DRUG REPURPOSING FOR COVID-19: A KNOWLEDGE GRAPH APPROACH

Xiaodong Li, Min Ou, Xuxiao Chris Ye, Wenya Sun

Supervisors: Prof. Reynold Cheng, Dr. Ruibang Luo, Dr. Esther Chan

Department of Computer Science & Pharmacology and Pharmacy

The University of Hong Kong

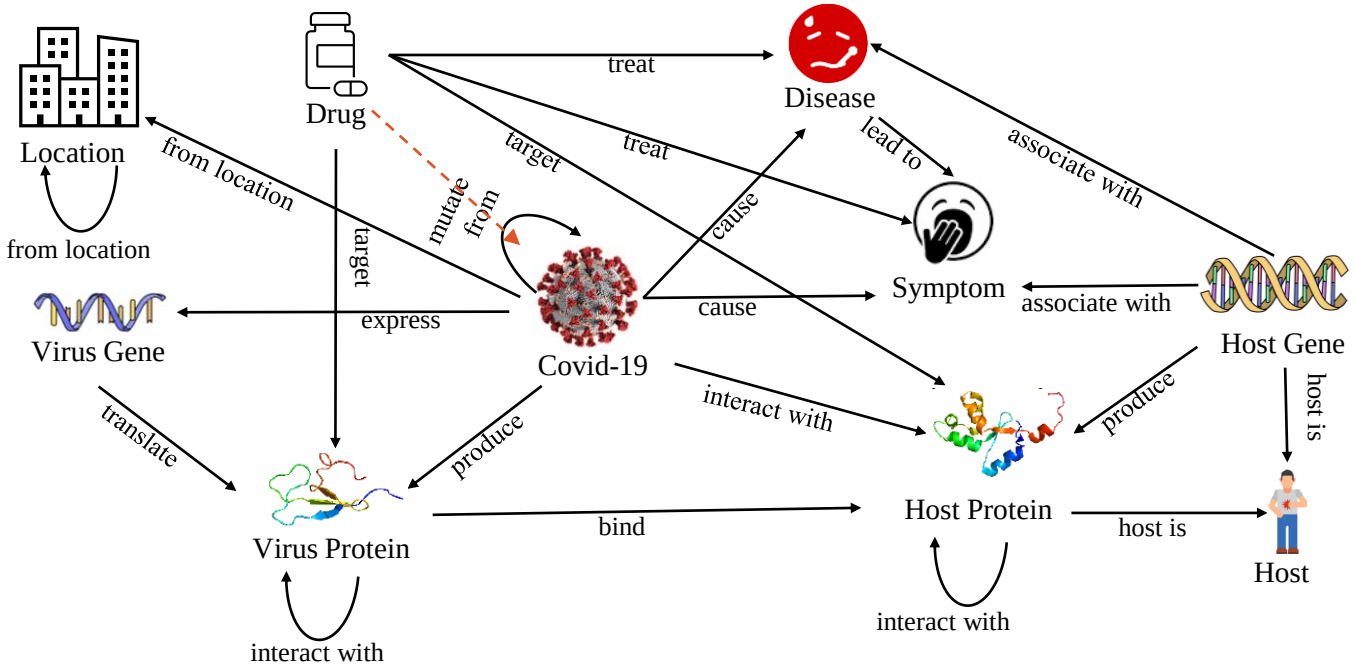
Section: Postgraduate, Group No. PG05

Special acknowledgement to Mr. Vincent Yan for the help and advice!

DRUG REPURPOSING: MOTIVATION

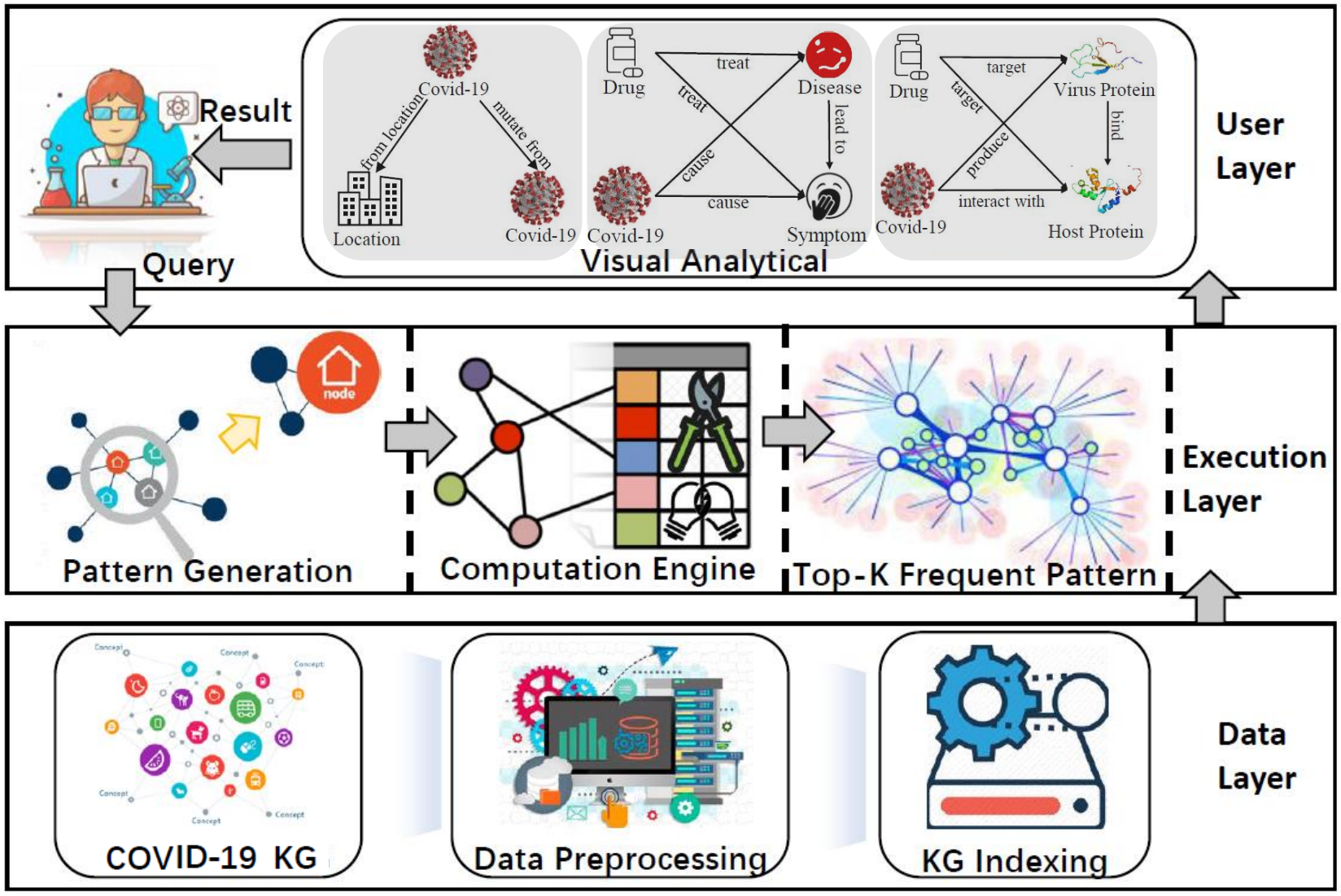
- SARS-CoV-2 outbreak (Covid-19) has become a global pandemic. No known effective Covid-19 drug treatment.
- Discovery of new drugs is time consuming + expensive.
- Traditional drug-repurposing methods (e.g., protein docking) does not consider complex interrelationship of drugs, proteins, genes, symptoms, diseases, etc.
- Therefore, we repurpose **existing drugs** using **network-based** strategies

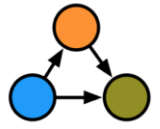
KNOWLEDGE GRAPH SCHEMA

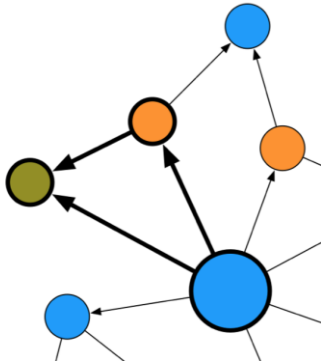


Integrated from OpenKG, HPO, NCBI, and DrugBank; over 48K nodes and 815K edges!

System Architecture

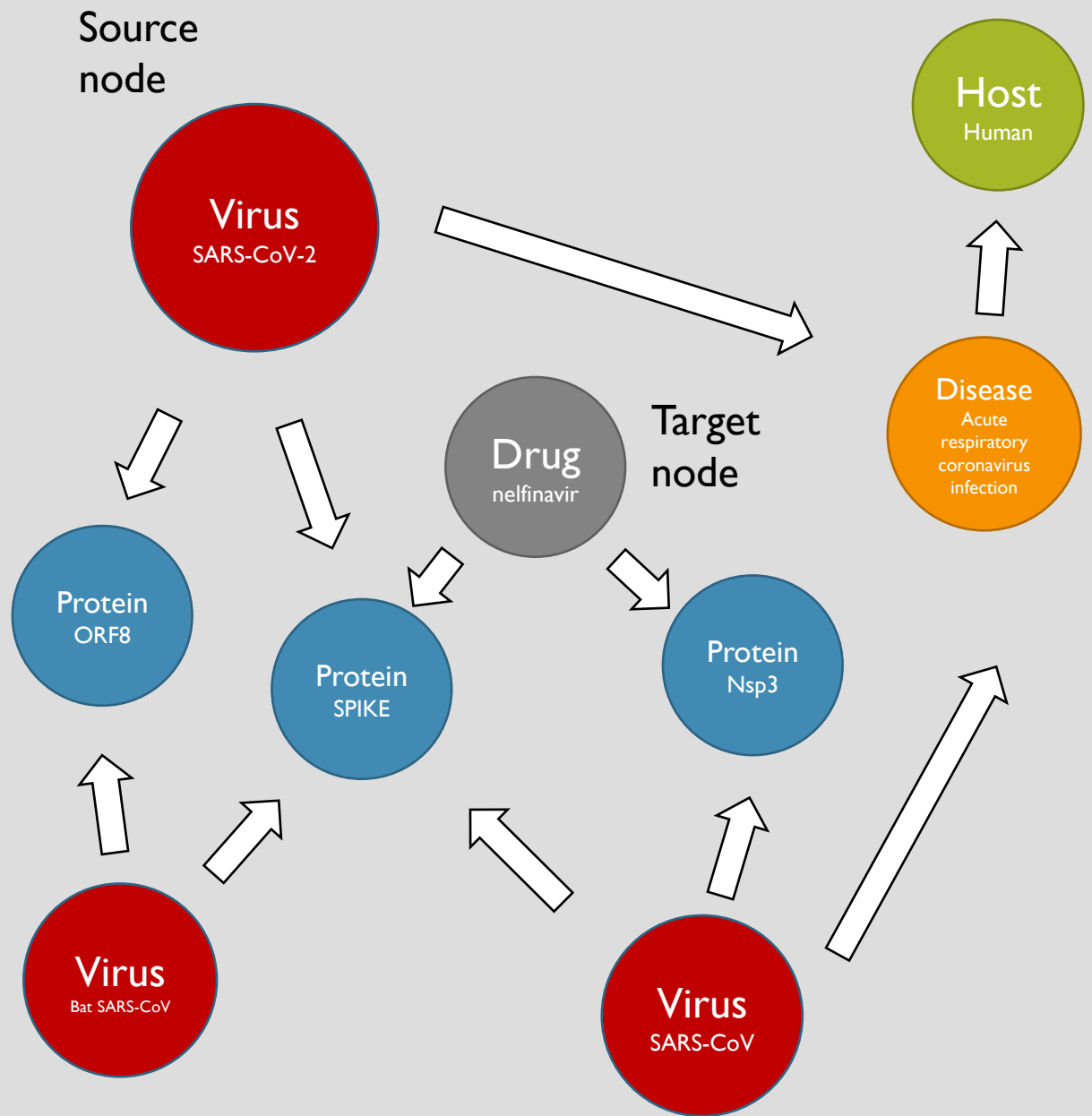






ALG.1 KNOWLEDGE GRAPH PPR

- Personalized Page Rank (PPR) is used by Twitter to present users with recommendations of other accounts that they may wish to follow.
- We adapt PPR into a knowledge graph version.
- Source node (SARS-CoV-2), target node (Each drug), with parameters: damping factor 0.85, iteration 2M.



ALG.2 MOTIF-BASED LINK PREDICATION

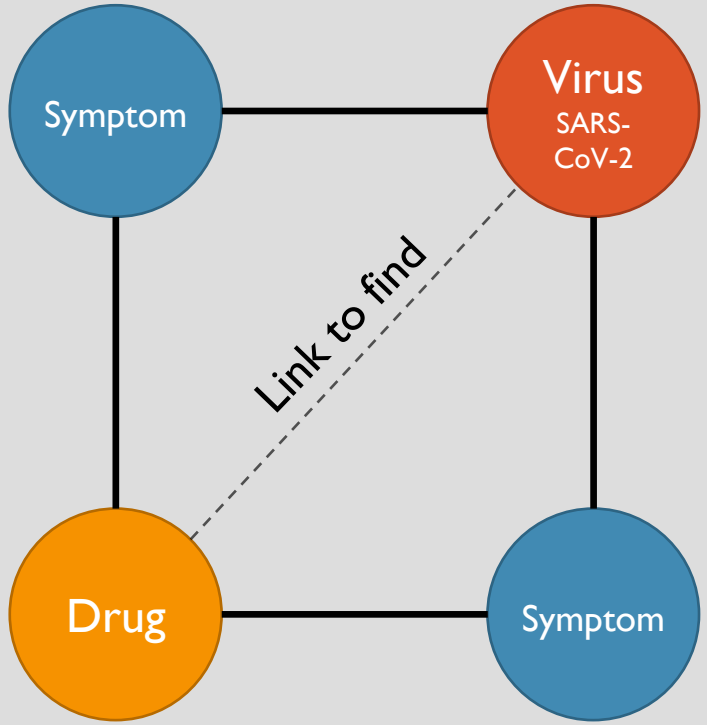
- Knowledge Graph too complex & large (hundreds of thousands of edges).
- Motifs: **small** & **frequent** graphlets of size *k* nodes, *k* = 1, 2, 3, 4, 5.
- Use "interesting" motifs M to generate Motif Feature Vector (MFV). For each drug D, MFV describes the frequency of M that contains s and SARS-CoV-2. Train a classifier with input (MFVs) to predict if the link (D, SARS-CoV-2) exists.
- Greater chance of D to serve in covid-19 treatment if the link has higher existential probability.

ALG.2 MOTIF BASED LINK PREDICTION

Consider the rectangle motif

- Occurs 174 times in KG
- Check frequencies of drugs occurring in motif:
  Drug DB05777 frequency: 10
  Drug DB00479 frequency: 1

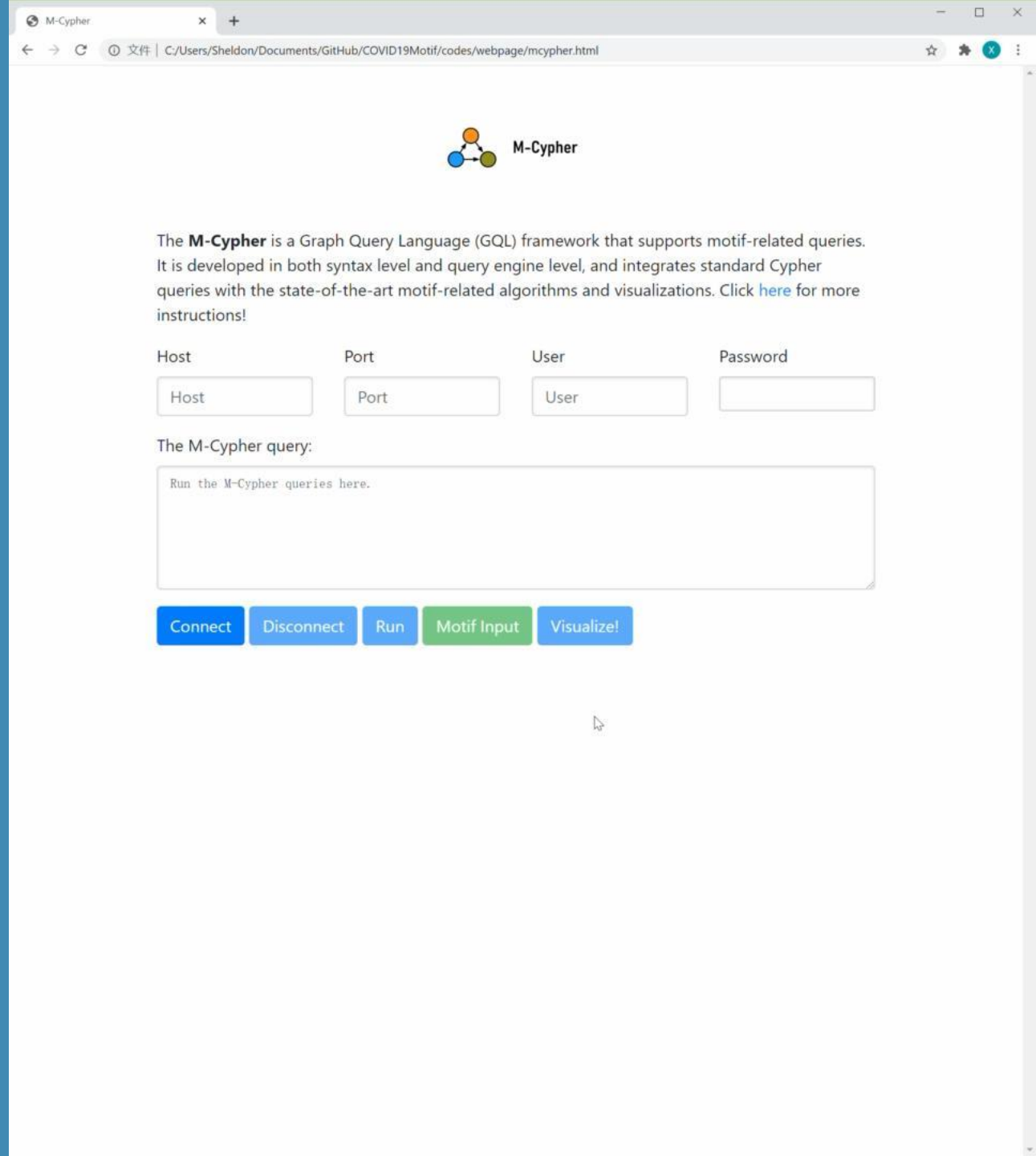
- Algorithm predicts higher likelihood of Drug-Virus link: DB05777.



M-CYPHER: POWERFUL TOOL TO ANALYZE KNOWLEDGE GRAPH

- M-Cypher: An efficient Graph Query Language, to help non-programmer users to explore the knowledge graph!

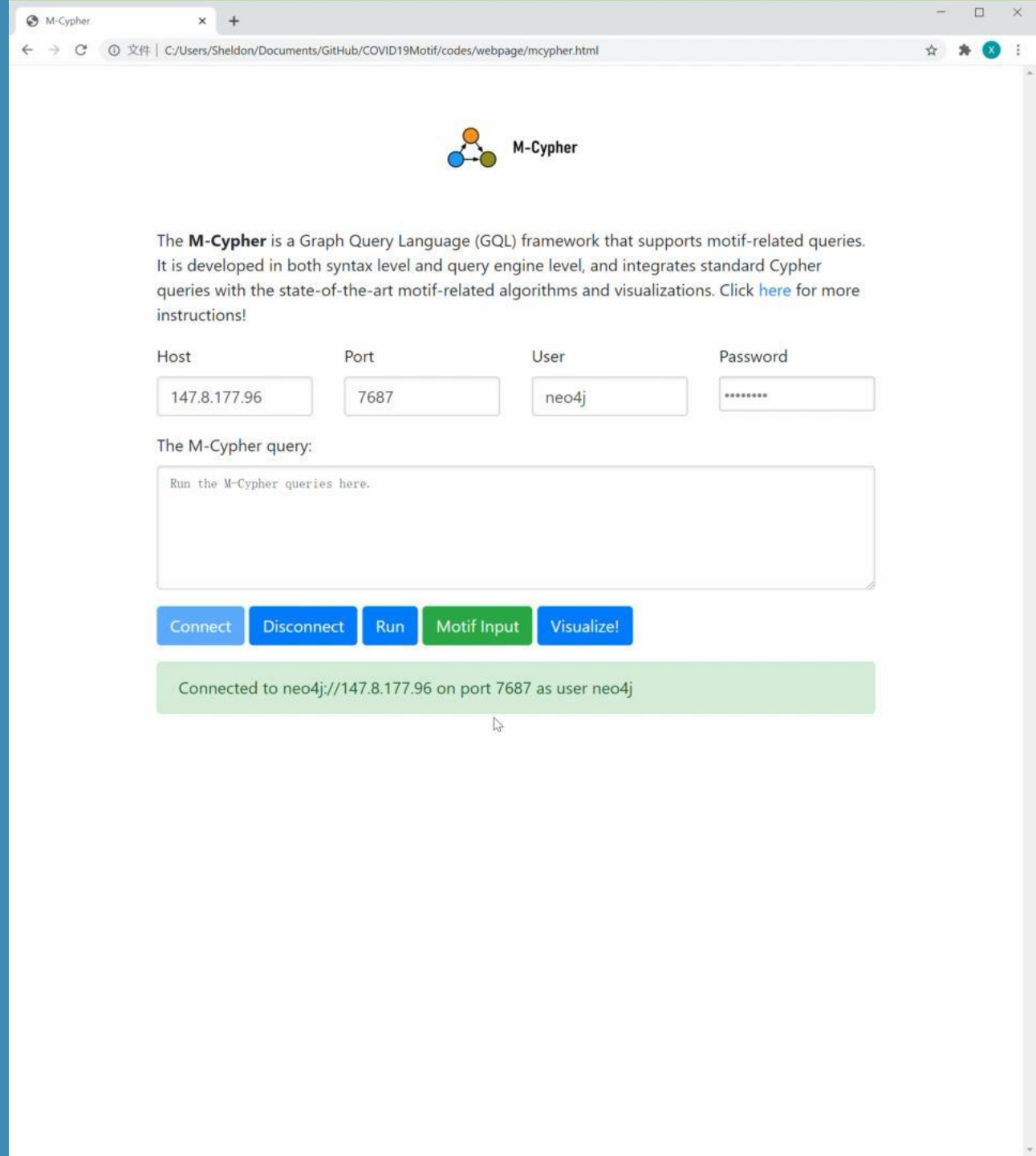
- M-Cypher efficiently collects statistics for drug re-purposing, using concept of **orbits** to avoid overcounting risk:
  - which motif is frequent?
  - which node/edge/path is interesting?
  - which subgraph worth exploration?



M-CYPHER: POWERFUL TOOL TO ANALYZE KNOWLEDGE GRAPH

- M-Cypher: An efficient Graph Query Language, to help non-programmer users to explore the knowledge graph!

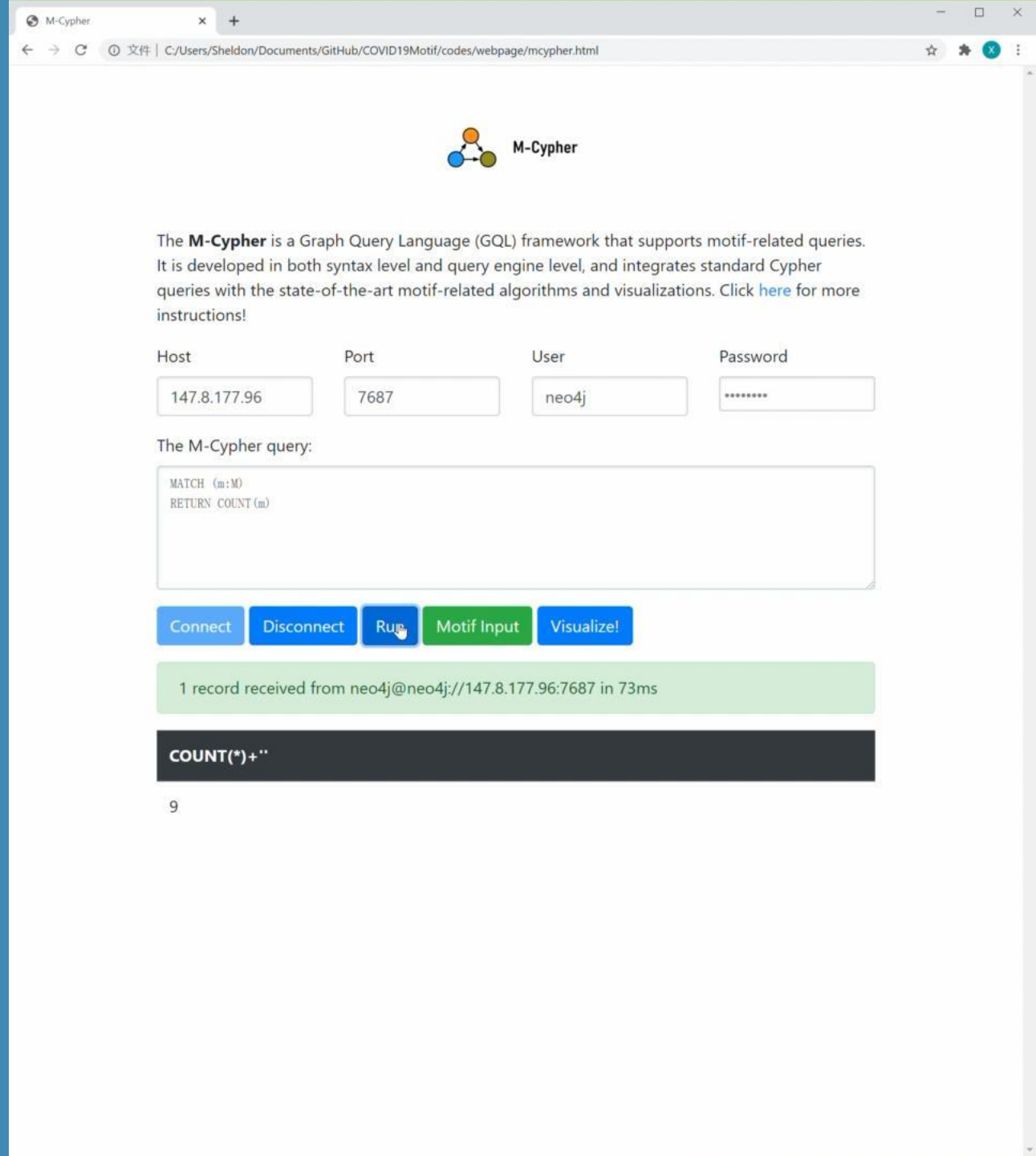
- M-Cypher efficiently collects statistics for drug re-purposing, using concept of **orbits** to avoid overcounting risk:
  - which motif is frequent?
  - which node/edge/path is interesting?
  - which subgraph worth exploration?



M-CYPHER: POWERFUL TOOL TO ANALYZE KNOWLEDGE GRAPH

- M-Cypher: An efficient Graph Query Language, to help non-programmer users to explore the knowledge graph!

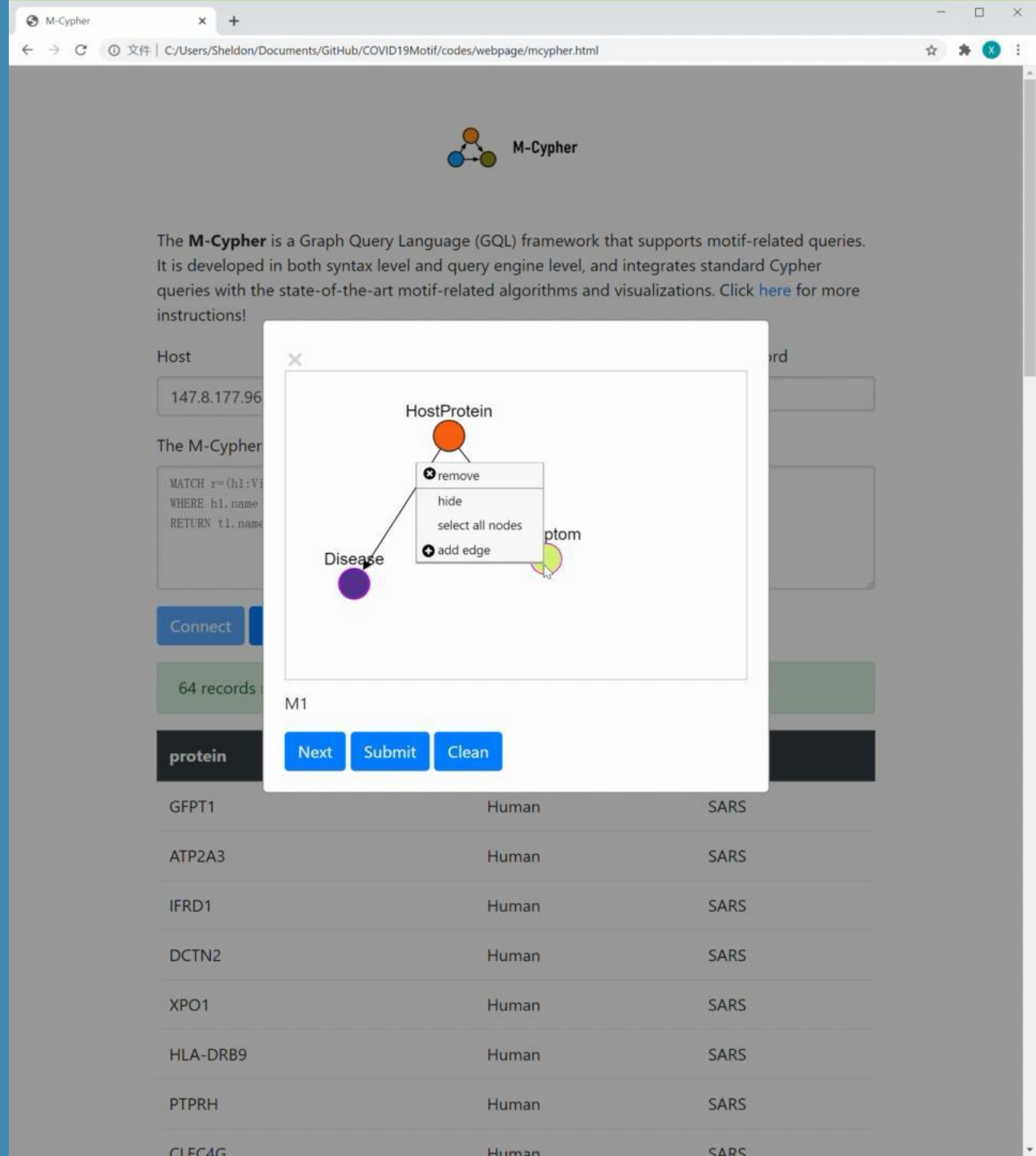
- M-Cypher efficiently collects statistics for drug re-purposing, using concept of **orbits** to avoid overcounting risk:
  - which motif is frequent?
  - which node/edge/path is interesting?
  - which subgraph worth exploration?



M-CYPHER: POWERFUL TOOL TO ANALYZE KNOWLEDGE GRAPH

- M-Cypher: An efficient Graph Query Language, to help non-programmer users to explore the knowledge graph!

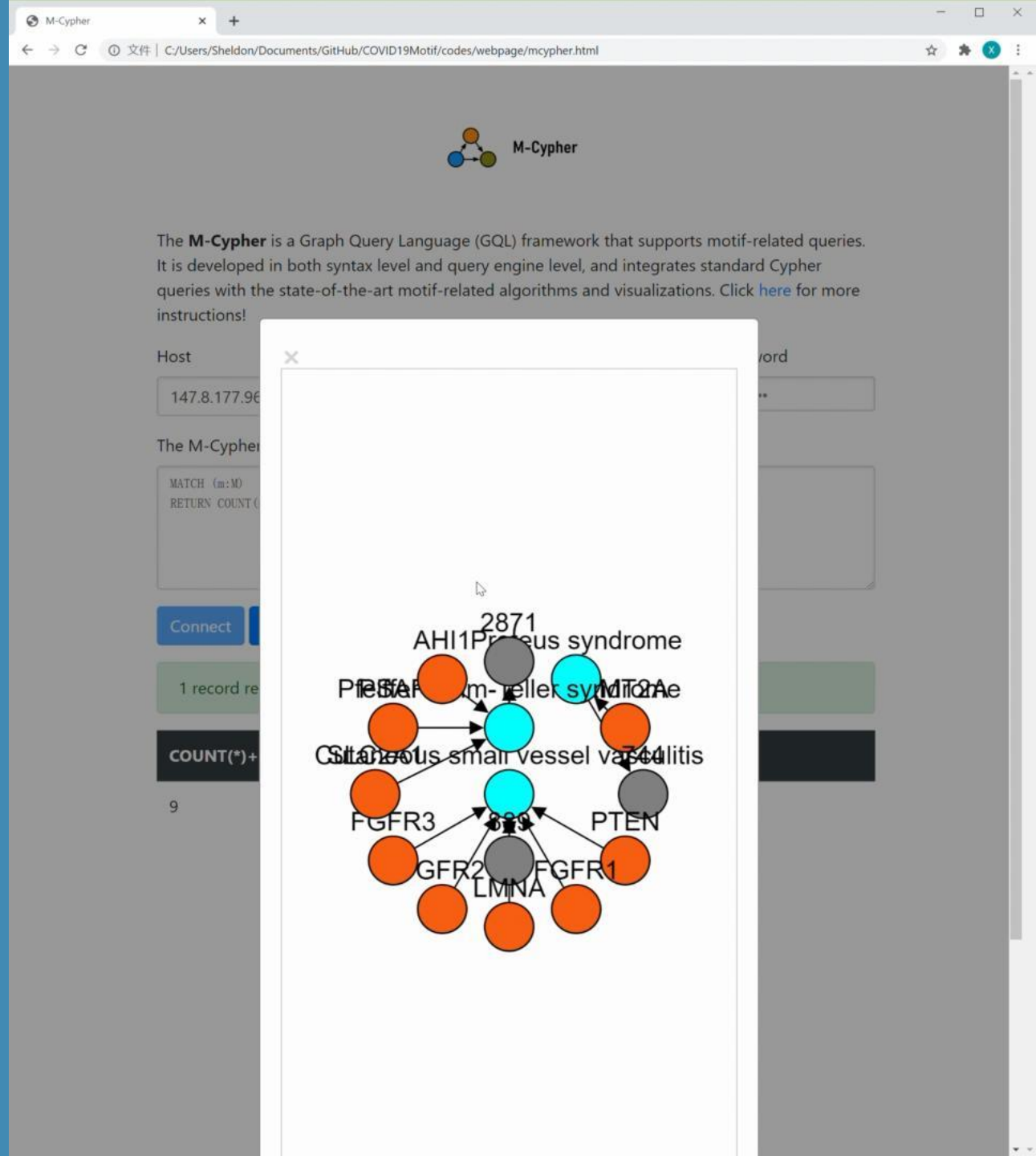
- M-Cypher efficiently collects statistics for drug re-purposing, using concept of **orbits** to avoid overcounting risk:
  - which motif is frequent?
  - which node/edge/path is interesting?
  - which subgraph worth exploration?

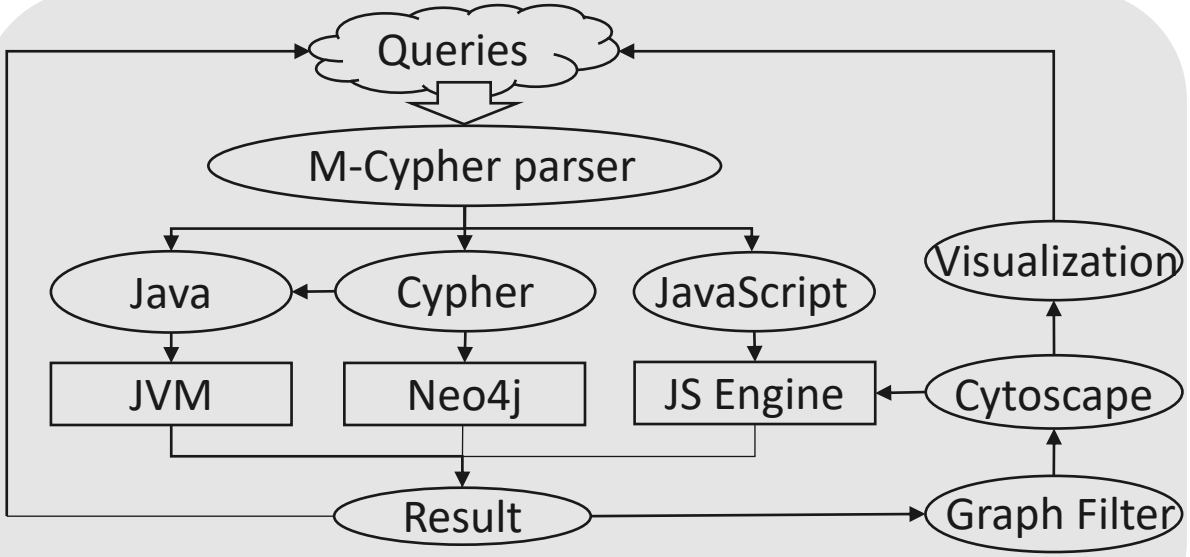


M-CYPHER: POWERFUL TOOL TO ANALYZE KNOWLEDGE GRAPH

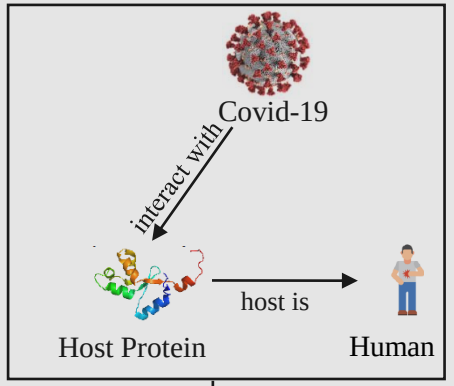
- M-Cypher: An efficient Graph Query Language, to help non-programmer users to explore the knowledge graph!

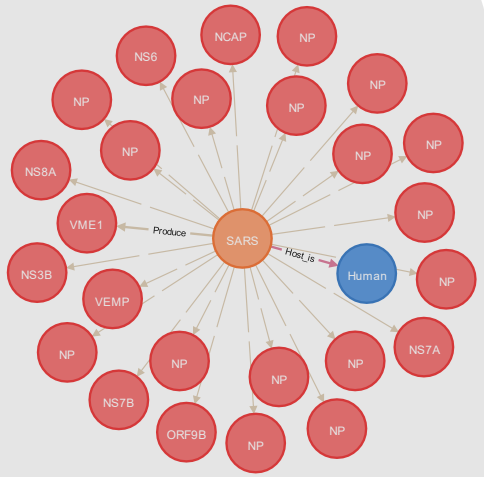
- M-Cypher efficiently collects statistics for drug re-purposing, using concept of **orbits** to avoid overcounting risk:
  - which motif is frequent?
  - which node/edge/path is interesting?
  - which subgraph worth exploration?





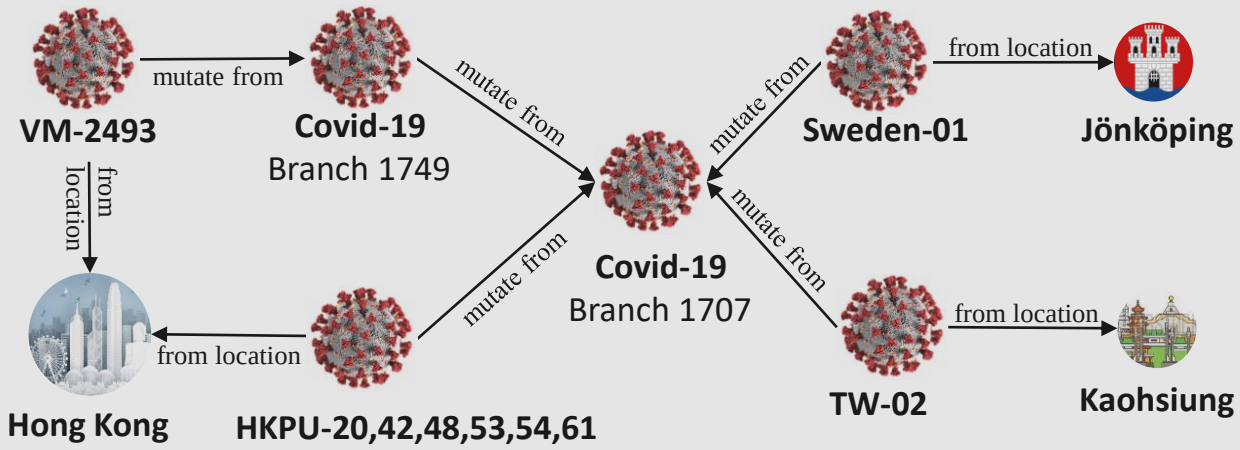
M-Cypher framework



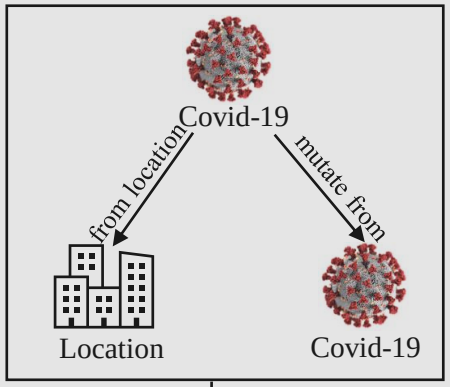



```
MATCH (m:M)
WHERE s.Virus = "SARS-CoV-2"
RETURN COUNT(m)
```

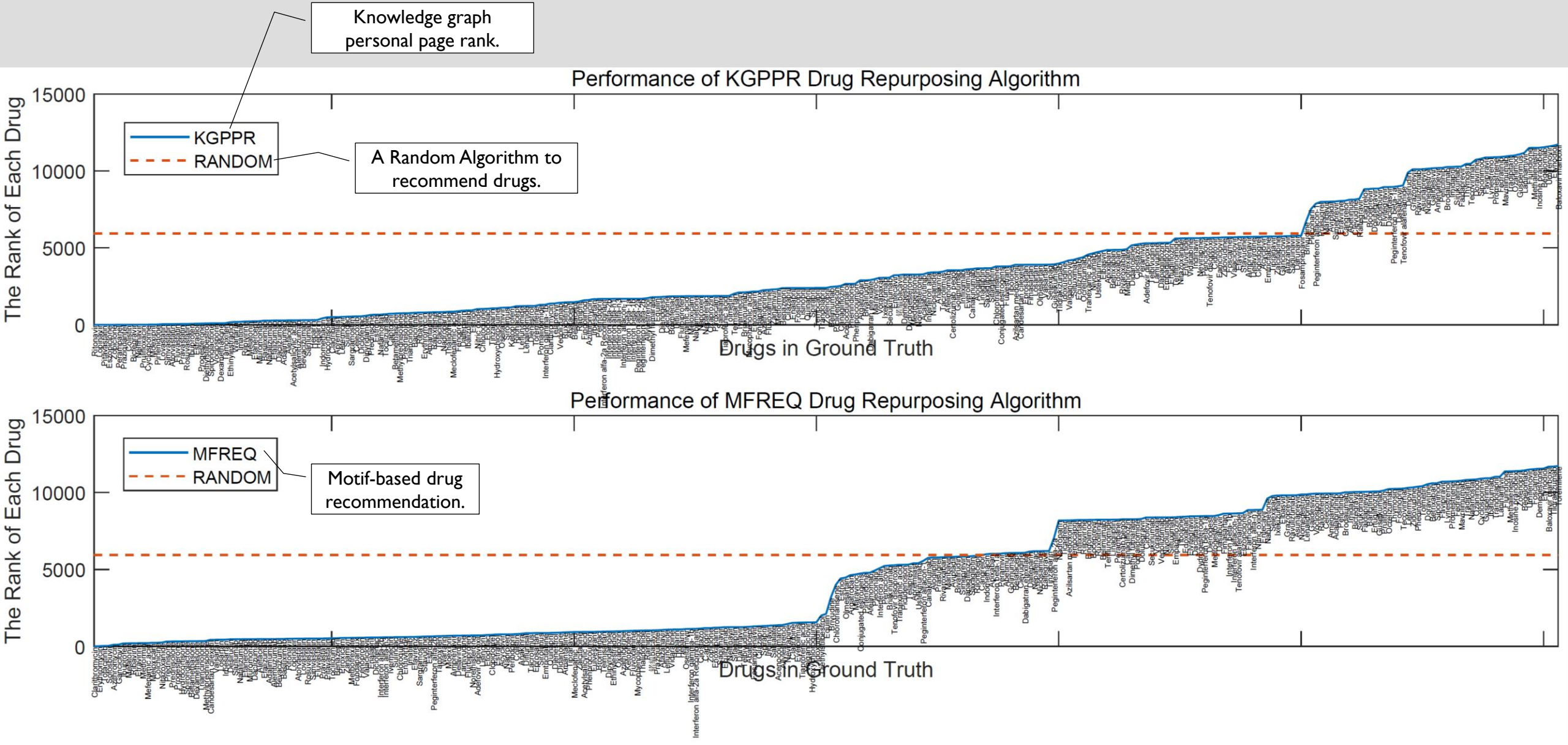
The top virus protein in SARS-CoV is NP!



Hong Kong, Kaohsiung and Jonkoping have identical branch 1707!




```
MATCH (s:Location)-[m:M*]->(t:Location)
WHERE s.name = "Hong Kong"
RETURN t.name
```


Knowledge graph personal page rank.
Performance of KGPPR Drug Repurposing Algorithm
KGPPR
RANDOM
A Random Algorithm to recommend drugs.
The Rank of Each Drug
15000
10000
5000
0
Drugs in Ground Truth
Performance of MFREQ Drug Repurposing Algorithm
MFREQ
RANDOM
Motif-based drug recommendation.
The Rank of Each Drug
15000
10000
5000
0
Drugs in Ground Truth

HIGHLIGHTED DRUG CANDIDATES FOR REPURPOSING

Drug_name	Clinical_trial	Drug_type	Rank
Ritonavir	Y	Antiviral agents	0.01%
Lopinavir	Y	Antiviral agents	0.02%
Pitavastatin	N	Statins	0.03%
Moexipril	N	ACEIs	0.12%
Lovastatin	N	Statins	0.35%
Simvastatin	N	Statins	0.36%
Atorvastatin	N	Statins	0.37%
Fluvastatin	N	Statins	0.44%
Pravastatin	N	Statins	0.46%
Rosuvastatin	N	Statins	0.47%
Dexamethasone	Y	Corticosteroids	0.84%
Sarilumab	Y	immunosuppressants	2.60%
Hydrocortisone	Y	Corticosteroids	4.10%
Prednisone	N	Corticosteroids	5.42%
Tocilizumab	Y	immunosuppressants	5.81%

Almost all drugs currently in clinical trial are recommended by our algorithms in top 10% from 11865 candidates in DrugBank!

FUTURE PLAN

- Enrich the covid-19 knowledge graph & collect more drugs which are recommended by domain experts.
- Apply deep learning algorithms to train new models.
- Analyse drugs recommended high but without evidence from literatures, e.g., **neuropsychiatric drugs**.
- Develop an user interface for medical experts to find personalized drugs.


THANK YOU