Graph Exploration w/ Neo4j

https://s3.amazonaws.com/dev.assets.neo4j.com/wp-content/uploads/graph-data-technologies-graph-databases-for-beginners.png

Our Project Partners

neo4j

HelmholtzZentrum münchen
Deutsches Forschungszentrum für Gesundheit und Umwelt

Graph exploration

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

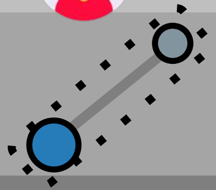
Users

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

Interactive algorithms

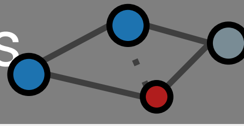
Importance-based Subgraph Mining EDBT '18
Exemplar Queries VLDB '14, VLDBJ '16, ICDE '18
Notable Characteristic Search EDBT '18

Intuitive queries

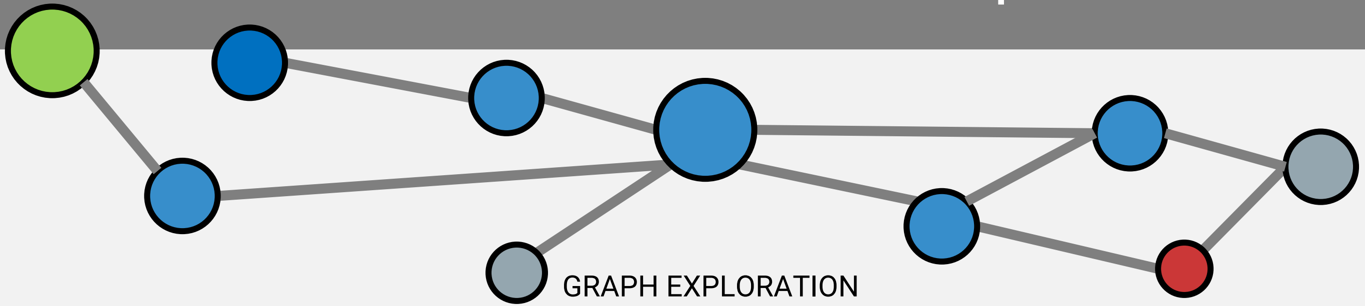


Graph Query Reformulation KDD '15
Faceted search on Graphs planned 2018

Adaptive Databases



Graph



Graph Exploration Stack

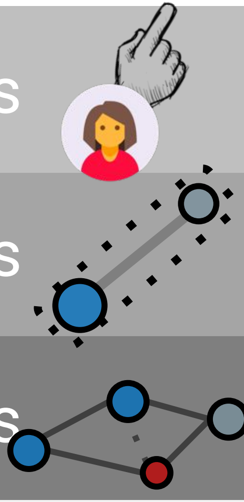
Users

This project is about ...

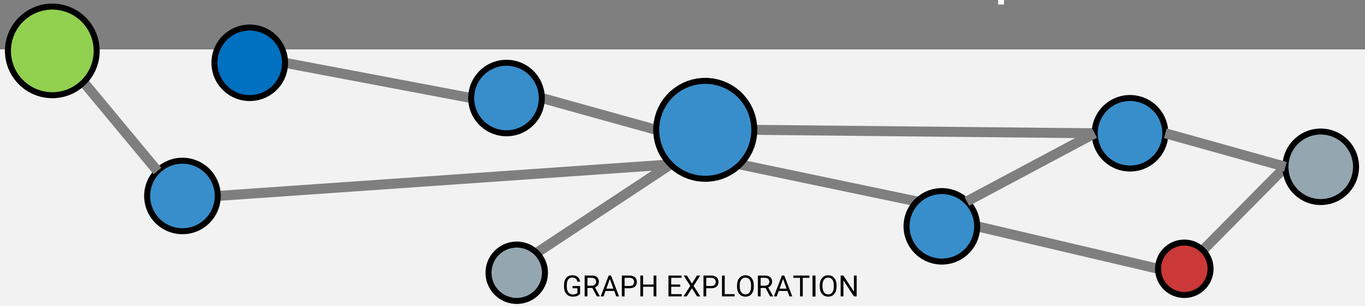
Interactive algorithms

Intuitive queries

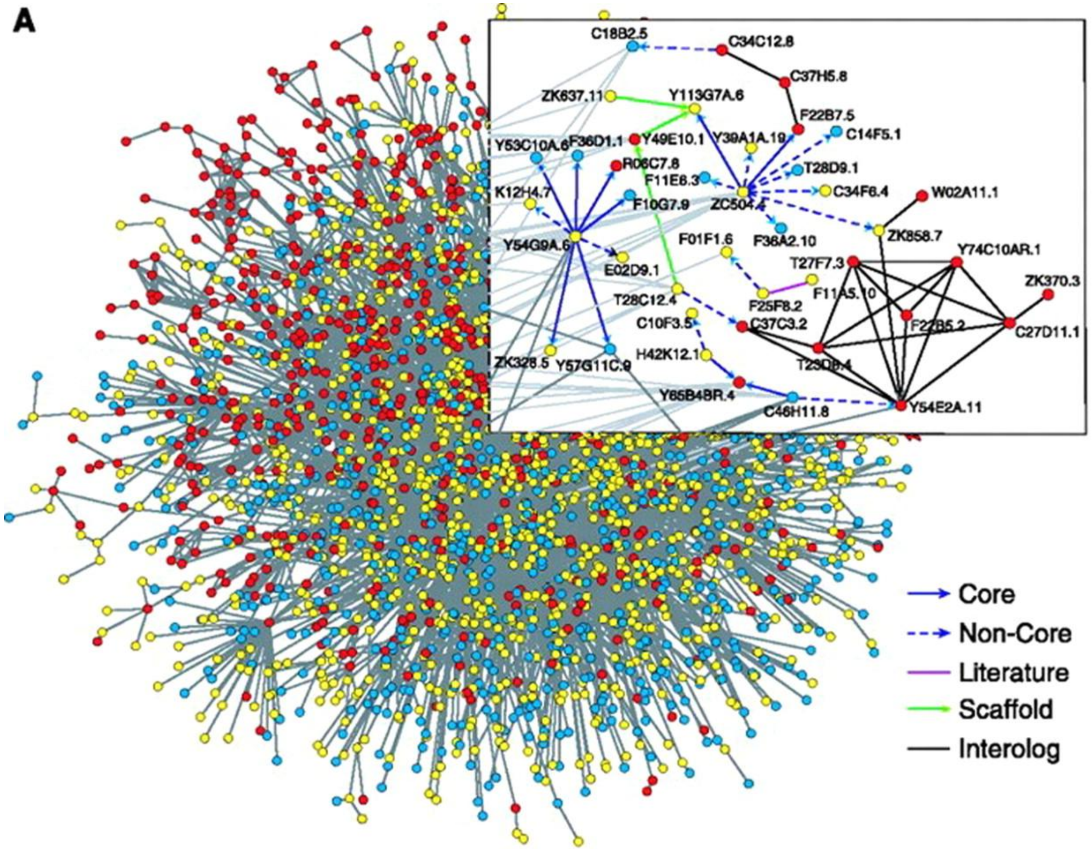
Adaptive Databases



Graph



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 = '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 gORDER BY g.name
MATCH (g)-[:CODES]-(:Transcript)-[:IS]-(ps:Probeset)-[:SIG]-(s:Sample)
WHERE s.name= 'mustafavi'
RETURN DISTINCT g.name
```

```
[:HAS]-(a2:Association)-[:HAS]-(p2:Phenotype)
```

```
WITH DIS
RETURN go.name, count(p)
LIMIT 10
```

Can we do better?

Problem

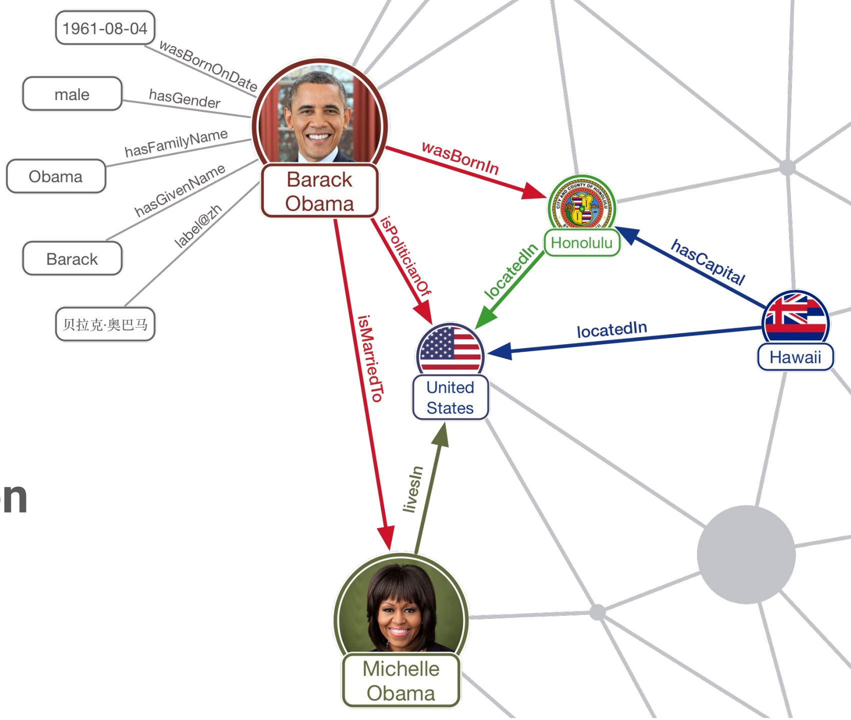
- Given two node sets:
  How similar are they in my understanding?

- *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?

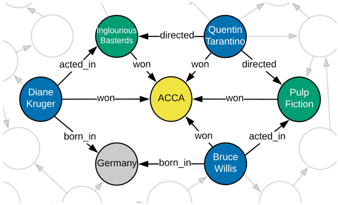
- (directed) graph G : ⟨V, E, φ, ψ⟩, where
  - V is a set of nodes,
  - $E \subseteq V \times V$ is a set of edges,
  - $\varphi : V \rightarrow L_V$ is an **edge labeling function** and
  - $\psi : E \rightarrow L_E$ is a **node labeling function**

We refer to the elements of L_V and L_E as node labels and edge labels



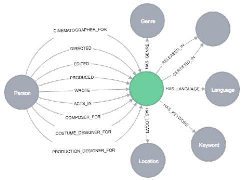
What are Meta-Paths?

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

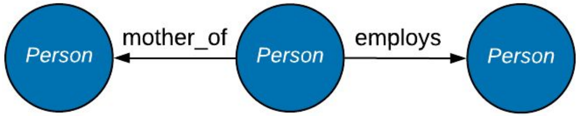


Graph

Path



Graph Schema



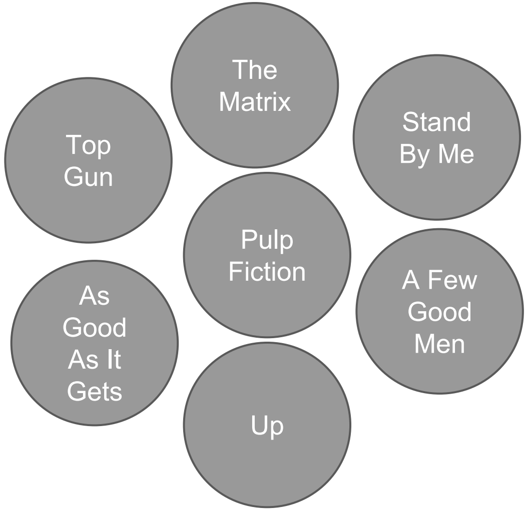
Meta-Path

Motivating Example

Q: How famous is Diane Kruger in America?

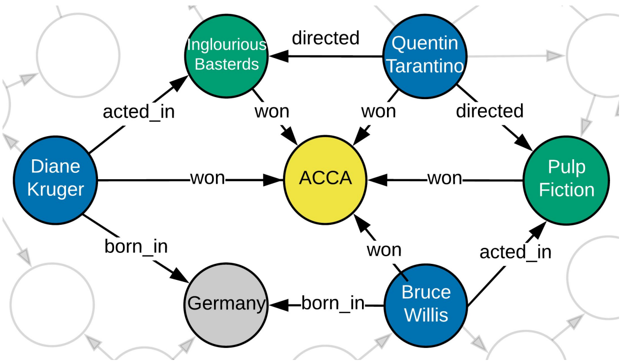
```
MATCH(n:Person)
WHERE n.name = "Diane Kruger"
RETURN n
```

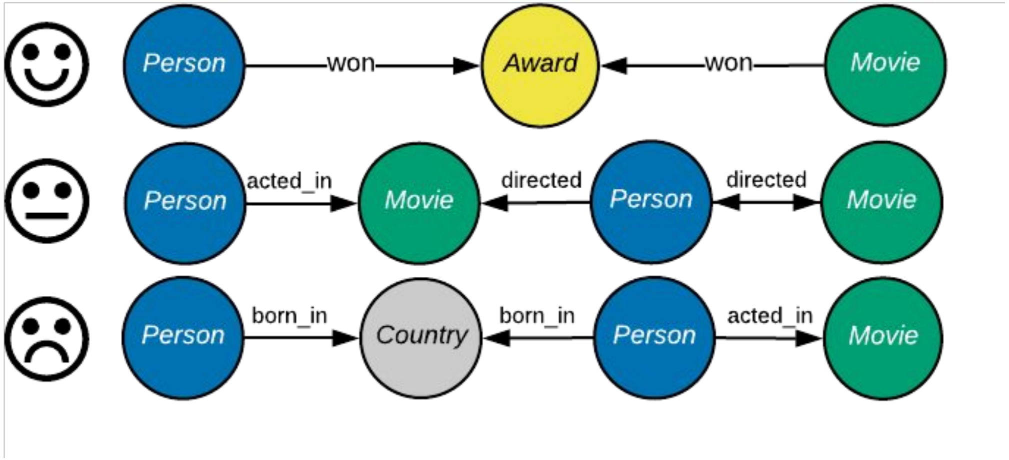
```
MATCH(m:Movie)
WHERE m.location = "America"
RETURN m
```



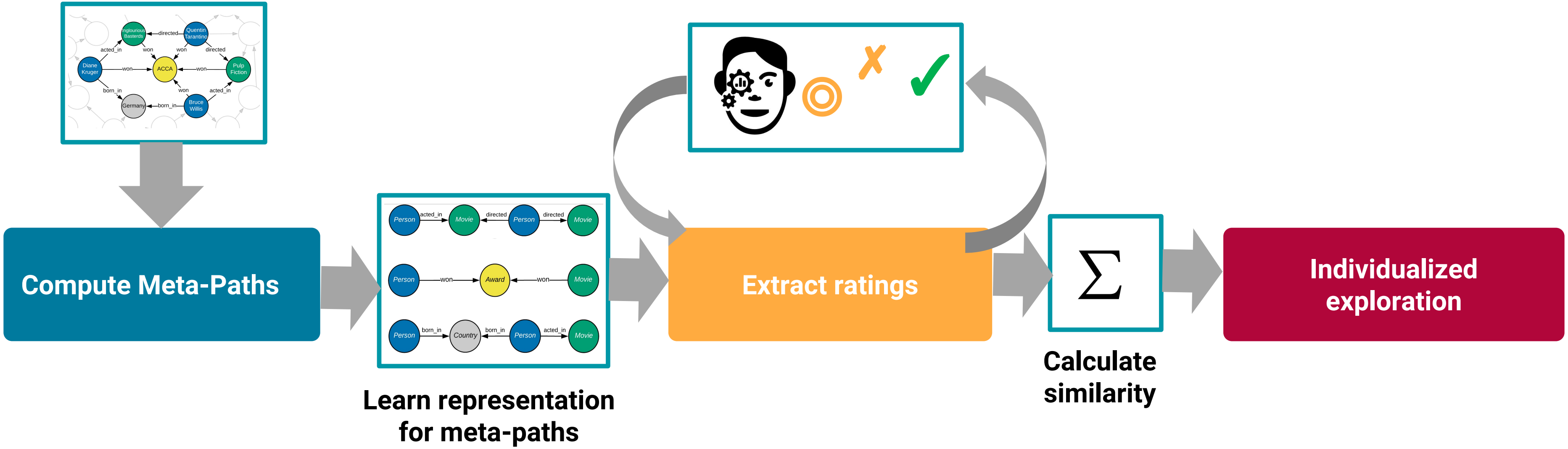
How similar are they?

- Similarity depends on
  - **expert knowledge**
  - **connections among nodes**



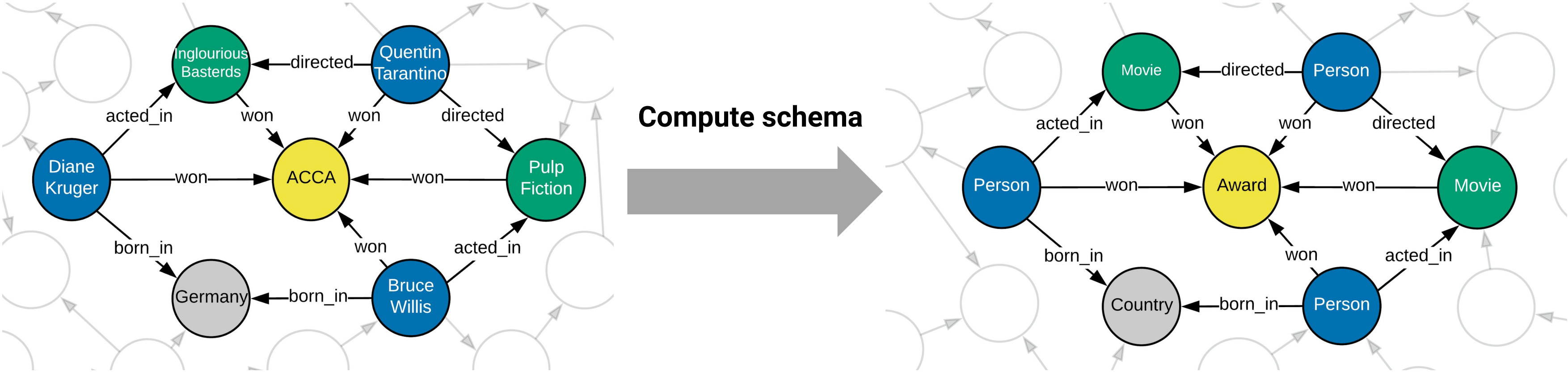


What does the System do and how?



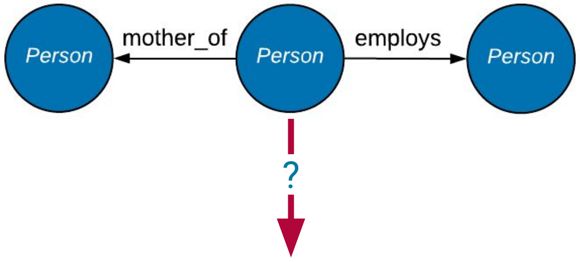
Approximate Meta-Paths

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.



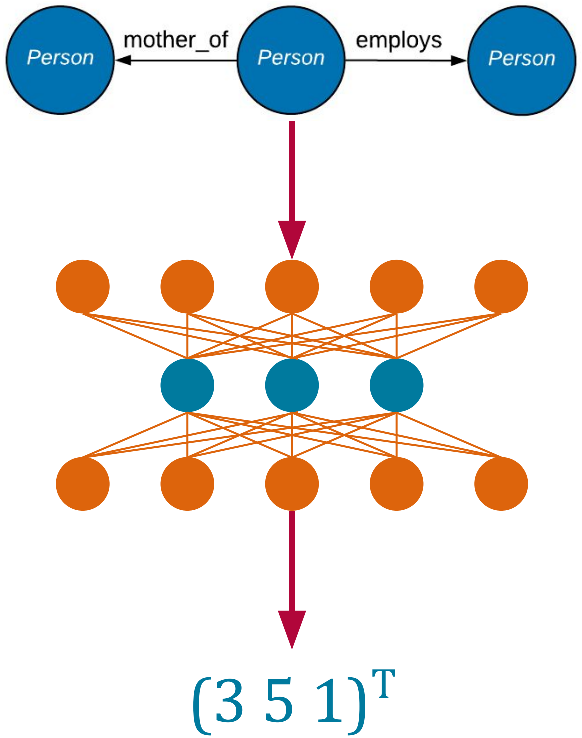
$$(3\ 5\ 1)^{T}$$

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.

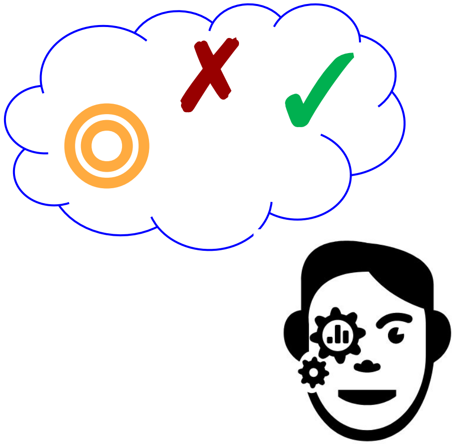
Our method: Transfer text embedding method to meta-paths.



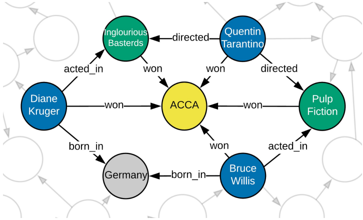
Learn the Domain Value of all Meta-Paths

- *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

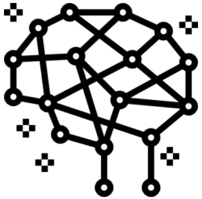


Use Learned Preferences for Graph Exploration

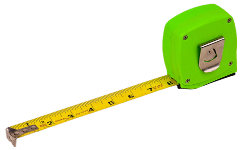


Graph (with meta-paths)

What is important in the graph?



Domain Knowledge



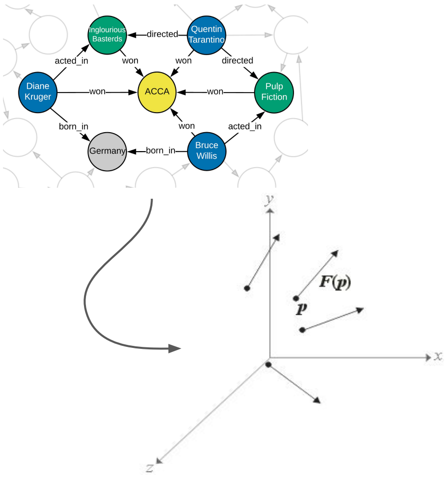
Personalized Exploration Tool

Similarity Measure

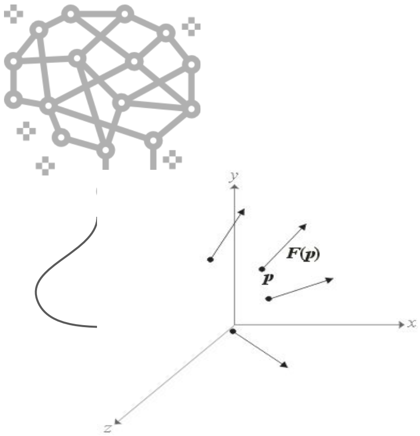
Related Nodes

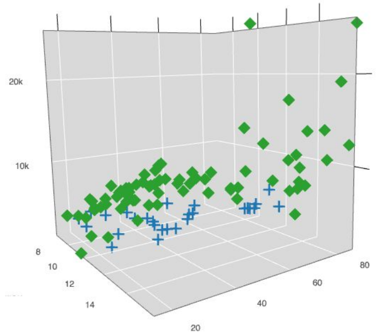
Stats

Personalized Node Embedding



precomputed





Transform Nodes to Vectors
(Graph-Embedding)

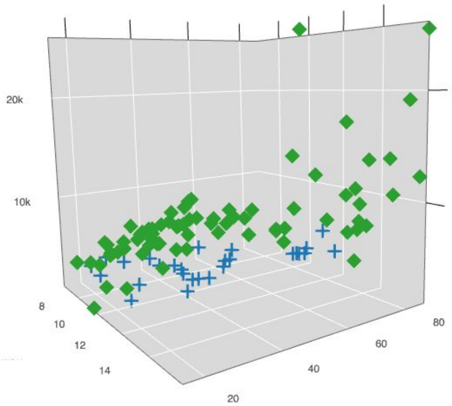
Adapt Vectors Using Domain-Knowledge

Personalized Vector Space

Personalized Exploration Tool

How close are my sets?

What nodes are close to my selection?

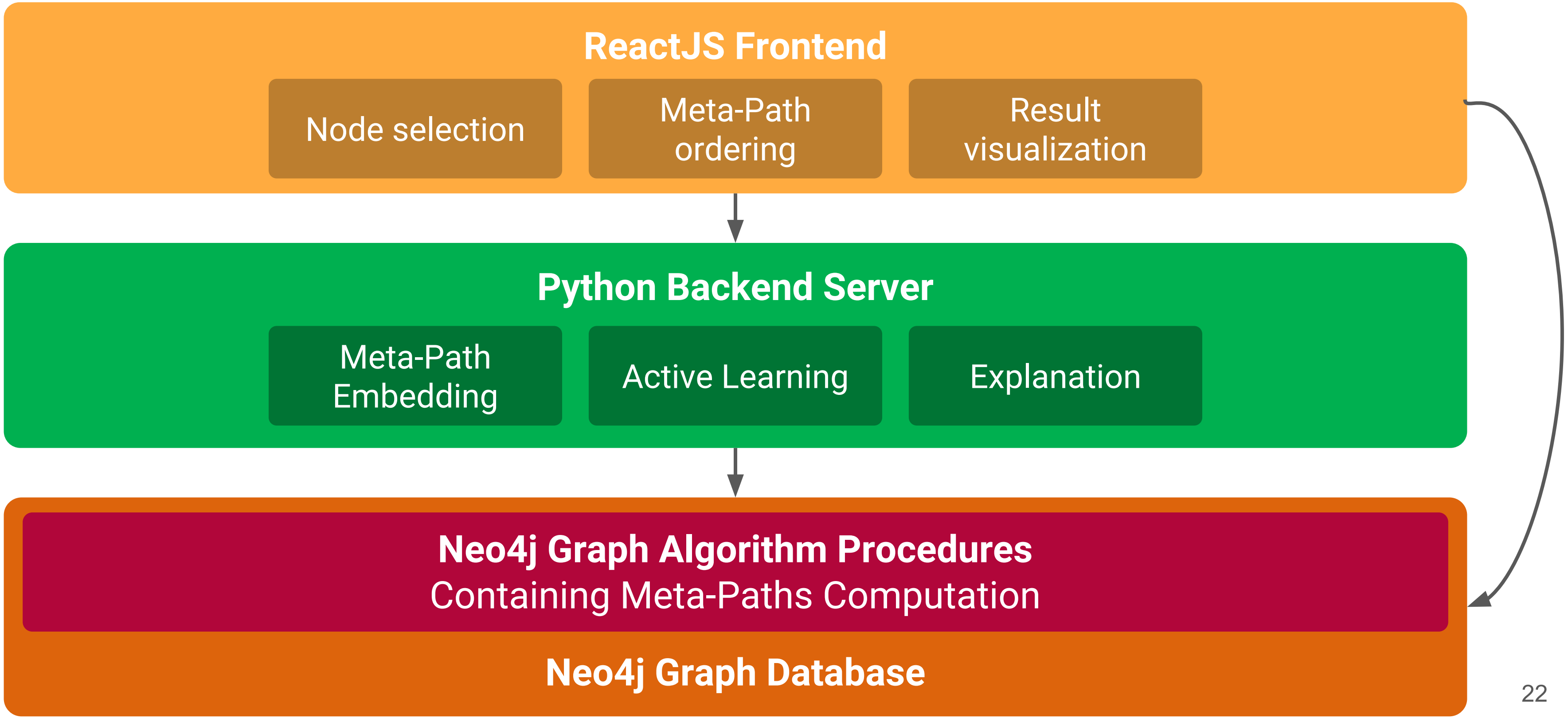


Find clusters!

Personalized Vector Space

What are outliers?

System Architecture - How does it work with Neo4j?



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

Sebastian @sebastian — Follow

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

6:34 PM - 25 June 2018

Adrian @mkuuwaujinga — Following

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

2:48 PM - 30 May 2018

Freya @feeds — Follow

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

2:48 PM - 6 May 2015

Pius Ladenburger @PiusL — Following

#KnowledgeGraphs are often incomplete, let‘s find missing edges and predict their edge type.

2:03 AM - 25 June 2018

Trending: #tweetyourthesis

Michael @BP — Following

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 @fswt — Following

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 — Following

Learning Similarities between Entities in a Heterogeneous Knowledge Base through incorporating the User's Domain Knowledge.

2:48 PM - 6 May 2018

Laurenz @laurenz — Following

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