Deep Drug Repurposer

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

THE DRUG COST DEBATE AT A GLANCE

$457 BILLION
The amount Americans spent on prescription drugs in 2015, up by about 8 percent over the previous year

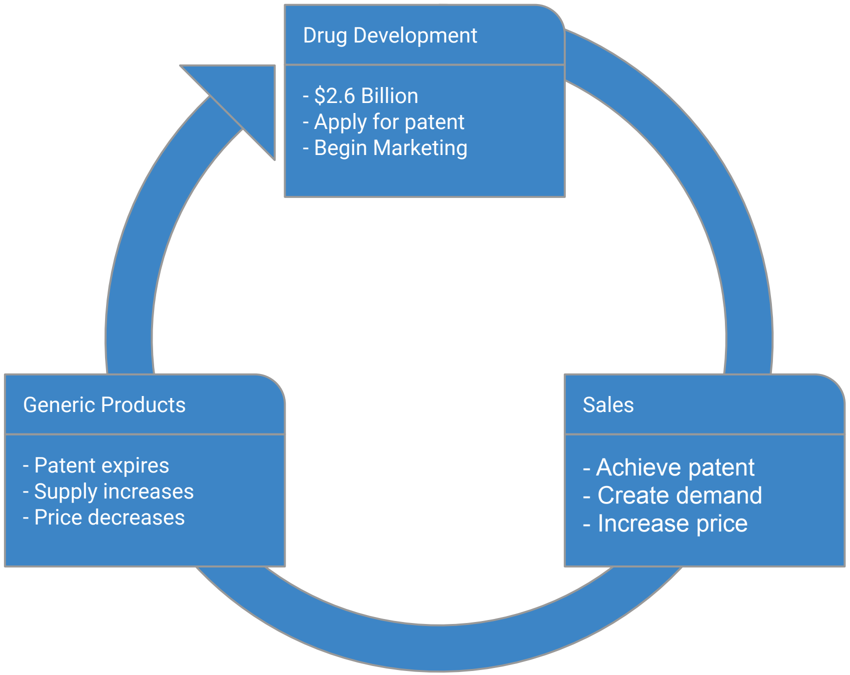
208%
The rise in prices for the most popular brand-name drugs from 2008 to 2016

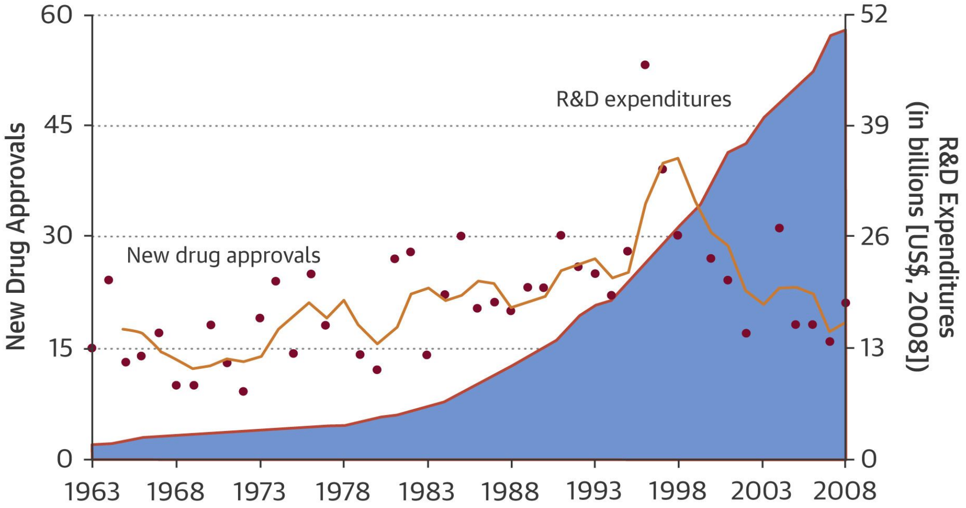
$14.5 MILLION
Median salary of a pharmaceutical firm CEO in 2015, more than any other industry

$6.4 BILLION
Amount drug companies spend advertising directly to consumers in the U.S. annually

$24 BILLION
Amount drug companies spend per year marketing to doctors

Motivation


Drug Development
- $2.6 Billion
- Apply for patent
- Begin Marketing
Sales
- Achieve patent
- Create demand
- Increase price
Generic Products
- Patent expires
- Supply increases
- Price decreases

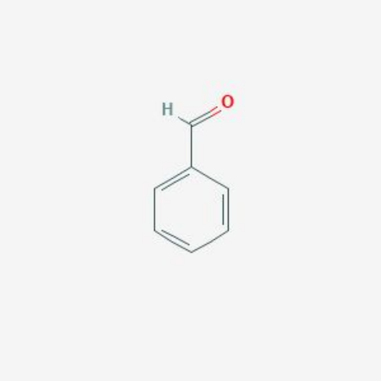

New Drug Approvals
R&D Expenditures (in billions [US$, 2008])
R&D expenditures
New drug approvals
60
45
30
15
0
52
39
26
13
0
1963
1968
1973
1978
1983
1988
1993
1998
2003
2008

Data

- DrugBank: Drug-target interaction data
  - 6,835 drugs and 4,217 targets
- PubChem: retrieve SMILES (molecular structure in machine readable strings)

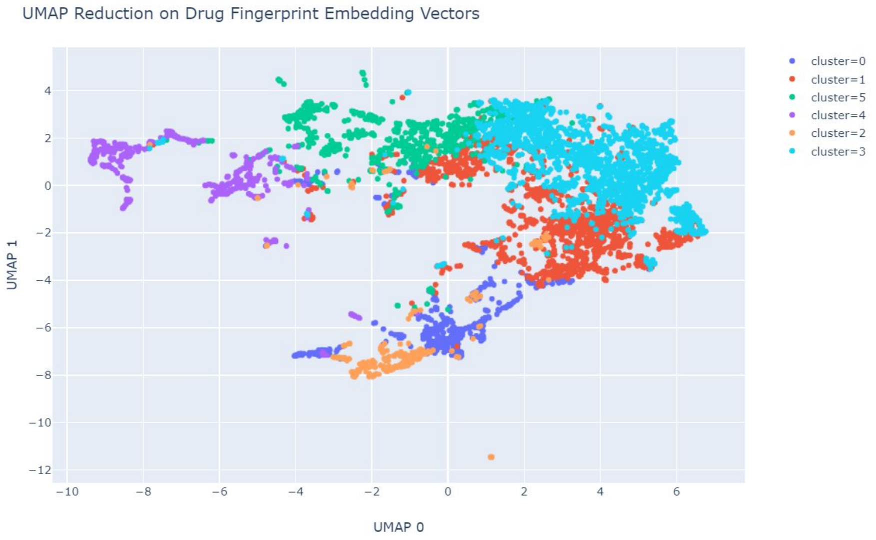
Benzaldehyde SMILE

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

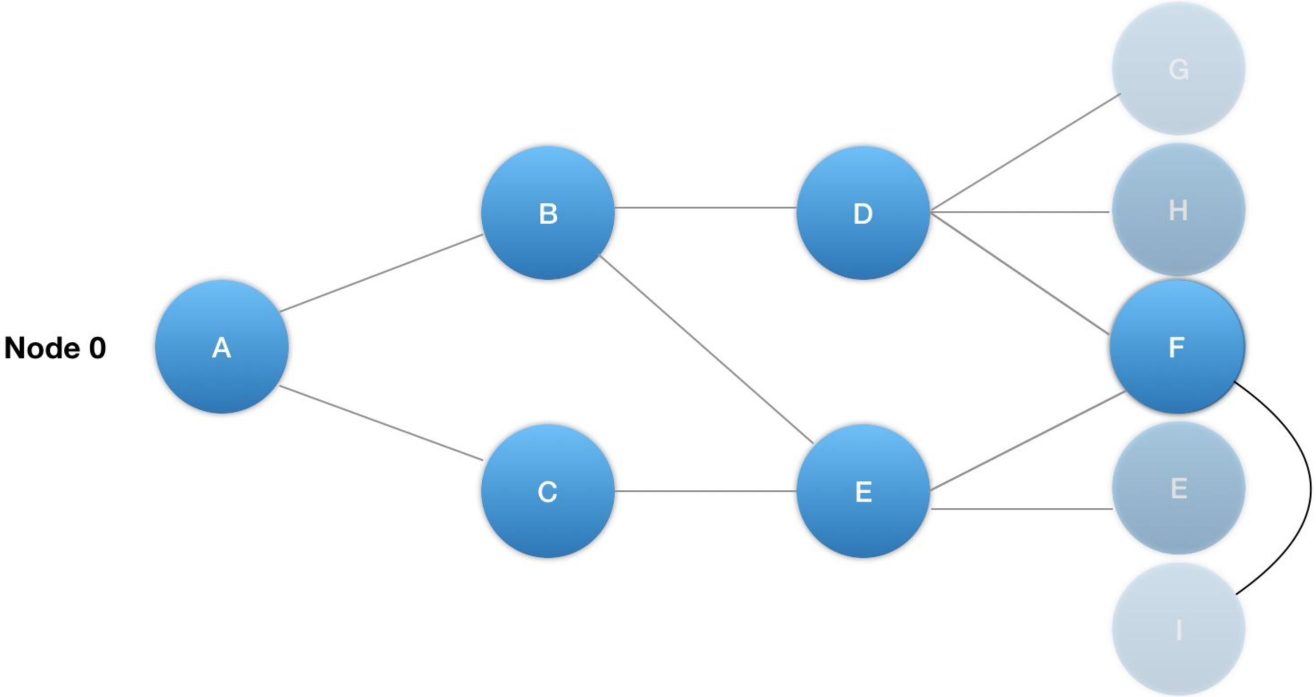


Embeddings for Model Inputs

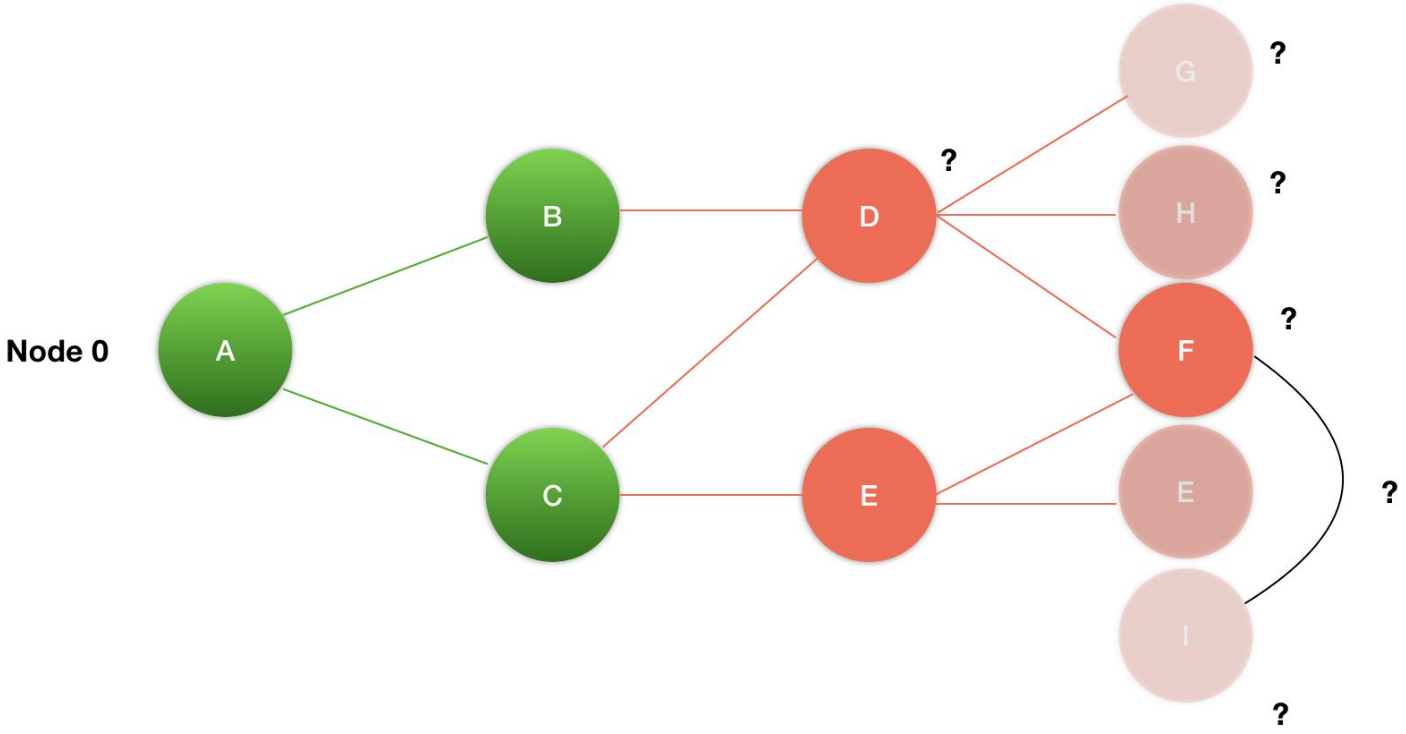
- Word2Vec for drug fingerprints
- Fasttext for target gene sequences



Negative Samples

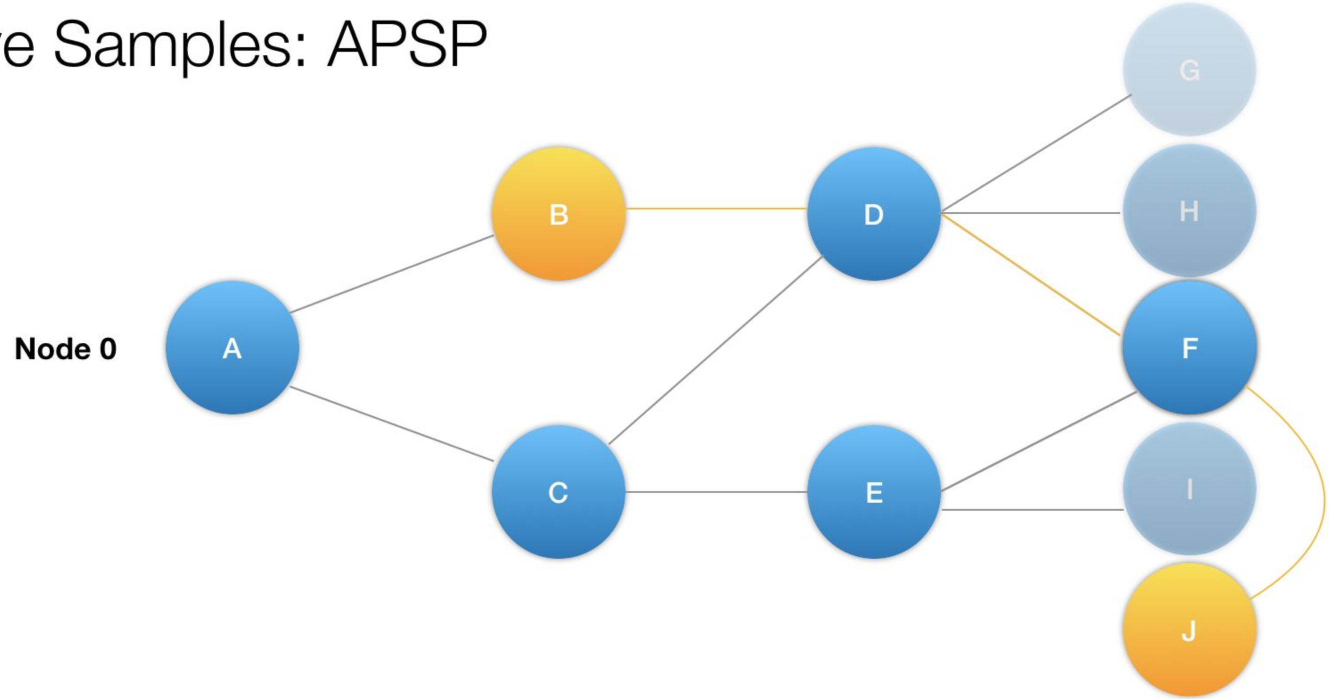


Negative Samples: Problem Statement

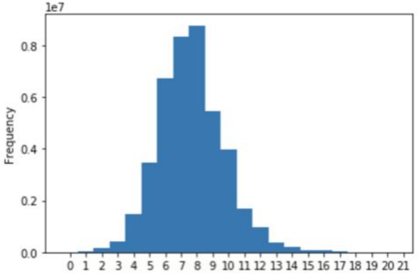


- If we consider all edges, not directly connected, to be negative, we risk overfitting the model
- Hypothesis is that we haven't yet discovered all edges (Drug target interactions)
- Therefore, negative samples must only be edges that we are absolutely certain aren't connected.

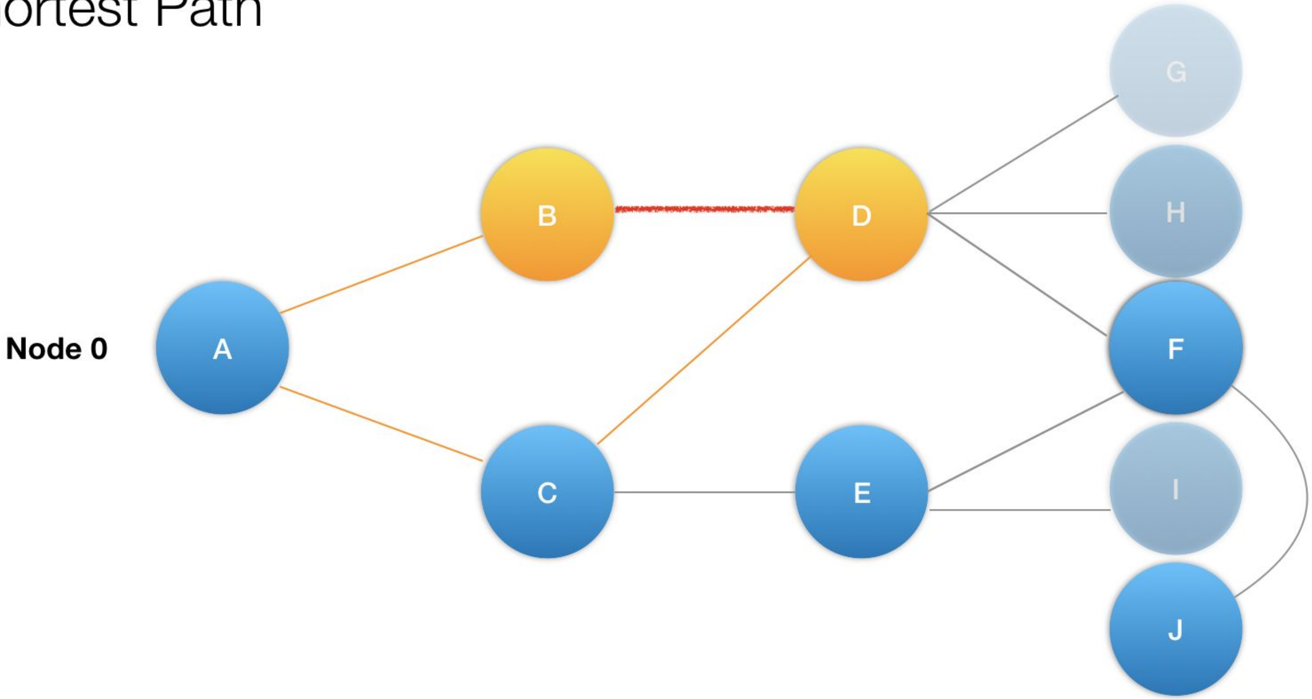
Negative Samples: APSP



Edge	Shortest Path
A - E	2
B - J	3

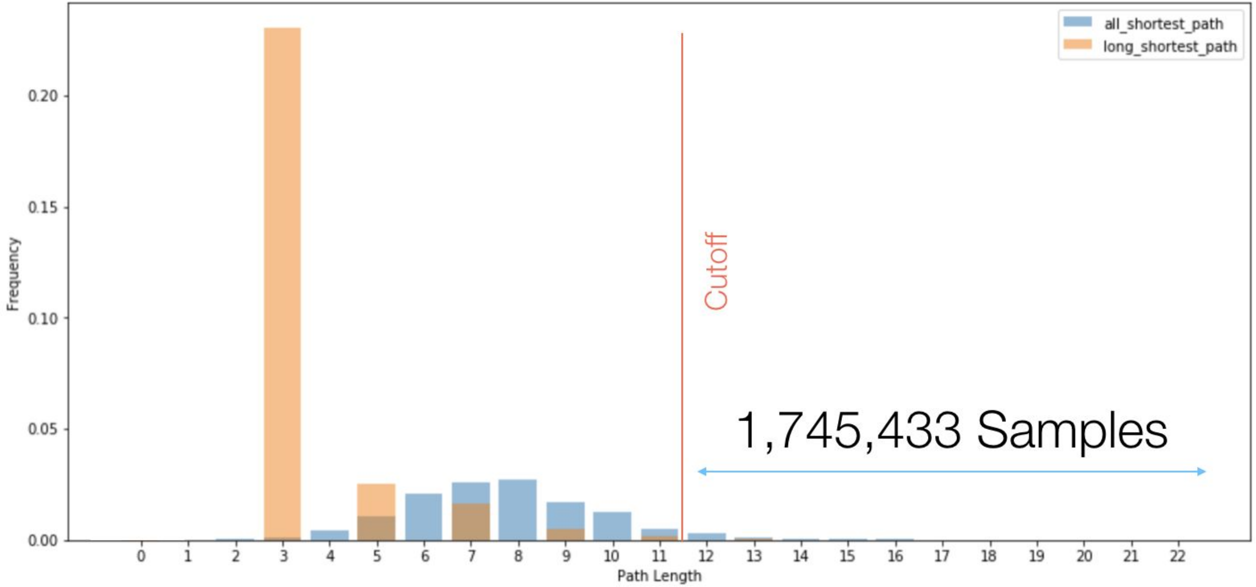


Negative Samples: Indirect Shortest Path



- **Deleting edge B - E, increased its distance to 3**
- **This is an indication of what the path length of a undiscovered edge might look like**

Negative Samples:
Comparing Distributions & Takeaways



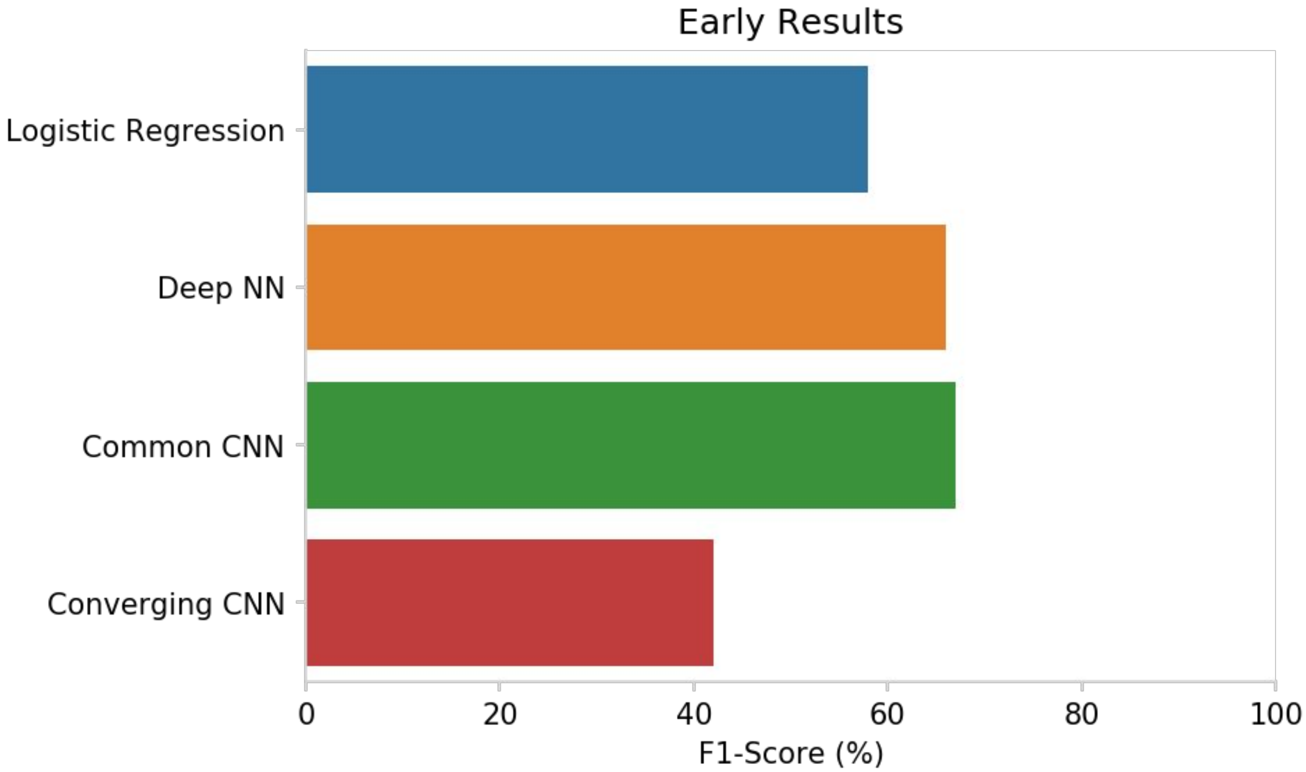
- **Most shortest path have a distance of 3**
- **Most indirect shortest paths have a distance of 8**
- **There are no indirect shortest paths longer than 11**

Therefore, our assumption: **Beyond a distance of 12, there is no evidence of an actual interaction**

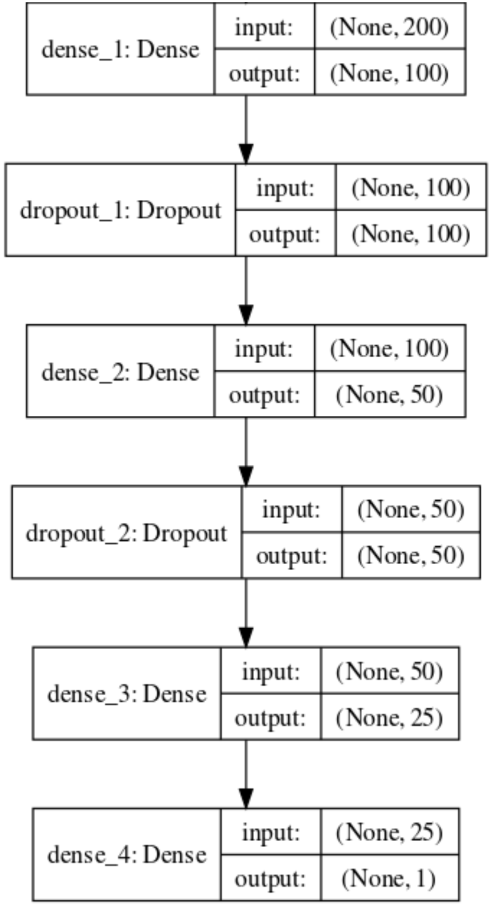
Early Model Results

- Deep Neural Networks : 3 Models that we want to try inspired from WideDTA deep learning model. (https://arxiv.org/abs/1902.04166)
    - Deep Learning Model.
    - Convoluting Drug and Target Together before giving it to Dense Network.
    - Convoluting Drug and Target separate before giving it to Dense Network.

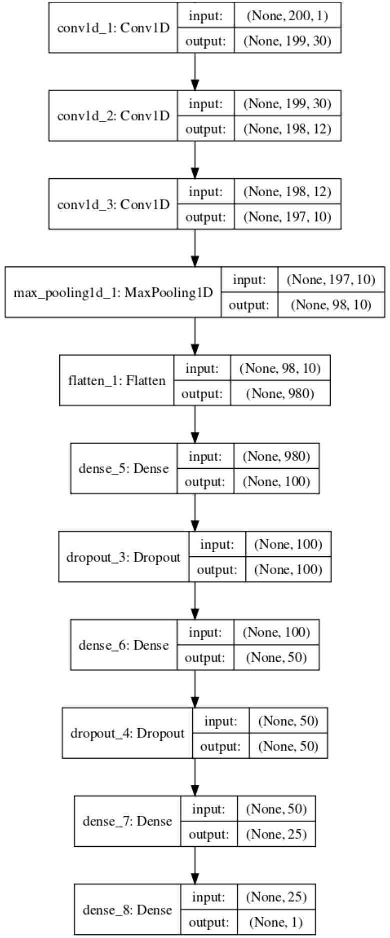
Early Model Results



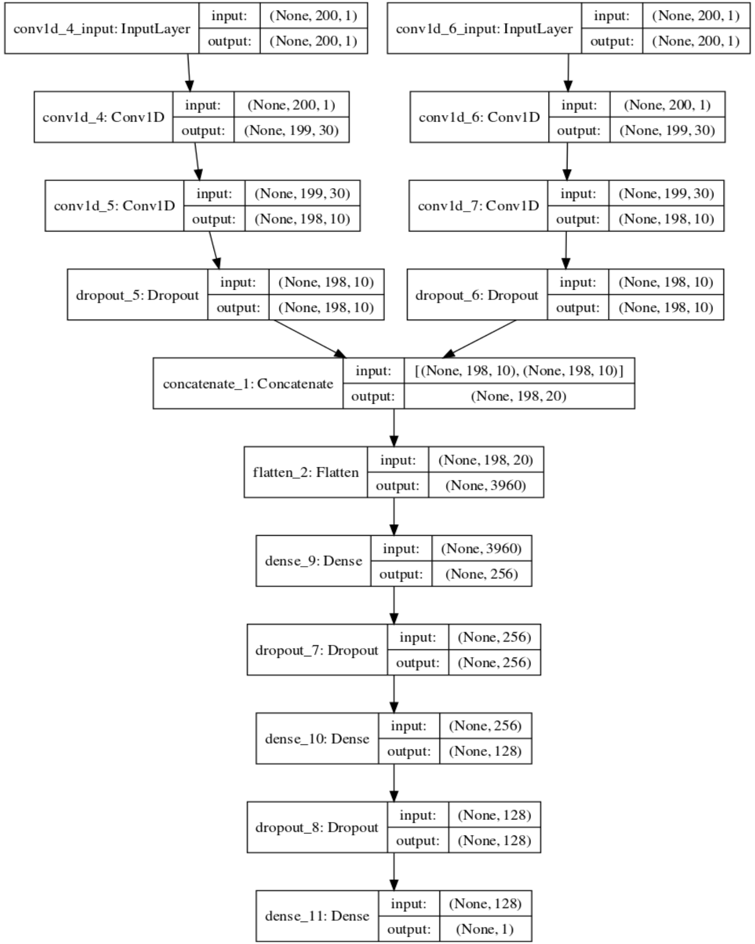
Dense Model Structure



CNN Model Structure

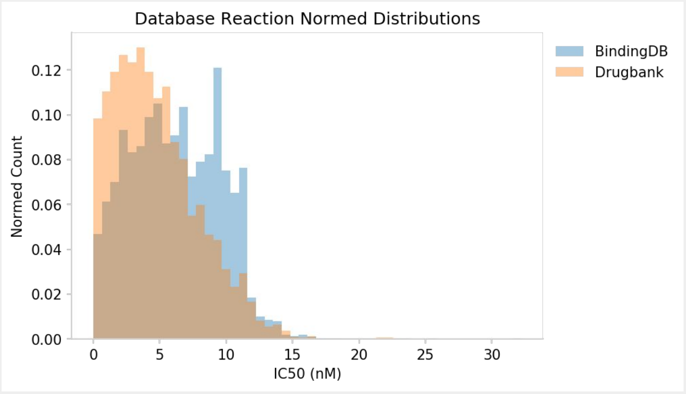


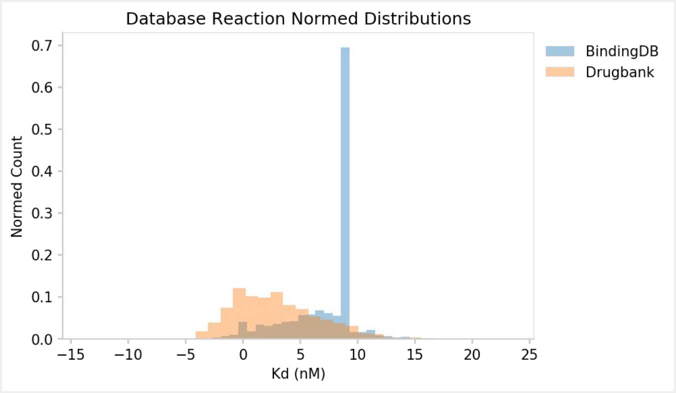
Drug-Target CNN Model Structure

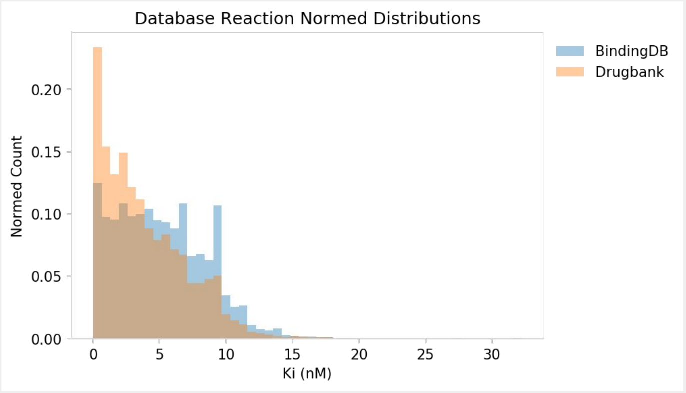


Validation Data

- Different distributions between Drugbank drugs and other chemicals
- 13,336 / 65,472 positive samples if threshold set to 5 log Kd (nM)







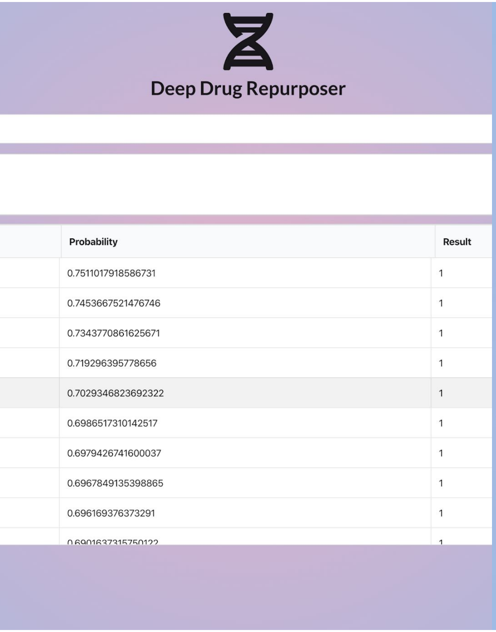
Moving Forward

1. Model Enhancement
2. Model validation on external data
3. Develop Web Portal

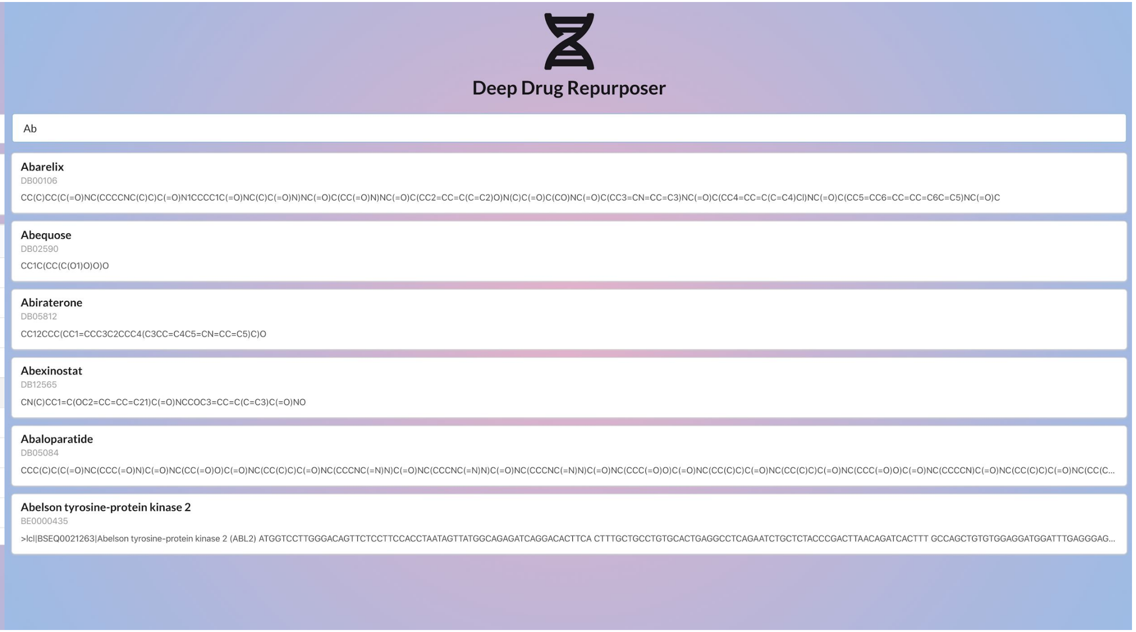
Questions

Citations

https://www.aarp.org/health/drugs-supplements/info-2017/rx-prescription-drug-pricing.html



Deep Drug Repurposer
Probability
Result
0.7511017918586731
1
0.7453667521476746
1
0.7343770861625671
1
0.719296395778656
1
0.7029346823692322
1
0.6986517310142517
1
0.6979426741600037
1
0.6967849135398865
1
0.696169376373291
1
0.6901637315750122
1



Deep Drug Repurposer
Ab
Abarelix
DB00106
CC(C)CC(C(=O)NC(CCCCNC(C)C)C(=O)N1CCCC1C(=O)NC(C)C(=O)N)NC(=O)C(CC(=O)N)NC(=O)C(CC2=CC=C(C=C2)O)N(C)C(=O)C(CO)NC(=O)C(CC3=CN=CC=C3)NC(=O)C(CC4=CC=C(C=C4)Cl)NC(=O)C(CC5=CC6=CC=CC=C6C=C5)NC(=O)C
Abequose
DB02590
CC1C(CC(C(O1)O)O)O
Abiraterone
DB05812
CC12CCC(CC1=CCC3C2CCC4(C3CC=C4C5=CN=CC=C5)C)O
Abexinostat
DB12565
CN(C)CC1=C(OC2=CC=CC=C21)C(=O)NCCOC3=CC=C(C=C3)C(=O)NO
Abaloparatide
DB05084
CCC(C)C(C(=O)NC(CCC(=O)N)C(=O)NC(CC(=O)O)C(=O)NC(CC(C)C)C(=O)NC(CCCNC(=N)N)C(=O)NC(CCCNC(=N)N)C(=O)NC(CCCNC(=N)N)C(=O)NC(CCC(=O)O)C(=O)NC(CC(C)C)C(=O)NC(CC(C)C)C(=O)NC(CCC(=O)O)C(=O)NC(CCCCN)C(=O)NC(CC(C)C)C(=O)NC(CC(C...
Abelson tyrosine-protein kinase 2
BE0000435
>lcl|BSEQ0021263|Abelson tyrosine-protein kinase 2 (ABL2) ATGGTCCTTGGGACAGTTCTCCTTCCACCTAATAGTTATGGCAGAGATCAGGACACTTCA CTTTGCTGCCTGTGCACTGAGGCCTCAGAATCTGCTCTACCCGACTTAACAGATCACTTT GCCAGCTGTGTGGAGGATGGATTTGAGGGAG...