

# Investigating Ontology Use in Artificial Intelligence and Machine Learning for Biomedical Research

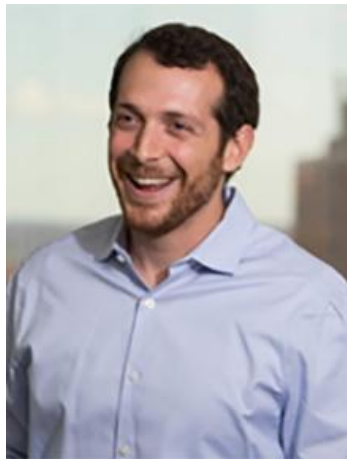
- A Preliminary Report from A Comprehensive Literature Review

---

Asiyah Yu Lin.  
ROBI workshop  
Sep. 25, 2022

# Acknowledgement

- The Biomedical Informatics Research and Data Science for Youth  
Highschool Seniors:  
Andrey I. Seleznev, Tianming “Danny” Ning, Paulene Grier  
Highschool Juniors:  
Lalisa M. Lin, Christopher Travieso
- The NIH ODSS DATA Scholar Program



Dr. Jaleal Sanjak  
(the DATA scholar  
2021 cohort)  
NCATS, NIH



Dr. Ansu Chatterjee  
(the DATA scholar  
2021 cohort)  
All of Us Research  
Program, OD, NIH



Dr. Asiyah Y. Lin  
(the DATA scholar  
2020 cohort)  
NHGRI, NIH

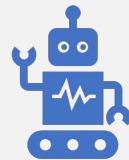
# Motivation



Machine learning, Artificial Intelligence, and Big Data in biomedical research

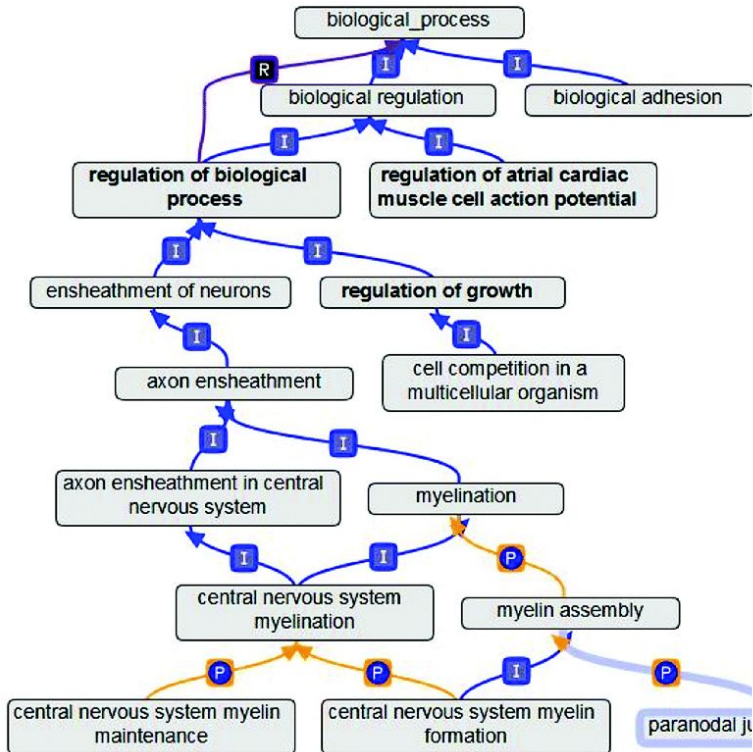


Founders and investors interest



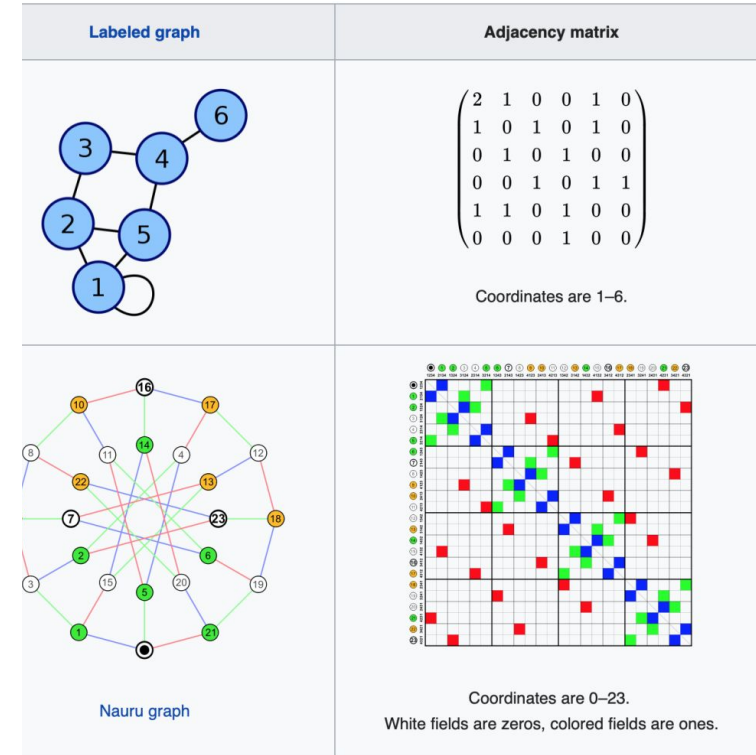
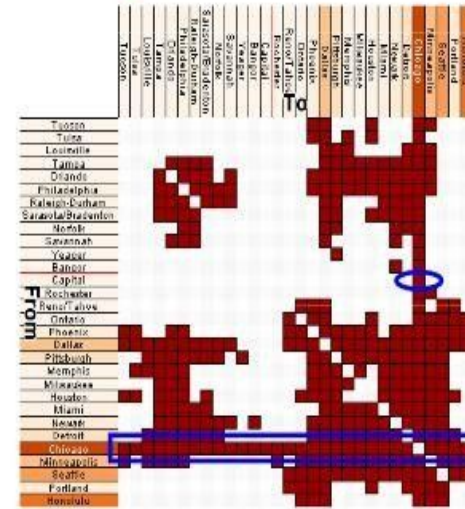
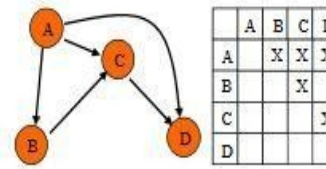
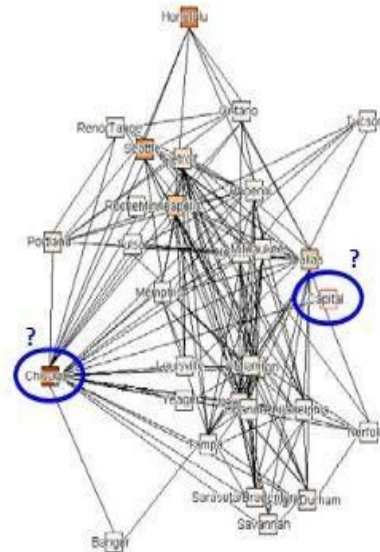
Can mathematically grounded machine learn from semantically grounded ontologies? And how ?

# Gene Ontology




# Matrix, graph theory, and visualization

## Matrix Visualization



# Literature Review Method

- **Question:** How was ontology used in AI/ML to solve biomedical problems?
- **Database:** PubMed Central 
- **Cut-off date:** 9/4/2022
- **Filter:** within 5 years (2017-2022)
- **Search strategy:**  
“ontology” AND (“machine learning” OR “artificial intelligence” OR “deep learning” OR “neural network” OR “embedding”) in Title and Abstract
- **Result:** 503 articles retrieved (ordered by relevance)
- **Screening result:** 108 papers included in this study.

# Highlights of the 6 Findings

1. The papers published shows **exponential growth** within recent 5 years, and it will continue growing.
2. The researchers in this field are **isolated and separated** in different countries:USA, China, Saudi&UK, Dr. Rob Hoehndorf is a hub, however within a niche of Saudi and UK group.

3. **Gene Ontology (GO)** dominated the research (57%)! A sum of **disease/condition/phenotypes** related ontologies made up to
4. Most papers solves **limited biological, clinical problems**. Top 1: protein/gene/function prediction.



5. ML methods: **Neural Network, Deep Learning, Vector (?2Vec)** related methods to numericize the semantics of ontologies.
  
6. Ontology is used as data category for training datasets and output target label. Or methods to reduce the dimensions of semantics. **Methods using full power (logic, axioms, full structure) of ontology is still lacking!**

# What we concluded so far?

- More ROBI.
- Communicate, collaborate, and generate ideas and solutions.
- Build the communities.
- Solve the real problems!.

Let a Thousand Flowers Bloom

百花齊放，百家爭鳴

