Cytokine Graph Database

Borna Bešić, Dilmurat Yusuf

Bioinformatics Group

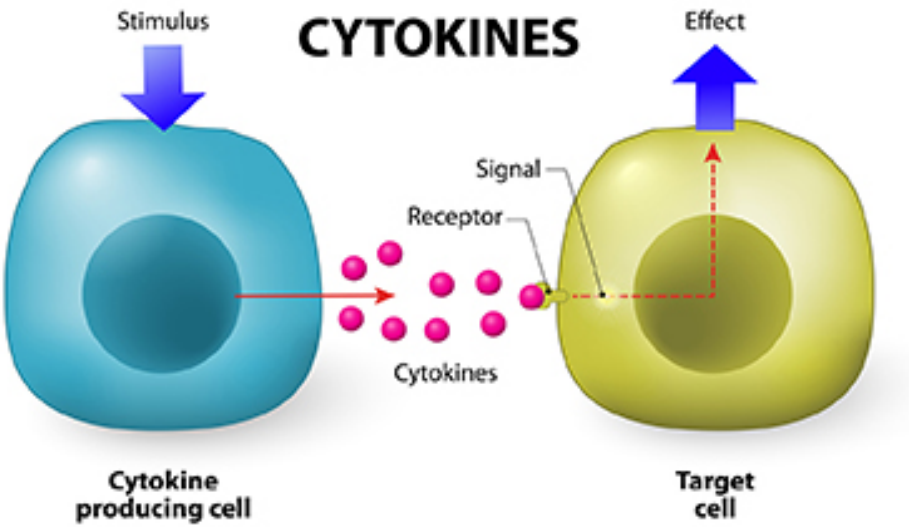
Albert-Ludwigs-Universität, Freiburg

October 30, 2018

Overview

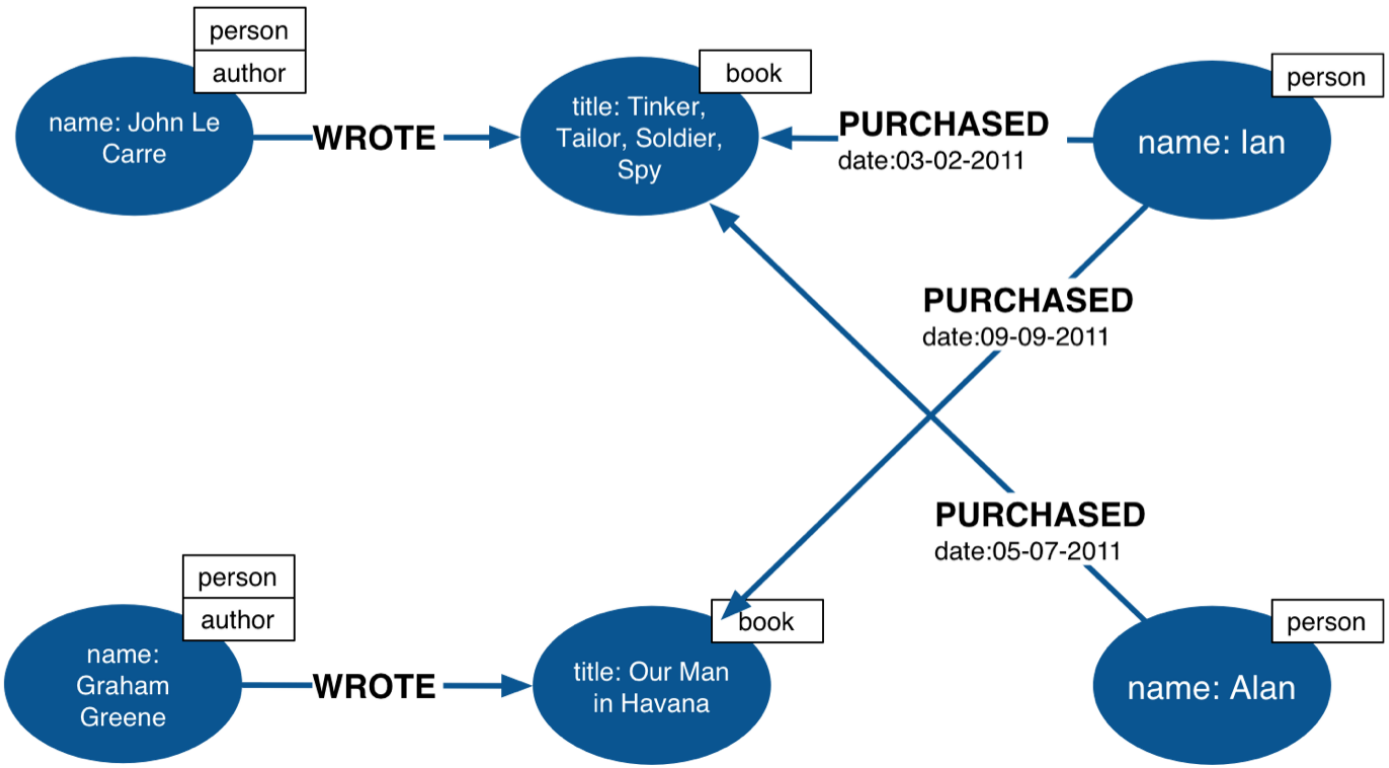
Cytokines

- Molecular messengers between cells
- Interact with cells of the immune system
- Regulate the body's response to disease and infection
- Mediate normal cellular processes in the body



Graph databases

- Nodes
- **Directed** edges / relationships
- Properties



Neo4j & Cypher

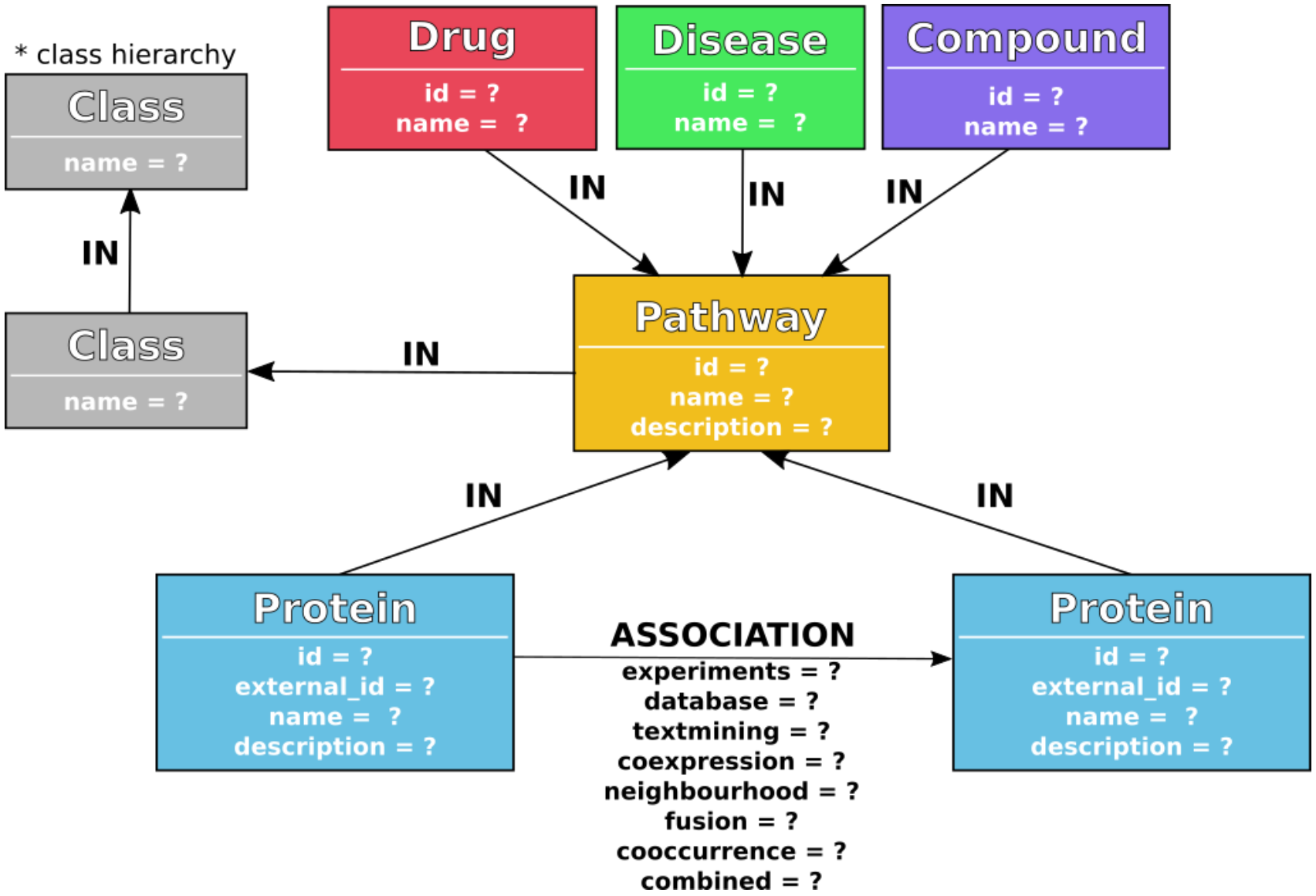
Figure: Neo4j - a graph database management system

Example of a Cypher query

```
MATCH (:Person { name: "Alan" })-[p:PURCHASED]->(b:book)
RETURN p.date, b
```

Overview

Database scheme



Statistics

- Nodes
  - Protein:
  - Pathway:
  - Drug:
  - Disease:
  - Compound:
  - Class:

- Relationships
  - **ASSOCIATION**:
  - **IN**:

Overview

Sources of data

- STRING
  - proteins
  - protein - protein associations
- KEGG PATHWAY
  - pathways
    - classes
    - compounds
    - drugs
    - diseases

STRING score channels

- 7 score channels → combined score
  1. Experiments
  2. Database
  3. Textmining
  4. Co-expression
  5. Neighbourhood
  6. Fusion
  7. Co-occurence

Evidence suggesting a functional link:

Neighborhood in the Genome:	none / insignificant.
Gene Fusions:	none / insignificant.
Cooccurence Across Genomes:	none / insignificant.
Co-Expression:	yes (score 0.152). In addition, putative homologs are coexpressed in other species (score 0.043).
Experimental/Biochemical Data:	yes (score 0.360).
Association in Curated Databases:	yes (score 0.900).
Co-Mentioned in PubMed Abstracts:	yes (score 0.962). In addition, putative homologs are mentioned together in other species (score 0.127).
Combined Score:	0.997

Figure: Scores for CCR5 and CCL5

Workflow

1. Collect the data
   - STRING: SQL dumps
   - KEGG PATHWAY: flat text files
2. Extract the useful information
   - Query the STRING SQL database
   - Parse the KEGG flat files
3. Merge the databases
   - Map KEGG identifiers to STRING identifiers
4. Build a graph database
   - Translate the associative data into a graph

Overview

What is the point?

- Drug discovery
- ...

Overview

(demo)

Overview

Next steps

- Extend the database
  - Protein - protein actions
  - KEGG DRUG
  - KEGG DISEASE
  - KEGG COMPOUND
- Other species
- All proteins
- Web server
- Machine learning

Thanks :)

https://backofenlab.github.io/cytokine-graph-db/
https://github.com/BackofenLab/cytokine-graph-db