Identifying Potential Drug-Drug Interactions

Presented by Amy Zhou

Impact

- 4th leading cause of death (estimated)
- 23% of U.S. take 3+ prescription drugs

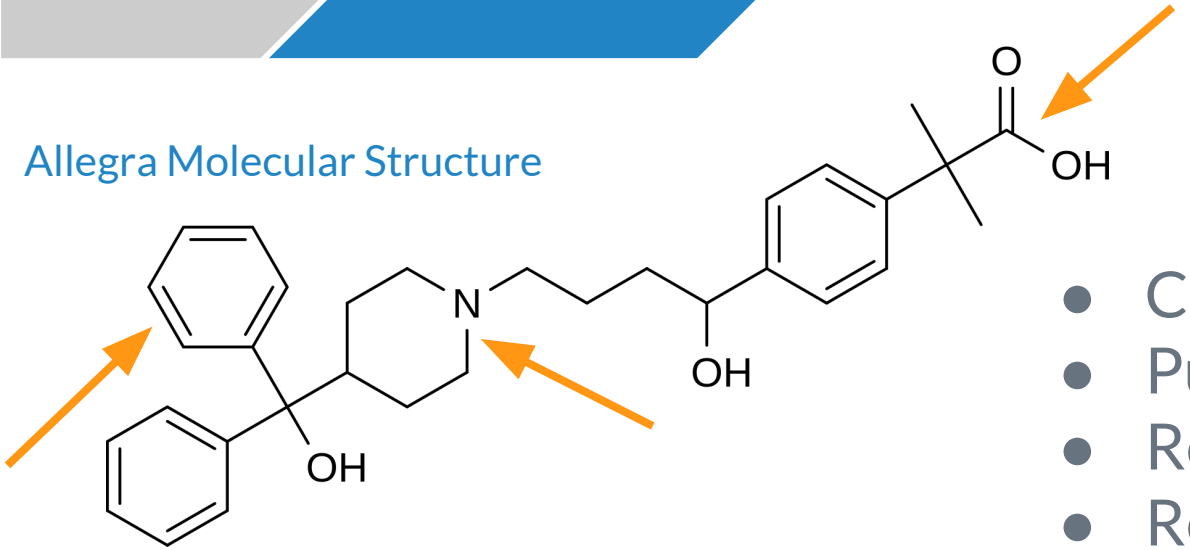
Goal: discover potential drug-drug interactions (DDIs) based on information about similar drugs

Data

- 5,806 unique drug IDs
- 581,055 documented interactions

Data **Features**

Allegra Molecular Structure



- Composition
- Purpose
- Receptors
- Receptor interaction

~6500 → 183 categorical features
via feature engineering, PCA

Pandas

scikit learn

Data > Features > **Cluster**

- Similar drugs based on features
- HDBSCAN
- 631 clusters - mostly <25 drugs each
- 93% of drugs clustered

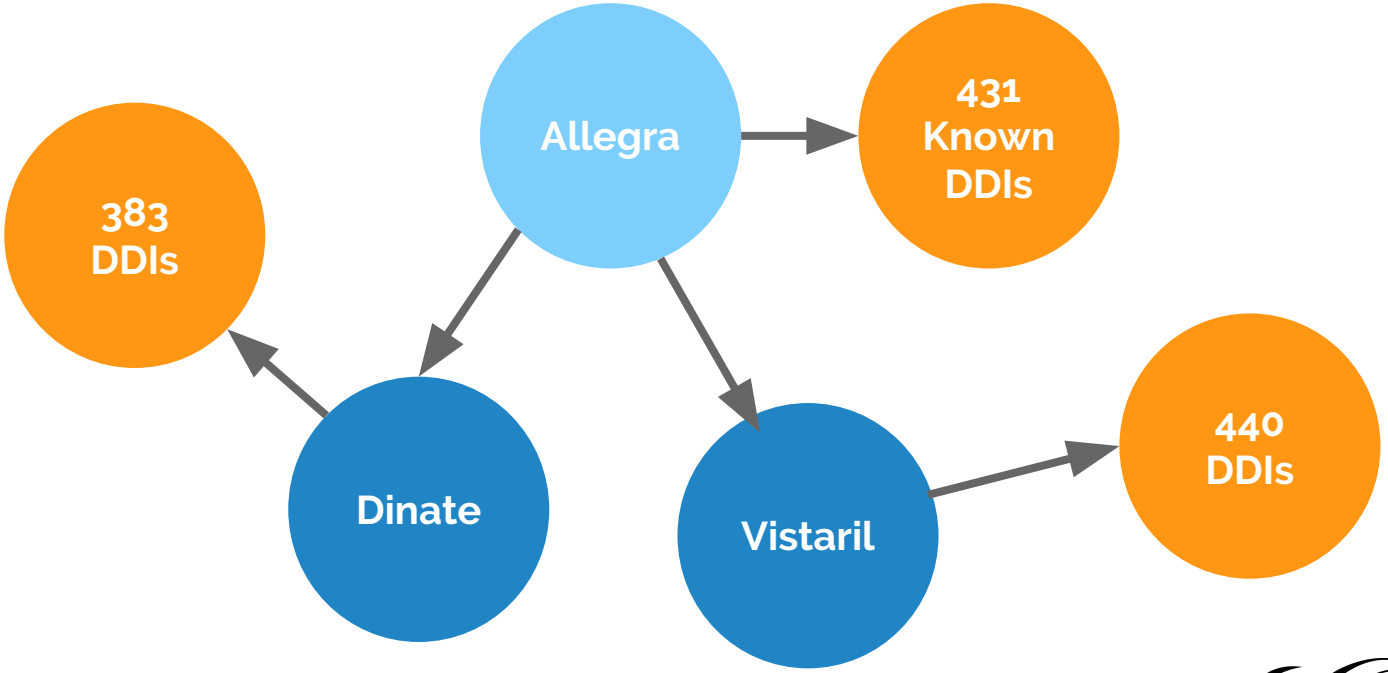
Drug Name	Use	Jaccard Score
Vistaril	Antihistamine, anesthetic	1.000
Dinate	Anti-nausea	1.000
Bromodiphen-hydramine	Antihistamine	1.000
Ulone	Cough suppressant	1.000
Xyzal	Antihistamine	1.000
Ahist	Antihistamine	1.000
Zyrtec	Antihistamine	0.071
Wal-hist	Antihistamine	0.027

- Within cluster
- Jaccard similarity

$$J(A,B) = \frac{|A \cap B|}{|A \cap B|}$$

PostgreSQL

Data > Features > Cluster > Rank > **Identify**



Gephi

Future Considerations

- ➔ Text analysis of severity
- ➔ Cross reference reported ADRs
- ➔ Identify next nearest cluster(s)

Beyond DDIs

- Clustering

NETFLIX

- Network analysis

FedEx

- Clustering + network analysis

Thanks!

Amy Zhou
zhouza@gmail.com
linkedin.com/zhouzamy/
github.com/zhouza