Graph Exploration w/ Neo4j
Our Project Partners

neo4j

HelmholtzZentrum München
Deutsches Forschungszentrum für Gesundheit und Umwelt
Efficiently extracting knowledge from graph data even if we do not know exactly what we are looking for

Graph Exploration: From Users to Large Graphs. CIKM 2016, SIGMOD 2017, KDD 2018
Graph Exploration Stack

Easy Active Search in Graphs  UAI’18*
Graph Embeddings from similarities  WWW’18, KDD’18

Importance-based Subgraph Mining  EDBT’18
Exemplar Queries  VLDB ’14, VLDBJ’16, ICDE’18
Notable Characteristic Search  EDBT’18

Graph Query Reformulation  KDD ’15
Faceted search on Graphs  planned 2018

Interactive algorithms
Intuitive queries
Adaptive Databases

Users

Graph
This project is about ...
Graph Exploration in Biology - Complex Graphs

http://jcs.biologists.org/content/joces/118/21/4947/F3.large.jpg
Graph Exploration in Biology - Status Quo

MATCH (p1:Phenotype)-[:HAS]-(a1:Association)-[:HAS]-(snp:Snp)-[:HAS]-(a2:Association)-[:HAS]-(p2:Phenotype)
WHERE p1.name = 'foo1' AND p2.name = 'foo2p' AND a1.p < 0.01 AND a2.p < 0.01
WITH DISTINCT snp
ORDER BY snp.sid
RETURN collect(snp.sid)

MATCH (p1:Phenotype)-[:HAS]-(a1:Association)-[:HAS]-(snp:Snp)-[:HAS]-(a2:Association)-[:HAS]-(p2:Phenotype)
WHERE p1.name = 'foo1' AND p2.name = 'foo2'
AND a1.p < 0.01 AND a2.p < 0.01
WITH DISTINCT snp
MATCH (snp)-[:IN]-(pw:PositionWindow)<-[[:IN]-(l:Locus)--(g:Gene)
WHERE l.feature= 'gene'
RETURN collect(DISTINCT g.name)

MATCH (p1:Phenotype)-[:HAS]-(a1:Association)-[:HAS]-(snp:Snp)-[:HAS]-(a2:Association)-[:HAS]-(p2:Phenotype)
WHERE p1.name = 'foo1' AND p2.name = 'foo2'
AND a1.p < 0.01 AND a2.p < 0.01
WITH DISTINCT snp
MATCH (snp)-[:IN]-(pw:PositionWindow)<-[[:IN]-(l:Locus)--(g:Gene)
WHERE l.feature= 'gene'
WITH DISTINCT g ORDER BY g.name
MATCH (g)-[:CODES]-(:Transcript)-[:CODES]-(p:Protein)-[:MEMBER]-(go:Goterm)
WHERE go.namespace= 'biological_process'
WITH DISTINCT go,p
RETURN go.name, count(p) ORDER BY count(p)DESC
LIMIT 10

MATCH (p1:Phenotype)-[:HAS]-(a1:Association)-[:HAS]-(snp:Snp)-[:HAS]-(a2:Association)-[:HAS]-(p2:Phenotype)
WHERE p1.name = 'foo1' AND p2.name = 'foo2'
AND a1.p < 0.01 AND a2.p < 0.01
WITH DISTINCT snp
MATCH (snp)-[:IN]-(pw:PositionWindow)<-[[:IN]-(l:Locus)--(g:Gene)
WHERE l.feature= 'gene'
WITH DISTINCT g ORDER BY g.name
MATCH (g)-[:CODES]-(:Transcript)-[:IS]-(ps:Probeset)-[:SIG]-(s:Sample)
WHERE s.name= 'mustafavi'
RETURN DISTINCT g.name
Can we do better?
Problem

- Given two node sets: **How similar are they in my understanding?**

- **Example**
  - Set of movies I like
  - Set of movies I don’t know
  - Will I like the movies I don’t know?
What is a Knowledge Graph?

- (directed) graph $G : \langle V, E, \phi, \psi \rangle$, where
  - $V$ is a set of nodes,
  - $E \subseteq V \times V$ is a set of edges,
  - $\phi : V \rightarrow L_V$ is an edge labeling function
    and
  - $\psi : E \rightarrow L_E$ is a node labeling function

We refer to the elements of $L_V$ and $L_E$ as node labels and edge labels.
What are Meta-Paths?

A meta-path for a path $\langle n_1, ..., n_t \rangle$, $n_i \in V$, $1 \leq i \leq t$ is a sequence $P : \langle \varphi(n_1), \psi(n_1, n_2), ..., \psi(n_{t-1}, n_t), \varphi(n_t) \rangle$ that alternates **node- and edge-types** along the path.
Motivating Example

Q: How famous is Diane Kruger in America?

MATCH(n:Person)
WHERE n.name = "Diane Kruger"
RETURN n

MATCH(m:Movie)
WHERE m.location = "America"
RETURN m

Diane Kruger

Top Gun
Pulp Fiction
As Good As It Gets
The Matrix
Stand By Me
A Few Good Men
Up
How similar are they?

- Similarity depends on
  - expert knowledge
  - connections among nodes
What does the System do and how?

Overview

Learn representation for meta-paths

Compute Meta-Paths

Extract ratings

Calculate similarity

Individualized exploration

✓

✗

◎
Approximate Meta-Paths

**Problem:** How to compute all meta-paths fast?

**Approx. Solution:** Mine meta-paths using the graph’s schema and learn classifier on real meta-paths.
Learning a Meta-Path Embedding

*Problem:* Vector representation required for active learning and preference prediction.
Learning a Meta-Path Embedding

*Problem:* Vector representation required for active learning and preference prediction.

*Solution:* Embed meta-paths
  → Similar meta-paths should have similar vectors.

*Our method:* Transfer text embedding method to meta-paths.
Learn the Domain Value of all Meta-Paths

- **Problem**: Users don’t want to rate all meta-paths
  → too many
  → time-consuming
  → tedious and boring

- **Solution**: Label only a few, but very informative paths
Use Learned Preferences for Graph Exploration

Graph (with meta-paths)

What is important in the graph?

Domain Knowledge

Personalized Exploration Tool

Similarity Measure

Related Nodes

Stats
Personalized Node Embedding

Transform Nodes to Vectors (Graph-Embedding)

Adapt Vectors Using Domain-Knowledge

Personalized Vector Space

Result Explanation
Personalized Exploration Tool

How close are my sets?

Find clusters!

Personalized Vector Space

What nodes are close to my selection?

What are outliers?
System Architecture - How does it work with Neo4j?

- ReactJS Frontend
  - Node selection
  - Meta-Path ordering
  - Result visualization

- Python Backend Server
  - Meta-Path Embedding
  - Active Learning
  - Explanation

- Neo4j Graph Algorithm Procedures
  - Containing Meta-Paths Computation

- Neo4j Graph Database
What about neo4j?

- Easy to get your code running in neo4j.
- Neo4j-graph-algorithms: efficiency vs convenience.
- Sometimes no stack-trace when an error occurs.
- Great support and community. Always available.
- Cypher: Easy to begin with, hard to master.

(hpi)-[:LIKES]-(neo4j)
Trending: #tweetyourthesis

Learning a vector representation for meta-paths where ones with similar meaning are close to each other

Teacher who prepare lectures for the 7th grade naturally choose from different resources than professors. Why not in ActiveLearning?

Interactively understand how machines see your problems with machine explanations. Teach them to do things right, for the right reasons!

#KnowledgeGraphs are often incomplete, let’s find missing edges and predict their edge type.
Trending: #tweetyourthesis

- **Michael** (@BP)
  How to compute meta-paths. And how to do it sometimes quicker using the graph-schema and classification.
  2:48 PM - 6 June 2018

- **Fabian** (@fswi)
  Classifying meta-paths as existent / non-existent helps getting better results when computing them on a graph's schema. So, let us improve!
  2:48 PM - 6 May 2018

- **Julius** (@julius)
  Learning Similarities between Entities in a Heterogeneous Knowledge Base through incorporating the User's Domain Knowledge.
  2:48 PM - 6 May 2018

- **Laurenz** (@laurenz)
  Your opinion is important! Adapt #KnowledgeGraphs to your view of the world and gain insights faster than ever before!
  2:48 PM - 6 May 2015