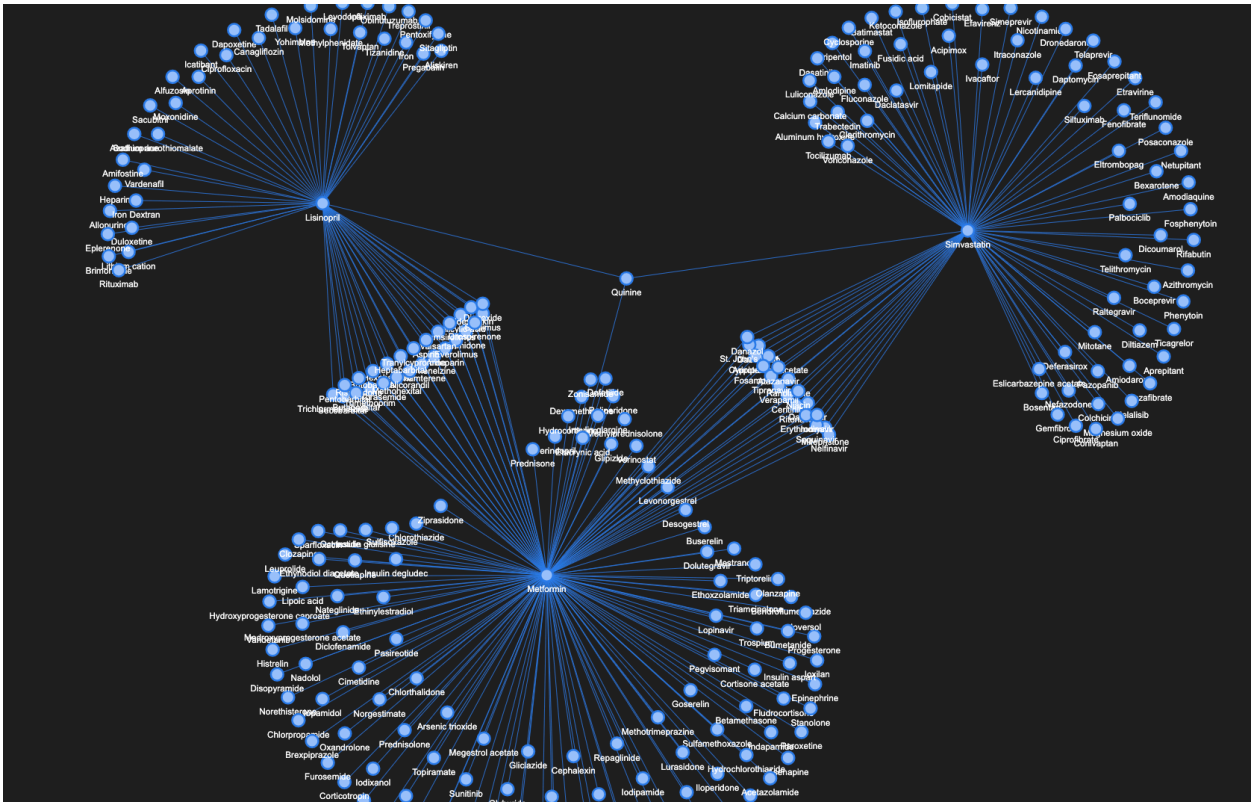
INFSCI 2415 Final REPORT

Sonish Sivarajkumar(sos86@pitt.edu)

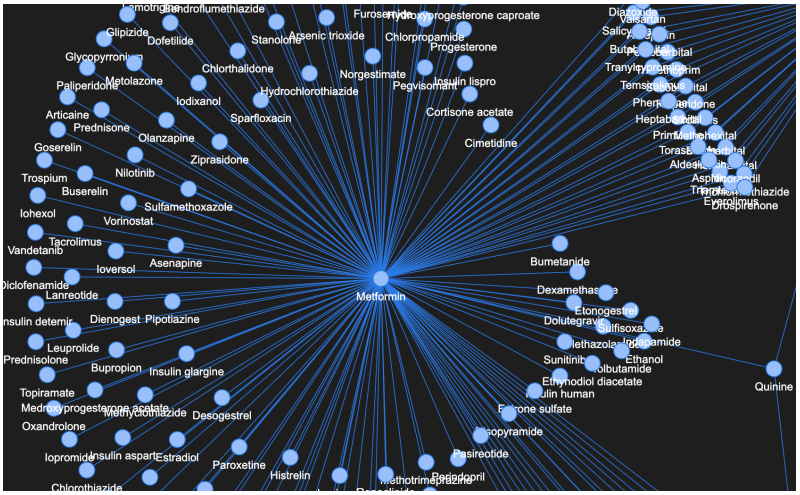
Title: Drug-Drug Interaction Analysis for accelerating Drug Discovery

Github: https://github.com/sonishsivarajkumar/InfViz_project

Main figure:



This is one of the main figures in the project used to discover how to explore and visualize drug interactions data. Legend is out of scope here, because all the nodes represent the drug entities and edges represent the drug-drug interactions. I tried using different colours for different nodes, but the graph is too big that the color get repeated. The repeated color can cause misinterpretation to the users that these nodes have common characteristics, hence I avoided using different colors. But all nodes are labeled here, but the labels are not visible as it is a zoomed-out version, which was done to accommodate the entire graph on this page.A zoomed-in portion of the image is shown below:



Data and Methods

drug_1_name	drug_2_name
Vardenafil	Telmisartan
Clonidine	Pentoxifylline
Clomipramine	Mirabegron
Desipramine	Perampanel
Levodopa	Hydralazine

This project utilized the publicly available ChCh-Miner dataset from the Stanford Network Analysis Project (SNAP). A network of 48,464 medications approved by the Food and Drug Administration (FDA) are included in this dataset. The values are represented by DrugBank (DB) codes, which must be matched with the names of the actual drugs. This is carried out manually by repeatedly searching the public DrugBank database for the corresponding drug names. The first 5 rows of the preprocessed dataset are shown on the right. I performed a network analysis on the dataset to find the interaction between the drugs.

Dataset link: https://snap.stanford.edu/biodata/

Significance:

Drug-Drug Interaction (DDI) prediction is one of the most critical issues in drug development and healthcare. In the field of network biology and network medicine, there is a particular interest in predicting results from drug–drug interactions to advance the speed of drug discovery. Existing data and modern computational methods allow to identify potentially beneficial and harmful interactions, and therefore, narrow drug trials ahead of actual clinical trials. Such automated data-driven investigation relies on machine learning techniques. However, traditional machine learning approaches require extensive preprocessing of the data that makes them impractical for large datasets.

The simple network visualization I used in this study can aid any clinical or drug expert, who may not be a computer scientist, to understand the drug-drug interactions easily. This allows them to accelerate the drug discovery process, without the external aid from a machine learning expert.

Findings – Larger Graph:

```
Name: Drug Interactions Network
Type: Graph
Number of nodes: 1505
Number of edges: 48224
Average degree:  64.0850
Network density: 0.042609740581041916
```

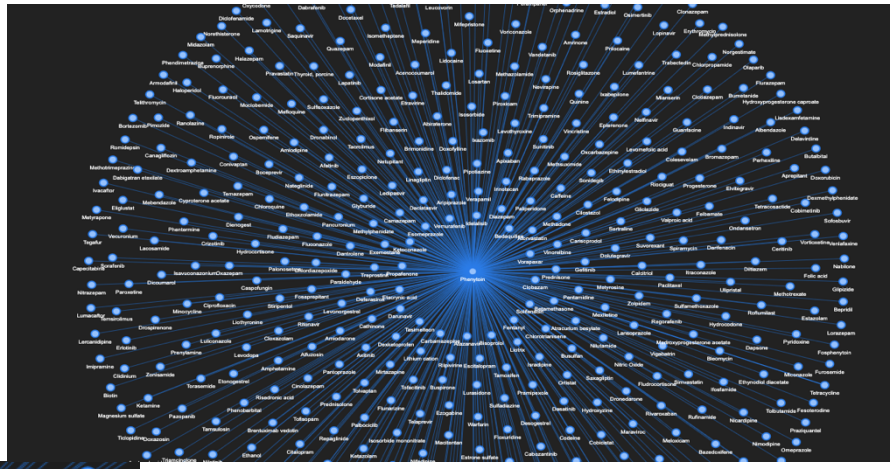
Here nodes represent the drug entities and edges represent the drug interactions.

```
Top 20 drugs by degree:
('Phenytoin', 442)
('Mifepristone', 377)
('Paroxetine', 368)
('Tranylcypromine', 358)
('Phenelzine', 342)
('Warfarin', 334)
('Deferasirox', 329)
('Nelfinavir', 321)
('Dronabinol', 312)
('Hydrocodone', 311)
('Clozapine', 311)
('Primidone', 307)
('Ritonavir', 306)
('Aripiprazole', 304)
('Fosphenytoin', 300)
('Saquinavir', 293)
('Acenocoumarol', 292)
('Carbamazepine', 290)
('Citalopram', 288)
('Dabrafenib', 285)
```

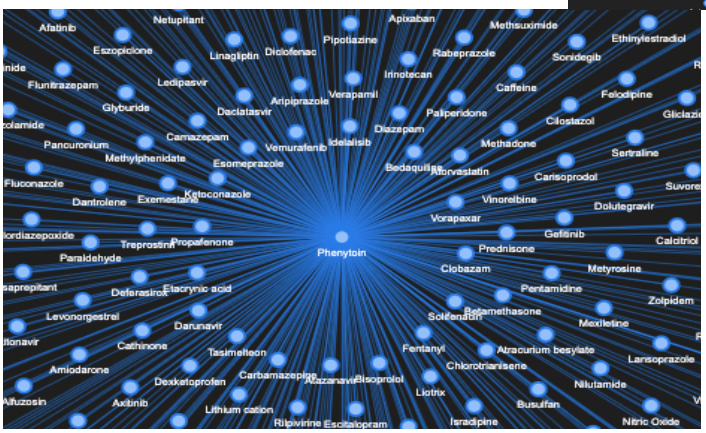
- There are a total of 1,505 drug entities and 48,224 interactions in the network.
- The average degree is 64, which indicates that each medicine interacts with 64 other drugs on an average.
- The low network density of 0.04 means that fewer drugs tend to interact with each other.
- By calculating the degree centrality of each node, we can get the top 20 drugs most frequently involved in drug interactions, as shown on the right.
- The output above reveals that the drug most frequently involved in drug interactions is **Phenytoin,** with a node degree of **442**. Phenytoin is a commonly used anti-epileptic medication for the treatment of seizure disorders. Its concentration in the blood must be kept carefully within consistent levels to ensure its effectiveness and safety

Subset data and Methods Part – 2

To make the graphs more meaningful and interpretable, working on a smaller subset of data would be better than the full dataset.



Exploring single drug: **Phenytoin =>**



=> cropped version for better visualization

The above visualization presents Phenytoin as the **central** node, surrounded by **442** other drug entities that it interacts with.

Findings – Smaller dataset:

- Exploring multiple drugs - I picked **three** drugs (Metformin, Lisinopril**,** and Simvastatin) for further exploration based on their key roles in the prevalent diseases of **diabetes**, **hypertension**, and **hyperlipidemia** respectively.
- I found that Quinine is the common interaction between the three drugs for the above diseases.
- Thus my hypothesis is that Quinine could replace the above three drugs in case of prevalence of the diseases- **diabetes**, **hypertension**, and **hyperlipidemia.** If that is clinically proven wrong by the domain experts, we can conclude that there is an overlap in the functions of the three drugs.