract the current 118 features (subsecond) and apply model for real-time answer.

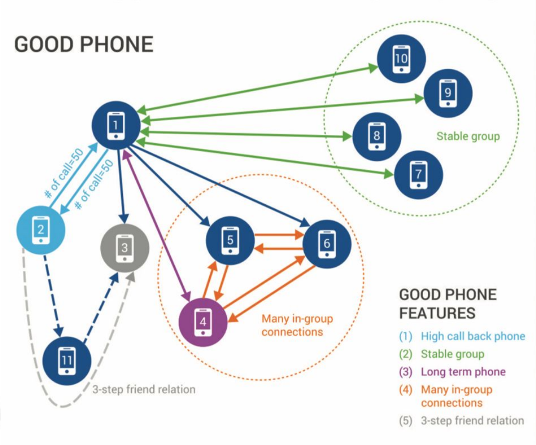
Results

- If unwanted call is predict, display alert on recipient's phone
- Process 2000+ calls/sec
- Improved customer satisfaction

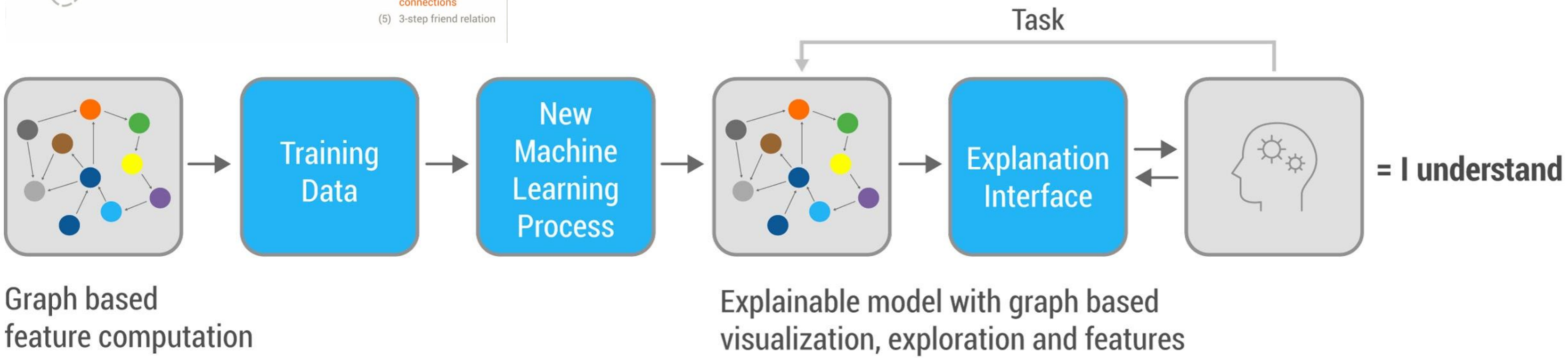
Graph Feature Example



Graph Features Support Explainable AI



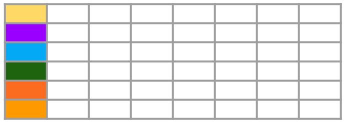
- Semantic relationships/patterns
- Visual relationships/patterns



Feature Extraction using Graph Embeddings

- ***Graph Embedding*** transforms graph structure into a compact set of vertex vectors.
  - Captures the essence of a vertex's "nature" as a set of *latent features*
  - Enables graph data to run efficiently on non-graph neural networks

perform embedding



vertex embeddings

Perform Analysis / ML

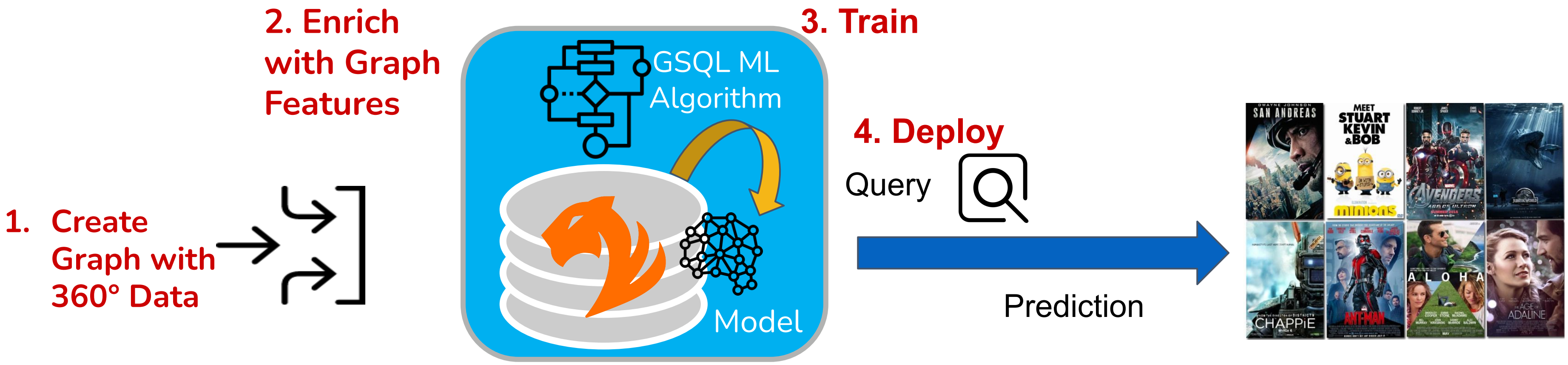
Works for numerous cases

- Recommendation (similarity)
- Fraud detection (classification)

Compact →
scalability

Set of vectors →
compatibility,
reduced complexity

(3) In-Database Machine Learning



- Simplified pipeline
- No need to export/import
- Fresh data, up to date
- DB needs to be scalable, with parallel processing

Applications:

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

In Database ML for Movie Recommendation

MARVEL'S THE AVENGERS

PG13, 2 hr.22 min.
Action & Adventure , Science Fiction & Fantasy
Directed By: Joss Whedon
In Theaters: May 4, 2012 Wide
On DVD: Sep 25, 2012
Walt Disney Pictures

The Avengers: Trailer 1
1 minute 55 seconds
Added: Apr 24, 2018

The Avengers: Trailer 2
2 minutes 22 seconds
Added: Apr 24, 2018

movie features

MARVEL'S THE AVENGERS REVIEWS

All Critics | Top Critics | **All Audience**

users

ratings

NEXT →

Danny D

5d ago

How many movies did it take to come up with this mundane plot ?

Benjamin C

Martyn K

Goals:

- Predict users' ratings for movies, based on previous ratings
- Recommend movies to users based on rating prediction

TigerGraph CLOUD

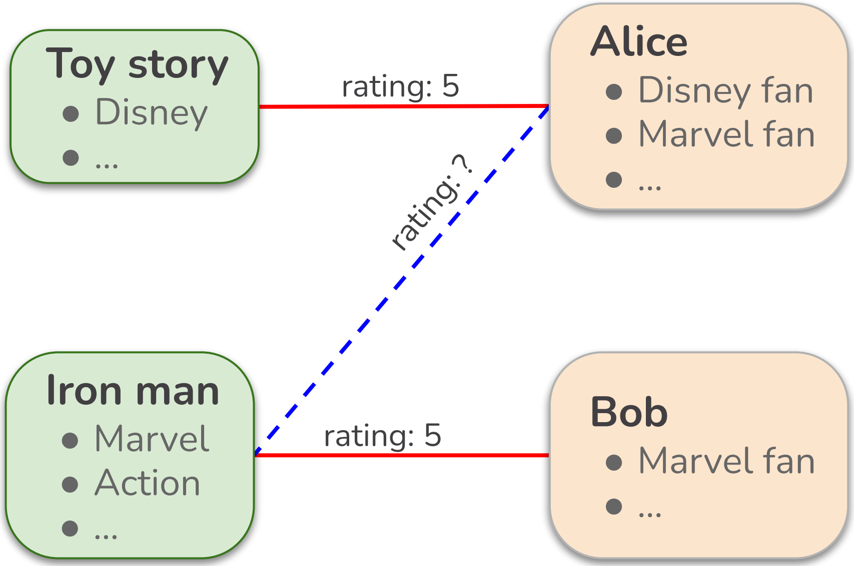
Low-Rank Approximation
Machine Learning v3

TigerGraph GraphGurus

EPISODE 28

An In-Database Machine Learning Solution For Real-Time Recommendations

User-Rating-Movie Graph



MovieLens dataset
https://grouplens.org/datasets/movielens/

- 100K ratings and 40K tags that 1K users gave to 17K movies
- Ratings are from 0 to 5 stars

Recommendation Approaches

- Collaborative filtering
- Content based method
- K-nearest neighbors
- **Latent factor (model-based)**
- Hybrid method
- ...

Movie Rating Prediction (Latent Factor Model)

$\theta^{(1)} = [5, 0]$ $\theta^{(2)} = [5, 0]$ $\theta^{(3)} = [0, 5]$ $\theta^{(4)} = [0, 5]$

romance

action

	Movie	Alice		Bob	Carol	Dave
$x^{(1)} = [0.9, 0]$	Love at last	5	4.5	5	0	0
$x^{(2)} = [1, 0.1]$	Romance forever	5	5.0	_	_	0
$x^{(3)} = [0.9, 0]$	Cute puppies of love	_	4.5	4	0	_
$x^{(4)} = [0.1, 1]$	Toy story	_	0.5	_	_	5
$x^{(5)} = [0.1, 1]$	Sword vs. karate	0	0.5	0	5	_
$x^{(6)} = [0, 0.9]$	Nonstop car chases	0	0.0	0	5	4

- Each movie has a latent factor vector: $\boldsymbol{\theta}^{(j)}$
- Each user has a latent factor vector: $\boldsymbol{x}^{(i)}$
- Predict the user j's rating to movie i by: $(\boldsymbol{\theta}^{(j)})^T\boldsymbol{x}^{(i)}$

Graph Machine Learning Techniques

Neural Networks for Graphs

Including the graph's connection information in the ML training produces more accurate models and/or reduces the need for explicit features.

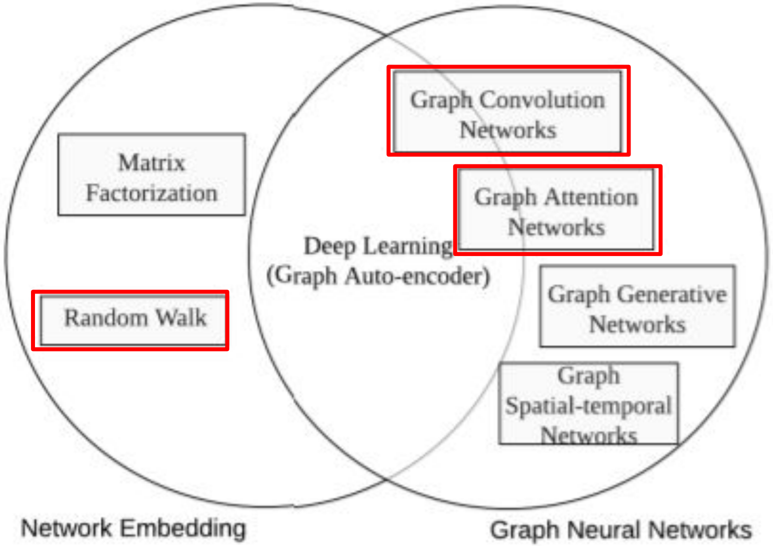
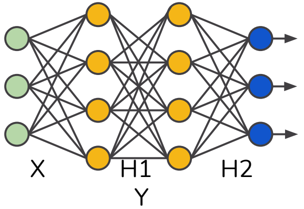


Fig. 2: Network Embedding v.s. Graph Neural Networks.

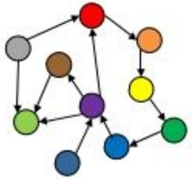
https://medium.com/@terngoodod/a-comprehensive-survey-on-graph-neural-networks-part-1-types-of-graph-neural-network-1dd93b823c70

Graph Neural Networks

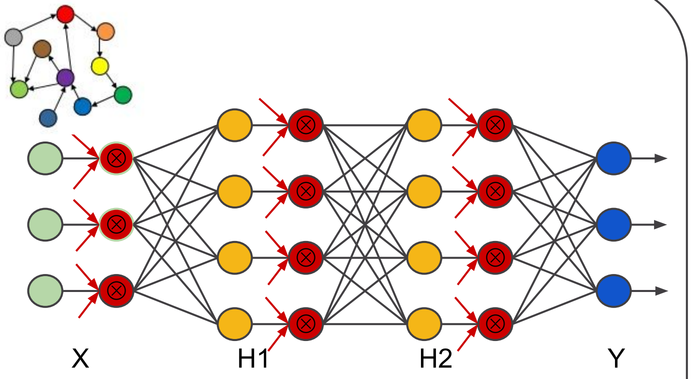


Neural Network

Powerful Machine Learning to Predict and Classify


Graph

Insight from Connected Data



Graph Convolutional Neural Network

- Combines the added insight from connected data with the modeling power of neural networks
- Uses the graph during training; in-database training is the ideal.

Summary

- **Natural Data Model** - Graph is how we think
- **Richer Data** - connections between entities, graph-based features
- Graphs have always had a **natural role in machine learning**:
    - Unsupervised learning through **graph algorithms**, frequent pattern mining
    - **Graph features** provide richer training data
    - **Graph embeddings** transform graph data into easier-to-use vectors
    - Learning through **neural networks** and deep learning
- Graph data models are uniquely qualified to provide **explanatory AI**.
- Native Graphs with Massively Parallel Processing like TigerGraph enable large scale feature extraction and in-graph analytics