

COVID-19

MEDICAL LITERATURE ANALYSIS

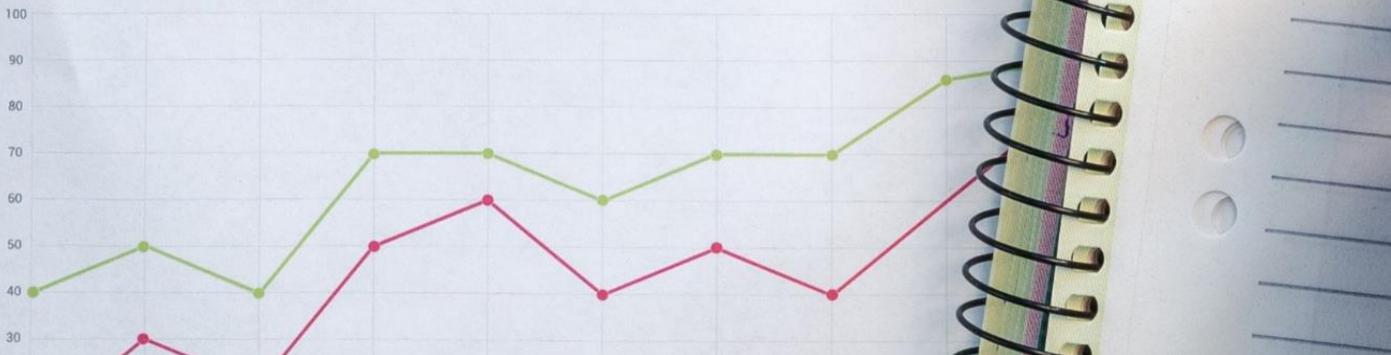
Malaysia Indicator

30th MARCH 2020

25%

50%

75%



THE CLOCK IS TICKING

The number of cases of COVID-19 continue to increase worldwide as scientists march on a race to identify the appropriate drugs treatment and to develop an effective vaccine.

While scientific research requires meticulous details and prolonged-time period, the growing pandemic that has threaten the lives of many requires immediate attention.

No of confirmed cases worldwide



KNOW THY ENEMY

Total Confirmed
529,591

Confirmed Cases in Top 10 Countries

- 83,836 US
- 81,782 China
- 80,589 Italy
- 57,786 Spain
- 43,938 Germany
- 29,551 France
- 29,406 Iran
- 11,812 United Kingdom
- 11,811 Switzerland
- 9,241 Korea, South

Total ACTIVE
297,705

Active Cases in Top 10 Countries

- 62,013 Italy
- 46,406 Spain
- 37,998 Germany
- 22,898 France
- 16,715 Iran
- 11,489 Switzerland
- 11,082 United Kingdom
- 7,027 Netherlands
- 6,748 Austria
- 5,340 Belgium

Total Death
23,970

Death Cases in Top 10 Countries

- 8,215 Italy
- 4,365 Spain
- 3,291 China
- 2,234 Iran
- 1,698 France
- 1,209 US
- 580 United Kingdom
- 435 Netherlands
- 267 Germany
- 220 Belgium

Total RECOVERED
122,150

Recovered in Top 10 Countries

- 74,181 China
- 10,457 Iran
- 10,361 Italy
- 7,015 Spain
- 5,673 Germany
- 4,955 France
- 4,144 Korea, South
- 681 US
- 675 Belgium
- 597 Diamond Princess

On March 11 (2020), the World Health Organisation (WHO) declared COVID-19 a worldwide pandemic. It is arguably the most disruptive global event in recent times. Unlike any other outbreak, COVID-19 is unprecedented in magnitude. The SARS (severe acute respiratory syndrome) outbreak in 2003 was never declared as worldwide pandemic, even when it affected 26 countries.

With the number of cases mounted to 529,591 worldwide, shifting the focus to the scientific community as the source of information that can equip us in confronting this new enemy.

DATA SOURCE

COVID-19 Open Research Dataset (CORD) is free, open resource of over 29000 full texts about **COVID-19** and the **coronavirus family of viruses**.

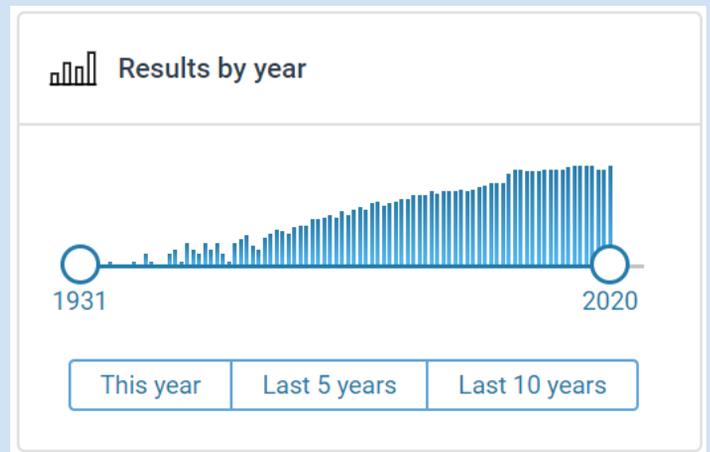
The dataset contains all COVID-19 and coronavirus-related research (e.g. SARS, MERS, etc.) from the following sources:

1. **PubMed's PMC open access corpus**
2. **Additional COVID-19 research articles from a corpus maintained by the WHO**
3. **bioRxiv and medRxiv pre-prints**

The dataset is updated weekly with the latest update is on 20 March 2020.



The screenshot shows the Semantic Scholar website with a search bar at the top. The main heading is "COVID-19 Open Research Dataset (CORD-19)" with a sub-heading "Access this dataset to help with the fight against COVID-19". Below this, there is a section titled "A Free, Open Resource for the Global Research Community" with a paragraph of text and a small graphic of a virus particle. The text describes the dataset's origin and purpose.



KEYWORDS SELECTION

Using the keywords: **COVID-19**, **SARS-CoV-2** and **2019-nCoV**, 974 medical research papers are retrieved from the database on 25th March 2020.

COVID-19

SARS-CoV-2

2019-nCoV

REFERENCES

A comprehensive clinical healthcare terminology directory is used to analyse the medical literature. With massive classification of medical terms and references captured from open source data intelligence, the lexicon serves as a learning platform to establish *Artificial Intelligence-based* data analytics result.

KEYWORDS HIERARCHY

Clinical finding
(finding)

Procedure
(procedure)

Organism
(organism)

Substance
(substance)

Pharmaceutical /
biologic product
(product)

Utilising the top five medical term hierarchies, the medical research paper is analysed to explore observable patterns in medical research related to COVID-19.

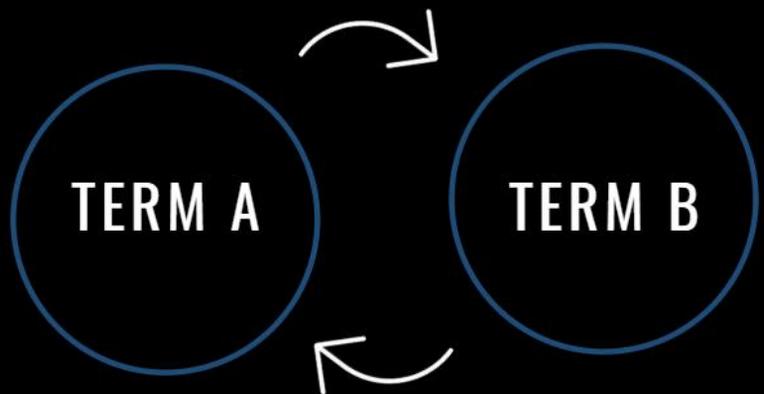
IDENTIFYING CORRELATION

ASSOCIATION RULES

SUPPORT A = Default popularity of keyword A.

CONFIDENCE A \rightarrow B = The likelihood of keyword B to appear if keyword A appeared.

RULE 1
(SUPPORT A / CONFIDENCE (A \rightarrow B))



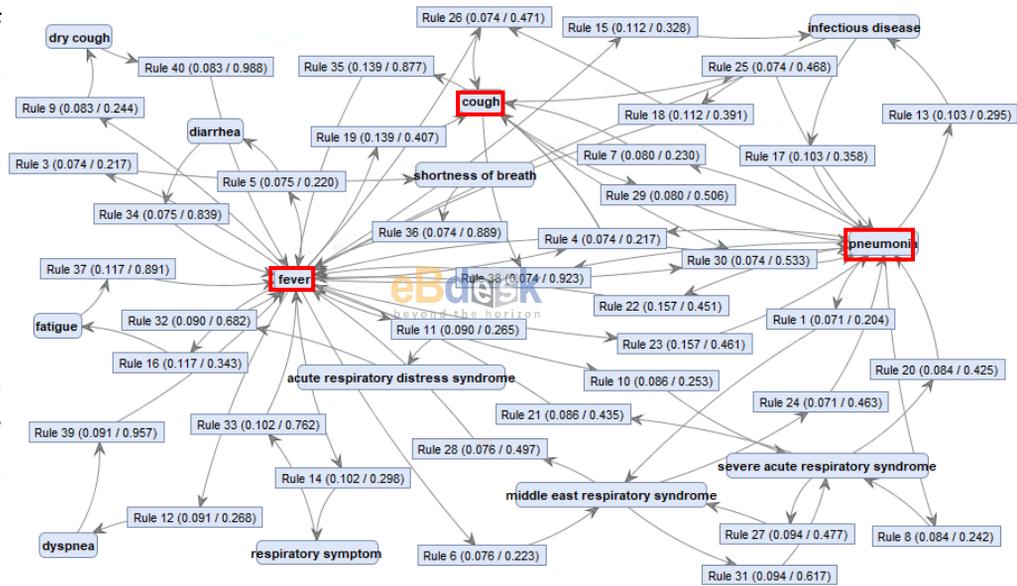
RULE 2
(SUPPORT B / CONFIDENCE (B \rightarrow A))

Association Rule is an analysis method for finding specific relation patterns between objects in large dataset. Using association rule, the likelihood of terms to appear together is quantified based on support and confidence rule of each term.

Based on 974 papers from CORD-19 database related to COVID-19 keywords, the association graph of each hierarchy is mapped and studied.

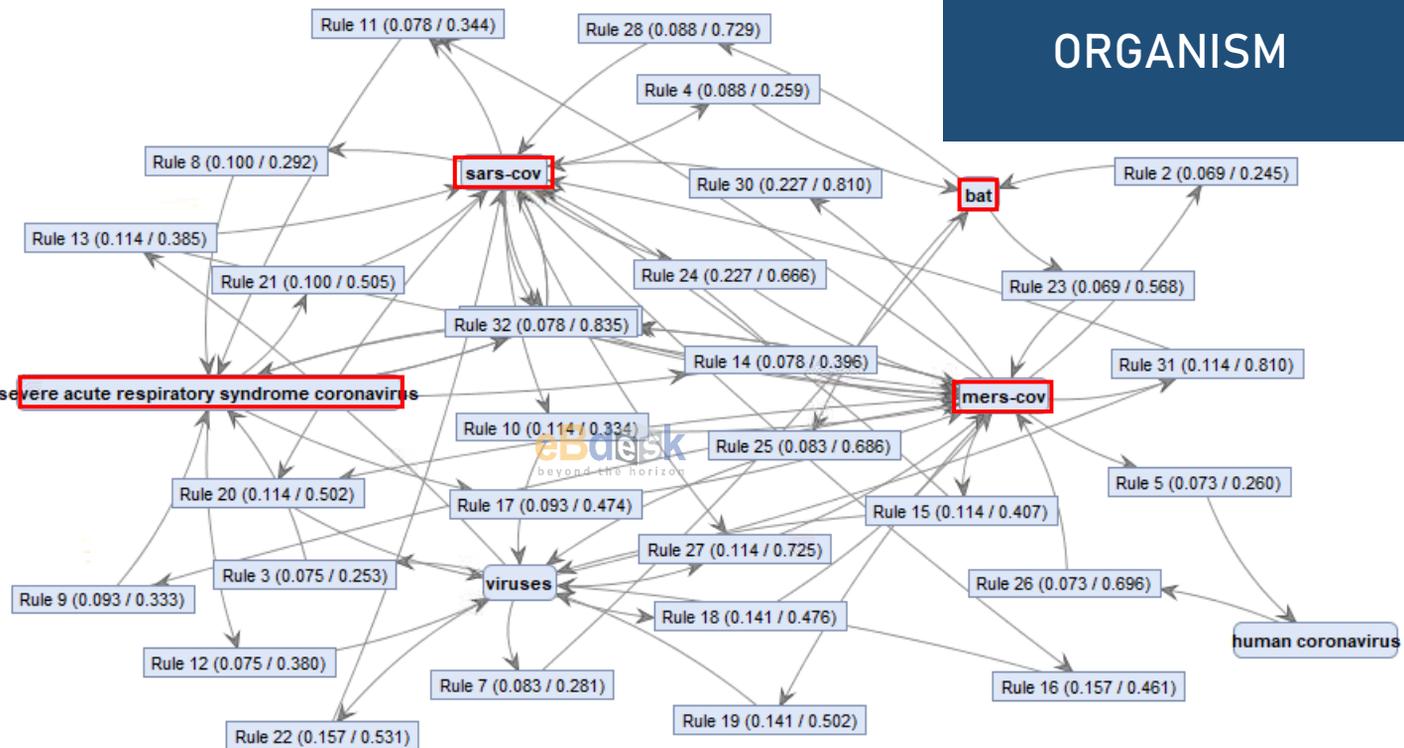
CLINICAL FINDING

As reflective of the clinical symptoms of COVID-19, the words **'fever'**, **'cough'** and **'pneumonia'** appear to be frequently related to many terms in the retrieved papers. Since most infection will appear as mild to moderate respiratory illness, other keywords as **'fatigue'** and **'respiratory symptoms'** also appear quite frequently.



Powered by **eBdesk** Big Data and Artificial Intelligence

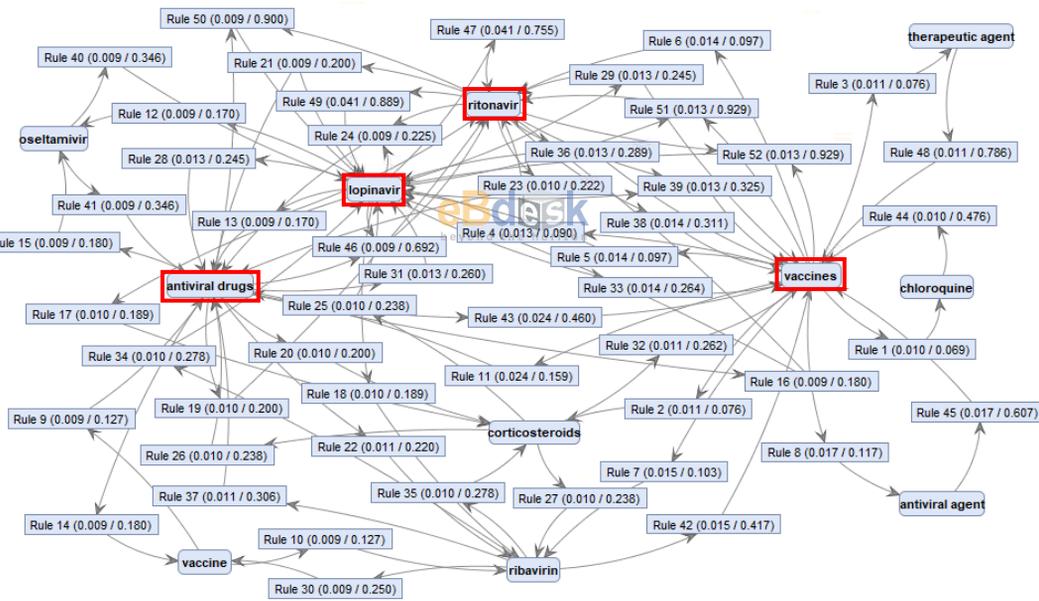
ORGANISM



From the retrieved papers the viruses mentioned are not only limited to COVID-19. Other coronaviruses that have contributed to **Severe Acute Respiratory Syndrome** as **'SARS-COV'** and **'MERS-COV'** appear frequently with multiple associations. Aside from the virus itself, the most likely reservoir of COVID-19, **'bat'** is also frequently mentioned and related to other keywords.

PHARMACEUTICAL

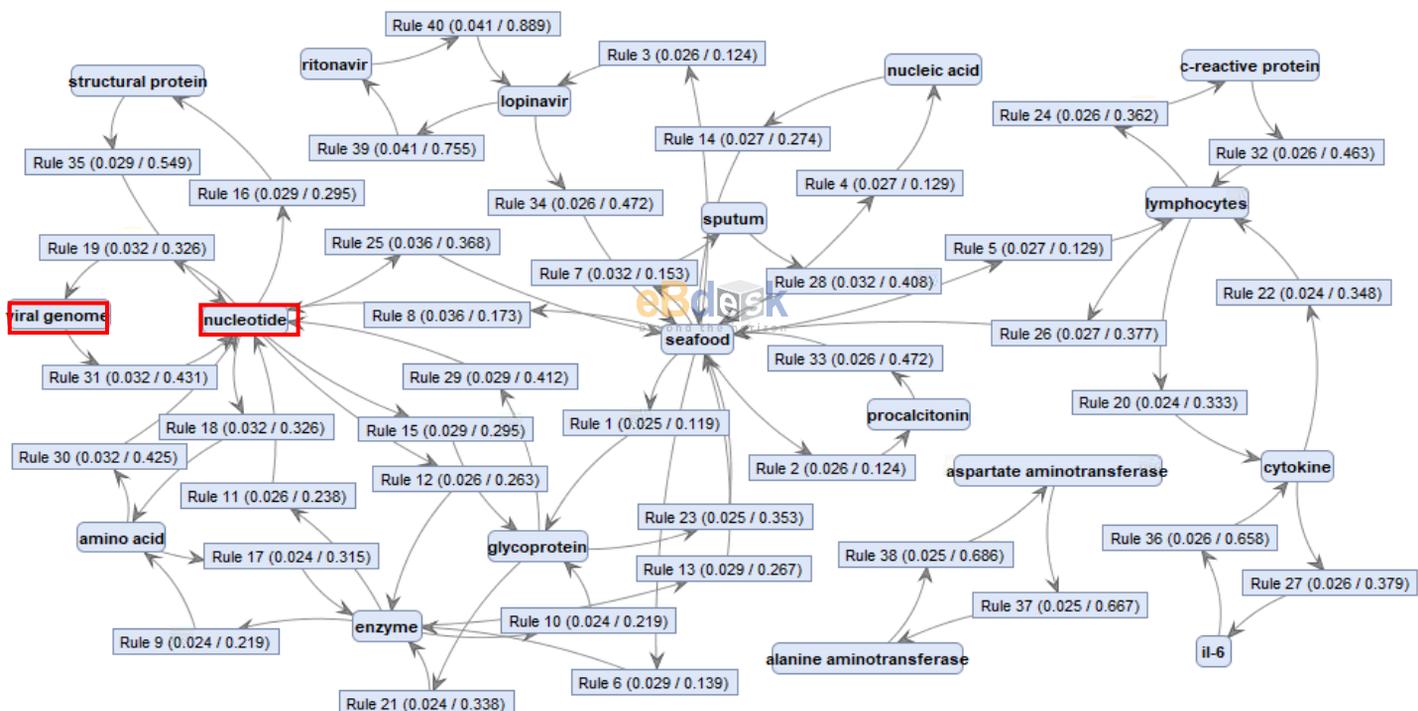
The scientific community continues to struggle to find the best drugs to treat COVID-19. Among the antiviral drugs recommended are 'lopinavir' and 'ritonavir', as seen by its high support value. Widely used for the treatment of HIV and AIDS, the recommendation remains debatable in the scientific sphere.



Powered by eBdesk Big Data and Artificial Intelligence

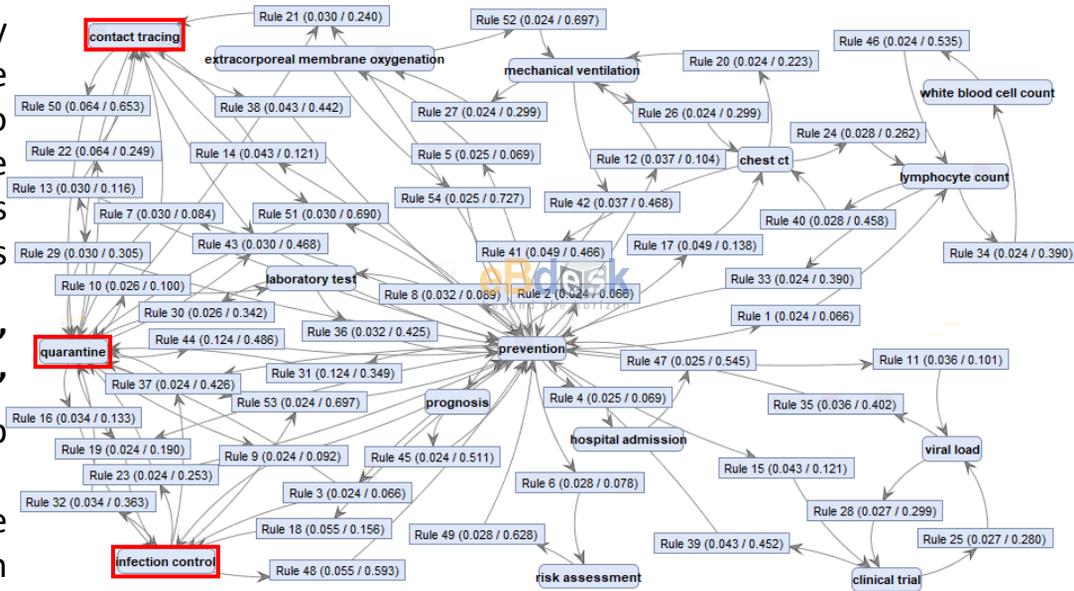
While the drug names also dominate the substance hierarchy as the most mentioned terms, 'nucleotide' and 'viral genome' keywords are also highly mentioned and relate to others. The knowledge on the virus's unique RNA sequence shared by China has helped other countries conduct relevant diagnostic tests in the community.

SUBSTANCE



PROCEDURE

The necessary measure to curb the pandemic is also discussed in the scientific papers with the terms 'quarantine', 'infection control' and 'contact tracing' among the top measures mentioned and have the most correlation to other measures.



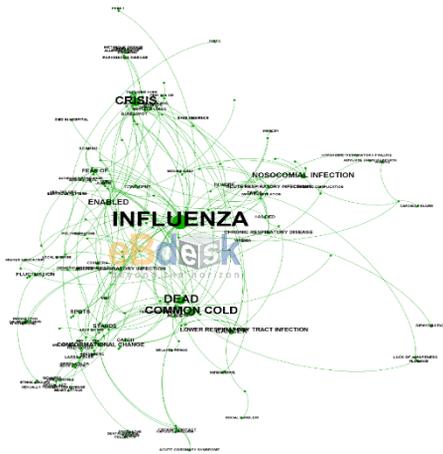
Powered by **eBdask** Big Data and Artificial Intelligence
BEYOND THE HORIZON

NETWORK ANALYSIS

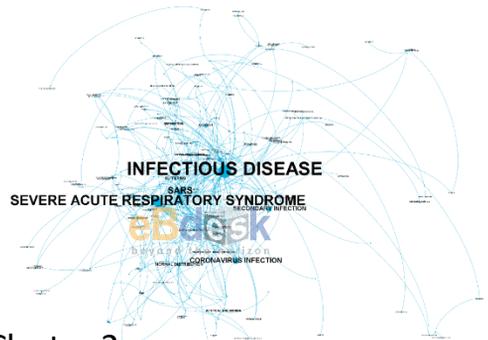
Network analysis will further illustrate the relationship between the research papers and the related medical terms. The most frequent medical term associated with a cluster of research paper is represented with larger size compared to others. The different cluster represents research papers that are group together by its associated keywords.

COVID-19 NETWORK ANALYSIS

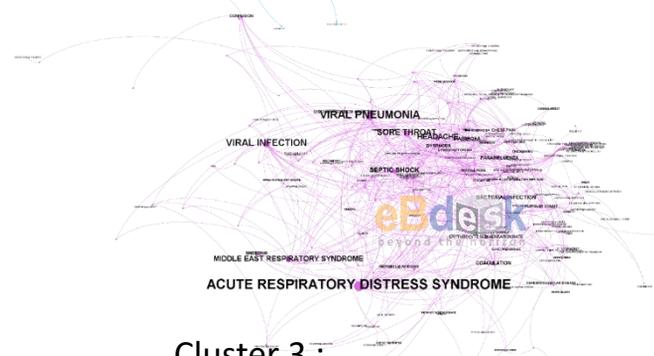
CLINICAL FINDING



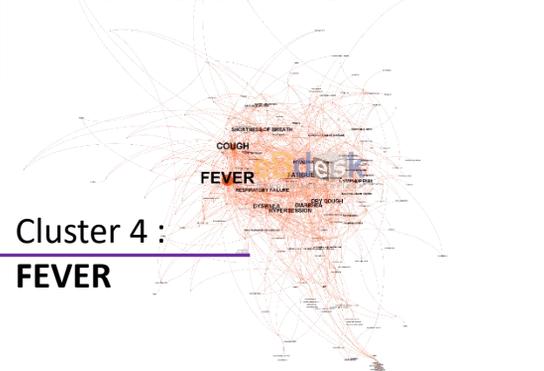
Cluster 1:
INFLUENZA



Cluster 2:
INFECTIOUS DISEASE



Cluster 3 :
**ACUTE RESPIRATORY
DISTRESS SYNDROME**



Cluster 4 :
FEVER

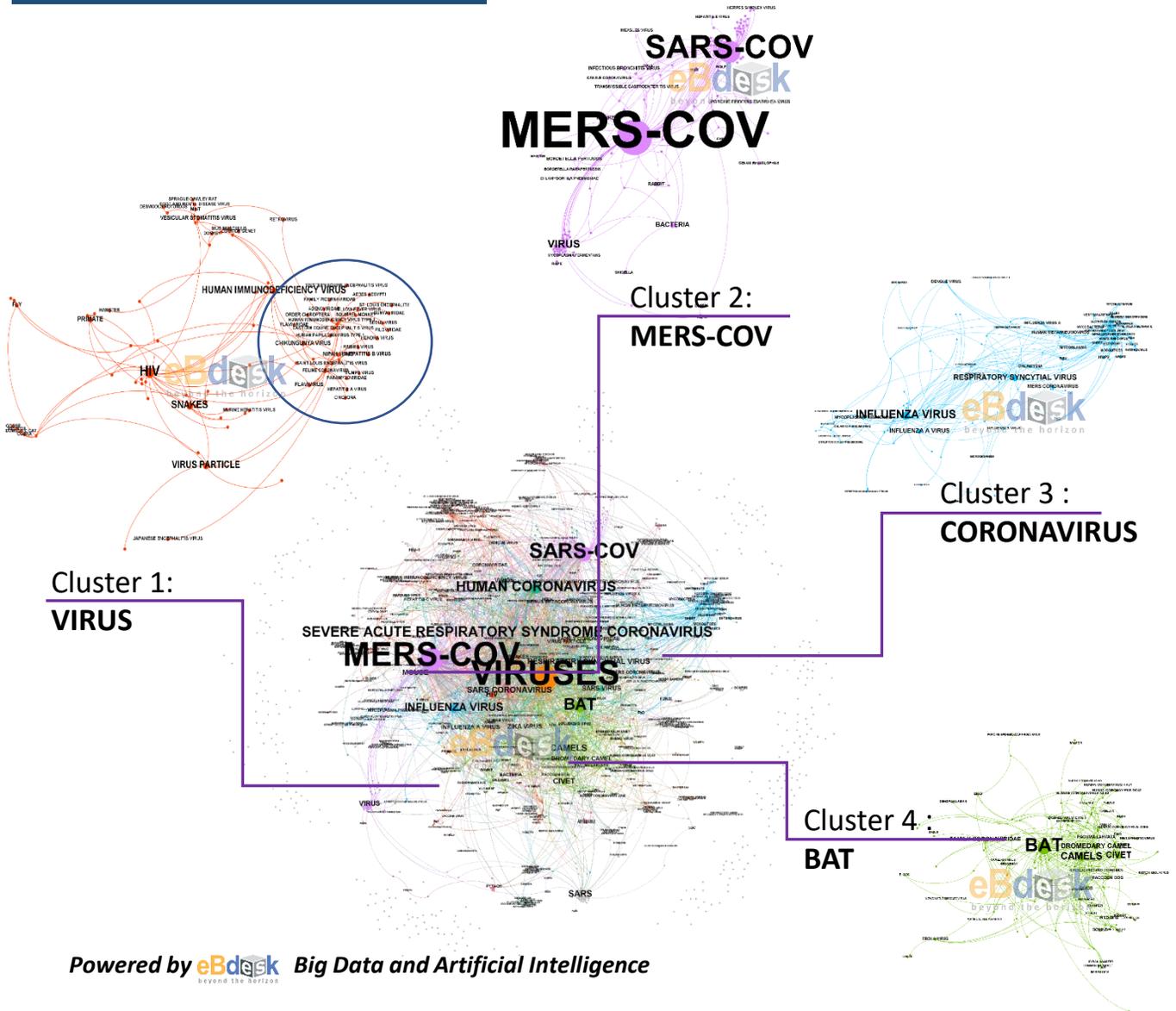
Powered by **eBdesk** Big Data and Artificial Intelligence

Based on 974 papers related to COVID-19, the network cluster of Clinical Finding terms are shown as follows. The overall dominating term is 'Fever' while the prominent terms in each cluster are as follow.

The first cluster represents the group research papers related to **influenza**. The second cluster mainly correlated on the COVID-19 status as infectious disease that cause **severe acute respiratory syndrome**. The third cluster emphasizes more on the clinical symptoms in the case of **acute respiratory distress** while the fourth cluster represents the correlation on **milder symptoms** as **fever**.

COVID-19 NETWORK ANALYSIS

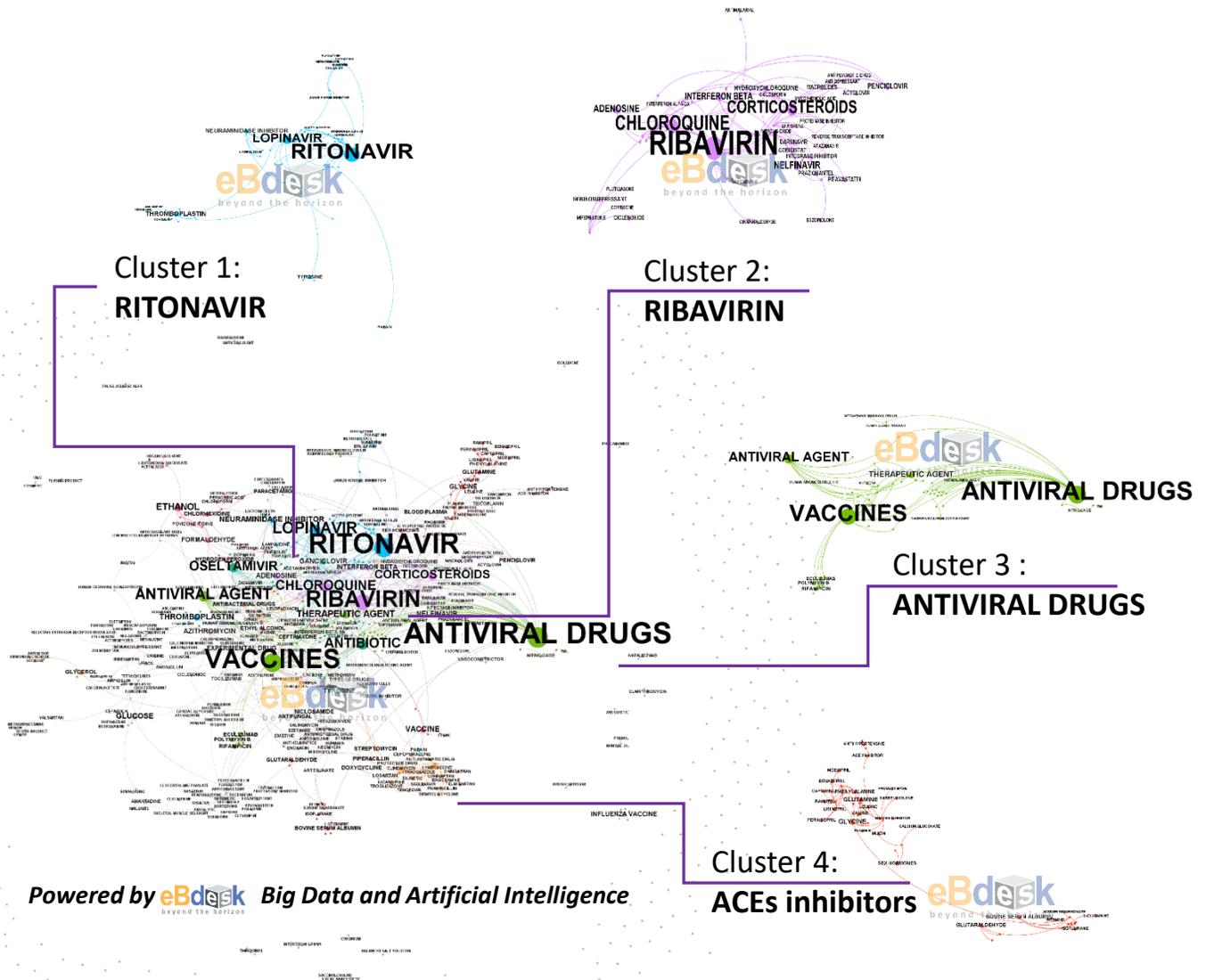
ORGANISM



The network of Organism terms based on the retrieved papers is shown as follows. Many **virus species** was mentioned in the first research papers cluster, perhaps in comparing the clinical presentation of the viral disease. The 2nd cluster was mainly preoccupied with **MERS-COV** and **SARS-COV**, the two coronaviruses that were responsible for Middle East Respiratory Syndrome and Severe Acute Respiratory Syndrome respectively. The third cluster on **coronavirus** represent medical research on the family of viruses that cause respiratory symptoms. Given **the animal origin** of the coronavirus outbreak, the fourth cluster is mainly correlated in term of the animal mentioned.

COVID-19 NETWORK ANALYSIS

PHARMACEUTICAL

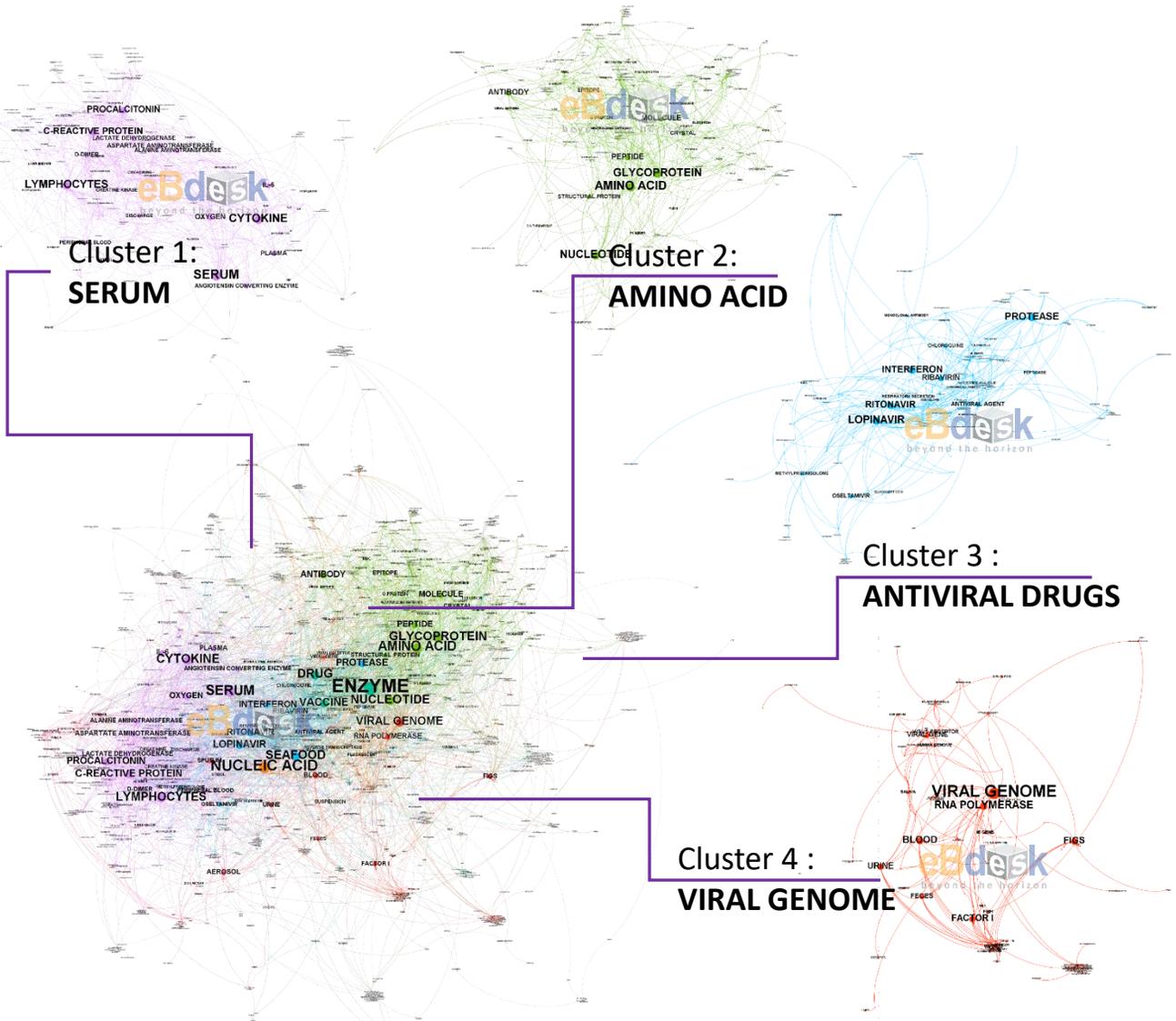


Based on the papers, the network of Pharmaceutical terms is shown above. While there is currently no specific treatment for coronaviruses, the symptoms can be treated. But the medical literature has been trying to identify potential drug treatment. Each cluster represents drugs experimented or recommended by the researchers. Among the most mentioned are **Ritonavir** and **Lopinavir** in the first cluster and **Ribavirin** in the second cluster. Until now, the quest remains open as conflicting medical literature available on the efficiency of each antiviral drugs.

The third cluster mentions more on **vaccines** alongside antiviral drugs. The fourth cluster, despite few numbers of paper, represents a vital discovery on **ACE inhibitors risk** for COVID-19 patients. Patients taking the drugs to treat high blood pressure might be at higher risk of death although it is still debatable.

COVID-19 NETWORK ANALYSIS

SUBSTANCE



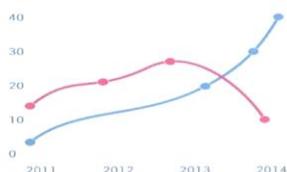
Powered by **eBdask** Big Data and Artificial Intelligence
beyond the horizon

The network of Substance terms based on the retrieved papers is shown as follows. Identifying the nature of the COVID-19 virus has been an important part of the research for diagnostic and treatment purpose. As such the keywords as **amino acid** and **viral genome** display the research surrounding the structure and subtype of the RNA virus. The virus genome was rapidly sequenced by Chinese researchers in the quest to help developing a diagnostic test. Yet there are still many rooms for research in the genomic nature of the virus.

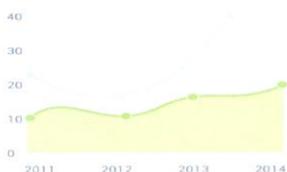
Malaysia Indicator

Morris Charts

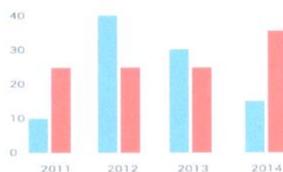
Line Chart



Area Chart



Bar Chart



Donut Chart



Sparkline Charts

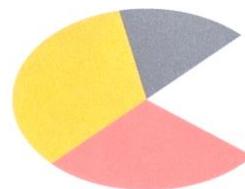
Line Chart



Bar Chart



Pie Chart



Easy Pie Charts



We are research-based entity focusing on open source data intelligence. Leveraging on Big Data and Artificial Intelligence, we establish various data storytelling in many context such as price intelligence, social listening, economic intelligence and media monitoring.

www.malaysiaindicator.com



FACEBOOK
[@malaysiaindicator](https://www.facebook.com/malaysiaindicator)



YOUTUBE
[@malaysiaindicator](https://www.youtube.com/malaysiaindicator)