Deep Drug Repurposer

By: Sarthak Kothari, Noah Demoes, Omair Shafi Ahmed, Forrest Hooton

Motivation

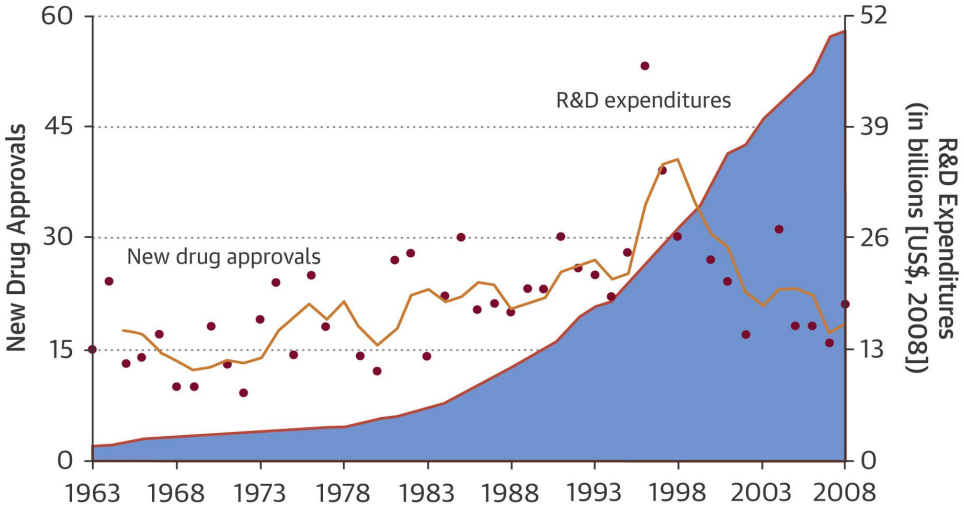
- Americans spent $324.6 Billion on prescription drugs last year
- Drug cost $2.6B and 17 years to develop on average
- Solving a small portion → big returns

The Present and Future

State-of-the-Art Review

Cardiovascular Drug Development: Is it Dead or Just Hibernating?

Christopher B. Fordyce MD, MSc *, Matthew T. Roe MD, MHS *, Tariq Ahmad MD, MPH *, Peter Libby MD †, Jeffrey S. Borer MD ‡, William R. Hiatt MD §, Michael R. Bristow MD, PhD §, Milton Packer MD ‖, Scott M. Wasserman MD ¶, Ned Braunstein MD #, Bertram Pitt MD **, David L. DeMets PhD ††, Katharine Cooper-Arnold MPH ‡‡, Paul W. Armstrong MD §§, Scott D. Berkowitz MD ‖‖, Rob Scott MD ‖, Jayne Prats PhD ¶¶, Zorina S. Galis PhD ‡‡, Norman Stockbridge MD, PhD ###, Eric D. Peterson MD, MPH *, Robert M. Califf MD *



Objectives

First Half

- Curate and join drug-receptor databases together from DrugBank, PubChem to establish drug structure to receptor relationships
- Develop a deep binary classifier model to identify potential drugs to repurpose for other receptors

Second Half

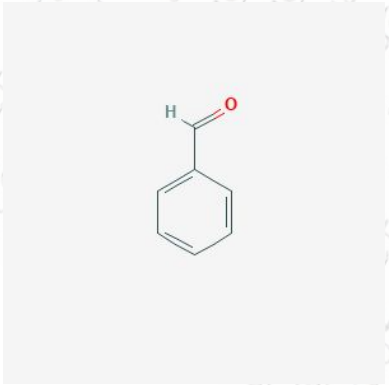
- Deeper validation of model
- Create and deploy a web interface for researchers to use that predicts potential receptors for a given drug

Research Plan: Data

- DrugBank: Drug-target interaction data
  - 6,835 drugs and 4,217 targets
- PubChem: retrieve SMILES (molecular structure in machine readable strings)
- Curated drug reactions
  - Dataset 1: continuous kinase inhibitor reactions with 72 drugs and 422 targets
  - Dataset 2: continuous protein binding scores with 2,116 drugs and 229 targets

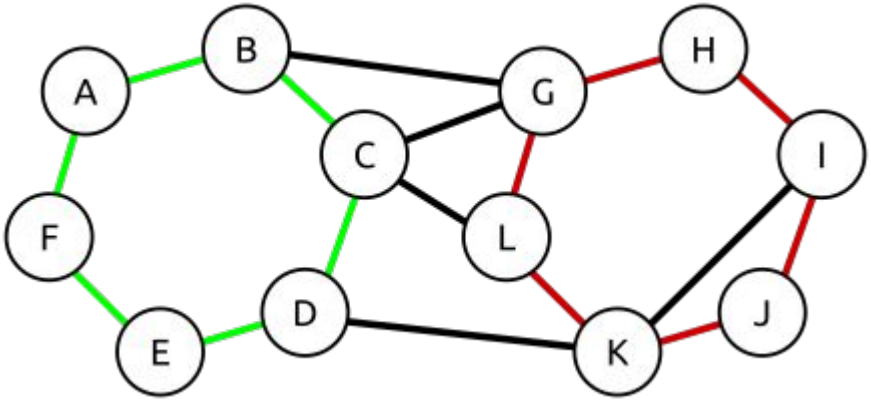
Benzaldehyde SMILE

C1=CC=C(C=C1)C=O



Research Plan: Negative Sampling

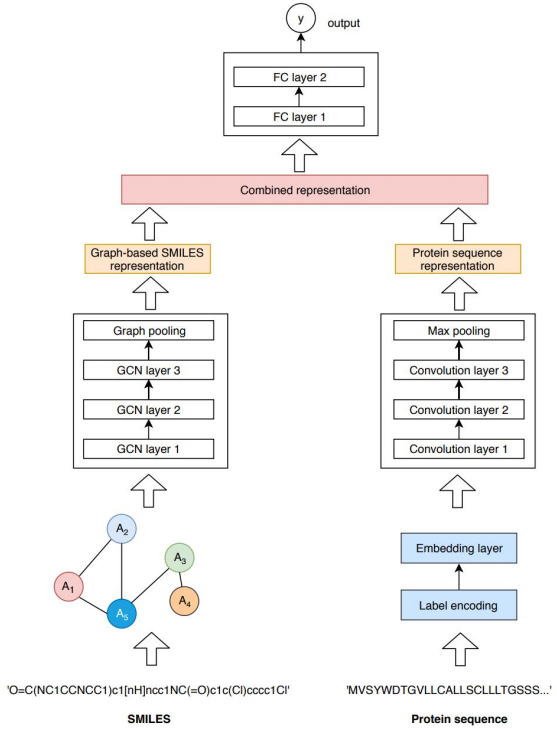
Use randomly sampled drugs with a path distance longer than the longest-shortest path



Research Plan: Model

Compound Structure as Strings

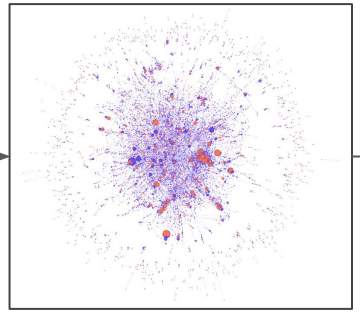
Compound Structure as Graph and String



Connecting the Dots

DrugBank

Drugs



Target/Receptors

SMILES

1 Dimensional Decoding

CNN

Molecular Representation in 3D

Non-Euclidean Modelling

NLP Based Techniques

Preliminary Results

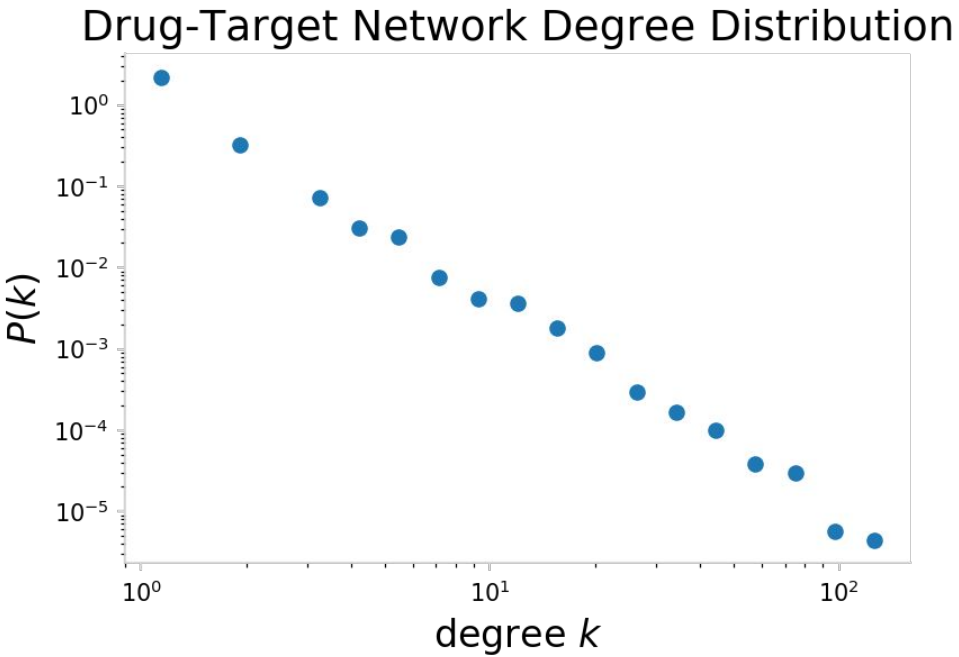
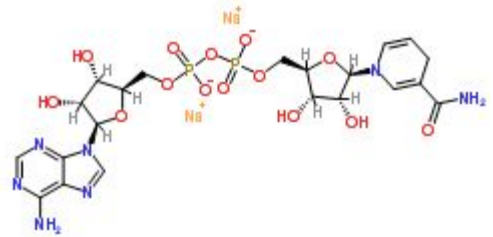


Figure 1: *The degree distribution for drugs the FoodBank drug-target interaction network.*

- Most drugs have fewer than 10 targets
- Drug with most targets is NADH, a supplement taken for improved cognition and dementia



Preliminary Results

- Large central component does not invalidate sampling technique
- Average path size: 7.66
- Network diameter: 22

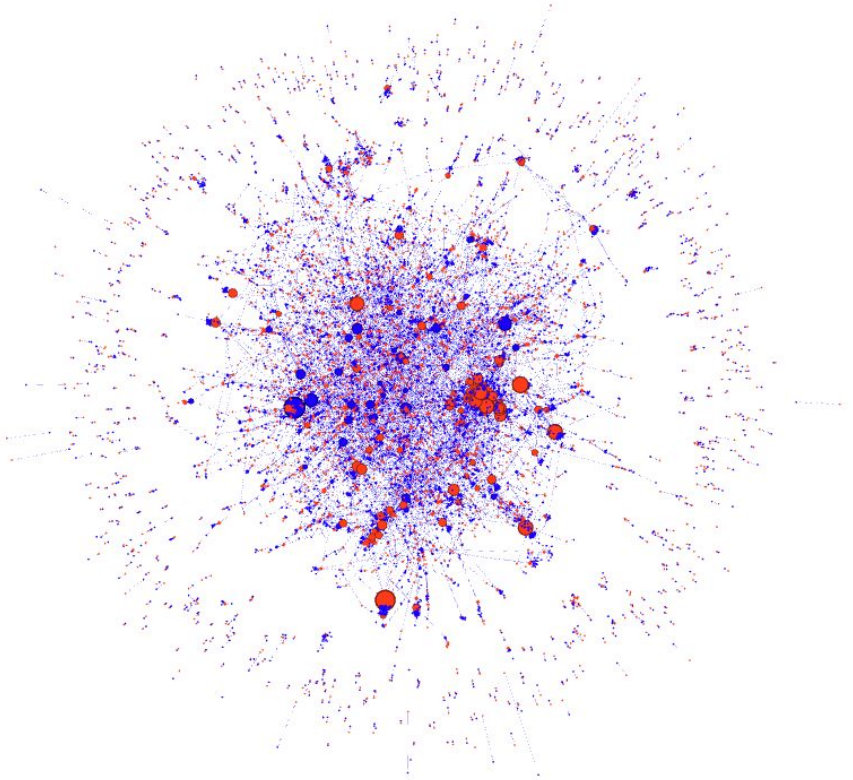


Figure 2: *The network layout of the FoodBank drug-target interaction network visualized using the ForceAtlas 2 algorithm. Red nodes represent targets, blue nodes represent drugs, and the node size corresponds with the respective node degree.*

References

[1] Business Research Company. The Growing Pharmaceuticals Market: Expert Forecasts and Analysis. *Market Research Blog* Available at: https://blog.marketresearch.com/the-growing-pharmaceuticals-market-expert-forecasts-and-analysis.

[2] Cheng, F. *et al.* Network-based approach to prediction and population-based validation of in silico drug repurposing. *Nature Communications* **9,** (2018).

[3] Commissioner, O. of the. Step 3: Clinical Research. *U.S. Food and Drug Administration* Available at: https://www.fda.gov/patients/drug-development-process/step-3-clinical-research.

[4] Davis, M. I. *et al.* Comprehensive analysis of kinase inhibitor selectivity. *Nature Biotechnology* **29,** 1046–1051 (2011).

[5] Dimasi, J. A., Grabowski, H. G. & Hansen, R. W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics* **47,** 20–33 (2016).

[6] Kim, S. *et al.* PubChem 2019 update: improved access to chemical data. *Nucleic Acids Research* **47,** (2018).

[7] Nadh: Uses, Side Effects, Interactions, Dosage, and Warning. *WebMD* Available at: https://www.webmd.com/vitamins/ai/ingredientmono-1016/nadh.

[8] Nguyen & Venkatesh, S. Prediction of drug–target binding affinity using graph neural networks. *bioRxiv* (2019). Available at: http://dx.doi.org/10.1101/684662.

[9] Tang, J. *et al.* Making Sense of Large-Scale Kinase Inhibitor Bioactivity Data Sets: A Comparative and Integrative Analysis. *Journal of Chemical Information and Modeling* **54,** 735–743 (2014).

[10] Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research* **46,** (2017).