Combining bottom up and top down approaches for Drug-Target-Interaction prediction

Tilman Hinnerichs

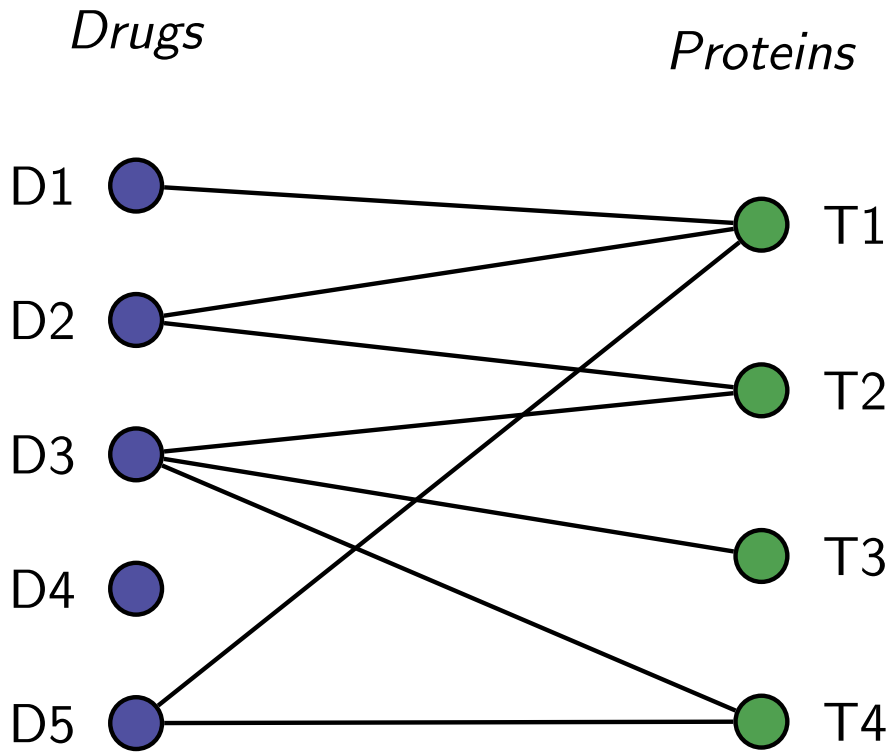
BORG - KAUST

December 09, 2019

Outline

Problem Description

Prediction over bipartite graph:



Classification of recent approaches[1][2]

	Drugs	Protein
bottom-up	▶ GCN over molecules ▶ drug similarity	▶ secondary structure prediction ▶ contact prediction ▶ convolution over amino acid sequences
top-down	▶ network approaches ▶ drug similarity	▶ protein similarity

[1] Chen Wang et al., Briefings in Bioinformatics, 2018
[2] Yu Ding et al., Briefings in Bioinformatics, 2019

Problems of recent approaches

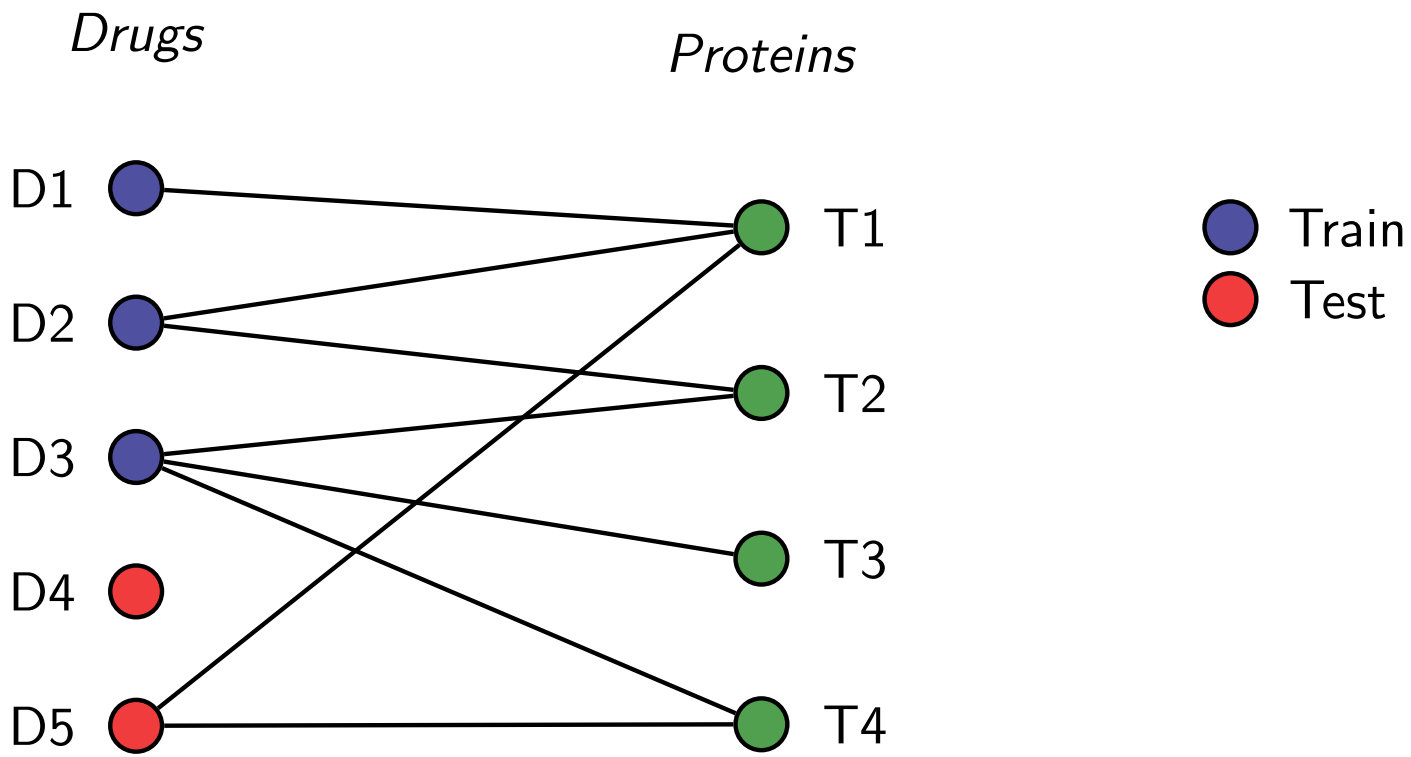
Main issues:

- Lack ability to generalize or are unable to spot small differences
- Usually only top-down or bottom-up
- Not making use of interaction networks

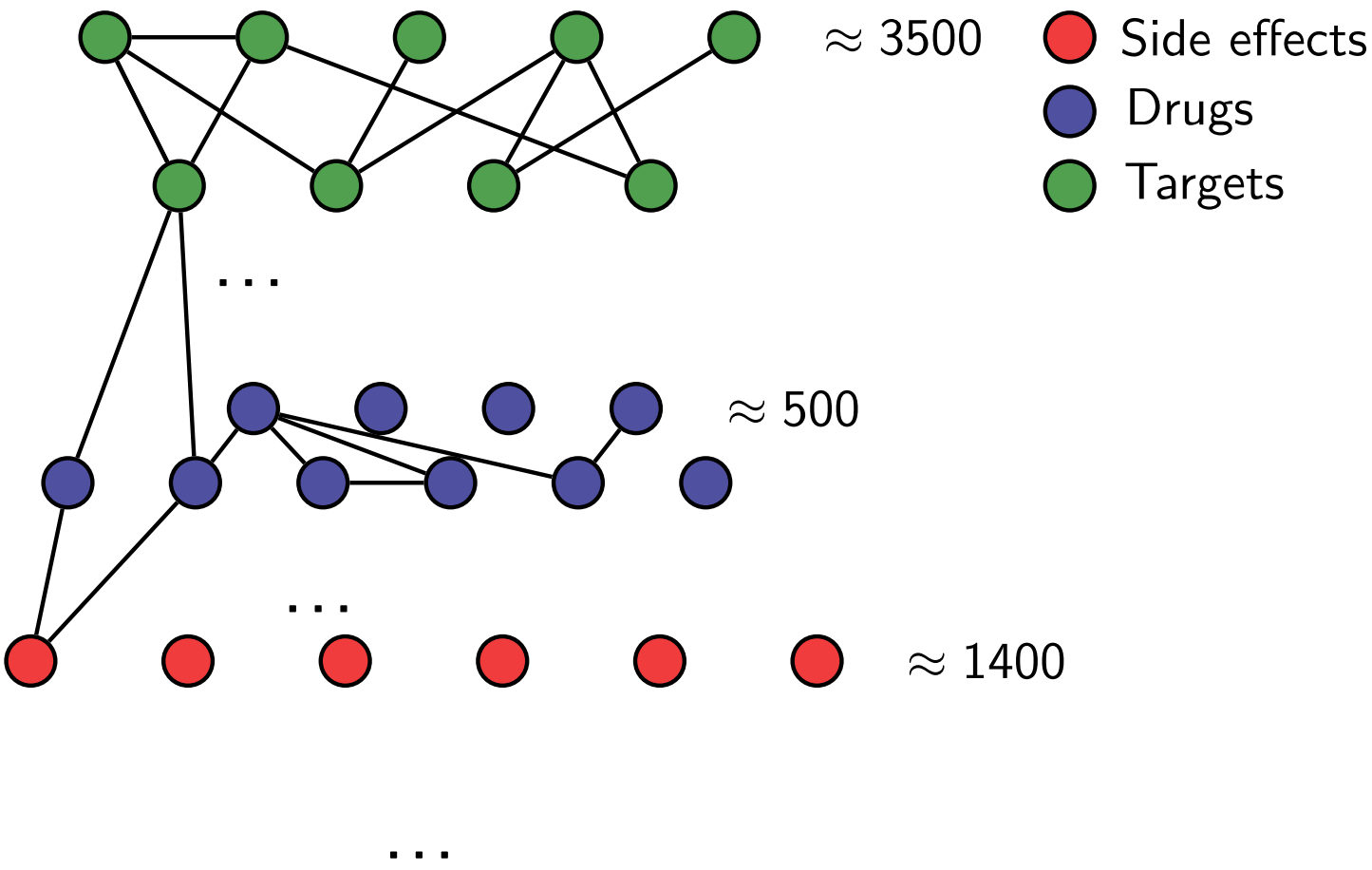
Problems of recent approaches

Main issues:

- Lack ability to generalize or are unable to spot small differences
- Usually only top-down or bottom-up
- Not making use of interaction networks



Features in context



Available data structures

→ Combine Bottom up and top down features for drugs and features
Data structures used:

- PPIs, DTIs
- Motifs along proteins
  - find motif along protein sequences with HMMs for each drug
- DeepGO embeddings
- DDIs, SIDER, MedDRA
- DL2vec
- Seq2seq for drugs

Drugs

- matrix[drug_index]
- semantic similarities over MedDRA
- Seq2seq from Smiles
- DL2vec over SIDER-MedDRA-HPO

Proteins

- DeepGO
- Motifs from HMM
- DL2vec Embeddings from Azza and Jun
- node degree percentile (Gray encoding)
- GNNs (node features):
  - Motifs from HMM
  - DDI induced targets
  - semantic similarity weighted targets
  - DL2vec embeddings from Azza and Jun
  - DL2vec over HPO/MP/MedDRA
  - DL2vec over PhenomeNET (GO,MP/HPO, SIDER)

Architectures

- every possible layer from PyG
- Node features with/without GNN
- cat([drug_feat, prot_feat]) as (node-) feature
- siamese over drugs and proteins/their (graph-)embeddings

Architectures

- every possible layer from PyG
- Node features with/without GNN
- cat([drug_feat, prot_feat]) as (node-) feature
- siamese over drugs and proteins/their (graph-)embeddings

Not included:

- Variations of all datasets