Graph Neural Networks (GNN) for predicting safe drug combinations

Amir H. Amanzadi
Guest lecturer
AI4DD course, September 2022

(2013 – 2019) **BSc. Double major in Chemistry and Mechanical Eng., Sharif University of Technology**

(2016 – 2019) **ML Researcher at the National Cognitive Research Institute**

(2019 – 2021) **MSc. Pharmaceutical Modelling, Uppsala University**

(2020 – 2021) **Thesis intern at Karolinska institute, Department of Oncology and Pathology**

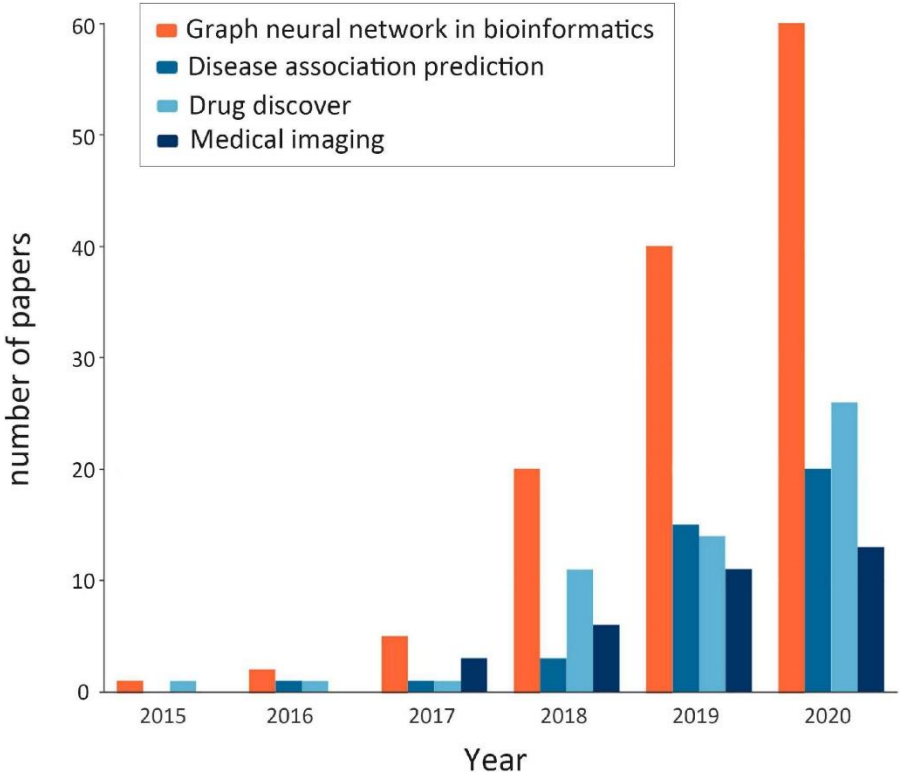
(2021 – 2022) **AI/ML Lead at Celeris Therapeutics**

(2022 – present) **ML researcher at Karolinska University Hospital**

Outline

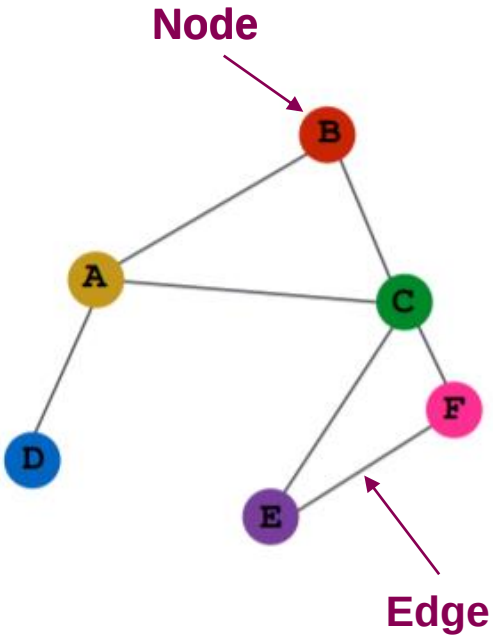
- Overview of Graph networks
- Introduction to drug combinations
- Application of GNNs in Drug Discovery
- Summary (pros and cons)
- Conclusion

Why should you care about GNN?



Ming-Jing Tang et al. 2021

What is a Graph?

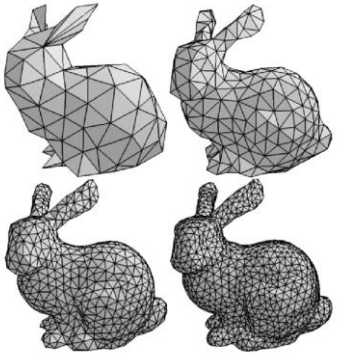


- Describes an interaction or relationships
- Contains Nodes and Edges
- Node: The object of interest
- Edge: The presence of a relation/interaction
- Features includes all useful information about Nodes and Edges

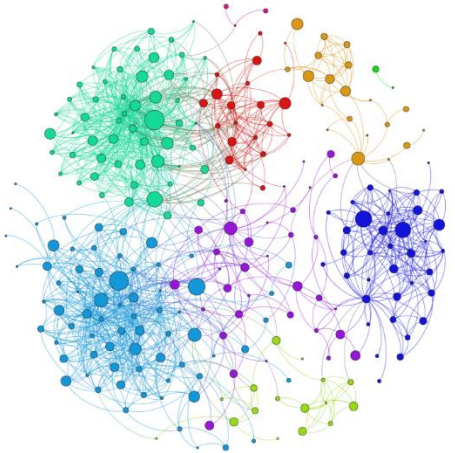
Jure Leskovec et al. 2018

Graphs (interactomes) are everywhere!

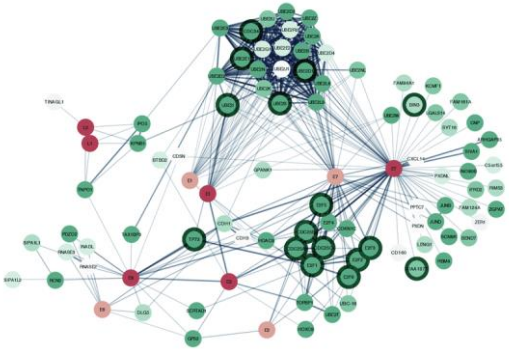
Social Networks



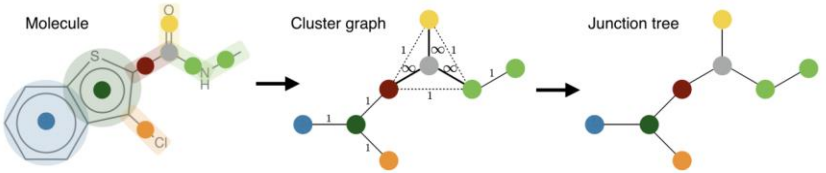
3D Surface/Mesh



Recommendation Systems

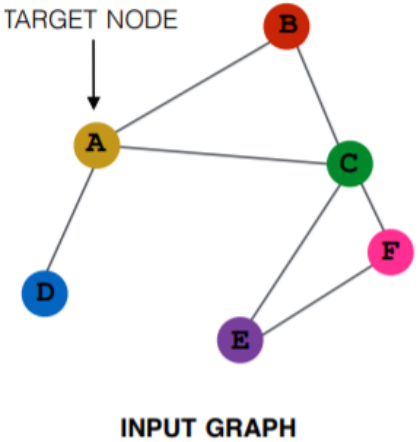


Biological Networks

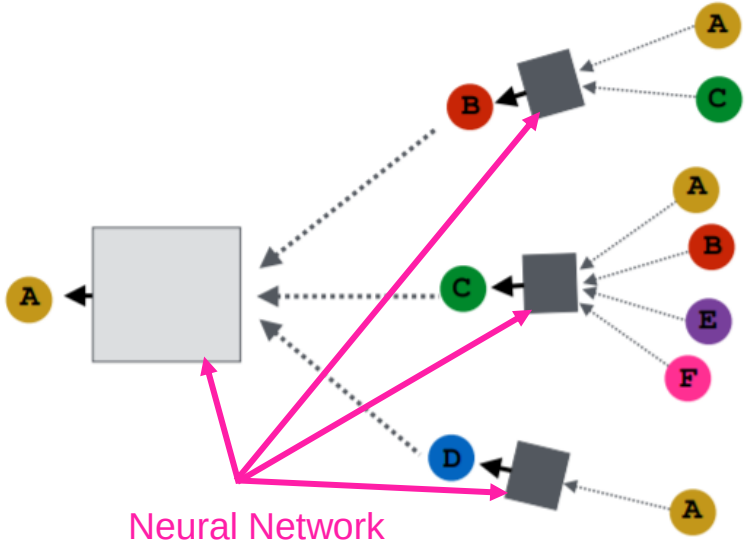


Molecules

What is a Graph Neural Network ?



INPUT GRAPH



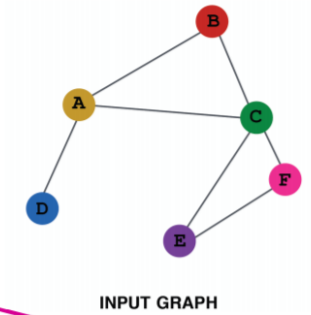
Neural Network

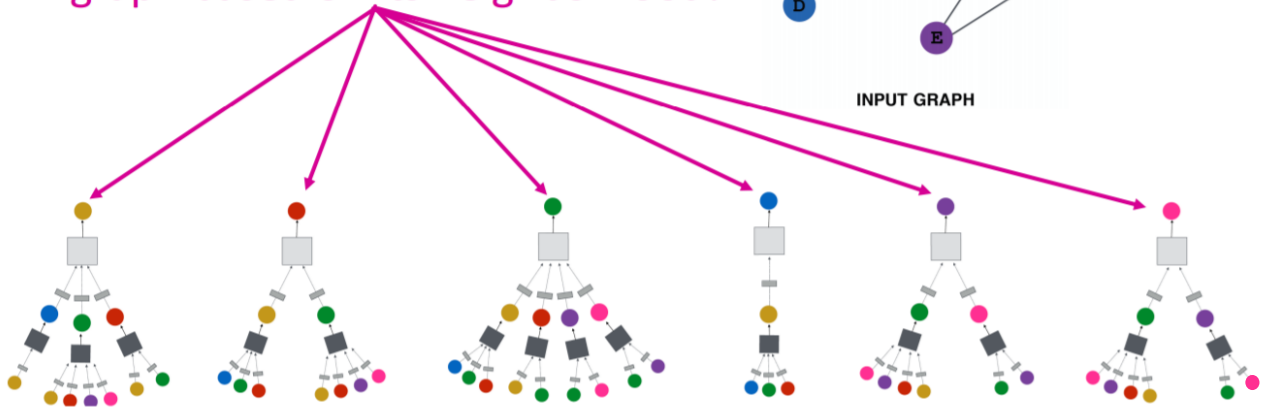
Jure Leskovec et al. 2018

What is a Graph Neural Network ?

- **Intuition:** Network neighborhood defines a computation graph

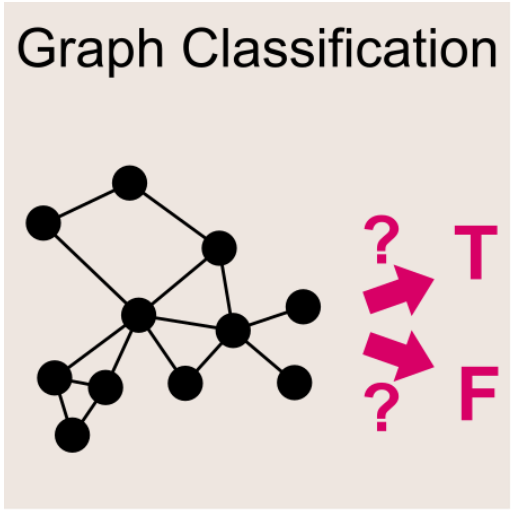
Every node defines a computation graph based on its neighborhood!

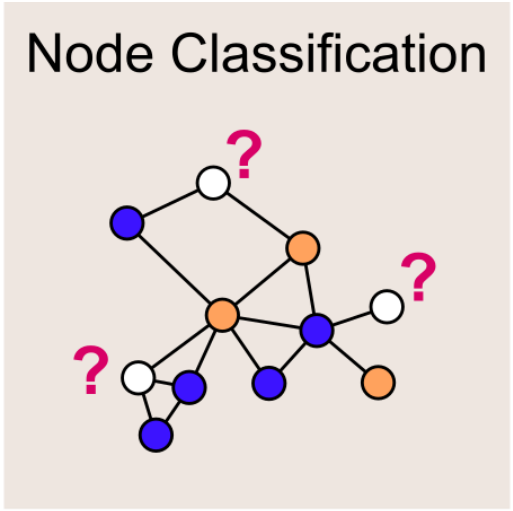


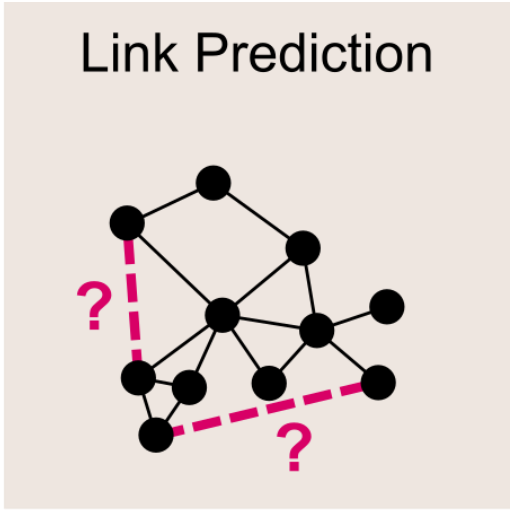


Jure Leskovec et al. 2018

Different types of GNN models.







Importance of drug combinations

- Many patients take multiple drugs to treat complex or co-existing diseases.

  → Some take more than 20 drugs to treat heart disease, depression, etc.

- COVID-19: 70%-80% of people were on 3+ medications *(McQueenie et al., 2020)*.
- WHO reports an 18 billion US dollars global expenditure caused by unsafe drug combinations per year.

Saftey of drug combinations

- High magnitude of possible drug combinations
- Limited clinical research on Polypharmacy Side Effects (PSE)
- Very expensive and time consuming to test in vitro/vivo
- Lack of general and standard procedures (different diseases)

Solution → **In silico**

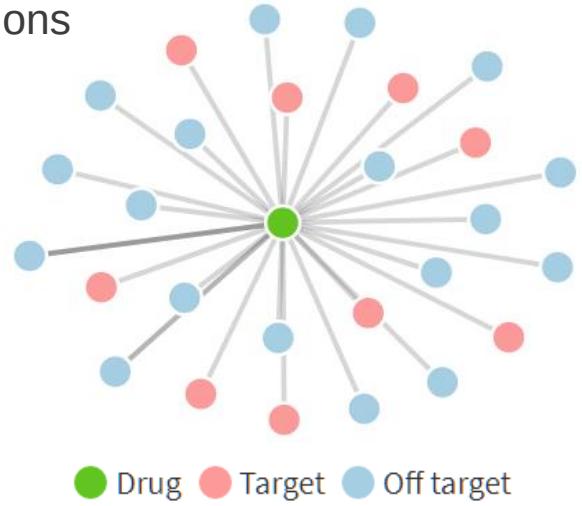
- Heterogenous datasets
  - → Protein-Protein interactions (PPI)
  - → Drug-Target interactions (DTI)
  - → Drug-Drug interactions (DDI)
- Incorporating all information in a single entity (holistic view)

Data structure → **Graph networks**

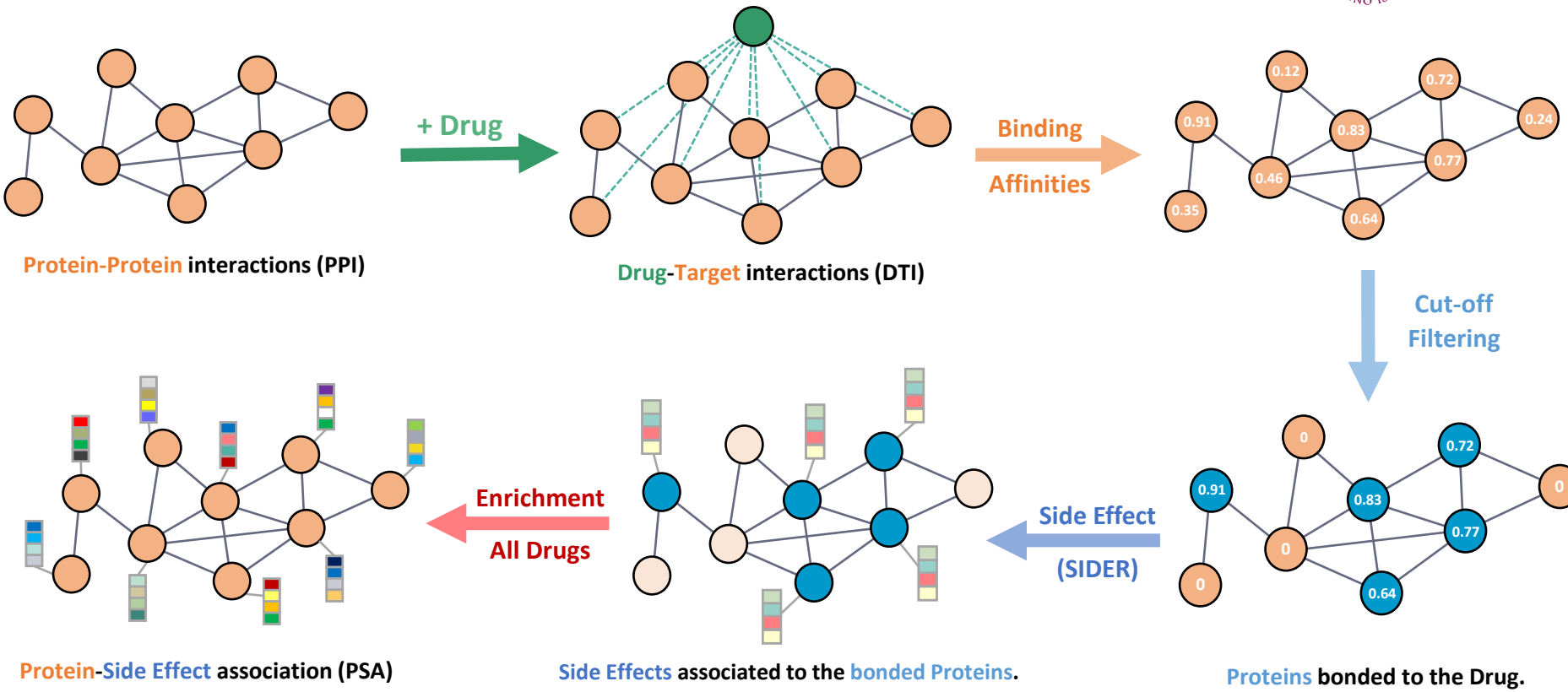
Aim: To predict and interpret PSE of any given drug pairs.

Addressing a data bias

- Drug-Target interactions (DTI)
  - → Useful for target identification, novel biomarker, drug repurposing, …
  - → **Not** useful for predicting side effects
- Side effects are mainly caused by off-target drug interactions
- All current models use the same DTI for SE prediction.
- Solution: Computing the entire drug-protein interactions



Overview of the pipeline



Protein-Protein interactions (PPI)

Drug-Target interactions (DTI)

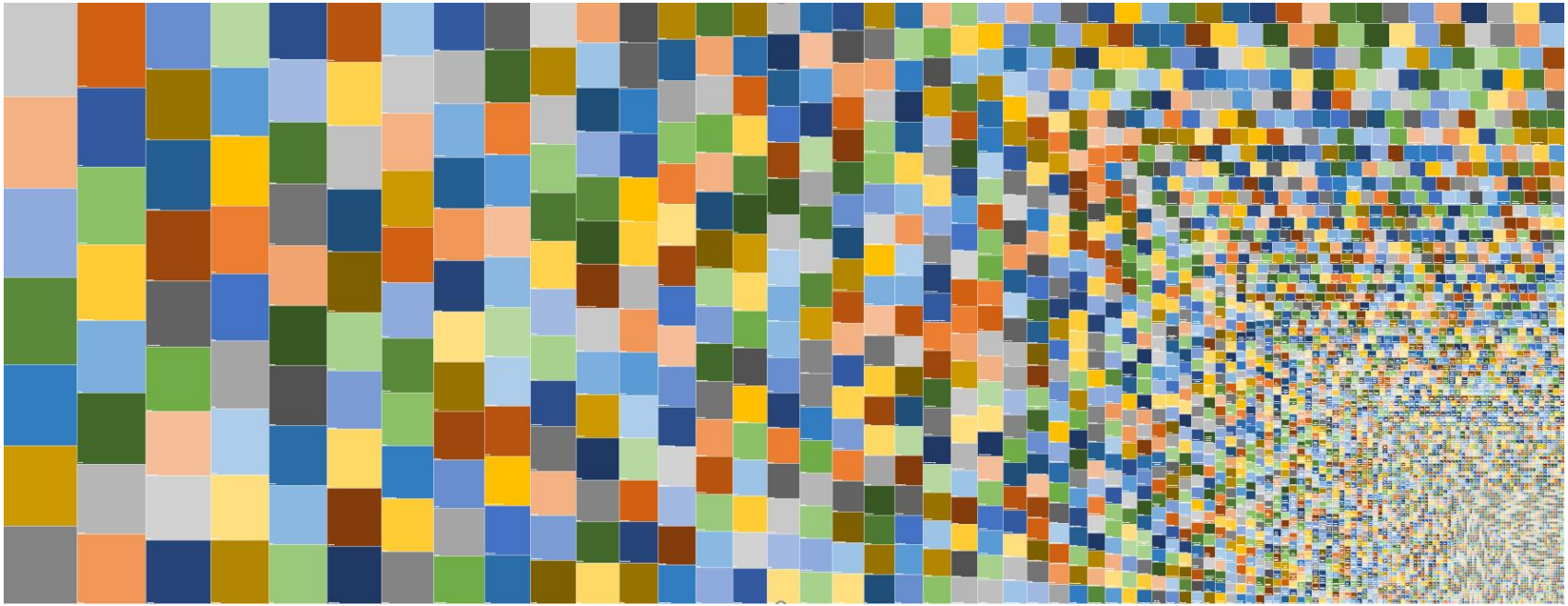
Proteins bonded to the Drug.

Side Effects associated to the bonded Proteins.

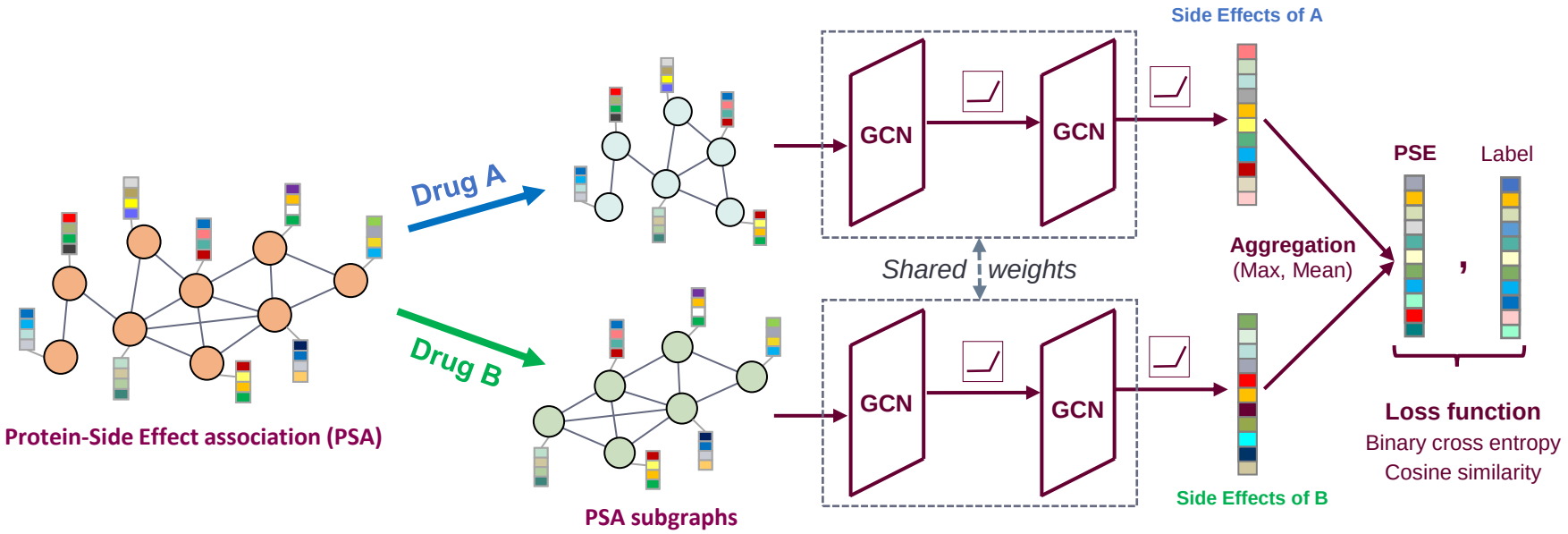
Protein-Side Effect association (PSA)

Protein-Side effect Association (PSA)

- 5085 unique side effects
- 964 adverse drug reaction (ADR) ⟶ GNN



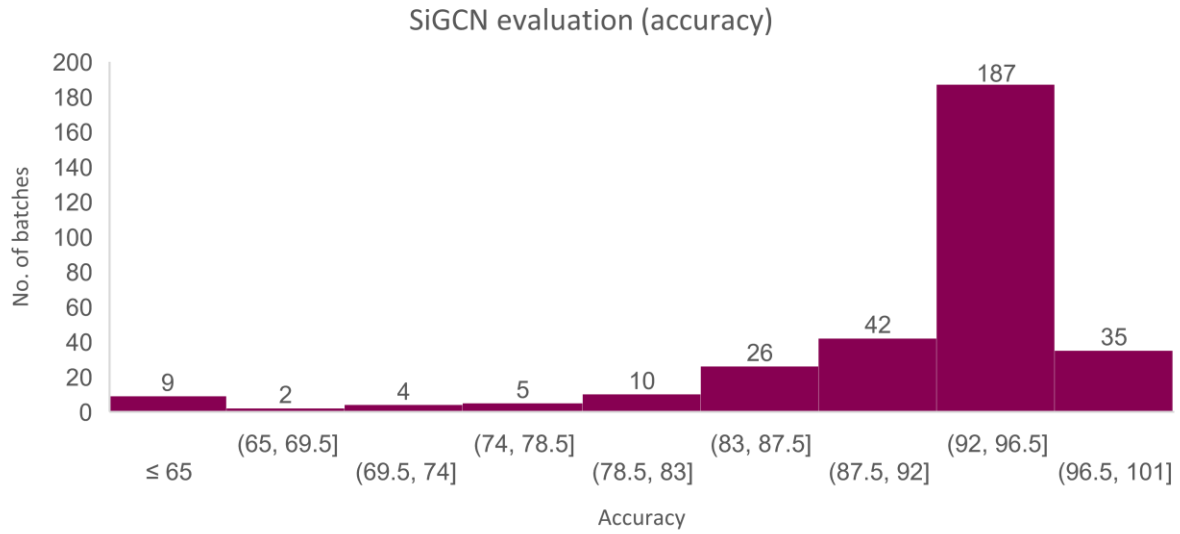
Siamese Graph Convolutional neural network (SiGCN)



SiGCN is trained on the aggregation of the outputs.

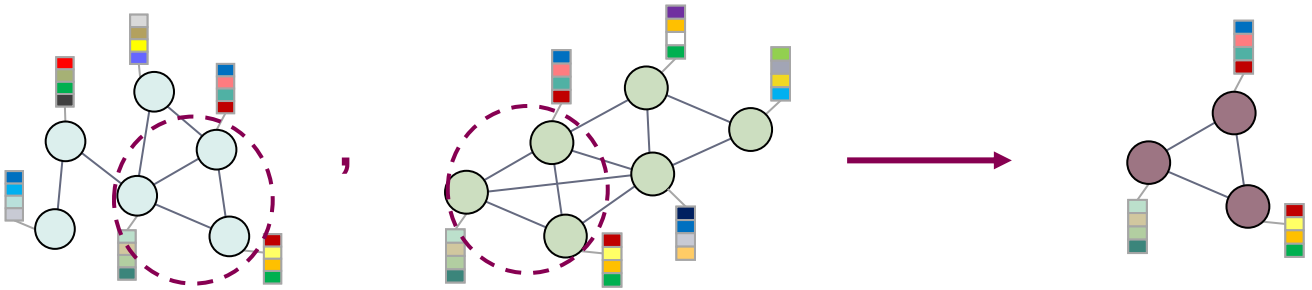
Model evaluation

- The pretrained model was saved and evaluated (batches of 35)
- 604 unseen drugs with their corresponding 10’000 paired drug combinations
- 91.43% accuracy ⟶ 9143 correct PSE predictions from 10000 drug combinations
- Concentrated distribution on 92-96.5% accuracy



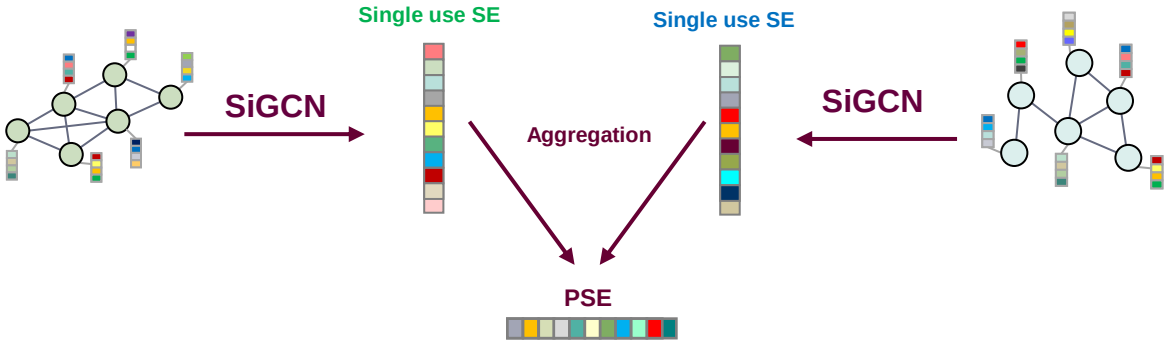
Interpretability of the model

- The outcome of SiGCN is also highly explainable.
  - → Every prediction can be backtracked to the associated proteins.
  - → Proteins that appear in both drug's PSA subgraphs can be interpreted as the source of the PSE.
- Can be examined future in terms of their tissue expression and pathways (DECISION)



Scalability of the model

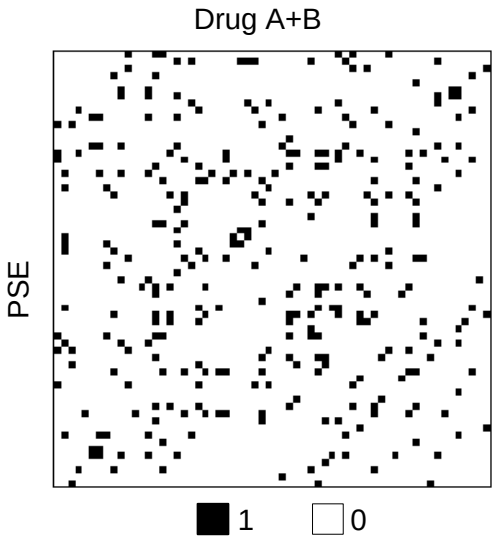
- Siamese enables scaling to any number of drug combinations.
  - → Model weights were trained based on the aggregation of the inputs
  - → mean pooling is a non-hieratical operation (not sensitive to change in order or scale)
- Model first produces single use SE, then aggregates them to predict PSE.
  - → Inductive reasoning ⟶ we can infer the of the scaled prediction of SiGCN is meaning full and reliable.



Limitation of the data

- Model suffers from a data limitation, can't control it but can be improved.
- It is rare for a drug combination to have 964 PSEs ⟶ **data sparsity**
- Despite high accuracy, model has a mediocre precision of 48.03%.
- Predicts non-occurring PSEs (0), better than occurring PSEs (1)
- Absence of regularization

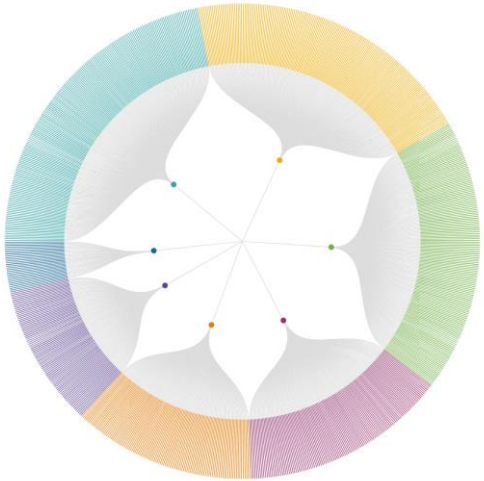
  → Complexity of implementing a drop out in dynamical graphs



Applied PSA

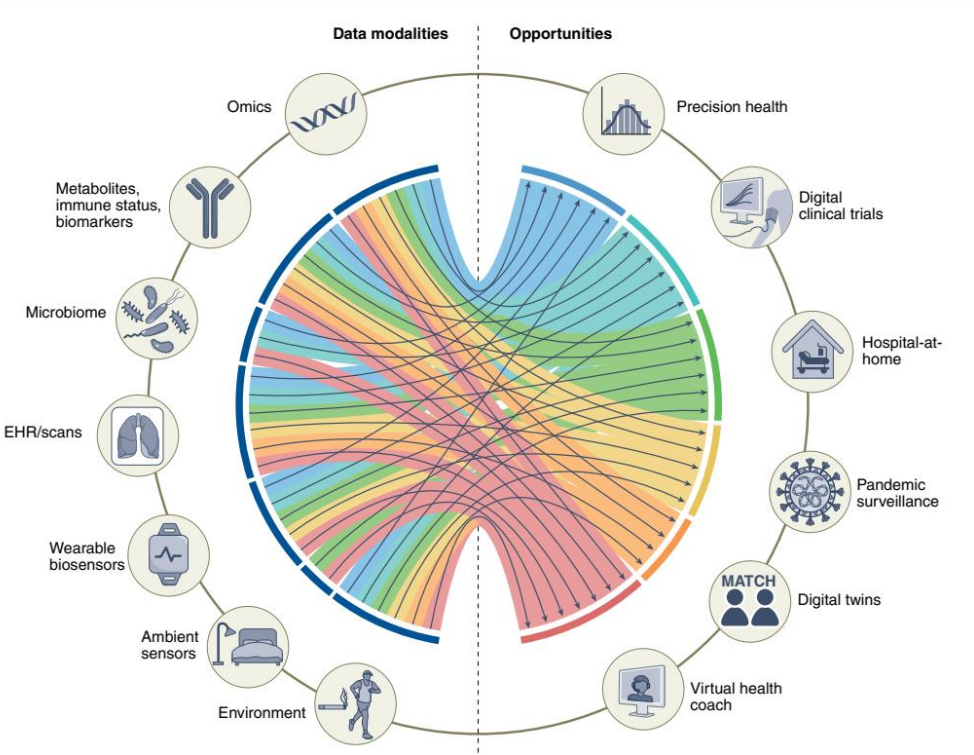
- DECISION:
  - Identifying optimal combinatorial therapies for decompensated cirrhosis
  - Cause and disease progression is unknown, 81% mortality rate in 9 month
  - Combination therapy of 1200+ patients from 30 countries

- Concerning side effects:
  - Ascites
  - Encephalitis
  - Decompensated cirrhosis
  - Acute hepatic failure
  - Liver and other organ failures



Multimodal learning

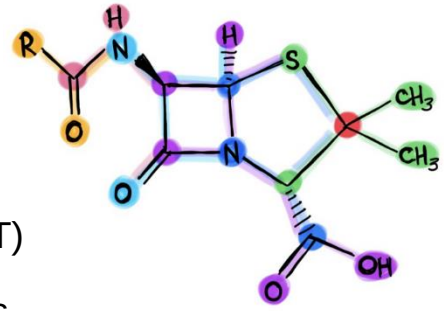
- Multi-Omics
  - → Transcriptomics
  - → Metabolomics
  - → Genomics
  - → Proteomics
- Electronic health records (MRIs, Blood samples, …)
- Drug combinations
- Lifestyle and environmental factors
- 3 drug combinations are now being in animal models



Pros and cons of GNNs

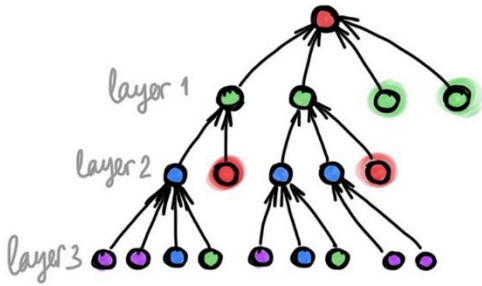
Pros:

- Works well with relational or interactome data structures
- High interpretability if implemented correctly
- Best method for learning structural properties (QSAR, Docking, ADMET)
- Can reach an optimum predictive capacity with relatively small datasets



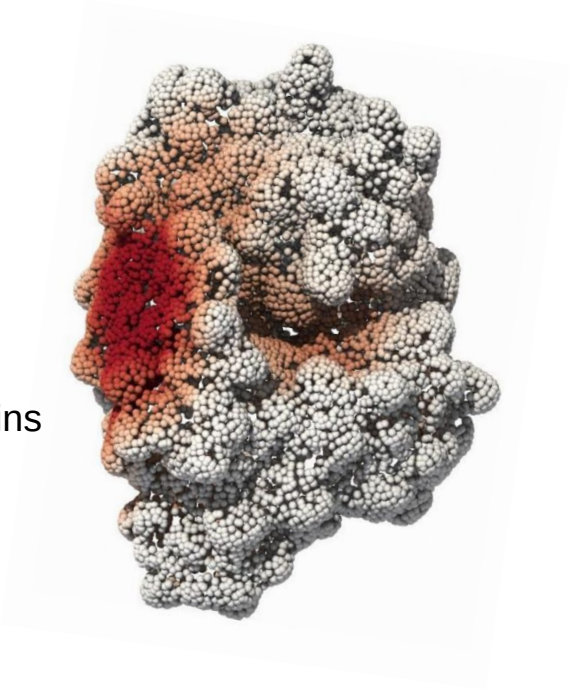
Cons:

- Not easy to implement in code
- Required lots of computational power
- Overfitting is a problem (Absence of regularization)!
- Lack of pretrained transforms, Transfer learning not an option



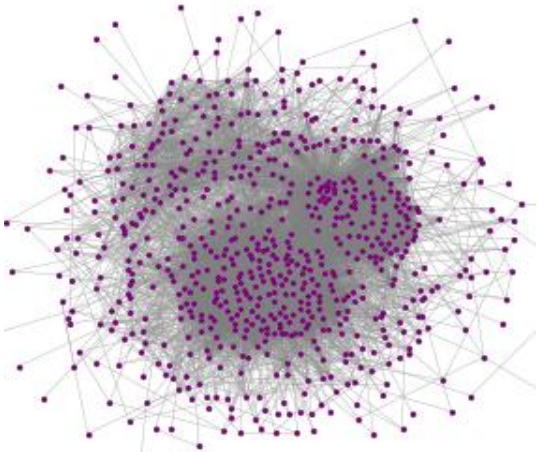
When to use GNNs?

- At least Medium sized but a well-balanced dataset
- Working with interactomes and network data
  - Protein-Protein interactions (PPI)
  - Drug-Target interactions (DTI)
  - Drug-Drug interactions (DDI)
  - Biological Pathways
- Linking Structural features to chemical and physical properties (proteins and molecules).
- When have enough time and computational power
- Not applicable when working with images or text



Conclusion

- Example of how GNNs can be used in Drug Discovery
- When to use GNN ⟶ Interactomes, networks and structural data
- When not to use GNN ⟶ Images, videos and text data
- Graphs are one the best representation of structure in ML
- Graphs also integrates the Multi-omics data well with ML
- GNNs are hard and time consuming to implement and optimize
- No ImageNet moment yet!
- It's a very hot topic both in the industry and the academia



PSA sub-graph of "Glyburide"

Karolinska Institutet

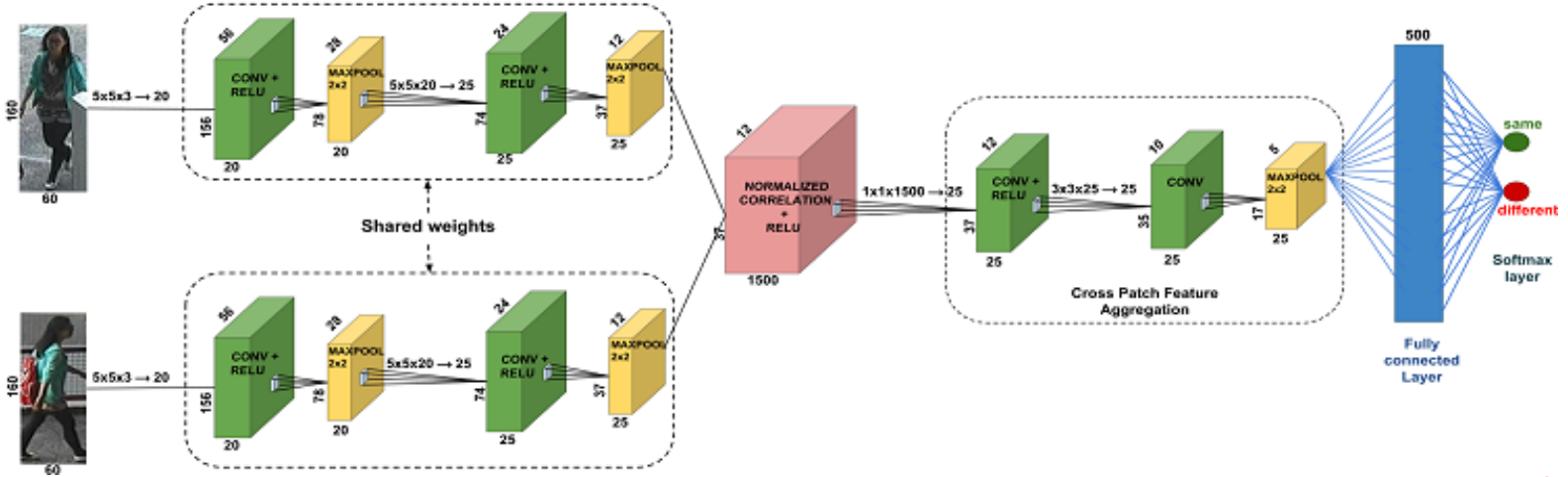
Thank you for your attention.

Any questions?

Amir.Amanzadi@gmail.com

Appendix - Siamese neural network

- Less data driven (one-shot learning)
- Able to handle double inputs (drugs)
- Compatible with graph networks but needs optimization



Subramaniam et al. 2016

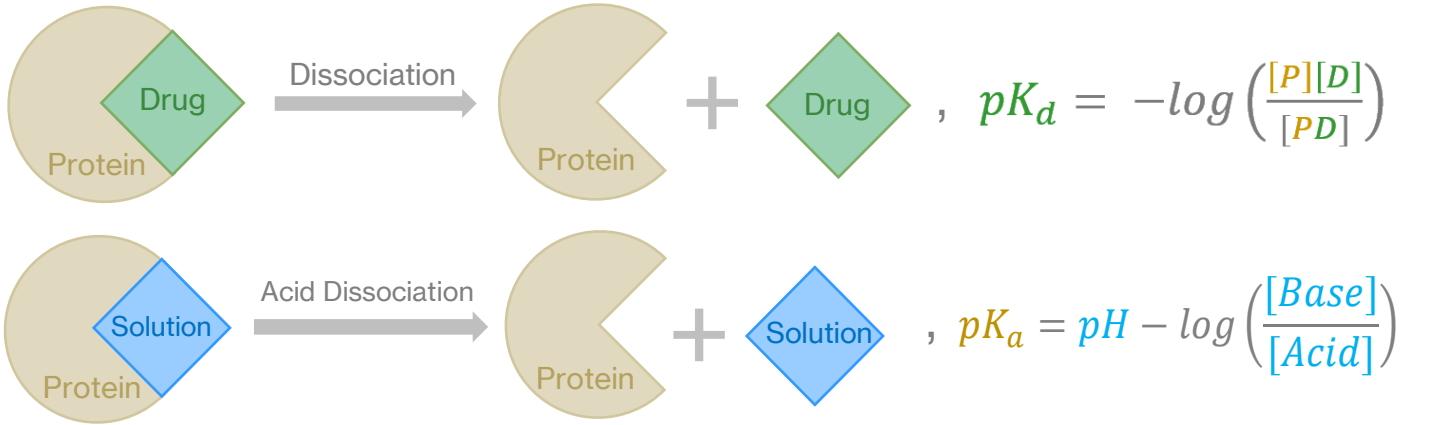
Appendix - Isoelectric point (pI)

- **Isoelectric point** is the pH at which a protein carries no net electrical charge or is electrically neutral in the statistical mean.
- pI is widely used in biochemistry for protein purification.
- Can be calculated from the protein sequence unlike the pKa
- Individualizes pKa values

$$pH(I) = pI = pK_a + log\left(\frac{[Base]}{[Acid]}\right) \xrightarrow[[Base] \approx [Acid]]{\text{neutral}} pI = pK_a$$

Cut-off: *pKd* > *pI*

Appendix - Cut-off values



Cut-off:
pKd > *pKa*

But in which *pH* ?

Isoelectric point (pI) ⟶ pH at which a protein is electrically neutral.

$$pH(I) = pI = pK_a + log\left(\frac{[Base]}{[Acid]}\right) \xrightarrow[[Base] \approx [Acid]]{\text{neutral}} pI = pK_a$$

pKd > pI

Appendix - Enrichment

- The simplest way to establish Protein-Side effect Association (PSA):
  - → If protein interacts with a drug & also the drugs have the side effect ⟶ Association
  - → Matrix multiplication
- Versatility ⟶ Modules: gene and tissue expressions, pathways, drug dose and SE frequency
- It may not be accurate ⟶ Neural network for optimization
- Simplest form for this project ⟶ Can add modules later

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

protein x *drug* *drug* x *SE* *protein* x *SE*

0x0 + 1x1 + 1x1 = 2

Appendix - PSA modularity

- Affinity calculations are derived from chemical potential
- Chemical potential only calculates 1:1 (mole) interactions

$$\mu_d = \left(\frac{\partial G}{\partial N_d}\right)_{T,P,N_p}$$

- Different modules
  - Dose $\longrightarrow \mu^* = \mu \times N_d$
  - Gene expressions $\longrightarrow \mu^* = \mu \times N_p$

$$pK_d{}^* = pK_d \times N_d \times N_p$$

  - SE frequency $\longrightarrow$ Changing *drug x SE from binary to decimal (probability)*

- Simplest form for this project $\longrightarrow$ Can add modules later

$$\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

protein x drug *drug x SE* *protein x SE*

Appendix - Kd and binding affinity

$$\Delta G = RT \ln \frac{K_d}{c^{\emptyset}} \text{ , standard reference concentration } c^{\emptyset} = 1 \text{ mol/L.} \longrightarrow pK_d = -\frac{c^{\emptyset} \log e}{RT} \Delta G$$

$$\mu_d = \left(\frac{\partial G}{\partial N_d}\right)_{T,P,N_p} \longrightarrow \partial G = \int_0^{N_d} \mu_d \times N_p \, \partial N_d$$

$$\Delta G = \mu_d \; N_p(N_d - 0) = \mu_d \, N_p N_d$$

$$pK_d = -\frac{c^{\emptyset} \log e}{RT} \mu_d \, N_p N_d$$

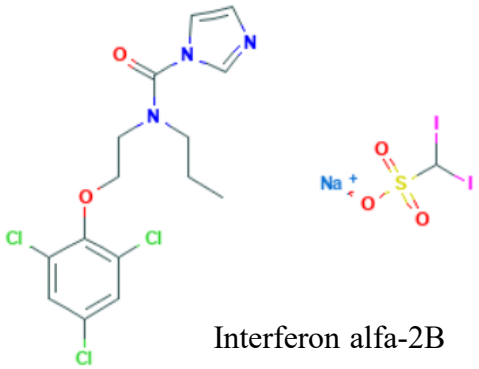
$$pK_d^* = pK_d N_p N_d$$

Tools and datasets

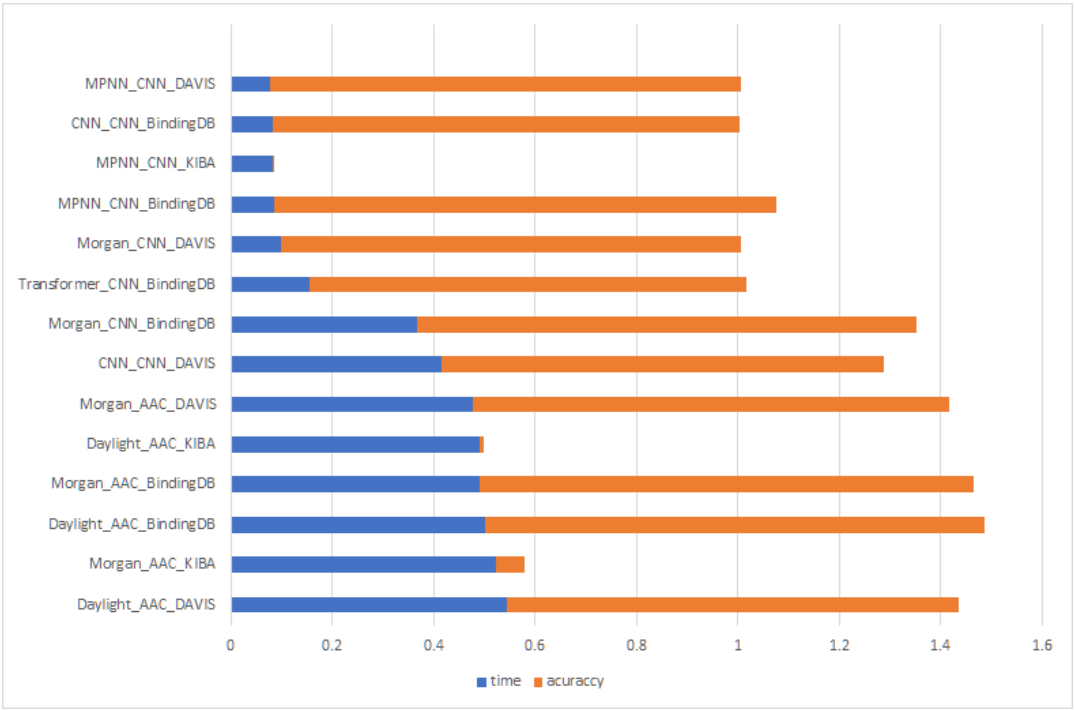
- **SIDER** ⟶ 1430 drugs, 5000+ side effects
- **STRING** ⟶ 19567 human protein (+ sequence), PPI network
- **SIB-Expasy** ⟶ pI calculation
- **DeepPurpose** ⟶ Binding Affinity prediction (15 pretrained ML models)
- **PubChem** ⟶ Drug SMILES
- **Data handling** ⟶ Pandas, Numpy
- **GNN** ⟶ Pytorch, DGL, PyG, NetworkX
- **Computation** ⟶ Docker, Singularity, CUDA, tmux

Appendix - Binding affinity predictions

- DeepPurpose: 14 pre-trained models
- Tested on 1000 affinity values (Kd)
- MPNNs: High accuracy, short run time
- CNN and Morgan as alternatives



Interferon alfa-2B



Protein-Side effect association (PSA)