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

Tilman Hinnerichs

BORG - KAUST

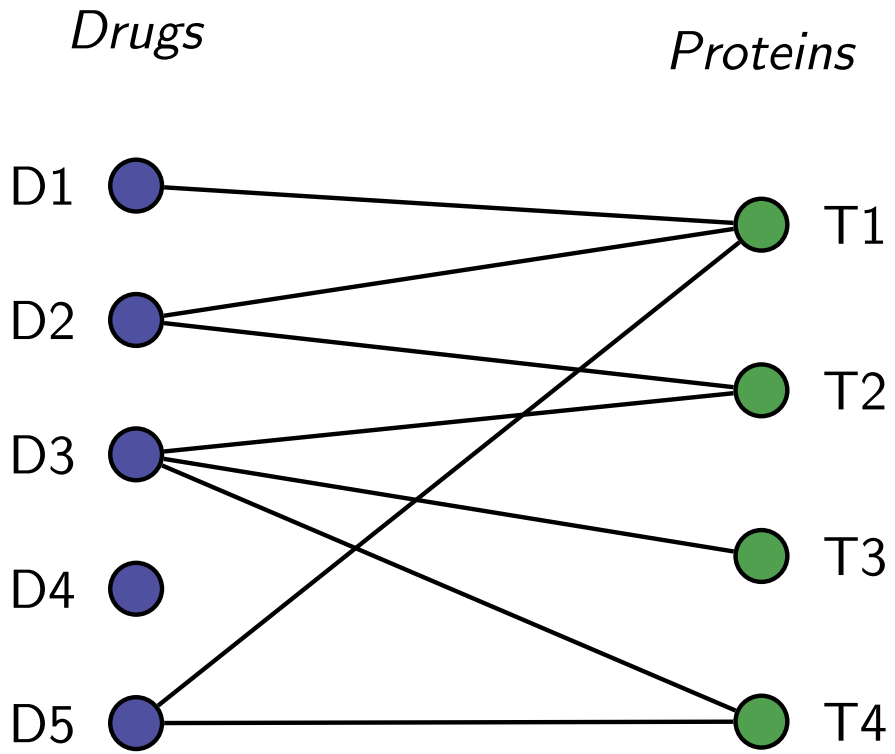
December 09, 2019

Outline

1. Problem Description
2. Recent approaches
3. A medium approach to DTI prediction
   Filtering possible targets

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

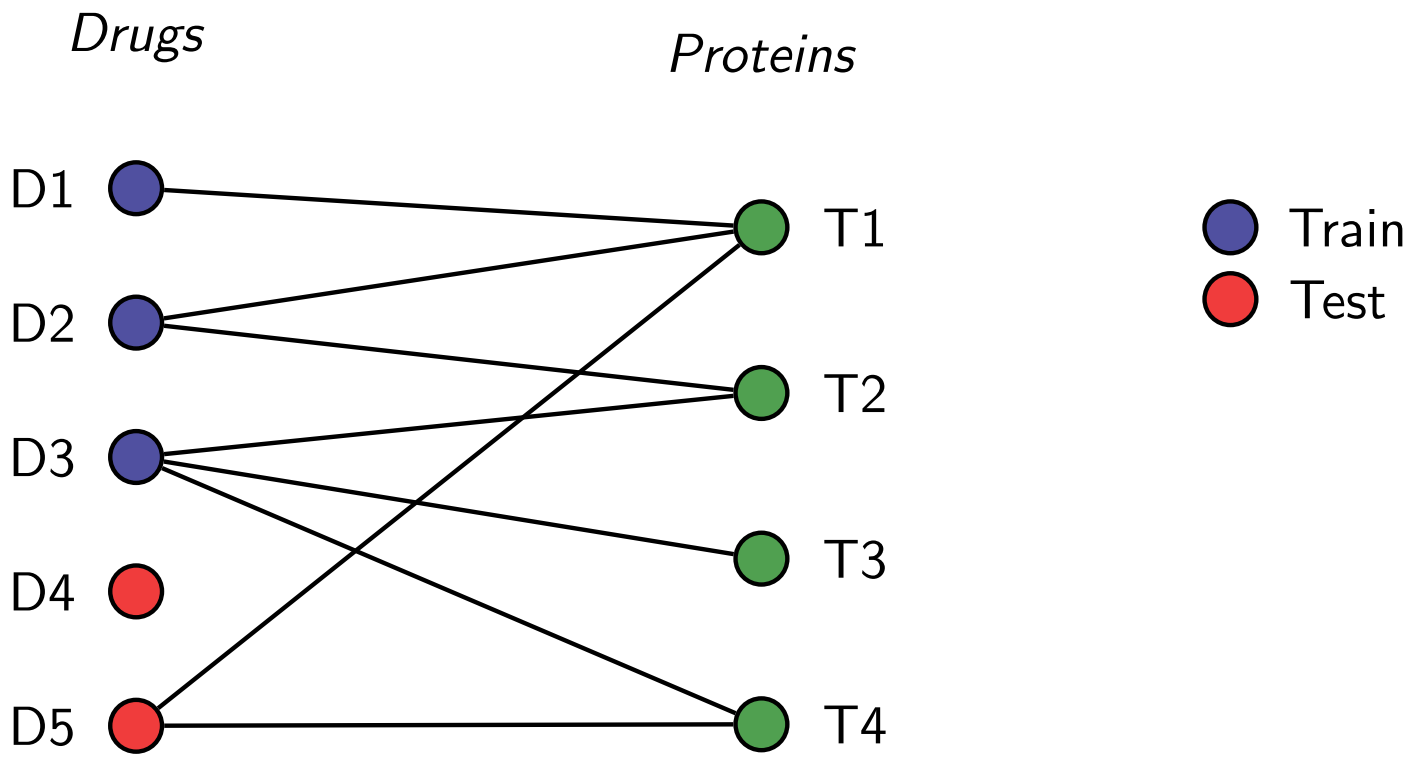
Main issues:

- 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



How to bring both approaches together?

1. For each drug, filter possible targets according to protein based measurement
2. Bring those together with the interaction networks for the prediction

How to build such a filter?

STITCH database got

- 390 000 chemicals, and
- 3.6 million proteins
- from over 2000 organisms

How to build such a filter?

STITCH database got

- 390 000 chemicals, and
- 3.6 million proteins
- from over 2000 organisms

For each drug, find a motif in the target amino acid sequences and query that against all possible proteins and use the results as features for the prediction

How to find motifs?

1. Build multi sequence alignment with alignment tool of choice over non-human proteins
   - FAMSA, MAFFT, Kalign, MSA Props, Muscle, ...
2. Build representation of that alignment
   - Hidden Markov model with HMMER
   - PWM
3. Query representation against database

How to find motifs?

1. Build multi sequence alignment with alignment tool of choice over non-human proteins
   - FAMSA, MAFFT, Kalign, MSA Props, Muscle, ...
2. Build representation of that alignment
   - Hidden Markov model with HMMER
   - PWM
3. Query representation against database

This is very expensive to compute ... (10000 hours for alignments of 614 drugs + IBEX queue times for such big jobs)

Full citations

- Yu Ding, Hong Wang, Hewei Zheng, Lianzong Wang, Guosi Zhang, Jiaxin Yang, Xiaoyan Lu, Yu Bai, Haotian Zhang, Jing Li, Wenyan Gao, Fukun Chen, Shui Hu, Jingqi Wu, Liangde Xu, Evaluation of drug efficacy based on the spatial position comparison of drug–target interaction centers, Briefings in Bioinformatics, , bbz024, https://doi.org/10.1093/bib/bbz024
- Chen Wang, Lukasz Kurgan, Review and comparative assessment of similarity-based methods for prediction of drug–protein interactions in the druggable human proteome, Briefings in Bioinformatics, , bby069, https://doi.org/10.1093/bib/bby069