

# 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

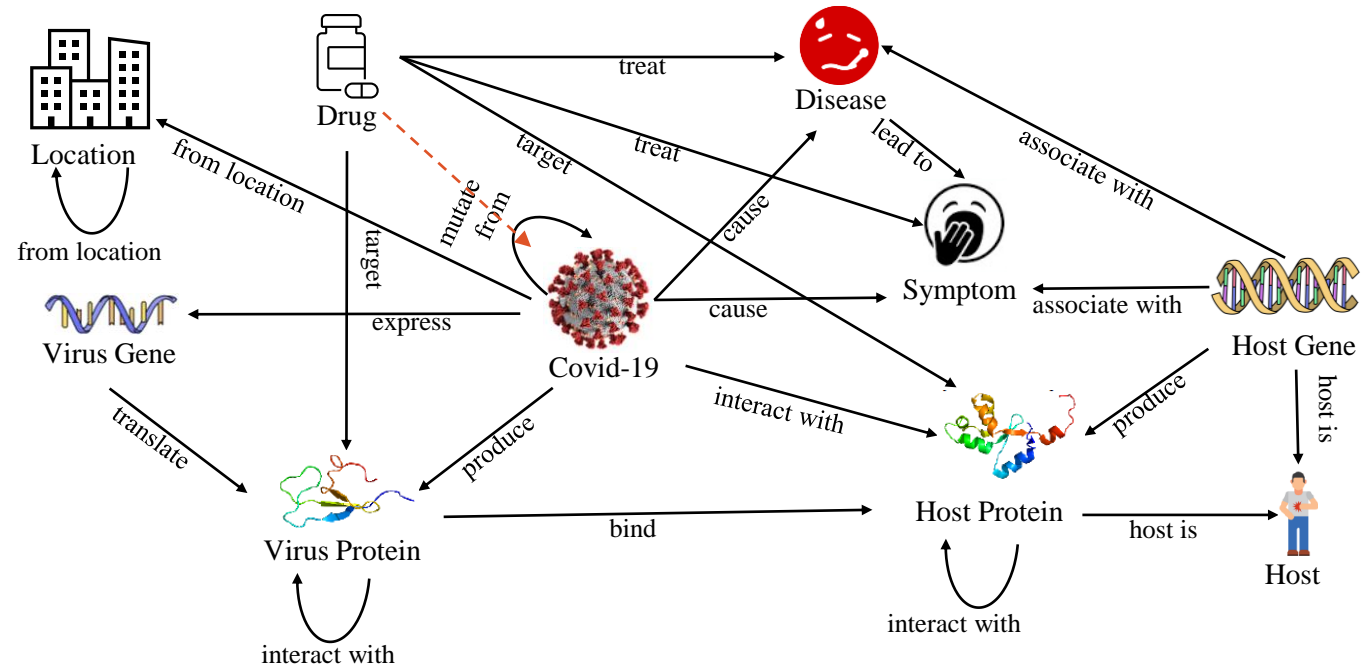
Section: Postgraduate, Group No. PG05

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

# DRUG REPURPOSING: MOTIVATION

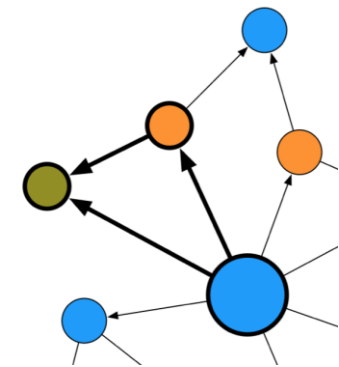
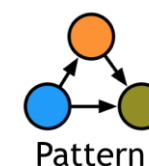
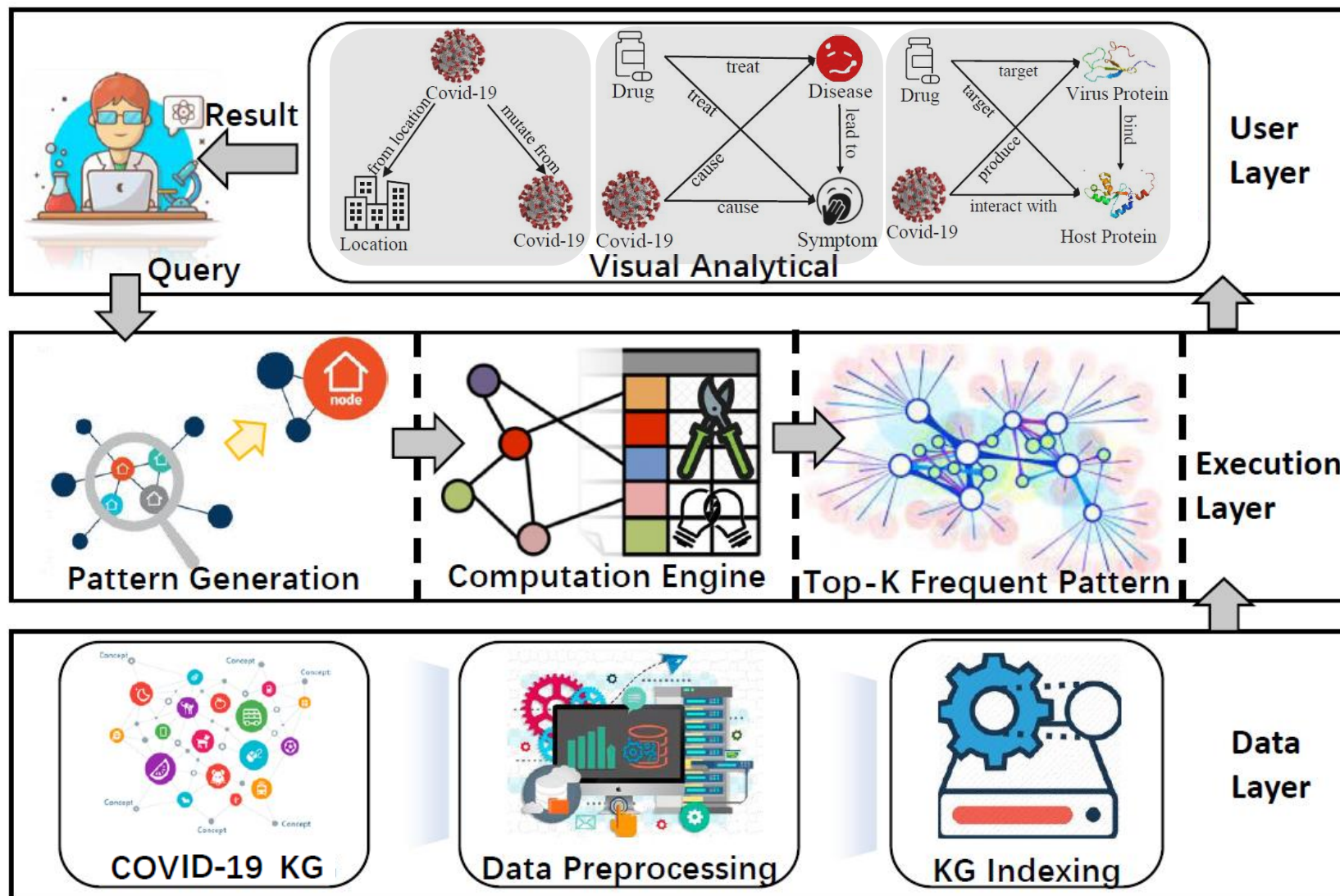
- SARS-CoV-2 outbreak (Covid-19) has become a global pandemic. No known effective Covid-19 drug treatment.
- Discovery of new drugs is time consuming + expensive.
- Traditional drug-repurposing methods (e.g., protein docking) does not consider complex interrelationship of drugs, proteins, genes, symptoms,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

# KNOWLEDGE GRAPH SCHEMA



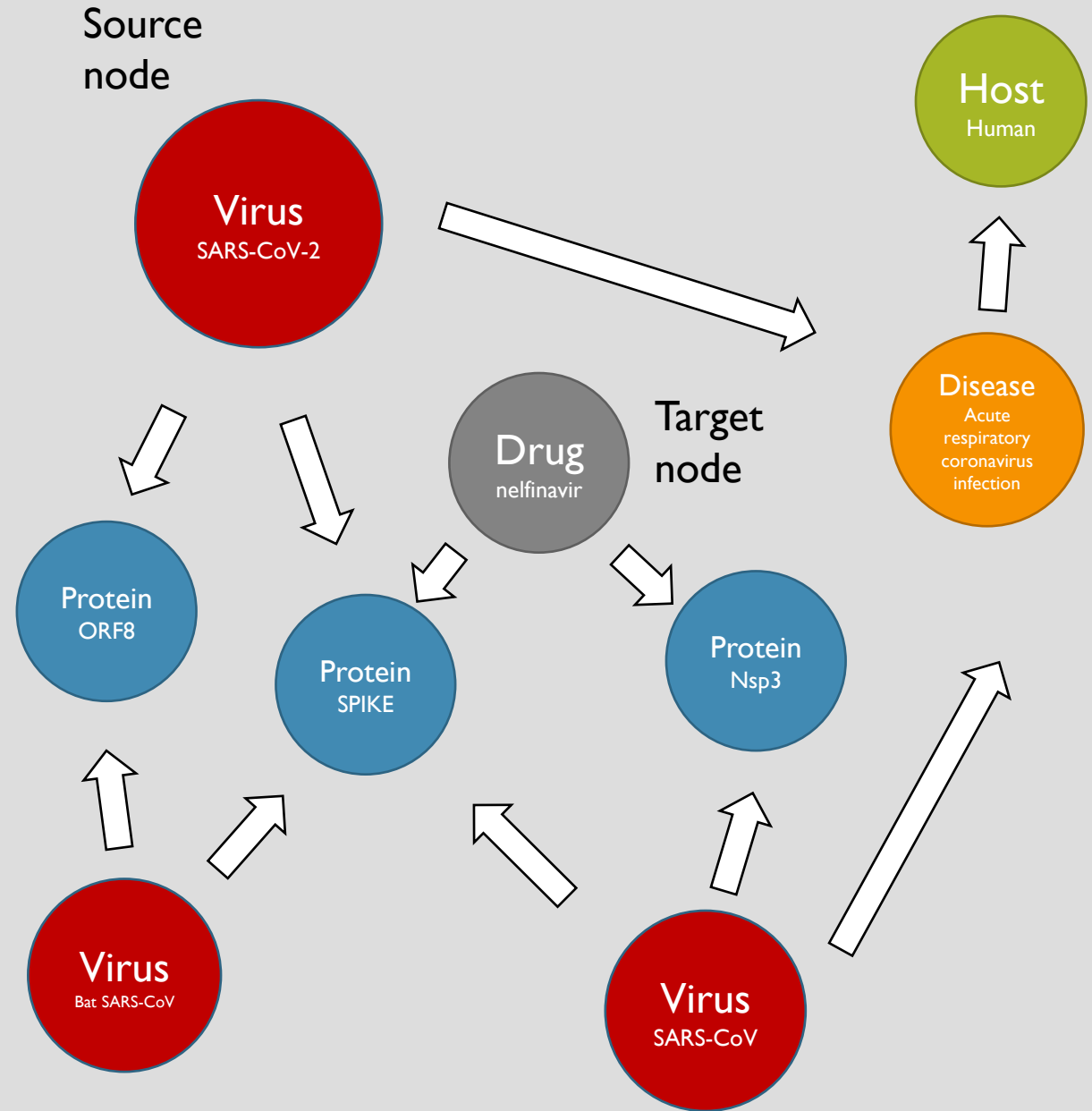
Integrated from OpenKG, HPO, NCBI, and DrugBank; over 48K nodes and 815K edges!

# 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.



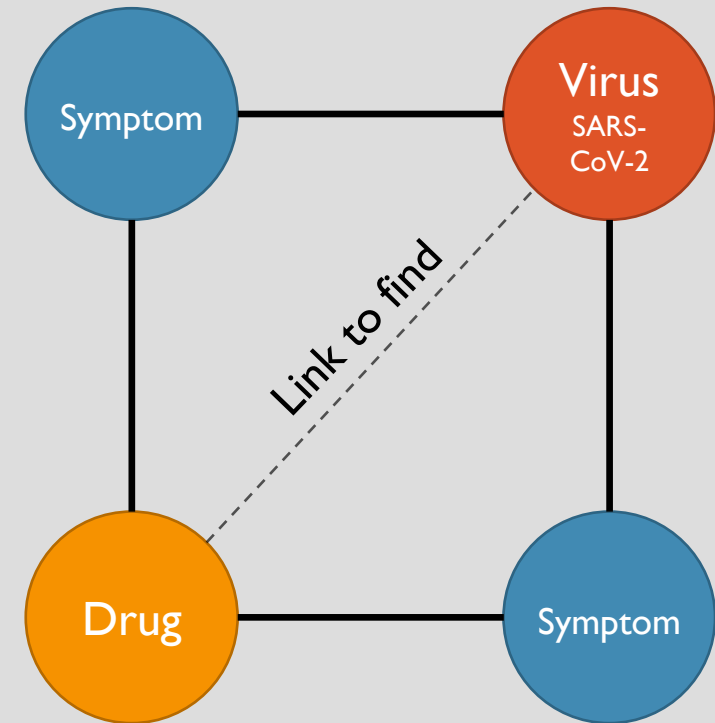
## ALG.2 MOTIF-BASED LINK PREDICATION

- 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  $(D, \text{SARS-CoV-2})$  exists.
- Greater chance of  $D$  to serve in covid-19 treatment if the link has higher existential probability.

## ALG.2 MOTIF BASED LINK PREDICTION

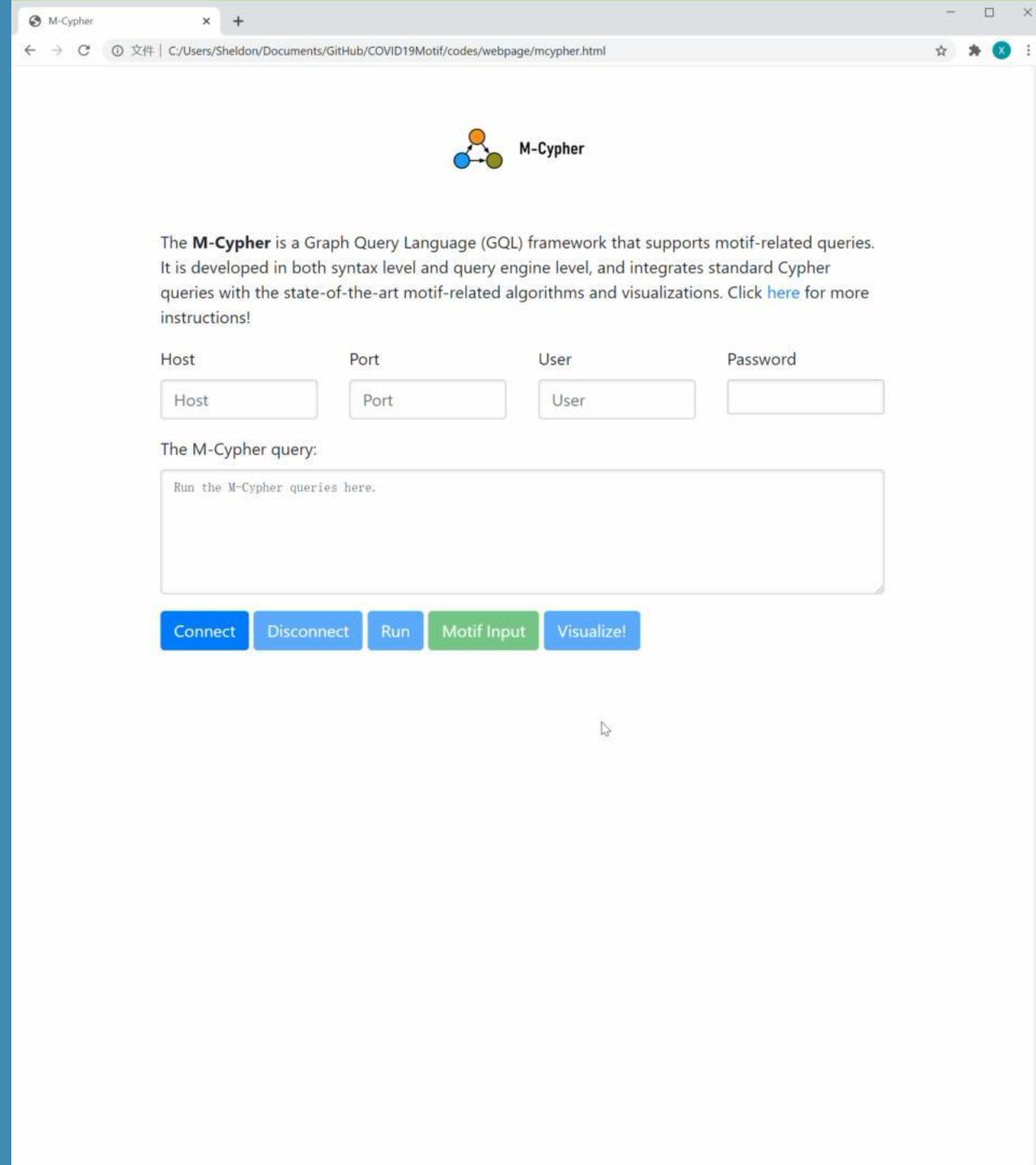
Consider the rectangle motif

- 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

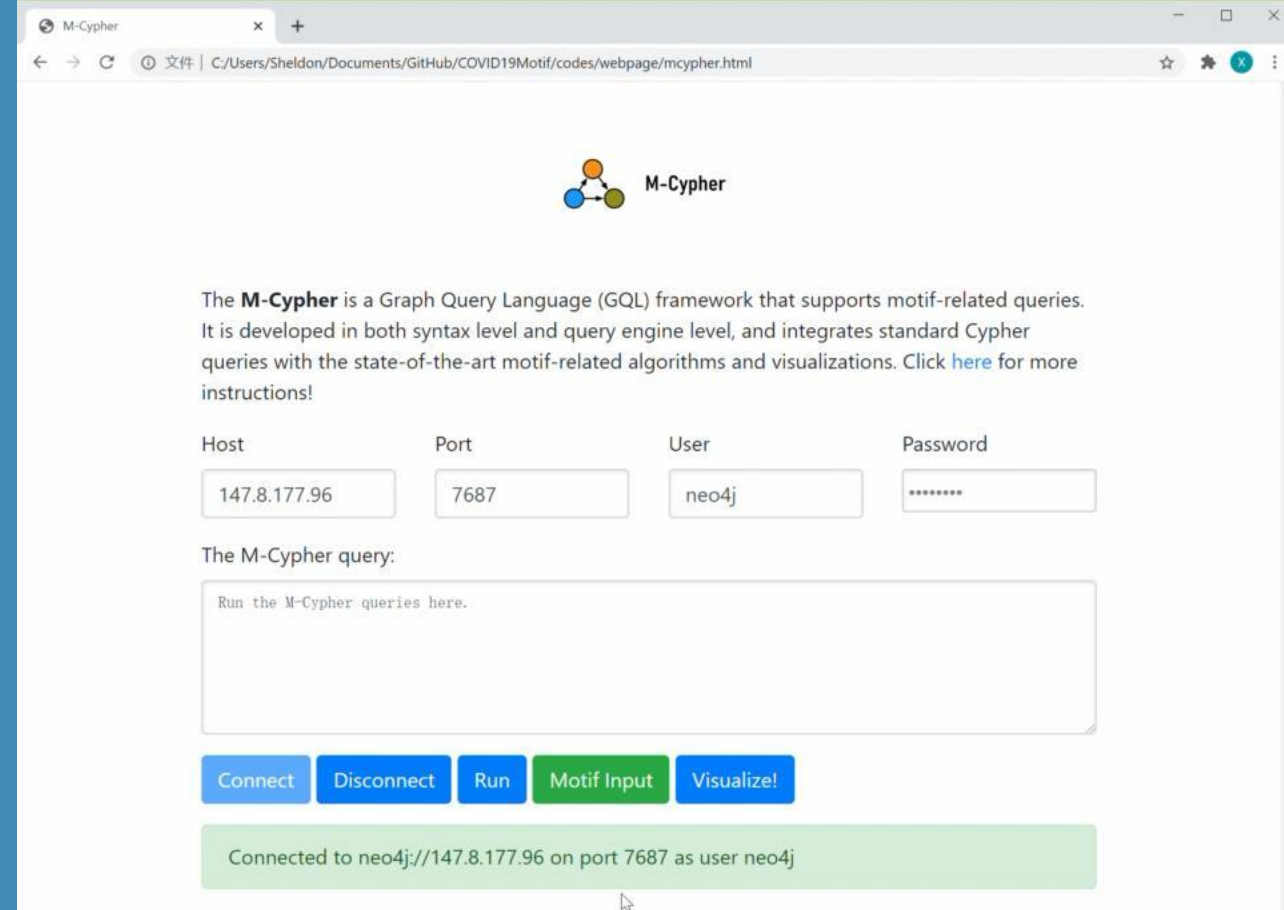
- M-Cypher: An efficient Graph Query Language, to help non-programmer users to explore the knowledge graph!
- M-Cypher efficiently collects statistics for drug re-purposing, using concept of orbits to avoid overcounting risk:
  - which motif is frequent?
  - which node/edge/path is interesting?
  - which subgraph worth exploration?





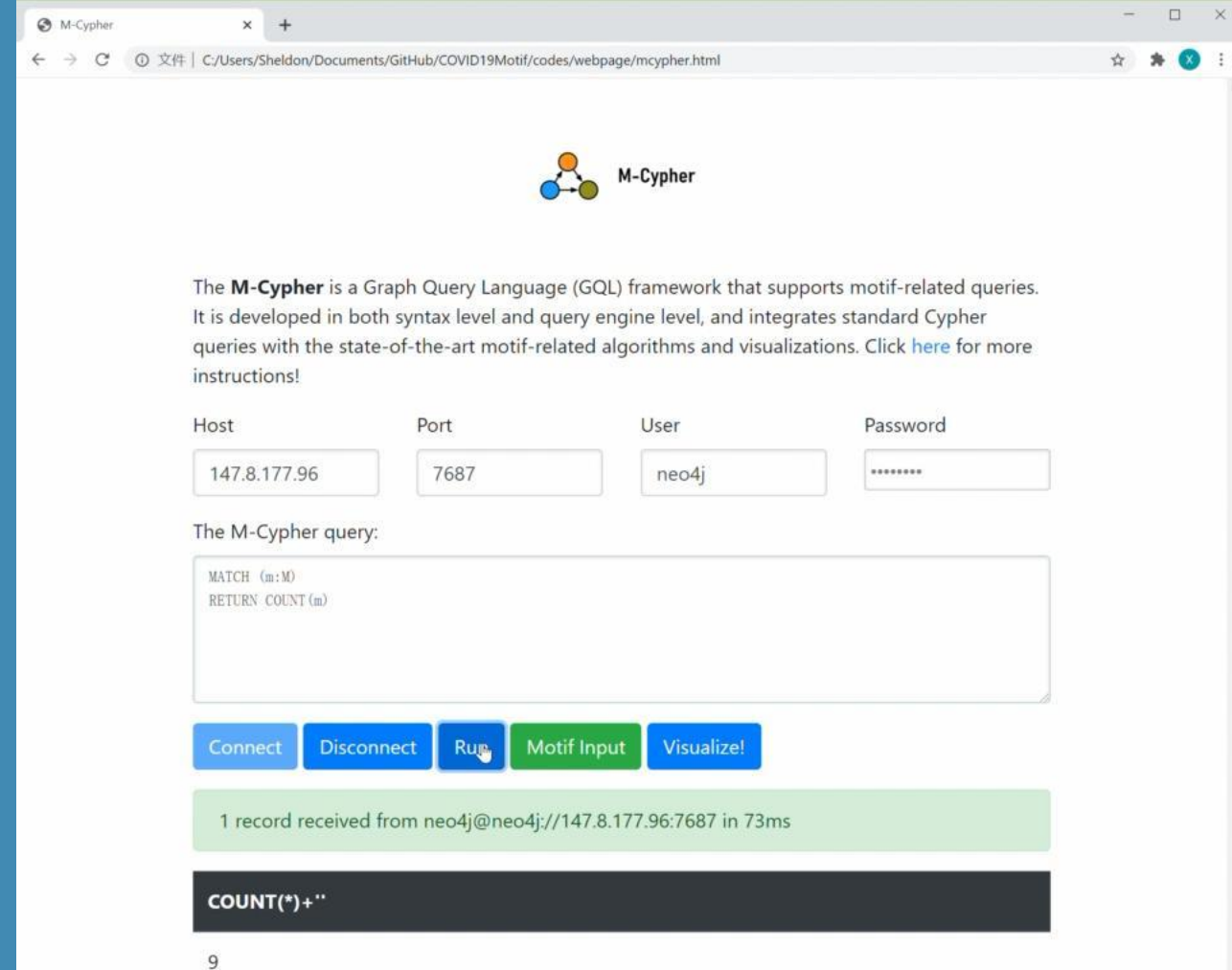
# M-CYPHER: POWERFUL TOOL TO ANALYZE KNOWLEDGE GRAPH

- 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?



The screenshot displays the M-Cypher web application interface. At the top, the browser address bar shows the URL: `C:/Users/Sheldon/Documents/GitHub/COVID19Motif/codes/webpage/mcypher.html`. The page features the M-Cypher logo, which consists of three colored nodes (orange, blue, green) connected by lines, followed by the text "M-Cypher".

Below the logo, a descriptive paragraph states: "The **M-Cypher** is a Graph Query Language (GQL) framework that supports motif-related queries. It is developed in both syntax level and query engine level, and integrates standard Cypher queries with the state-of-the-art motif-related algorithms and visualizations. Click [here](#) for more instructions!"

The interface includes a form for connection details with four fields: "Host" (containing "147.8.177.96"), "Port" (containing "7687"), "User" (containing "neo4j"), and "Password" (containing "\*\*\*\*\*").

Below the form, a text area labeled "The M-Cypher query:" contains the following query:

```
MATCH (m:M)
RETURN COUNT(m)
```

Underneath the query area are five buttons: "Connect", "Disconnect", "Run" (which is highlighted with a mouse cursor), "Motif Input", and "Visualize!".

A green status bar below the buttons displays the message: "1 record received from neo4j@neo4j://147.8.177.96:7687 in 73ms".

At the bottom, a dark gray box shows the query result: `COUNT(*)+"`. Below this box, the number "9" is displayed.

# 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?

The M-Cypher is a Graph Query Language (GQL) framework that supports motif-related queries. It is developed in both syntax level and query engine level, and integrates standard Cypher queries with the state-of-the-art motif-related algorithms and visualizations. Click [here](#) for more instructions!

Host: 147.8.177.96

The M-Cypher

```
MATCH r=(h1:V1)
WHERE h1.name
RETURN t1.name
```

Connect

64 records

protein

M1

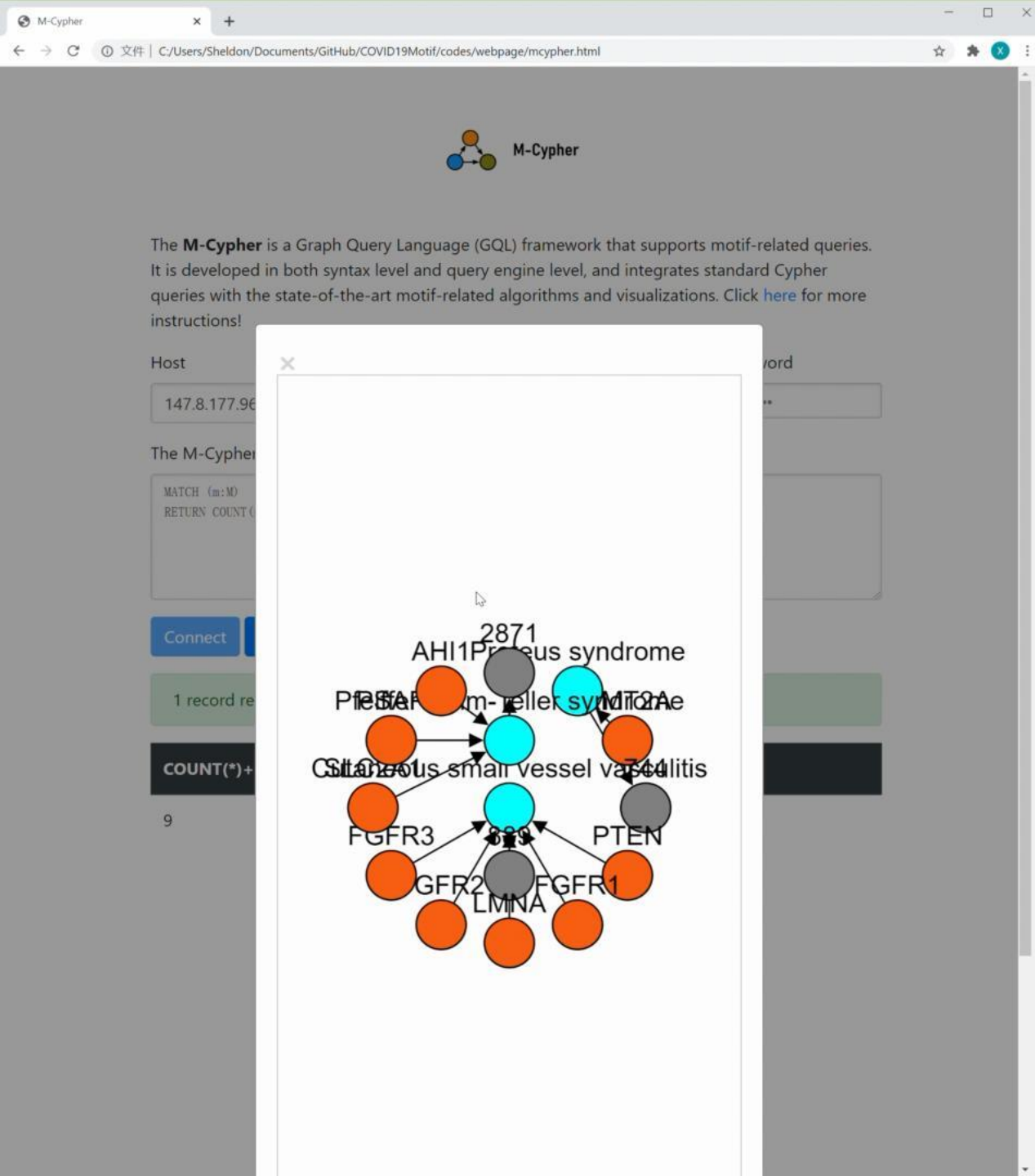
Next Submit Clean

GFPT1	Human	SARS
ATP2A3	Human	SARS
IFRD1	Human	SARS
DCTN2	Human	SARS
XPO1	Human	SARS
HLA-DRB9	Human	SARS
PTPRH	Human	SARS
CLEC4G	Human	SARS

The screenshot shows the M-Cypher web interface. At the top, there's a logo and a description of the tool. Below that, a host address is displayed. A graph query is shown in a text area, and a 'Connect' button is next to it. Below the query, it says '64 records'. A table of results is shown with columns for 'protein', 'Host', and 'Disease'. A modal window is open over the table, showing a graph visualization with nodes 'HostProtein' (orange) and 'Disease' (purple) connected by an edge. A context menu is open over the 'HostProtein' node, showing options: 'remove', 'hide', 'select all nodes', and 'add edge'. The table lists proteins like GFPT1, ATP2A3, IFRD1, DCTN2, XPO1, HLA-DRB9, PTPRH, and CLEC4G, all associated with 'Human' and 'SARS'.

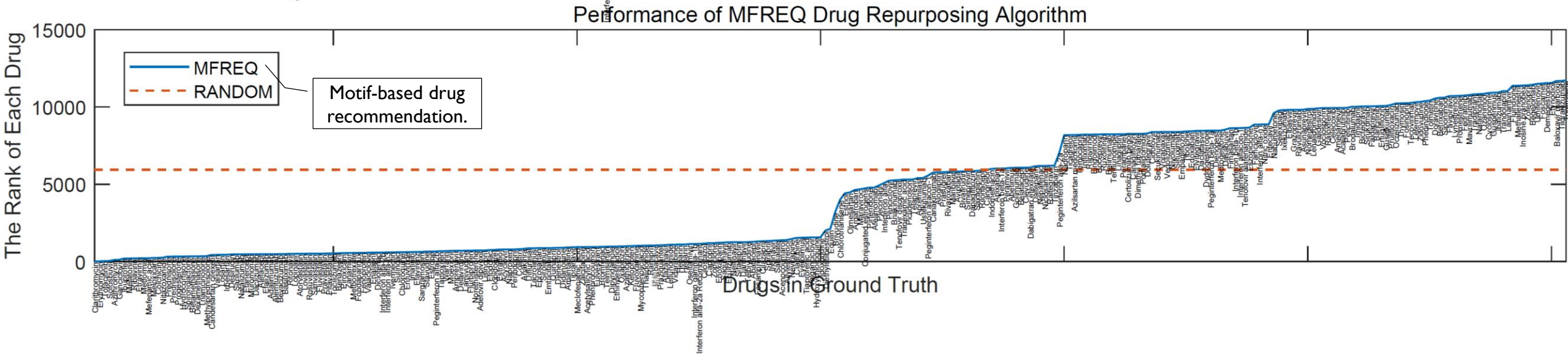
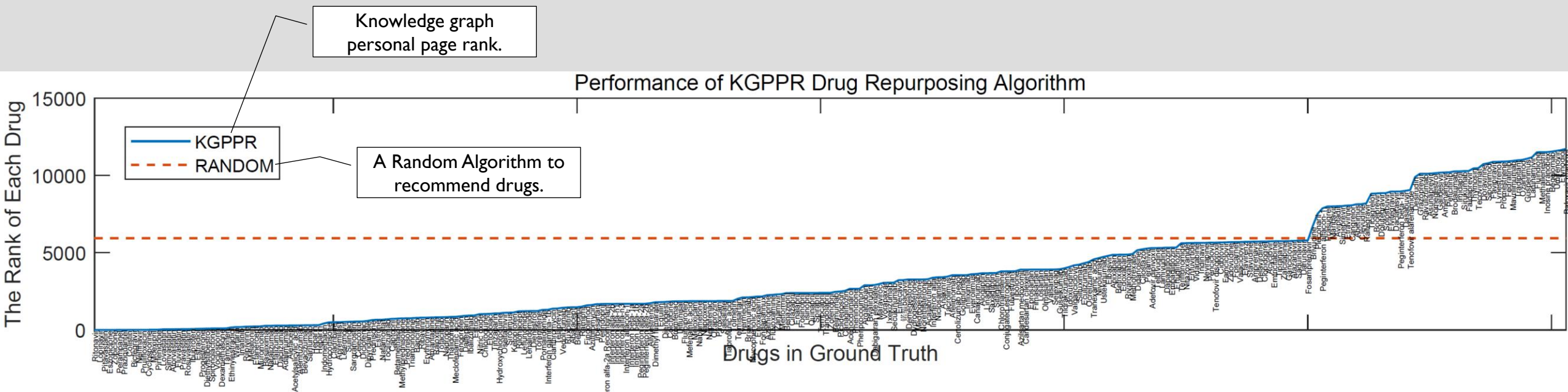
## M-CYPHER: POWERFUL TOOL TO ANALYZE KNOWLEDGE GRAPH

- 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?









# 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**