

# Cytokine Graph Database

Borna Bešić, Dilmurat Yusuf

Bioinformatics Group

Albert-Ludwigs-Universität, Freiburg

November 8, 2018

# Overview

- 1 Introduction
- 2 Methodology
- 3 Results
- 4 Next steps

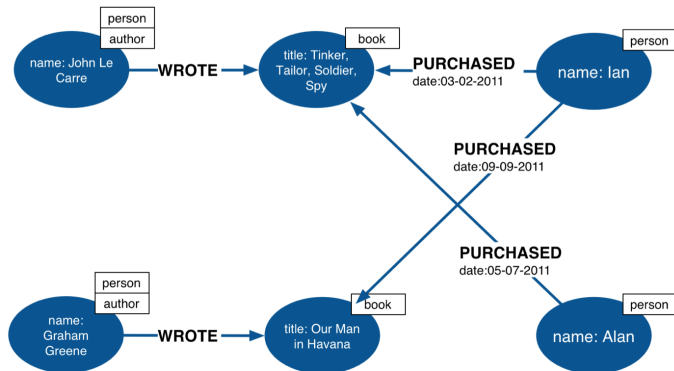
# What do we want?

- Easier access to data
  - More flexible searches
  - Faster queries
  - Dynamic user interfaces
- Associated data in one place
  - Unification of databases
- This leads to
  - Accessible protein - disease data
  - Faster drug discovery
- ...

- Database of known and predicted protein-protein interactions
- 7 score channels → combined score
  - ① Experiments
  - ② Database
  - ③ Textmining
  - ④ Co-expression
  - ⑤ Neighbourhood
  - ⑥ Fusion
  - ⑦ Co-occurrence
- Best overall performance for gene set recovery by propagation ([Huang et al., Systematic Evaluation of Molecular Networks for Discovery of Disease Genes, Cell Systems](#))
- Full database dumps available for download: 512.8 GB (compressed)

# Graph databases

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



## Example of a Cypher query

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

- Pros:
  - Outperforming RDBMS for associative data
  - No redundancy
  - Easier to change the design
- Cons:
  - Not standard; new query language (Cypher)
  - Harder to do summing queries and max queries efficiently

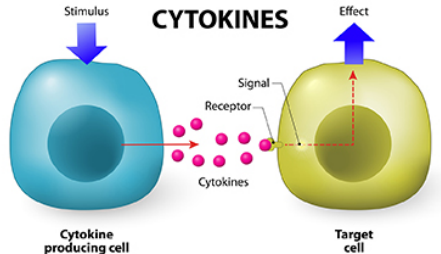
# Why Neo4j?

	Neighbor network	Best-scoring path	Shortest path
PostgreSQL	206.31 s	1147.74 s	976.22 s
Neo4j	5.68 s <sup>a</sup>	1.17 s	0.40 s
Speedup	36×	981×	2441×

Figure: Have CT, Jensen LJ. Are graph databases ready for bioinformatics?

# 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





# Overview

- 1 Introduction
- 2 Methodology
- 3 Results
- 4 Next steps



- STRING

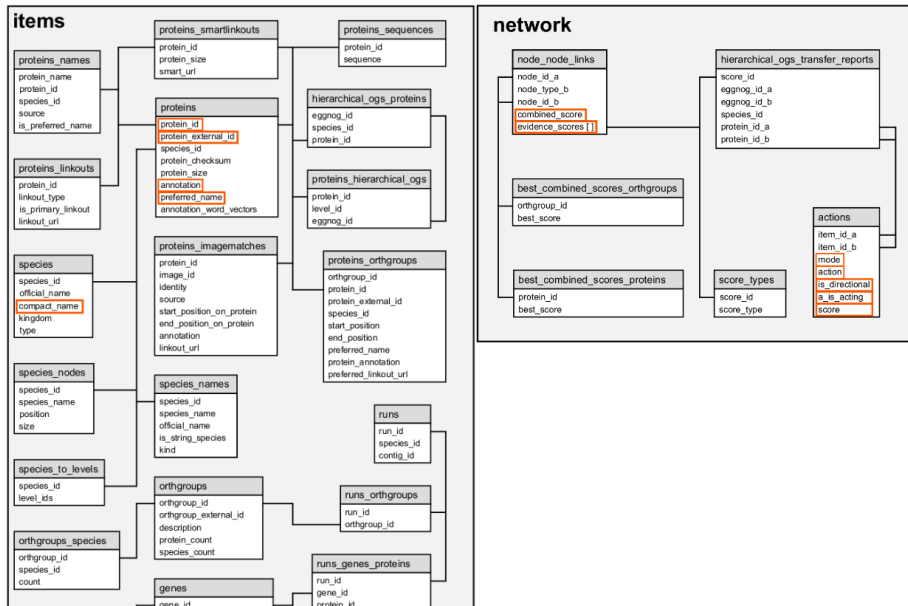
- proteins
- protein - protein associations



- KEGG PATHWAY

- pathways
  - classes
  - compounds
  - drugs
  - diseases

# STRING Schema

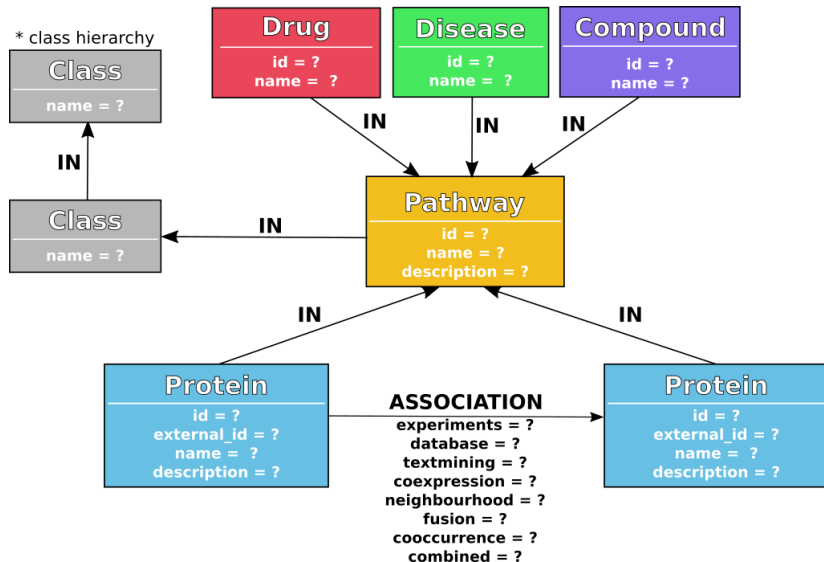


# KEGG PATHWAY Flat file

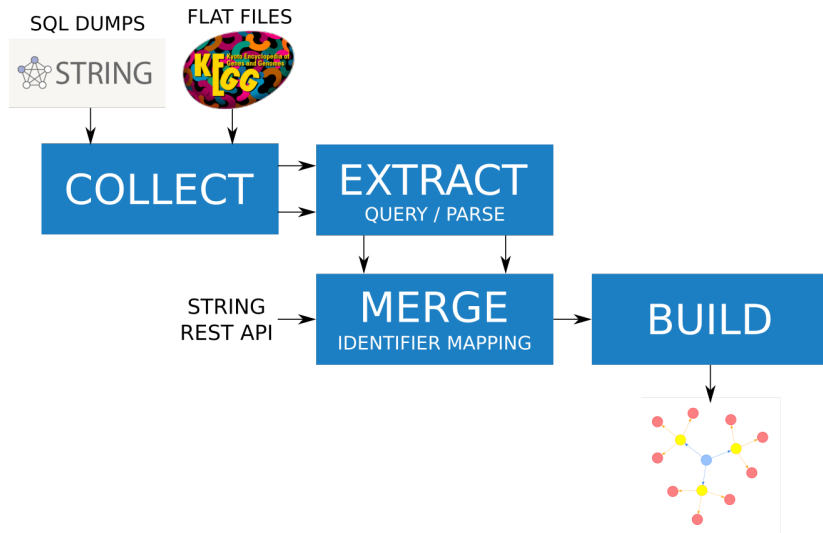
```

ENTRY          hsa00010          Pathway
NAME           Glycolysis / Gluconeogenesis - Homo sapiens (human)
DESCRIPTION    Glycolysis is the process of converting glucose into pyruvate...
CLASS          Metabolism; Carbohydrate metabolism
PATHWAY_MAP    hsa00010 Glycolysis / Gluconeogenesis
MODULE         hsa_M00001 Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate [PATH:hsa00010]
               hsa_M00002 Glycolysis, core module involving three-carbon compounds [PATH:hsa00010]
               hsa_M00003 Gluconeogenesis, oxaloacetate => fructose-6P [PATH:hsa00010]
               hsa_M00307 Pyruvate oxidation, pyruvate => acetyl-CoA [PATH:hsa00010]
DISEASE        H00069 Glycogen storage disease
               H00071 Hereditary fructose intolerance
               H00072 Pyruvate dehydrogenase complex deficiency
               ...
DRUG           D00123 Cyanamide (JP17)
               D00131 Disulfiram (JP17/USP/INN)
               D07257 Lonidamine (INN)
               D08970 Piraqliatin (USAN)
DBLINKS        BSID: 82926
               GO: 0006096 0006094
ORGANISM       Homo sapiens (human) [GN:hsa]
GENE           3101 HK3; hexokinase 3 [KO:K00844] [EC:2.7.1.1]
               3098 HK1; hexokinase 1 [KO:K00844] [EC:2.7.1.1]
               3099 HK2; hexokinase 2 [KO:K00844] [EC:2.7.1.1]
               80201 HKDC1; hexokinase domain containing 1 [KO:K00844] [EC:2.7.1.1]
               2645 GCK; glucokinase [KO:K12407] [EC:2.7.1.2]
               ...
COMPOUND        C00022 Pyruvate
               C00024 Acetyl-CoA
               C00031 D-Glucose
               C00033 Acetate
               ...
REFERENCE      (map 1)
AUTHORS        Nishizuka Y (ed).
TITLE          [Metabolic Maps] (In Japanese)
JOURNAL        Tokyo Kagaku Dojin (1980)
REFERENCE      (map 1)
AUTHORS        Nishizuka Y, Seyama Y, Ikai A, Ishimura Y, Kawaguchi A (eds).
TITLE          [Cellular Functions and Metabolic Maps] (In Japanese)
JOURNAL        Tokyo Kagaku Dojin (1997)
REFERENCE
AUTHORS        Michal G.
TITLE          Biochemical Pathways
JOURNAL        Wiley (1999)
    
```

# Cytokine Graph DB Scheme



# Workflow



# Overview

- 1 Introduction
- 2 Methodology
- 3 Results**
- 4 Next steps

- Cytokine-cytokine receptor interaction - Human (Homo sapiens)
- Nodes
  - Protein: 282
  - Pathway: 325
  - Drug: 3 731
  - Disease: 969
  - Compound: 3 465
  - Class: 49
- Relationships
  - **ASSOCIATION**: 13 900
  - **IN**: 17 752



(demo)

# Overview

- 1 Introduction
- 2 Methodology
- 3 Results
- 4 Next steps

# 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>