@datascisalon

DATASCIENCE SALON VIRTUAL

Smarter AI with Analytical Graph Databases

Dr. Victor Lee
Head, Product Strategy
TigerGraph

Gaurav Deshpande
VP, Marketing
TigerGraph

“Graph analysis is possibly the single most effective competitive differentiator for organizations pursuing data-driven operations and decisions after the design of data capture.”

Gartner

DATASCIENCE SALON VIRTUAL

@datascisalon

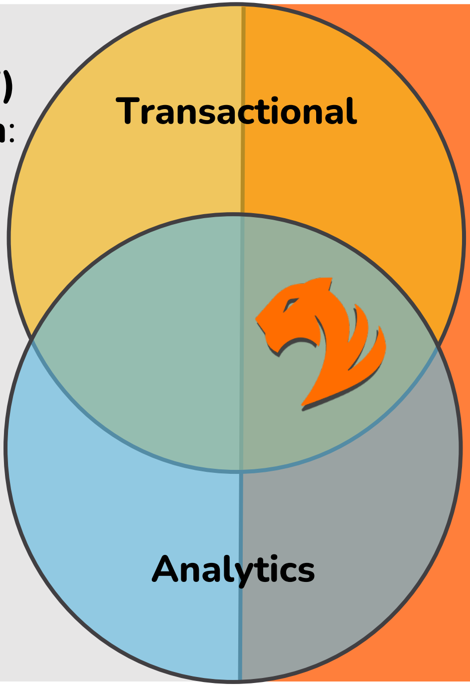
AGENDA

- What is a Graph Database?
  What is an Analytical 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

Types of Graph Databases

Semantic (RDF) Knowledge Graph:

- Collection of facts (RDF triples)
- Ontology to model concepts & rules
- **Pattern matching**
- Logical **inference**
- Standards-based

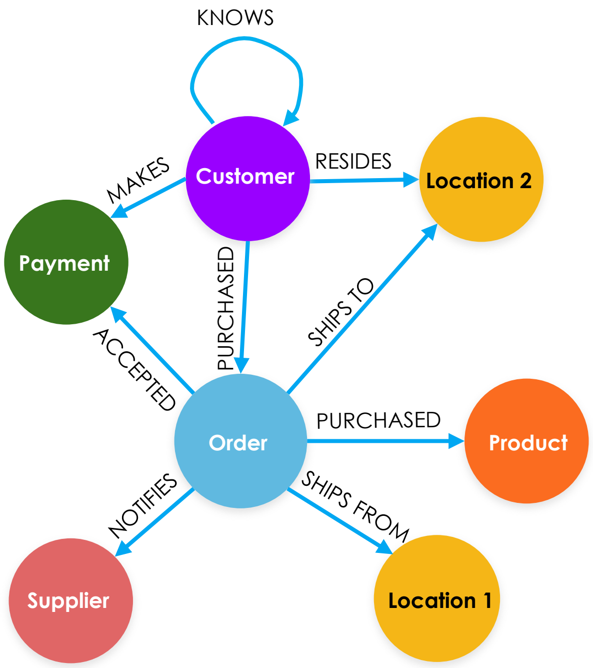


Property Graphs:

- Node and Edge objects
- **Higher performance** for queries, transactions, and **advanced analytics**
- **Pattern matching**
- Schema-free or Schema-based
- Schemas allows application-specific tuning

TigerGraph is a High-Performance and Scalable Property Graph, for both Analytics & Transactions.

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

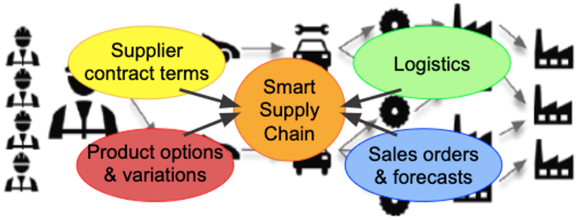
Real World Better Outcomes from Graph+AI

Healthcare: Real-time recommendations	Industrial Supply Chain: Analytics for decisions	Financial Services: Real-time fraud detection

Healthcare:
Real-time recommendations

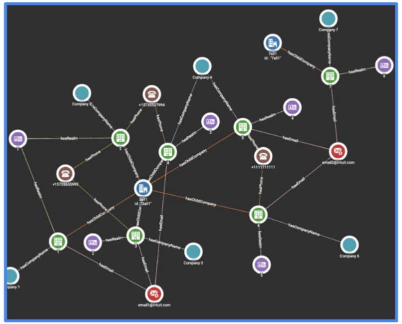
- 1.3TB graph brain
- Real-time care recommendations
- Improving healthcare, lowering cost

Industrial Supply Chain:
Analytics for decisions



- Analytics: weeks→ minutes
- Reveal opportunities, optimize tactical & strategic decisions
- Saving $25M+/yr

Financial Services:
Real-time fraud detection

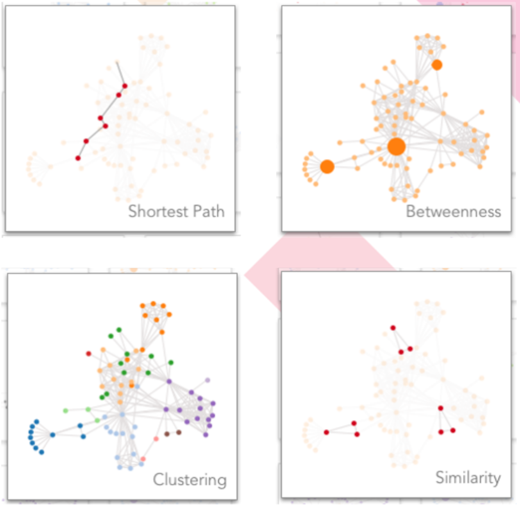


- Integrates multiple tools
- "Magical" real-time visual results for investigators
- Scalable for growth

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**

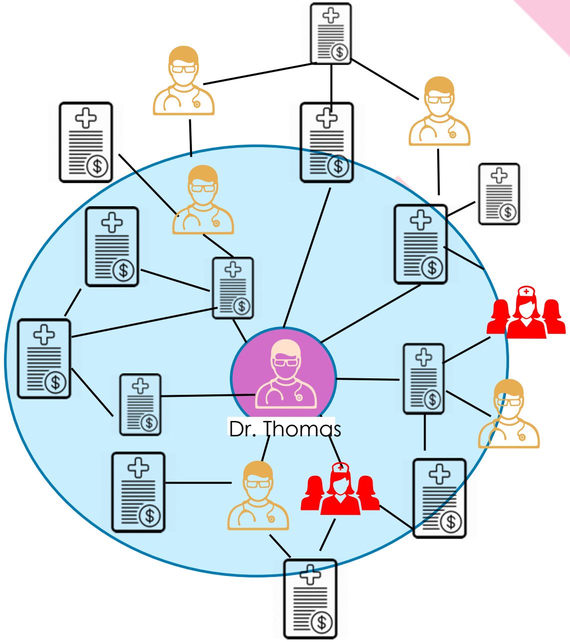
Finding the Most Influential Health Care Providers in a Community

- 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)?

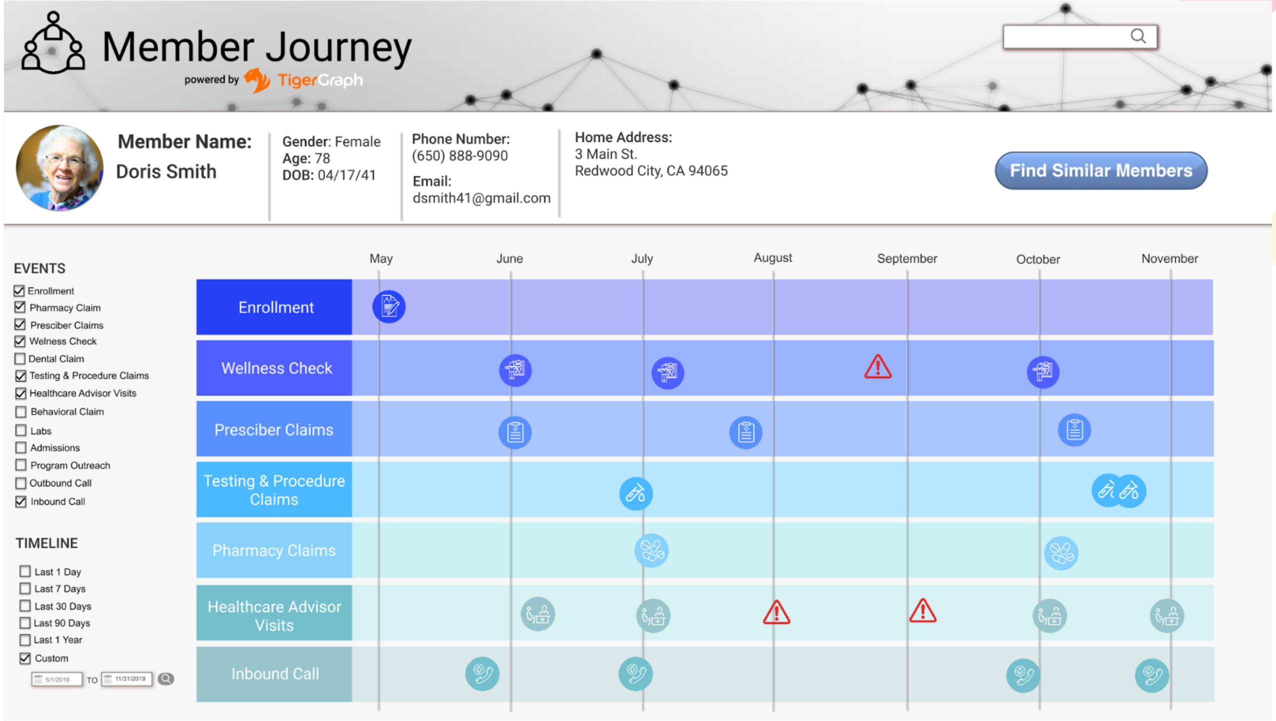
  ⇒ **Use Community Detection** to find the groups surrounding Influencers



Graph with Patients, Providers, and Service Claims

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



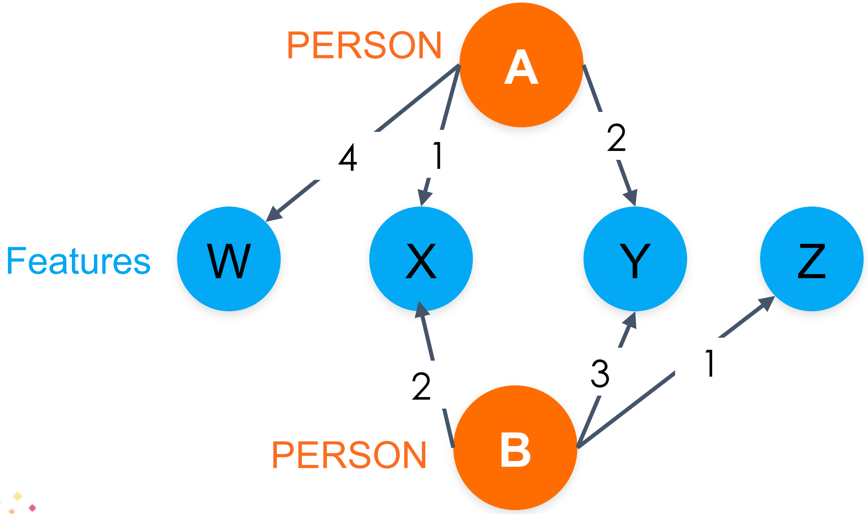
DATASCIENCE SALON VIRTUAL

@datascisalon

Graph-Based Structural Similarity

Use a vertex's neighbors as its feature set

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

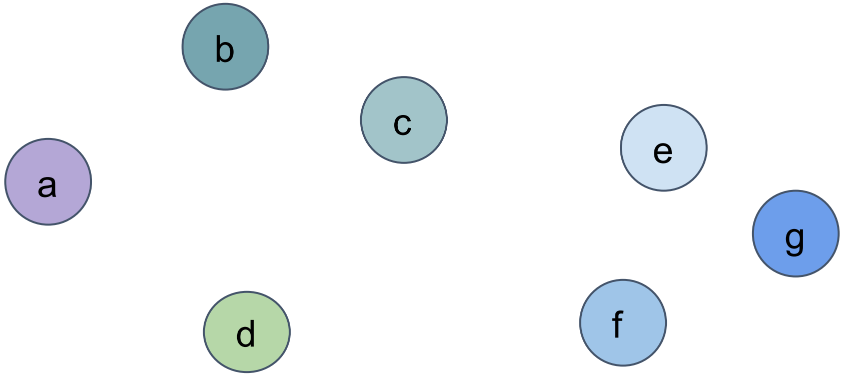


A's weighted neighbors = {4,1,2,0}
B's weighted neighbors = {0,2,3,1}

Cos(A,B) = 8 / [√21√14] = **0.4666**

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

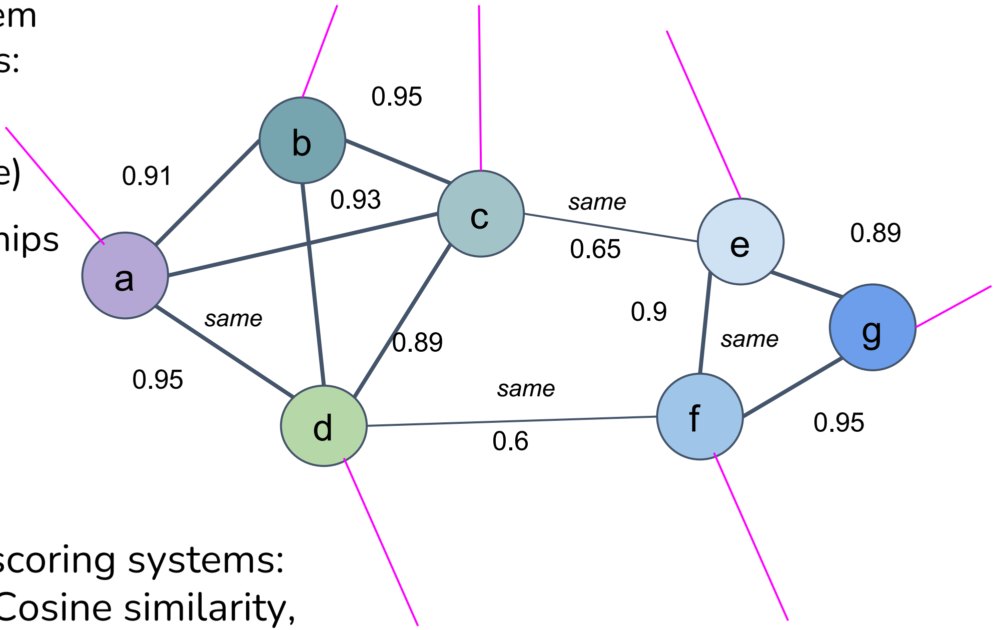
Entity Resolution using Similarity Scores



Entity Resolution using Similarity Scores

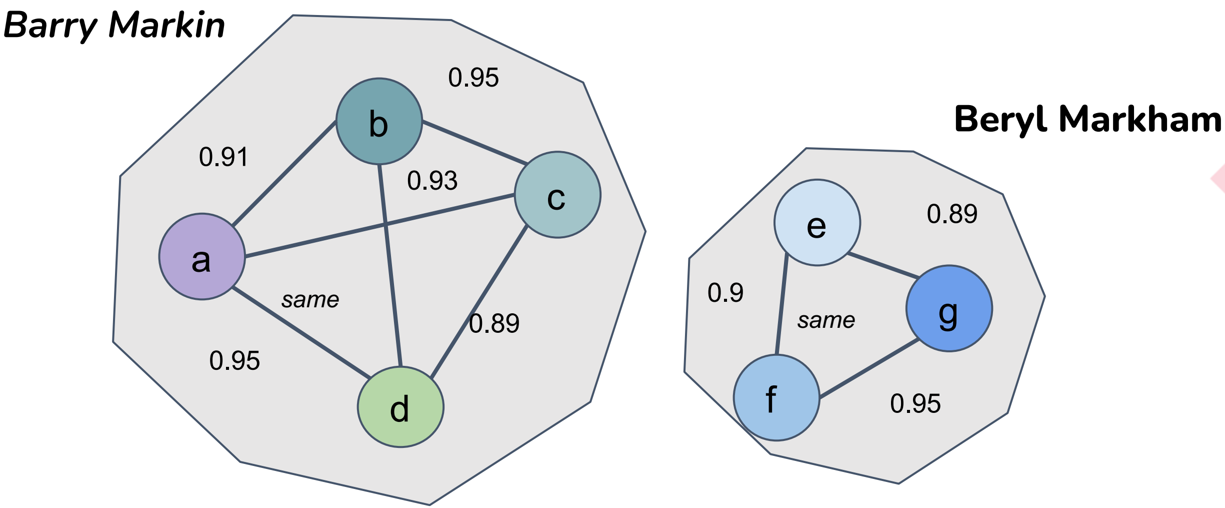
Apply a scoring system for comparing entities:

- 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

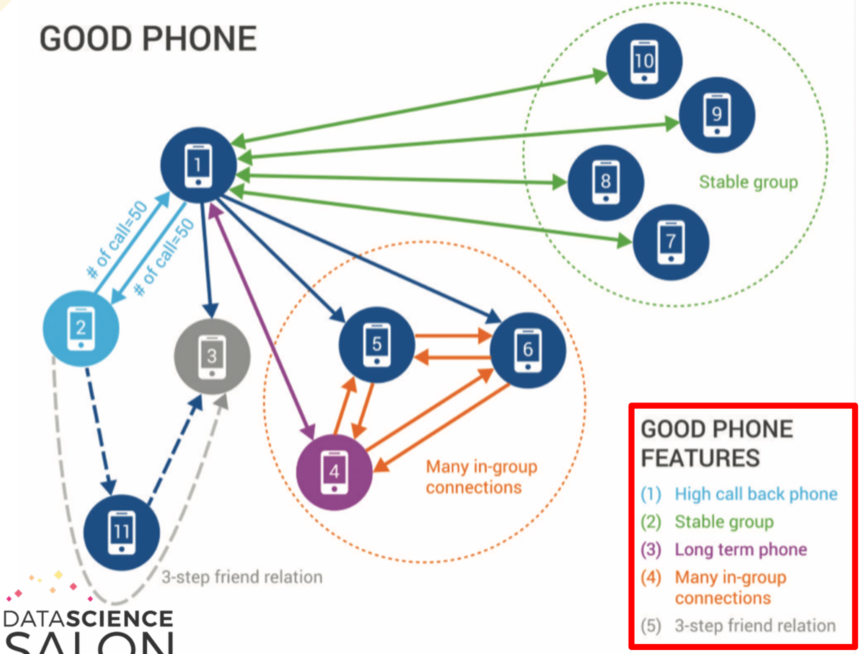


DATASCIENCE SALON VIRTUAL

@datascisalon

Case 2: Graph Feature Extraction

Customer: China Mobile



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.

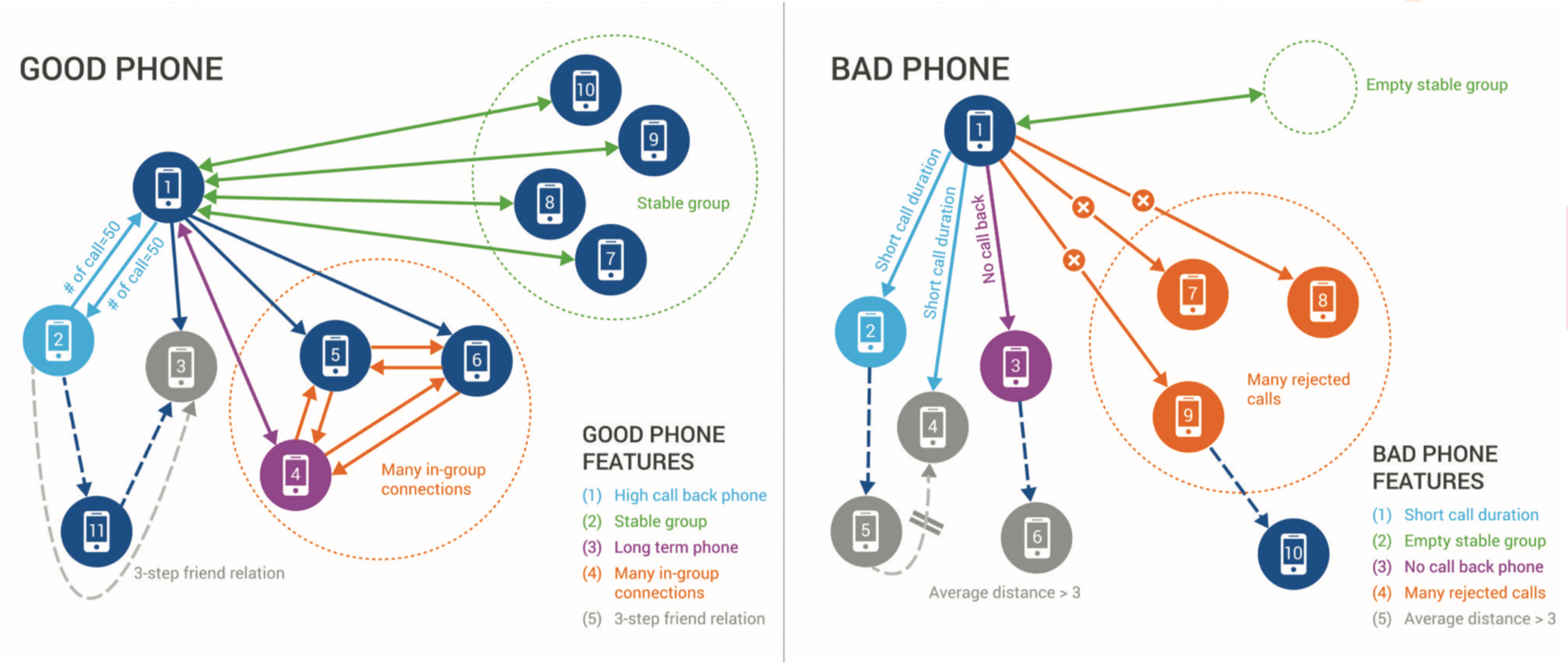
Results

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

DATASCIENCE SALON VIRTUAL

@datascisalon

Case 2: Graph Feature Extraction

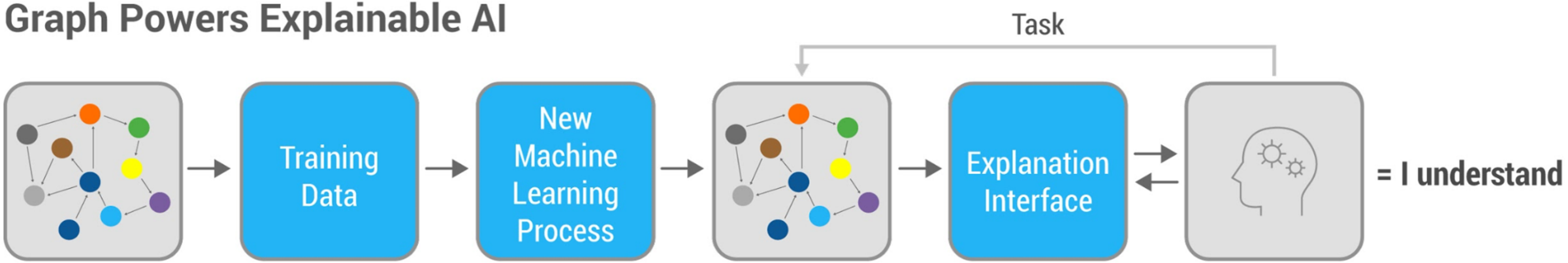


DATASCIENCE SALON VIRTUAL

@datascisalon

Powering Explainable AI with Graph Database

Graph Powers Explainable AI



Graph based feature computation

Explainable model with graph based visualization, exploration and features

Doctor "UptoNoGood" Features

Stable_group - None

Average_cost_of_care_for_opioid_addiction - 180%

Potential_prescriber_facility_collusion - YES

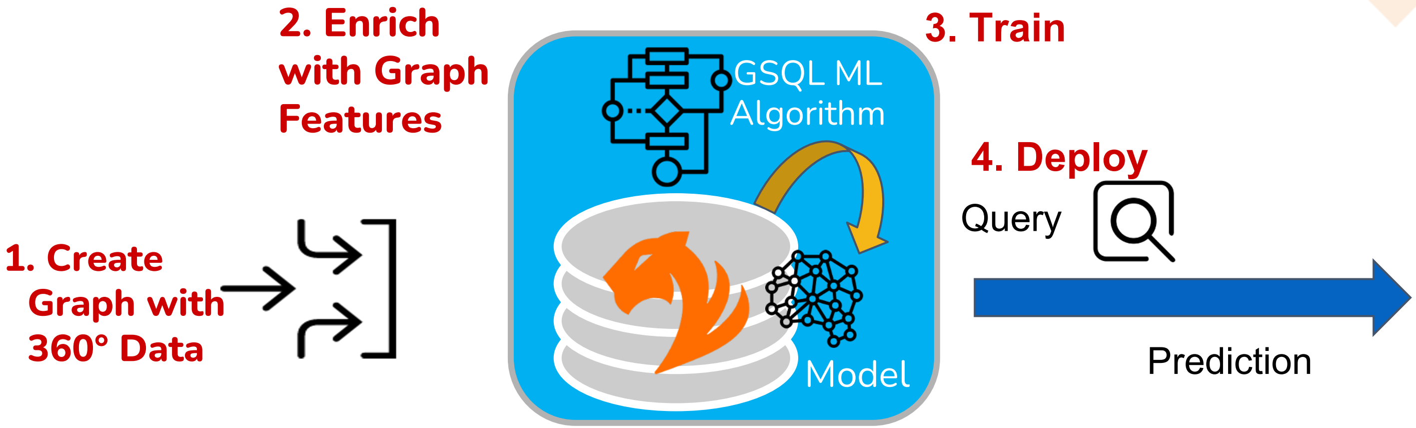
Human Readable Interface

Dr. UptoNoGood does not see patients for a common group of healthcare conditions despite being a general practitioner

Average cost of care for treating opioid addiction for this prescriber and their referral network is 180% of the average for the area

Dr. UptoNoGood shares an address with the owner & administrator of "New Day" opioid treatment facility

Case 3: In-Graph Database Machine Learning

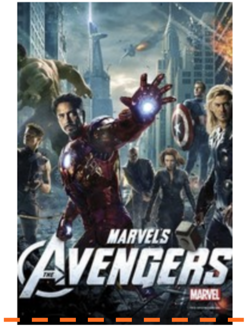


- Native graph storage
- Coded once, auto scale-out & scale-up
- Real-time updates
- GSQL Turing-complete language
  - Preprocess data
  - Training: flow-control, accumulator, pattern match
  - Model validation

Applications:

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

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

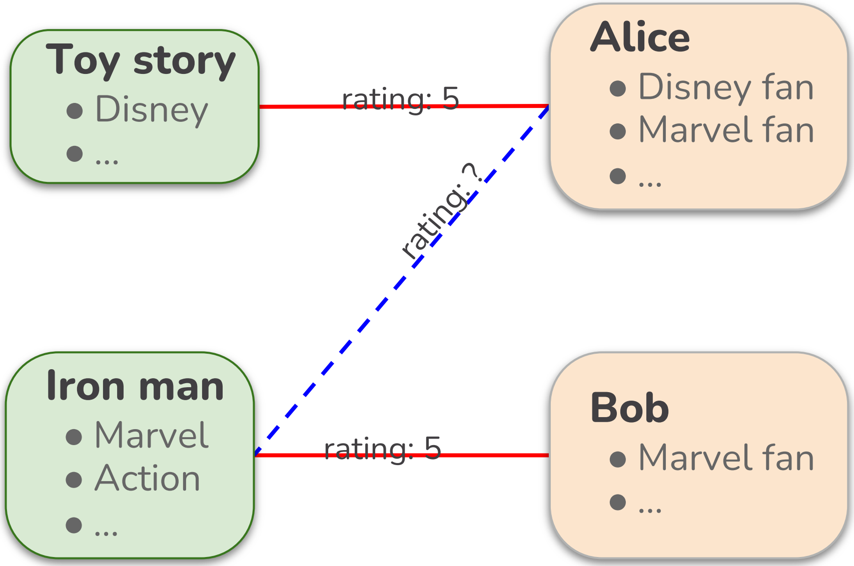
Low-Rank Approximation
Machine Learning v3

TigerGraph GraphGurus

EPISODE 28

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

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

The Future of In-Graph ML

Neural Networks for Graphs

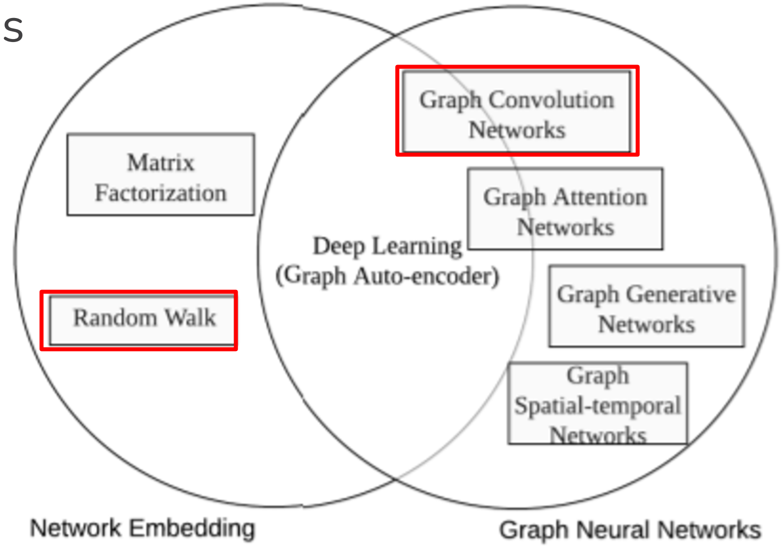
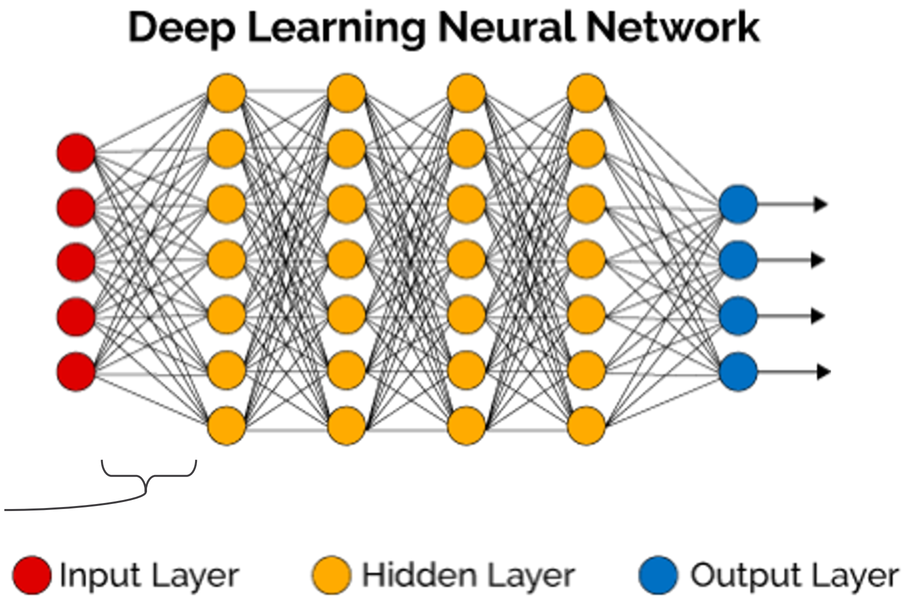


Fig. 2: Network Embedding v.s. Graph Neural Networks.

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

Basic Neural Network

Input layer = X
= Feature Vector
for each graph
vertex

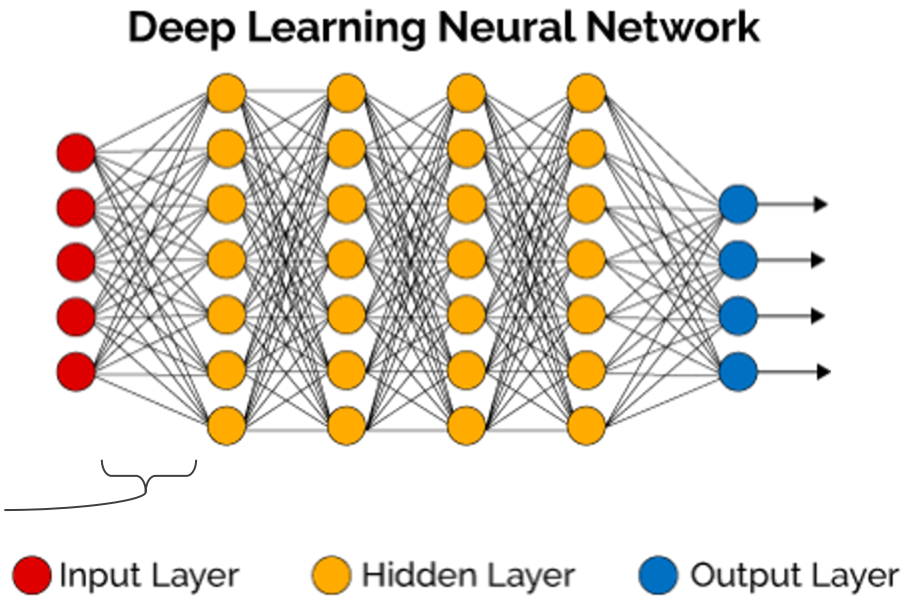


Propagation P =
weighted edges

Hidden Layer H1 = Activation_Function(X*P),
Hidden Layer H2 = Activation_Function(H1*P)

Graph Convolutional Neural Network

Input layer = X
= Feature Vector
for each graph
vertex



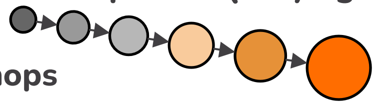
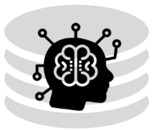
Propagation P =
weighted edges

A = Adjacency Matrix (Graph Edges)

Hidden Layer H1 = Activation_Function(**A***X*P),

Hidden Layer H2 = Activation_Function(**A***H1*P)

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 • Derive new-insights from legacy/external data stores

Summary for "Why Graph for ML/AI"?

- **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
  - Learning through **graph 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

Starter Kits and Developer Portal for Graph+ML

TigerGraph CLOUD

1. Content-based movie recommendation: *similarity*, k-nearest neighbor + latent factor
2. Entity resolution: Link & merge similar entities, based on *similar* properties and neighbors
3. Low-rank approximation of graph relationships
4. Graph feature engineering for anti-fraud ML

Select a Starter Kit *

Pick a Starter Kit with sample graph data schema, dataset, and queries (e.g. Fraud Detection, Recommendation Engine, Supply Chain Analysis, etc.).

Additional information including overview video at tigergraph.com/starterkits

All | Anti-fraud | Geospatial Analysis | Graph Algorithms | Healthcare | Knowledge Graph | Machine Learning | Recommendations

- In-Database Machine Learning Recommendation
- In-Database Machine Learning for Big Data Entity Resolution
- Low-Rank Approximation Machine Learning
- Machine Learning and Real-time Fraud Detection

Developer Portal

dev.tigergraph.com

Learn > Machine Learning

1. Unsupervised Learning with Graph Algorithms

2. Feature Set Extraction for Machine Learning

3. ML Enrichment with Graph Features

4. Graph Enrichment with Machine Learning

5. In-database ML Techniques for Graphs

Get Started

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.

tigergraph.com/cloud/

@TigerGraphDB /tigergraph /TigerGraphDB /company/TigerGraph