Graph Machine Learning Applications in Biomedicine

Maria Brbić
Leskovec group
Stanford University
Biological systems are naturally represented as networks!

- Protein interaction networks
- Cell networks
- Disease networks
Uncoupled Knowledge

**Challenge:** Many dissociated databases of biological entities

- **Drugs:** DrugBank, PubChem, ChEBI...
- **Disease:** MeSH, DiseaseOntology, DDB,...
- **Adverse events:** MedDRA, ADReCS,...

How can we integrate this knowledge and develop machine learning methods that can reason over it to discover new biology?
Integrate to Discover

Key: Integrate knowledge to capture complex underlying biological mechanisms
Integrate to Discover

**Key:** Integrate knowledge to capture complex underlying biological mechanisms
Key idea: Construct knowledge graph that models known biology and learn to reason over it.
Why Knowledge Graphs?

- We can explicitly store knowledge about underlying biology and chemistry
- Interpretable
- Easy to update and improve
Example: Bio-Knowledge Graph

- Represent facts as triples \((h, r, t)\)
  - (‘BRCA1’, ‘associated_with’, ‘Breast_Neoplasms’)
  - (‘Breast_Neoplasms’, ‘is_a’, ‘Breast_Disease’)
  - ...

- **Node types:** drug, disease, adverse event, protein, functions, ...

- **Relation types:** causes, assoc, treat, interact, ...

```plaintext
§

Node types: drug, disease, adverse event, protein, functions, ...
Relation types: causes, assoc, treat, interact, ...
```
Why Knowledge Graphs?

Discover new biology by predicting missing links in the knowledge graph.

Can drug A cause adverse event B?
Can drug A treat disease C?
Is disease C associated with protein P?
Building Predictive Models over KGs

How can we leverage knowledge graphs and build predictive models?

Key idea: Learn to embed nodes (e.g., drugs, diseases, protein)
Recap: Graph Convolutional Networks

Idea: Node’s neighborhood defines a computation graph

Learn how to propagate information across the graph to compute node features

[Kipf and Welling, ICLR 2017]
R-GCN: Learn relation-specific neural network weights

Weights $W_{r_1}$ for $r_1$

Weights $W_{r_2}$ for $r_2$

Aggregate information from relation-type specific node’s neighborhood

[Schlichtkrull, Kipf et al, Proceedings of ESWC 2018]
1) Given a graph, learn a low-dimensional vector \((\text{embedding})\) for every node.

2) Use the learned embeddings for downstream prediction task.
What Tasks Can We Solve?

Example applications:

1) Predicting polypharmacy side effects
2) Predicting drug-disease treatments
3) Predicting outcomes of clinical trials
Predicting drug polypharmacy side-effects

Zitnik, Agarwal, Leskovec. *Bioinformatics* 2018
Many patients take multiple drugs to treat complex or co-existing diseases

Task: Given a pair of drugs predict adverse side effects

[Zitnik, Agarwal, Leskovec. Bioinformatics 2018]
Approach: Build a Graph

Model:

- Polypharmacy side effects
- Drug-protein interactions
- Protein-protein interactions

[Zitnik, Agarwal, Leskovec. Bioinformatics 2018]
**Task:** Given a partially observed graph, predict labeled edges between drug nodes.

**Example query:** Given drugs $c, d$, how likely is an edge $(c, r_2, d)$?

Co-prescribed drugs $c$ and $d$ lead to side effect $r_2$.

[Zitnik, Agarwal, Leskovec. *Bioinformatics* 2018]
Approach: Drug Side Effects

1) Take the graph and learn a $d$-dimensional vector (embedding) for every node

2) Use the learned embeddings to predict side effects of drug pairs

[Zitnik, Agarwal, Leskovec. Bioinformatics 2018]
**De novo Predictions**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Drug $c$</th>
<th>Drug $d$</th>
<th>Side effect $r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pyrimethamine</td>
<td>Aliskiren</td>
<td>Sarcoma</td>
</tr>
<tr>
<td>2</td>
<td>Tigecycline</td>
<td>Bimatoprost</td>
<td>Autonomic neuropathy</td>
</tr>
<tr>
<td>3</td>
<td>Omeprazole</td>
<td>Dacarbazine</td>
<td>Telangiectases</td>
</tr>
<tr>
<td>4</td>
<td>Tolcapone</td>
<td>Pyrimethamine</td>
<td>Breast disorder</td>
</tr>
<tr>
<td>5</td>
<td>Minoxidil</td>
<td>Paricalcitol</td>
<td>Cluster headache</td>
</tr>
<tr>
<td>6</td>
<td>Omeprazole</td>
<td>Amoxicillin</td>
<td>Renal tubular acidosis</td>
</tr>
<tr>
<td>7</td>
<td>Anagrelide</td>
<td>Azelaic acid</td>
<td>Cerebral thrombosis</td>
</tr>
<tr>
<td>8</td>
<td>Atorvastatin</td>
<td>Amlodipine</td>
<td>Muscle inflammation</td>
</tr>
<tr>
<td>9</td>
<td>Aliskiren</td>
<td>Ticloprozole</td>
<td>Breast inflammation</td>
</tr>
<tr>
<td>10</td>
<td>Estradiol</td>
<td>Nadolol</td>
<td>Endometriosis</td>
</tr>
</tbody>
</table>
De novo Predictions

<table>
<thead>
<tr>
<th>Rank</th>
<th>Drug $c$</th>
<th>Drug $d$</th>
<th>Side effect $r$</th>
<th>Evidence found</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pyrimethamine</td>
<td>Aliskiren</td>
<td>Sarcoma</td>
<td>Stage et al. 2015</td>
</tr>
<tr>
<td>2</td>
<td>Tigecycline</td>
<td>Bimatoprost</td>
<td>Autonomic neuropathy</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Omeprazole</td>
<td>Dacarbazine</td>
<td>Telangiectases</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Tolcapone</td>
<td>Pyrimethamine</td>
<td>Breast disorder</td>
<td>Bicker et al. 2017</td>
</tr>
<tr>
<td>5</td>
<td>Minoxidil</td>
<td>Paricalcitol</td>
<td>Cluster headache</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Anagrelide</td>
<td>Azelaic acid</td>
<td>Cerebral thrombosis</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Atorvastatin</td>
<td>Amlodipine</td>
<td>Muscle inflammation</td>
<td>Banakh et al. 2017</td>
</tr>
<tr>
<td>9</td>
<td>Aliskiren</td>
<td>Tioconazole</td>
<td>Breast inflammation</td>
<td>Parving et al. 2012</td>
</tr>
<tr>
<td>10</td>
<td>Estradiol</td>
<td>Nadolol</td>
<td>Endometriosis</td>
<td></td>
</tr>
</tbody>
</table>

Case Report

Severe Rhabdomyolysis due to Presumed Drug Interactions between Atorvastatin with Amlodipine and Ticagrelor
Predicting drug-disease treatments

Ruiz, Zitnik, Leskovec. *Nature Communications* 2021
Disease-Drug Treatments

Goal: Understanding drug-disease mechanisms of action

Will the drug treat a disease?

What genes alter drug efficacy?

[Ruiz, Zitnik, Leskovec. Nature Communications 2021]
Multi-Scale Interactome

Model drug-disease treatment by integrating proteins and a hierarchy of biological functions

Covers:
- 1,661 drugs
- 840 diseases
- 17,660 proteins
- 9,798 biological functions

[Ruiz, Zitnik, Leskovec. *Nature Communications* 2021]
Explaining Drug Disease Mechanisms

**Approach:** Compute network diffusion profiles by biased random walks to explain how drug and disease effects propagate in a biological network.

[Ruiz, Zitnik, Leskovec. *Nature Communications* 2021]
Multi-Scale Interactome: Results

Significant improvement over baselines

<table>
<thead>
<tr>
<th>Method</th>
<th>Graph</th>
<th>AUROC</th>
<th>Avg Prec</th>
<th>Rec@50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein Overlap</td>
<td>🌟🌟</td>
<td>0.499</td>
<td>0.064</td>
<td>0.298</td>
</tr>
<tr>
<td>Functional Overlap</td>
<td>🌟🌟</td>
<td>0.558</td>
<td>0.050</td>
<td>0.237</td>
</tr>
<tr>
<td>Molecular-scale Interactome</td>
<td>🌟🌟🌟</td>
<td>0.620</td>
<td>0.065</td>
<td>0.264</td>
</tr>
<tr>
<td>Multiscale Interactome</td>
<td>🌟🌟🌟</td>
<td>0.705</td>
<td>0.091</td>
<td>0.347</td>
</tr>
</tbody>
</table>

Interpretability: Can reveal the proteins and biological functions relevant to treatment

Multiscale vs Molecular-scale Interactome

+13.7%  +40.0%  +31.4%
Predicting outcome of clinical trials

Joint work with:
Prabhat Agarwal, Michihiro Yasunaga, Jure Leskovec
Machine Learning for Clinical Trials

Can we use machine learning to guide clinical trials design?

Why is it hard?

- Unstructured and incomplete text
- Challenging to define success
- Multiple patients sub-groups and multiple phases
- How to model complex biological and chemical mechanisms?
Predicting Outcome of Clinical Trials

- Model and disentangle 3 key factors:
  - Biology of Disease
  - Chemistry of Drug
  - Study Protocol
**Challenge:** Clinical trials database consists of highly unstructured, unlabeled data

- **Intervention/treatment:**
  
  Drug: LY003003 (Rotigotine, extended-release microspheres)

  Patients to be enrolled to 70 mg dose group will receive 14 mg in the first week, 28 mg in the second week, 42 mg in the third week, 56 mg in the fourth week and then 70 mg in the next 5 weeks.

- **Inclusion Criteria:**
  1. Patient had Parkinson’s Disease that meet the clinical diagnostic criteria of the brain bank of the Parkinson’s Disease Association of the United Kingdom.
  2. Patient was Hoehn & Yahr stage ≤3 (excluding stage 0); 3. Patient was male or female aged 18 to 75 years;
  4. Patient had a Mini Mental State Examination (MMSE) score of ≥25;
  5. Patient had a Unified Parkinson’s Disease Rating Scale (UPDRS) motor score (Part III) of ≥10 but ≤30 at Screening.
  6. Patient who signed the informed consent form volunteered to participate in this clinical trial and could cooperate with the prescribed inspections.

- **Primary Outcome Measures:**

  1. Frequency of adverse events [Time Frame: From screening up to day 50]

  Adverse events to evaluate the safety and tolerability of LY03003

---

**How is drug applied?**

**Who is eligible to apply?**

**How is success measured?**

**How can we structure this data?**
Clinical Trials Knowledge Graph

- **Node types:** trial arm, drug, disease, primary outcome, population, adverse events
- **Relation types:** tests, investigates, inclusion, exclusion, has_outcome, has_AE, has_serious_AE
Grounding Clinical Trials KG

**Approach:** Ground clinical trials knowledge graph in underlying biology and chemistry

- **Biology of disease:**
  - Disease-associated genes, protein interactions, protein pathways,…

- **Chemistry of drug:**
  - Drug target and off-target proteins, chemical structure, side effects, …
TrialNet Knowledge Graph

- Covers 70k interventional clinical trials
- 300k nodes (clinical trials protocol entities, biological and chemical entities)
- 10.8 millions edges across 18 relation types
Prediction Tasks

- **Efficacy:** How likely will a new drug surpass existing treatments?

  \[ f(\text{z}_1, \text{z}_2) = \{1, 2\} \]

- **Safety:** Is the drug safe for use?

  \[ f(\text{z}) = \{0, 1\} \]

  Predict which adverse events will happen in a trial

  Predict which treatment will result in higher progression-free-survival
Our model significantly outperforms PubMedBERT with a +16% improvement. Never-before-seen drugs.
Takeaways

- Knowledge graphs are powerful representation for storing and capturing known biology
- Graph neural networks that operate on heterogeneous graphs can be used to build predictive models
- We can predict drug safety, drug efficacy, outcome of clinical trials
Acknowledgements

SNAP group
Jure Leskovec (PI)
Michihiro Yasunaga
Prabhat Agarwal
Camilo Ruiz
Antoine Bosselut
Adrijan Bradaschia
Kaidi Cao
Serina Chang
Weihua Hu
Hamed Nilforoshan
Alex Porter
Hongyu Ren
Yusuf Roohani
Rok Sosic
Tailin Wu
Rex Ying
Jiaxuan You

Collaborators:
Marinka Zitnik, Harvard
Russ Altman, Stanford Bioengineering
Hongjie Li, Baylor College of Medicine
Liqun Luo, Stanford Biology
Colleen McLaughlin, Stanford Biology
Angela Pisco, CZ Biohub
Stephen Quake, Stanford Biology
ChiChi Xie, Stanford Biology