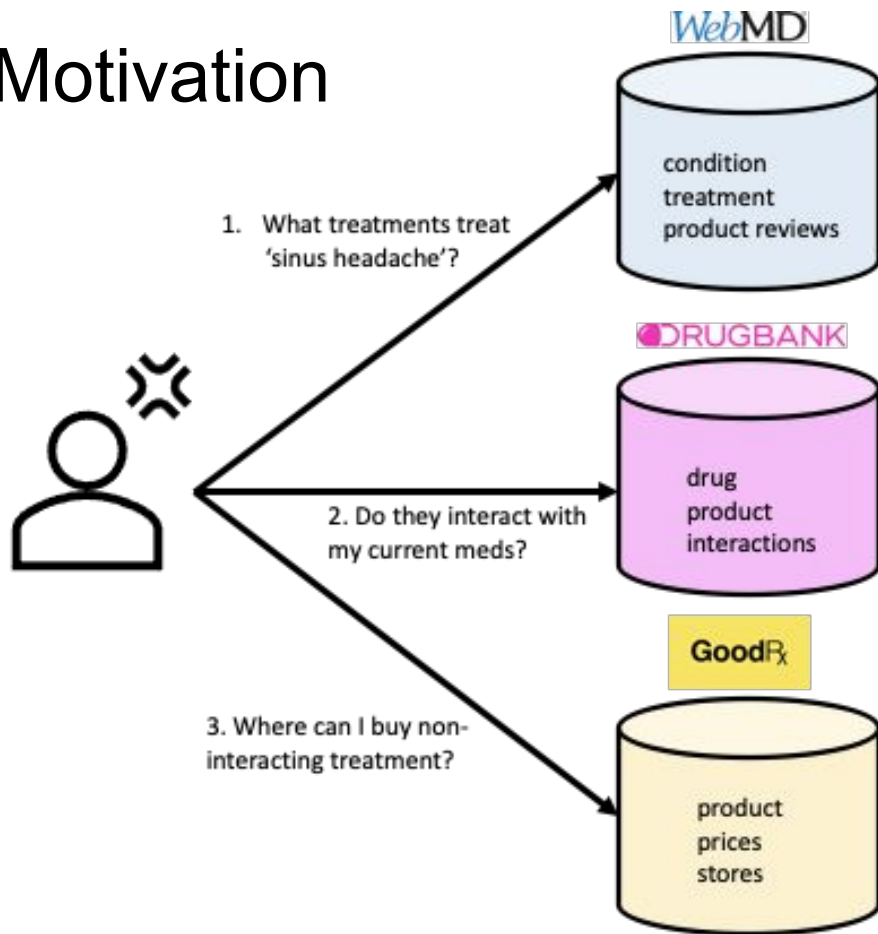# Open Drug Knowledge Graph

Knowledge Graph of Treatments, Prices, and Interactions

Mark Mann, Filip Ilievski, Mohammad Rostami, Aastha, Basel Shbita
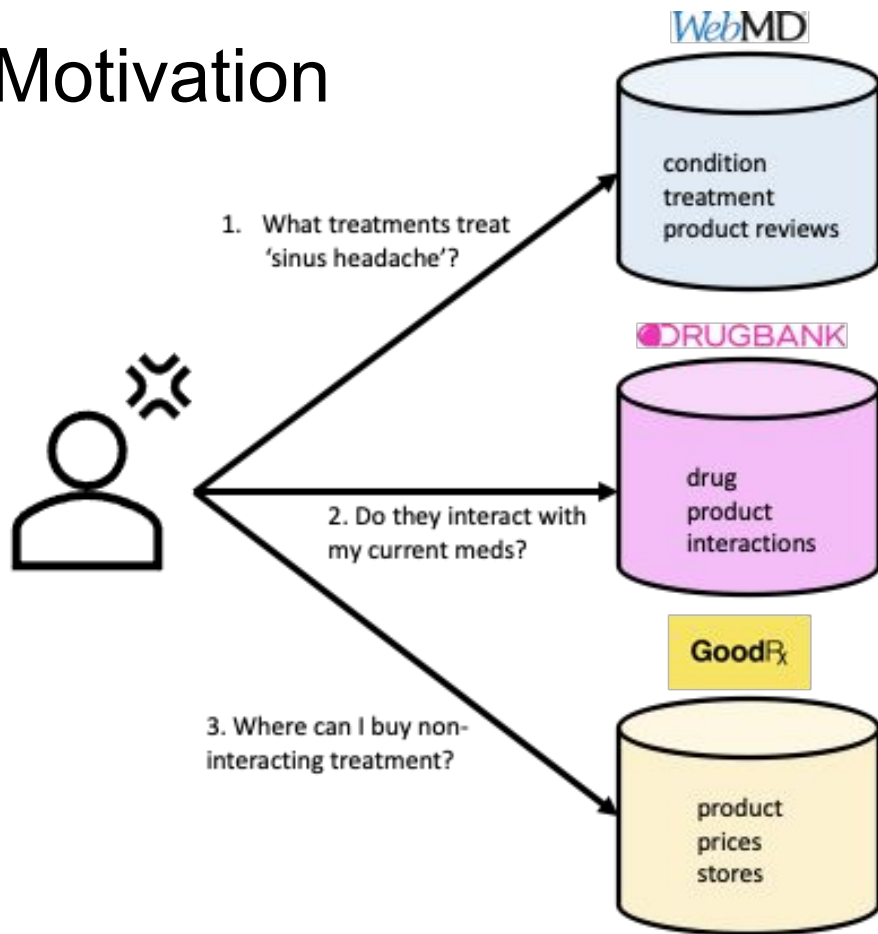
Motivation

1. What treatments treat 'sinus headache'?
2. Do they interact with my current meds?
3. Where can I buy non-interacting treatment?

PROBLEM

WebMD — condition, treatment, product reviews

DRUGBANK — drug, product, interactions

GoodRx — product, prices, stores

USER STORY: "I want to search for drugs that treat my condition. But I'm only interested in drugs within my price range, that don't interact with my current medication."
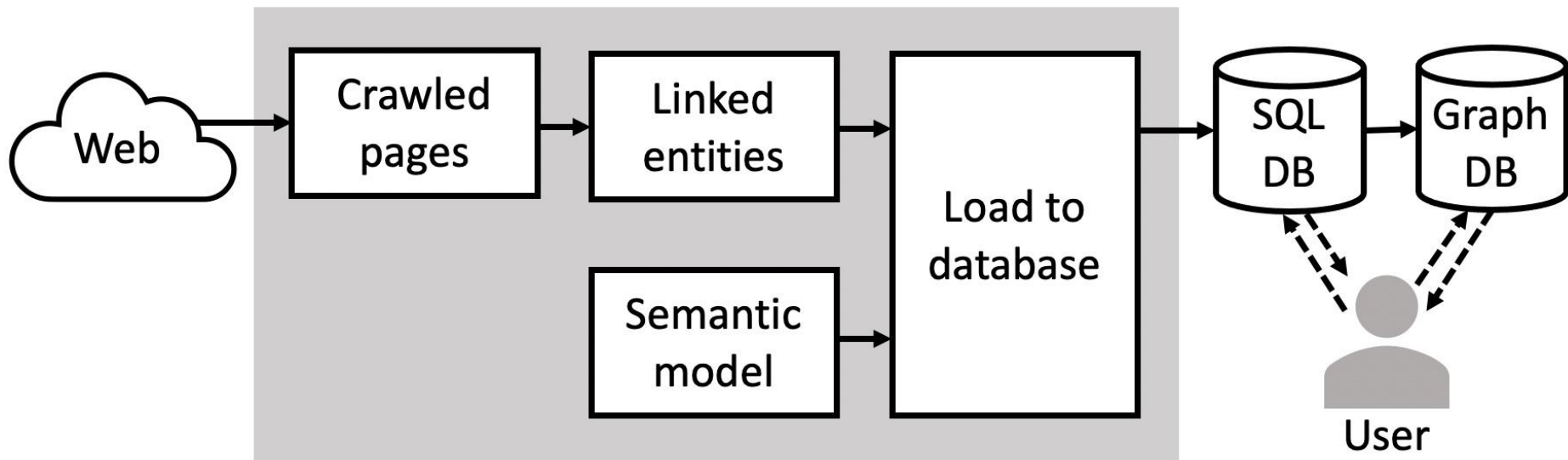
Motivation

1. What treatments treat 'sinus headache'?

WebMD
condition
treatment
product reviews

2. Do they interact with my current meds?

DRUGBANK
drug
product
interactions

3. Where can I buy non-interacting treatment?

GoodRx
product
prices
stores

PROBLEM

USER STORY: "I want to search for drugs that treat my condition. But I'm only interested in drugs within my price range, that don't interact with my current medication."

All (3) questions answered in one view

Intelligent Drug Shopper

WebMD
DRUGBANK
GoodRx
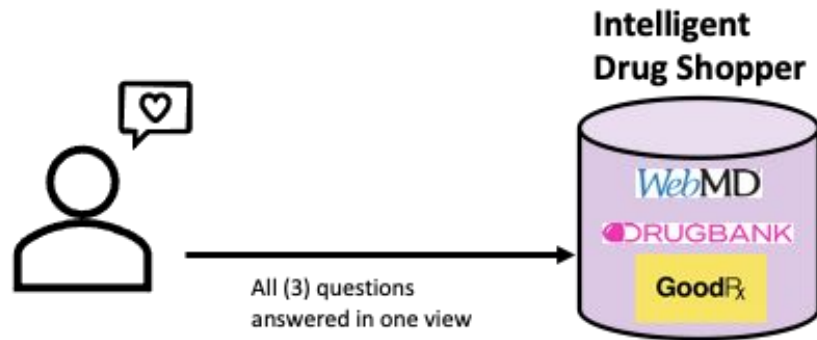
SOLUTION

KG Construction Pipeline

Data Acquisition

Source	Extraction Method	Record Volumes	features
WIKIDATA	SPARQL Endpoint	drug = 787 product = 1560 treatment = 9709	Drug(name, drugbank_id, wiki_url*) Product(drug_id*, name*) Condition(name*) Treatment(drug_id*, condition_id*)
WebMD	Scraper	product = 12857 treatment = 977	Product(drug_id, name, rx_type*, n_reviews*) Condition(name) Treatment(drug_id, condition_id)
DRUGBANK	Data Dump	product = 25637 interactions = 356254 manufacturer = 4092	Drug(name, drugbank_id*, drugbank_url*) Manufacturer(name*) Product(drug_id, name, manf_id*) Interaction(source_drug*, target_drug*)
GoodRx	Scraper	product = 974 price = 20688 store = 23	Store(name*) Price(product, store*, type*, price*, goodrx_link*)

*unique feature across all sources
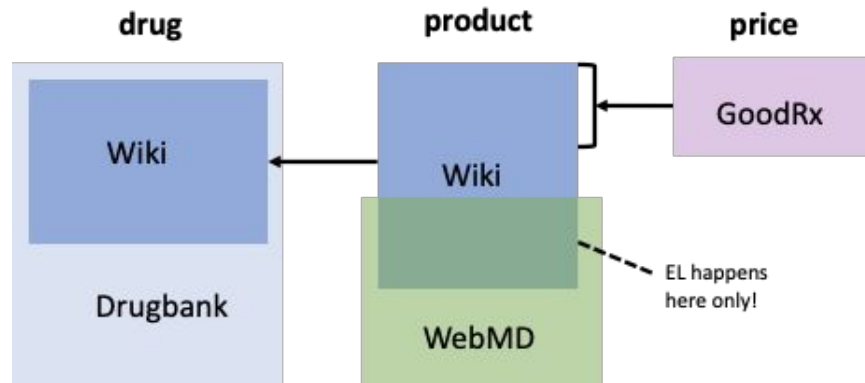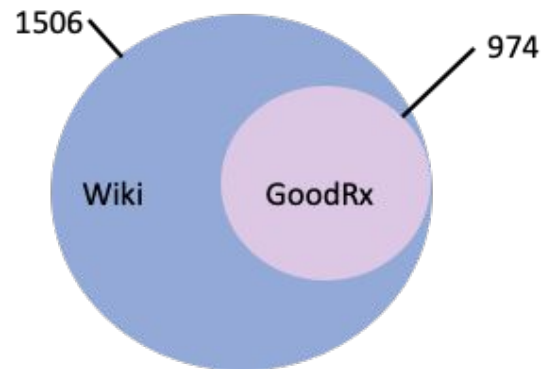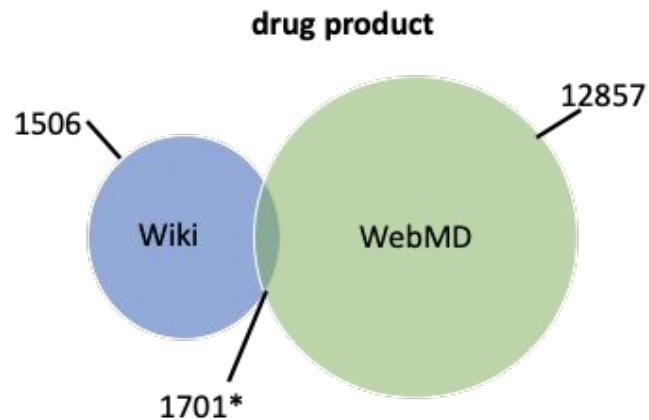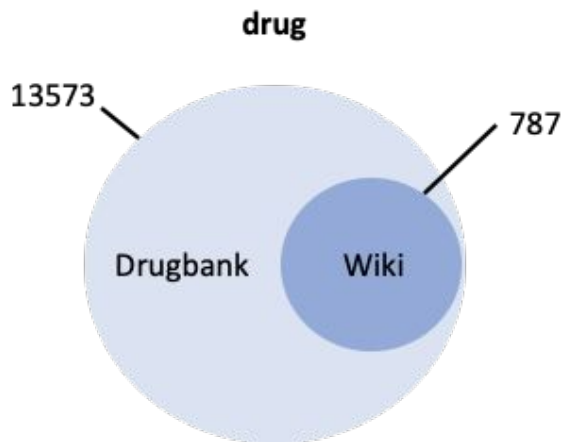
Semantic Model
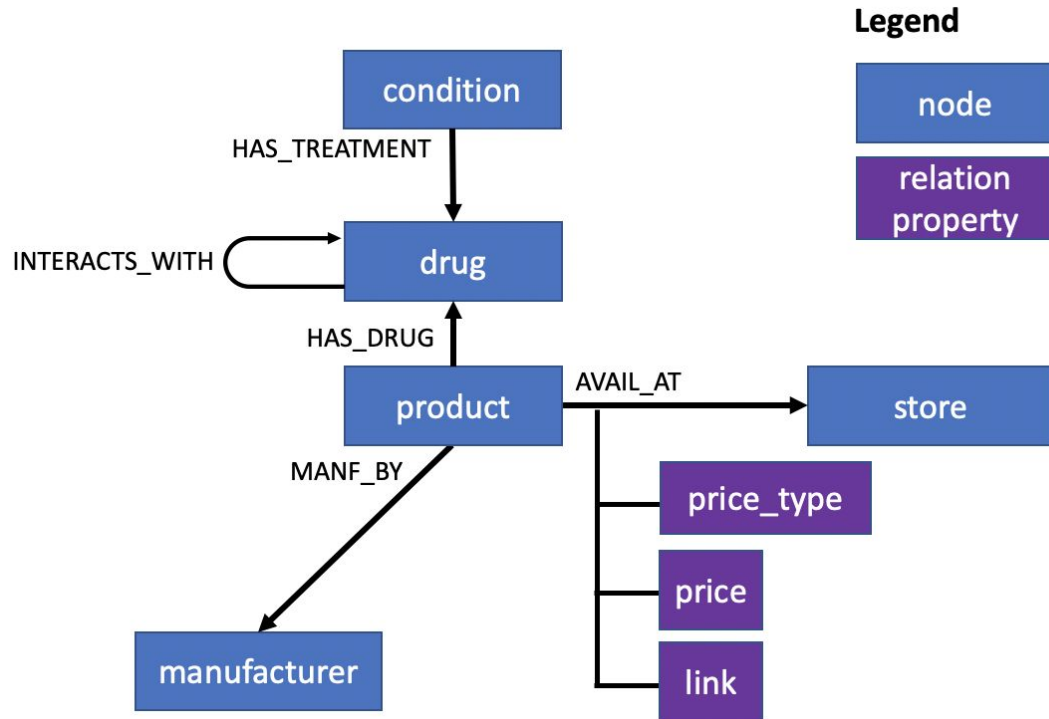
Entity Linking Overview

ENTITY ENRICHMENT

+ Use external **Drugbank ID** as source-of-truth drug identifier

+ **Hub-and-spoke** approach to link from all other sources to one central source (Wiki)

+ EL problem space **reduced** to just one task (Wiki-WebMD)

Linkage Task	Linkage Method	Effort
Wiki-Drugbank	Drugbank ID	Low
Wiki-GoodRx	Wiki products fed as seed list to GoodRx API	Low
Wiki-WebMD	Entity linking with the product name	High

Entity Linking Overview

drug

13573 — Drugbank

787 — Wiki

drug product

1506 — Wiki

12857 — WebMD

1701* (intersection Wiki/WebMD)

1506 — Wiki

974 — GoodRx

* one-to-many pairs allowed
from Wiki-to-WebMD

Entity Linking: Wiki→WebMD

METHOD

+ Match Wiki to WebMD drug products
+ One:many::Wiki:WebMD
+ Matched 1701 pairs

EVALUATION

+*Hash-based* blocking on first two
chars reduced pairs from 20M to 178k
pairs

+*Jaccard similarity* on drug names
finds 97% (37 of 38) dev-set pairs

+*Pairs quality* could be improved, but
is acceptable given memory budget

```
#Generate blocks > pairs, compute similarity, metrics, and output results
b = BlockPairSim(ds_wiki, ds_webmd)
b.create_hash_blocks()
b.make_score_pairs(groundtruth)
b.write_result()
```

```
b.metrics
```

```
{'all_pairs': 178519,
 'match_count': 1701,
 'no_match_count': 176818,
 'pairs_found': 38,
 'recall': 0.9736842105263158,
 'precision - pairs quality': 0.0002072608517860844}
```

```
trial.evaluate()
print('Trial statistics based on Ground-Truth from development set data:')
print(f'tp: {trial.true_positives:.06f} [{len(trial.true_positives_list)}]')
print(f'fp: {trial.false_positives:.06f} [{len(trial.false_positives_list)}]')
print(f'tn: {trial.true_negatives:.06f} [{len(trial.true_negatives_list)}]')
print(f'fn: {trial.false_negatives:.06f} [{len(trial.false_negatives_list)}]')
```

```
Trial statistics based on Ground-Truth from development set data:
tp: 0.973684 [37]
fp: 0.000000 [0]
tn: 1.000000 [798]
fn: 0.026316 [1]
```
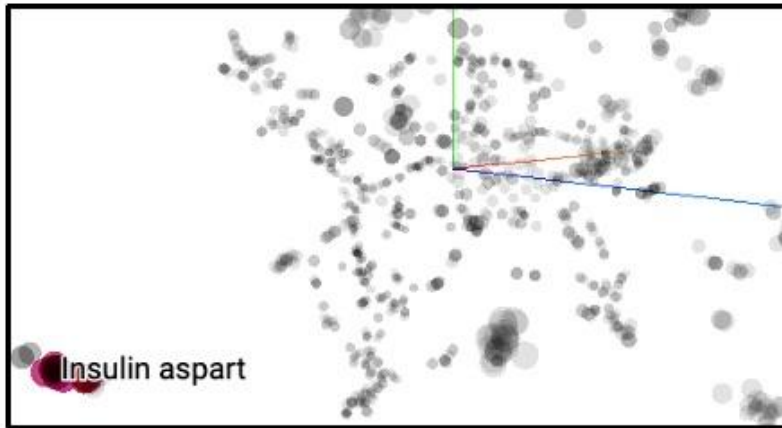
Open Drug Knowledge Graph embeddings

ComplEx and TransE

3,654 instances

Evaluation on a ranking task by Wang et al. (2019):

Embedding	MRR	Hits@1	Hits@3	Hits@10
TransE	0.25	0.50	0.29	0.13
Complex	0.32	0.57	0.36	0.19

Do drugs that treat similar symptoms cluster together ?

Drug: "Insulin aspart"	
Pramlintide	0.059
Insulin glargine	0.098
Insulin human	0.103
Insulin detemir	0.109
Metformin	0.161
Insulin lispro	0.175
Exenatide	0.231
Glyburide	0.232
Glimepiride	0.236
Acarbose	0.242

Do similar conditions cluster together?
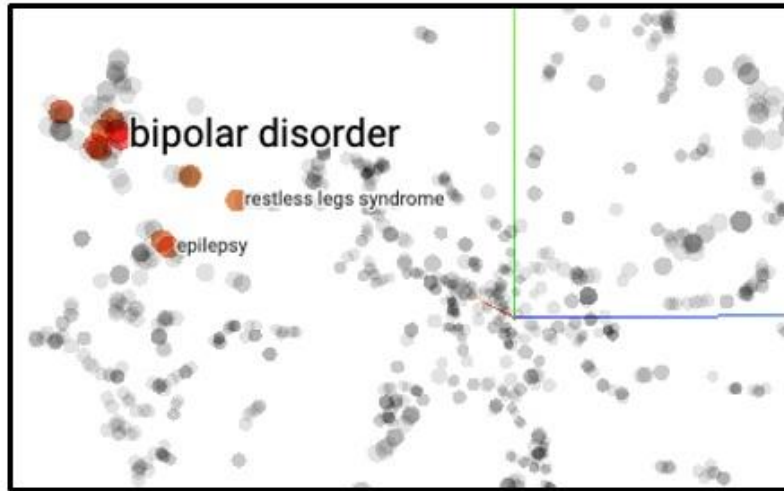
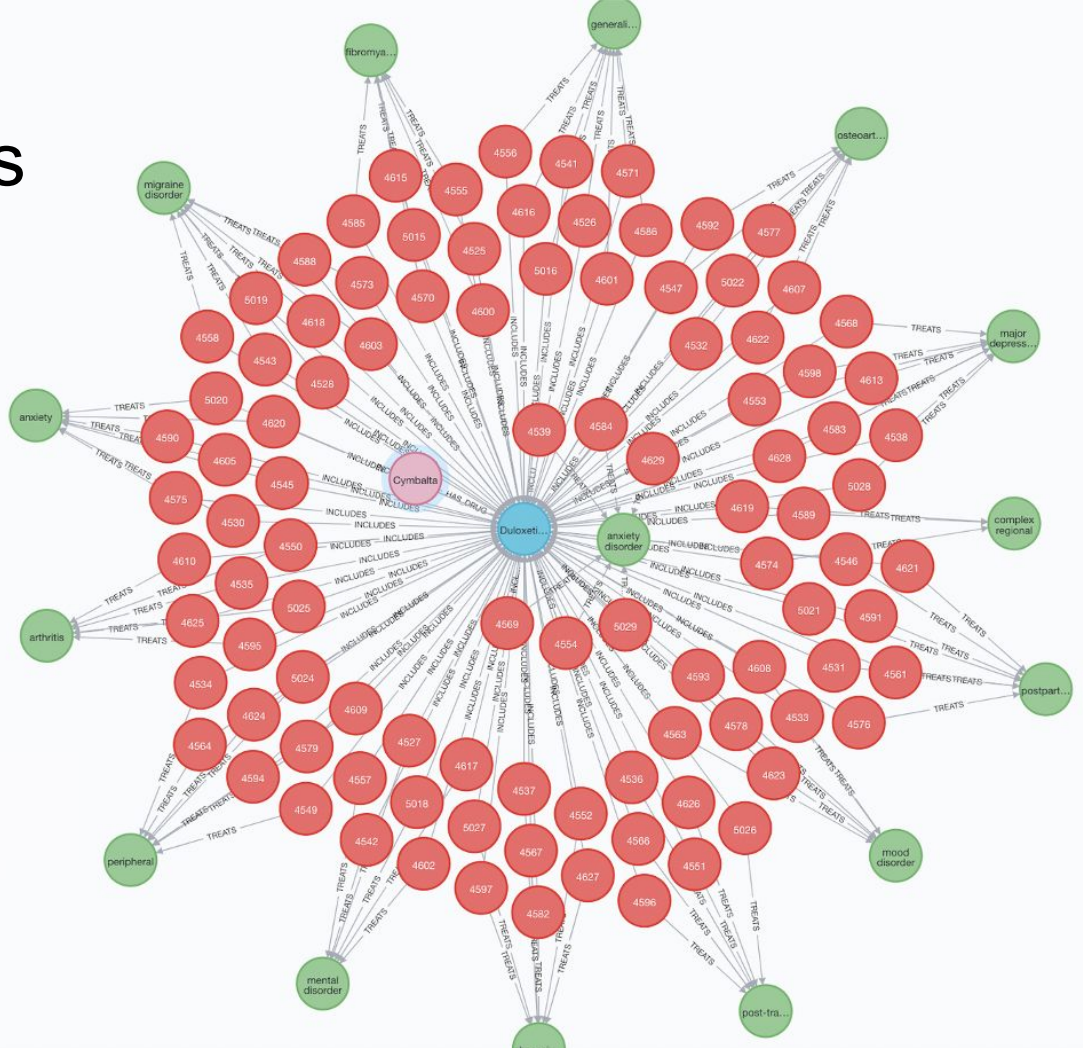| Condition: "Bipolar disorder"	
mood disorder	0.309
schizophrenia	0.399
insomnia	0.401
epilepsy	0.428
psychosis	0.433
post-traumatic stress disorder	0.446
trigeminal neuralgia	0.448
anxiety	0.460
restless legs syndrome	0.468
dysarthria	0.481

Where can I find a treatment?

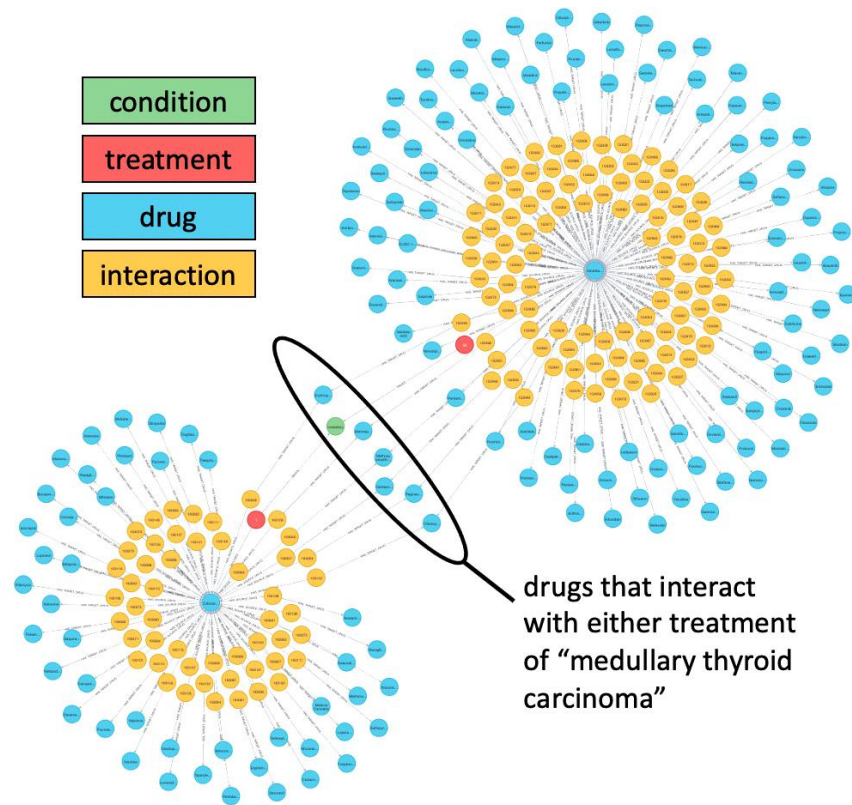For condition 'sinus headache', find treatments under $10 and where to purchase them
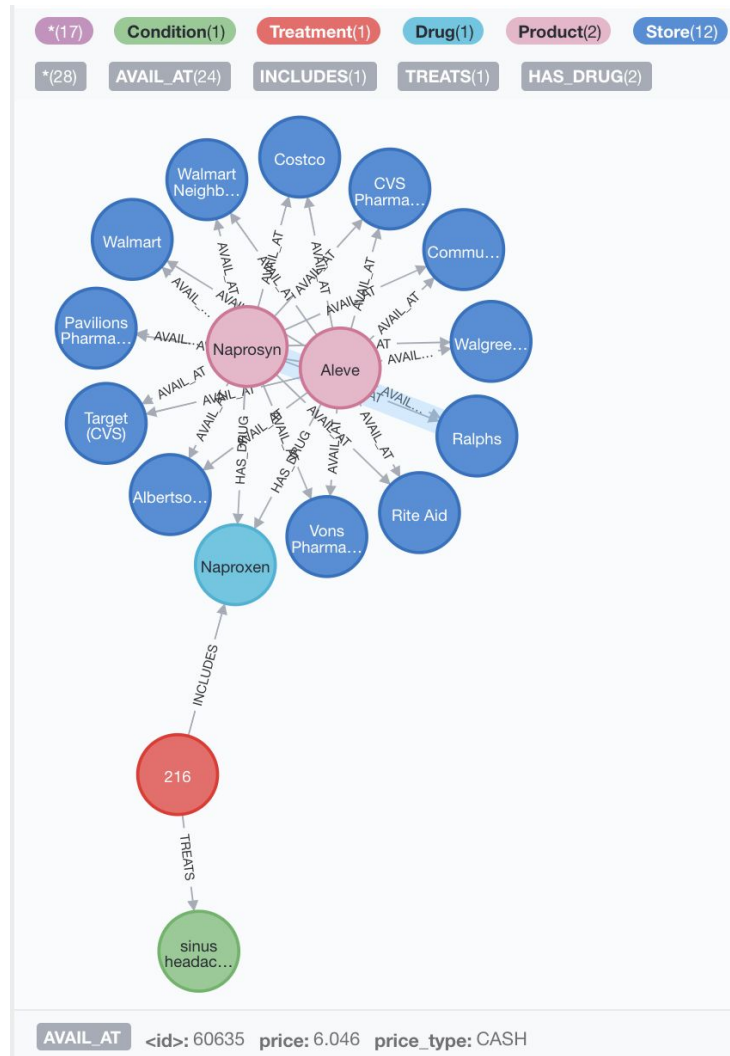
What conditions does this drug treat?

QUERY: For the drug with the most ratings 'Cymbalta', what conditions does it treat?

mainly mental health conditions: anxiety, mood disorder, major depression, ...

How do two treatments of a symptom interact?

condition
treatment
drug
interaction

drugs that interact with either treatment of "medullary thyroid carcinoma"

Allowing users to search

Open Drug Knowledge Graph

Consolidate medical knowledge to empower regular users

Pipeline: data acquisition + entity resolution + semantic modeling

Graph embeddings find similar treatments and symptoms

Structured queries provide more crisp explanation

Users can explore the data through a friendly interface

Thank you!