Sentiment and Emotion analysis with topic modelling of COVID-19 tweets – **PROJECT 1**

Chilukuri, Susmitha

Supervisor: Yu, Fei

Advisor: Gibson, Amelia

Introduction

- Infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
- First identified in December 2019 – Wuhan, Hubei, China
- 2/11/2020 – name announced as COVID-19
- 3/2020 – Pandemic declaration by WHO

Impact of COVID-19

https://covid19.who.int/

As per 11/8/2020, 4.30 pm CET,
49,578,590 confirmed cases
1,245,717 deaths

Effects of COVID-19

- Physiologically
- Psychologically
- Economically
- Environmentally
- Politically, etc.

Objective

To conduct sentiment analysis of information posted on a social media platform about COVID-19 pandemic.

Research questions

- What has been posted and shared among the public?
- What are the publics major interests and perceptions?
- Strategies to use social media to promote health education and fight misinformation and disinformation

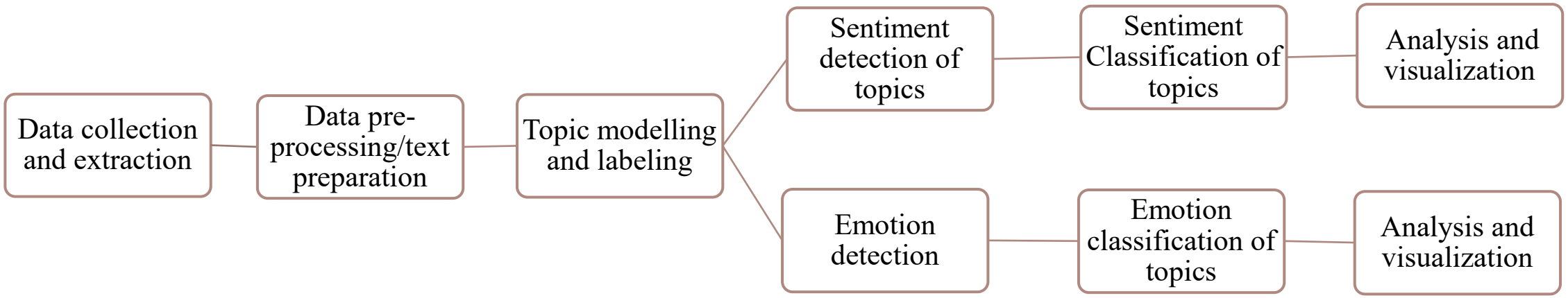
Why sentiment and emotion analysis?

Overview of wider public opinion

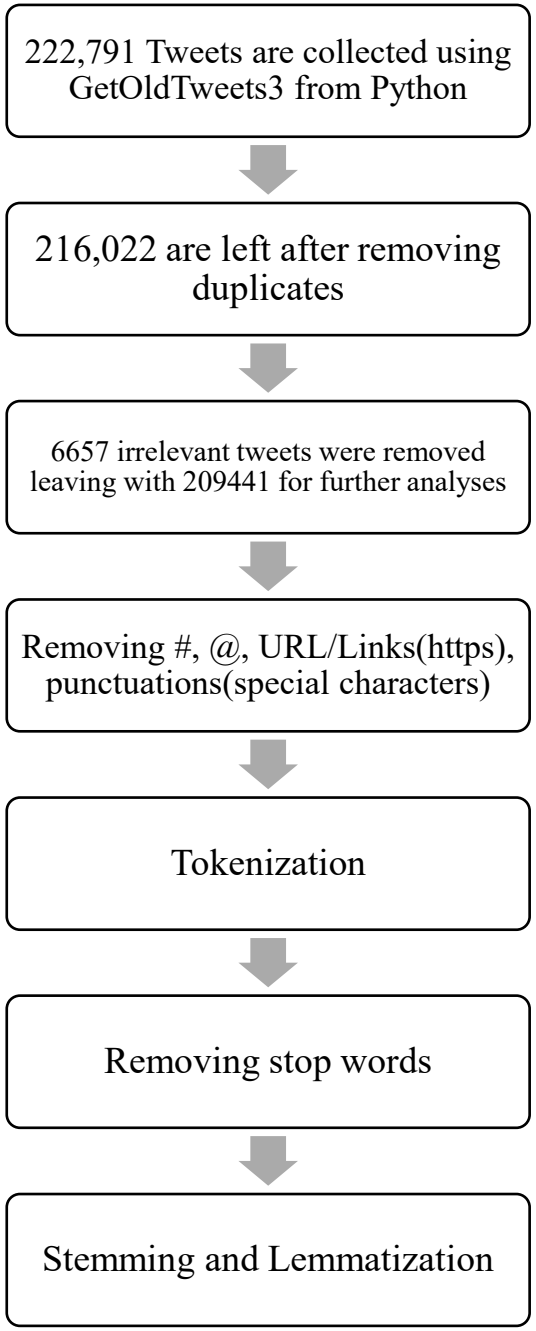
Method:

- Python
- Twitter - Twitter API - GetOldTweets3
- 1/1/2020 – 7/29/2020
- More than 200,000 tweets
- *“Coronavirus”* OR *“COVID”* OR *“Hubei”* OR *“SARS-CoV-2”* OR *“2019nCoV”* OR *“Wuhan”* OR *“SARS-Coronavirus-2”* OR *“SARSCoV2”* OR *“nCoV2019”* OR *“Pandemic”* OR *“nCoV-2019”*

Project workflow model



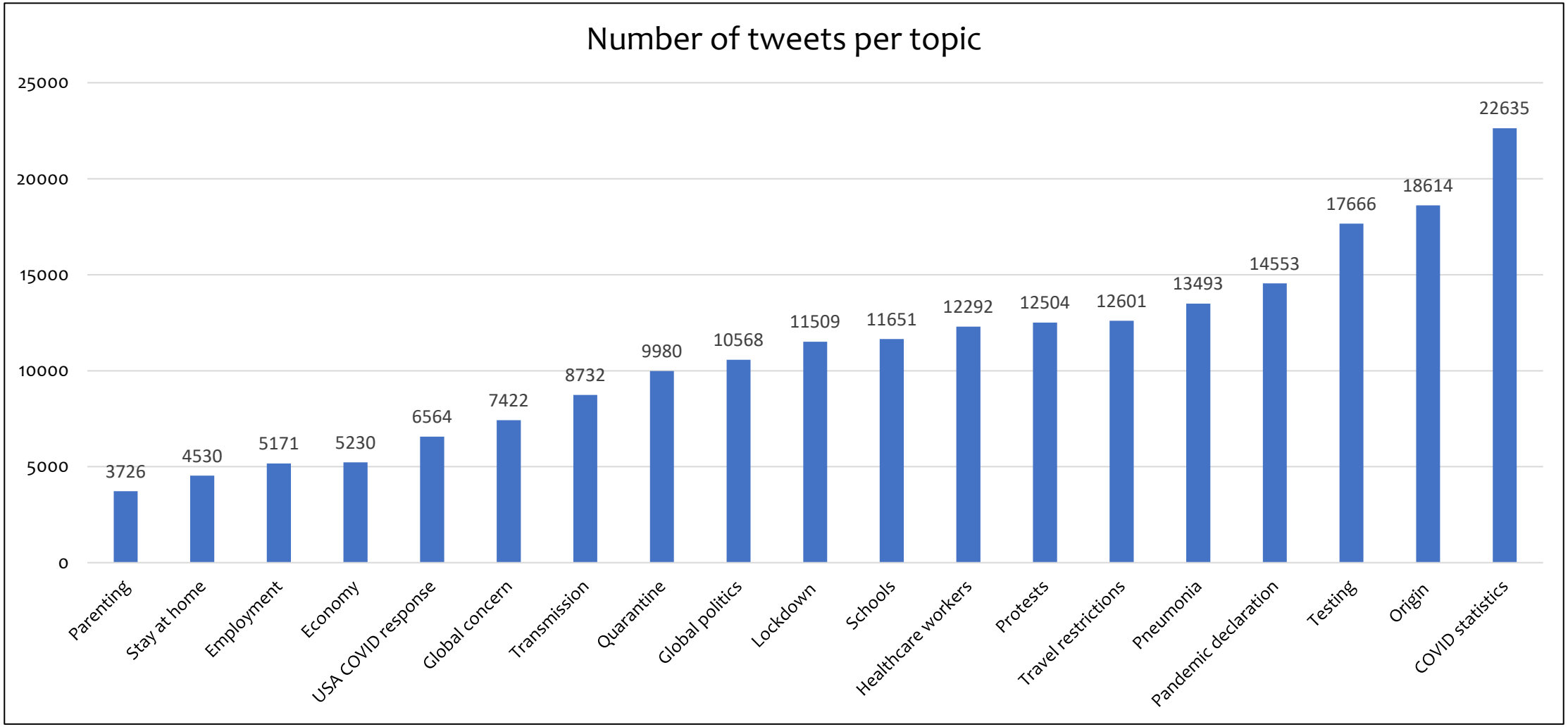
Data extraction and preprocessing

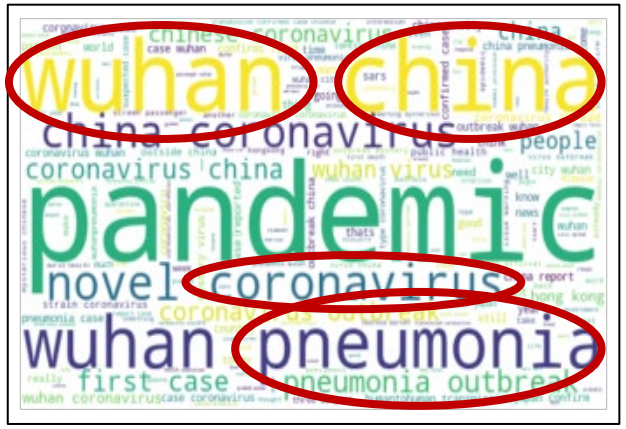


Topic modelling

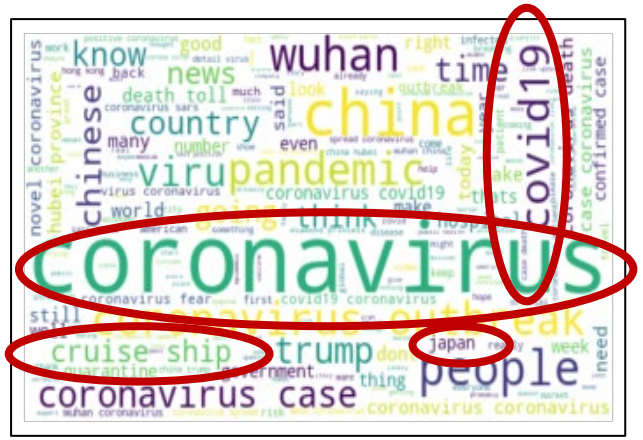
19 topics – grouped into 3 themes

CHINA	EFFECTS OF COVID-19 ON PEOPLE	GLOBAL EFFECTS OF COVID-19
• Outbreak/Pneumonia • Origin	• Travel restrictions • Healthcare workers • Schools • Stay at home • Employment • Parenting • Quarantine • Testing • Protests	• Lockdown • Economy • Global concerns • Global politics • USA COVID response • Pandemic declaration • Transmission • COVID statistics

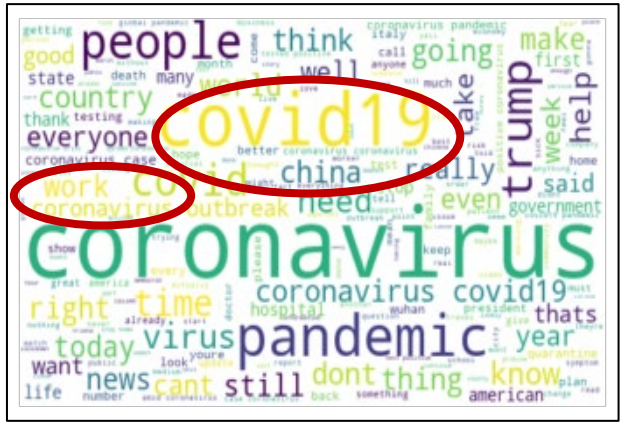

Number of tweets per topic
25000
20000
15000
10000
5000
0
3726
4530
5171
5230
6564
7422
8732
9980
10568
11509
11651
12292
12504
12601
13493
14553
17666
18614
22635
Parenting
Stay at home
Employment
Economy
USA COVID response
Global concern
Transmission
Quarantine
Global politics
Lockdown
Schools
Healthcare workers
Protests
Travel restrictions
Pneumonia
Pandemic declaration
Testing
Origin
COVID statistics



January 1st – 31st, 2020

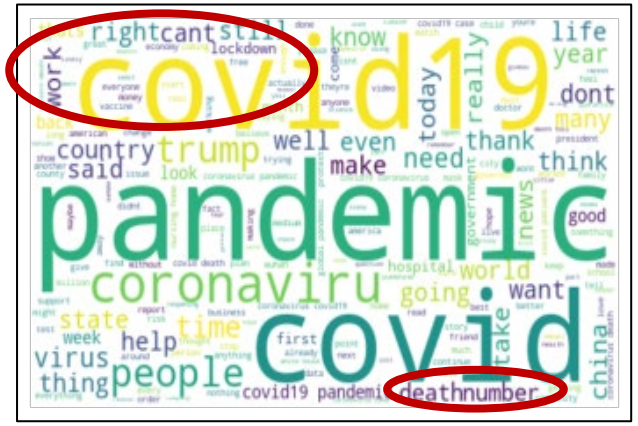


February 1st – 29th, 2020

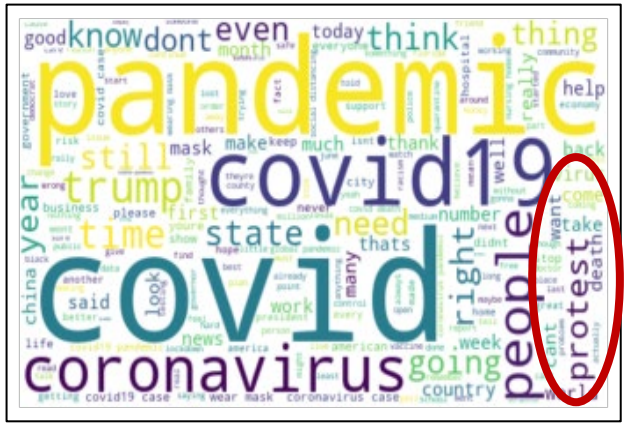


March 1st – 31st, 2020

April 1st – 30th, 2020



May 1st – 31st, 2020



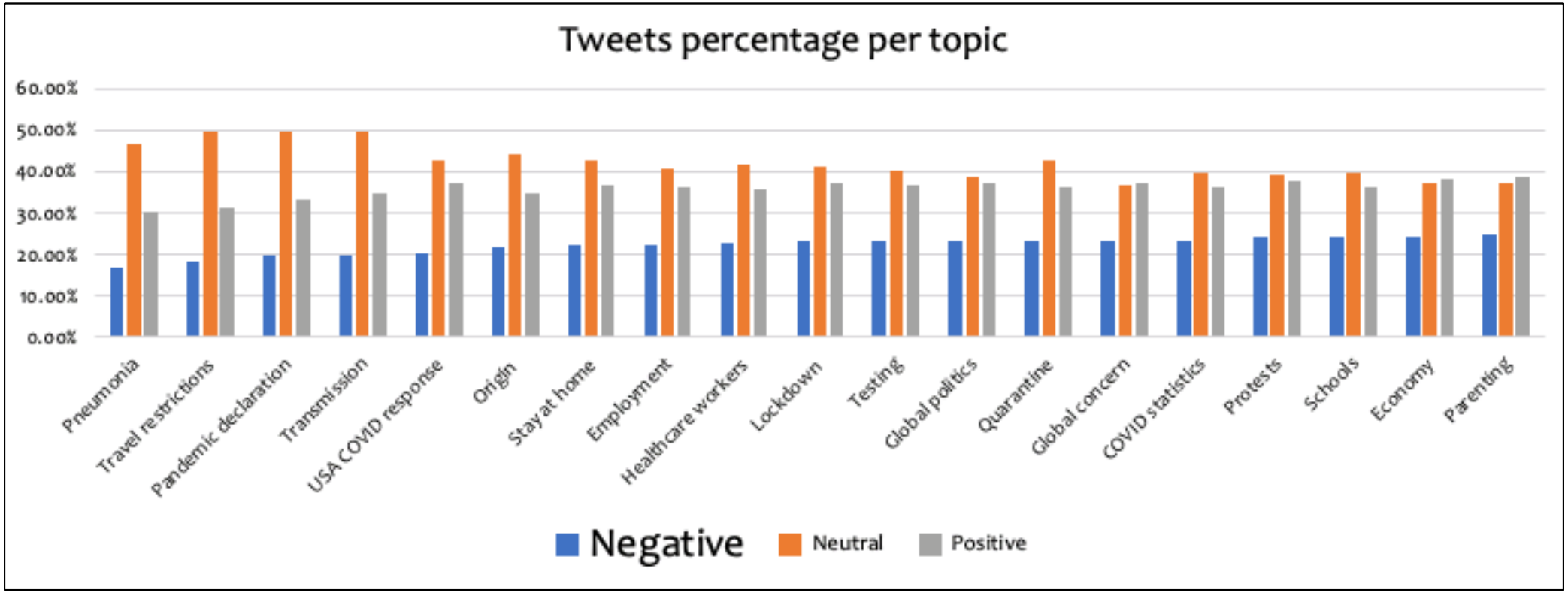
June 1st – 30th, 2020

July 1st – 29th, 2020

Sentiment analysis - TextBlob

3 classes – based on polarity score

- Positive (>1)
- Negative (<1)
- Neutral (0)

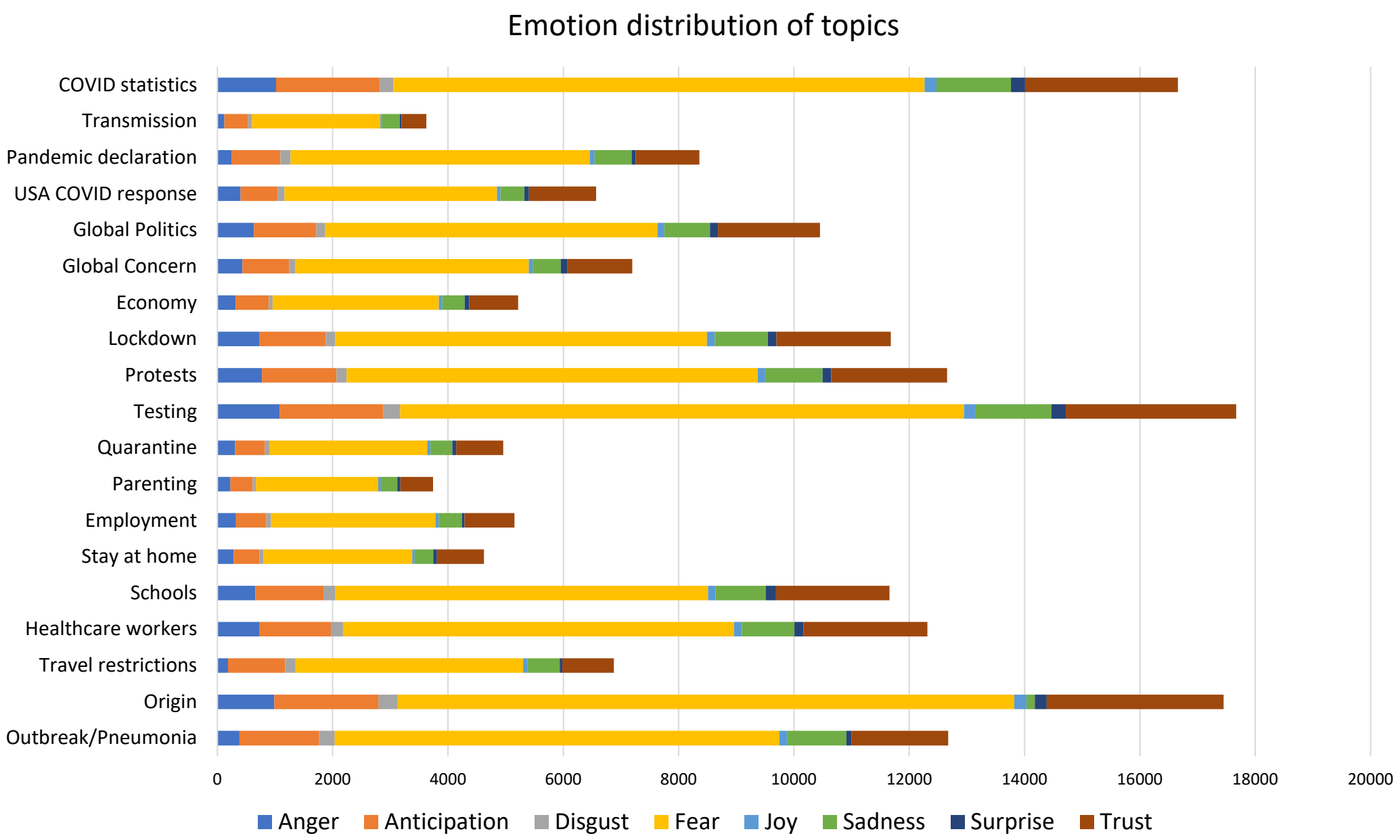

Tweets percentage per topic
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
Pneumonia
Travel restrictions
Pandemic declaration
Transmission
USA COVID response
Origin
Stay at home
Employment
Healthcare workers
Lockdown
Testing
Global politics
Quarantine
Global concern
COVID statistics
Protests
Schools
Economy
Parenting
Negative
Neutral
Positive

Emotion classification - NRCLex

- Fear
- Trust
- Anger
- Anticipation
- Surprise
- Sadness
- Disgust
- Joy

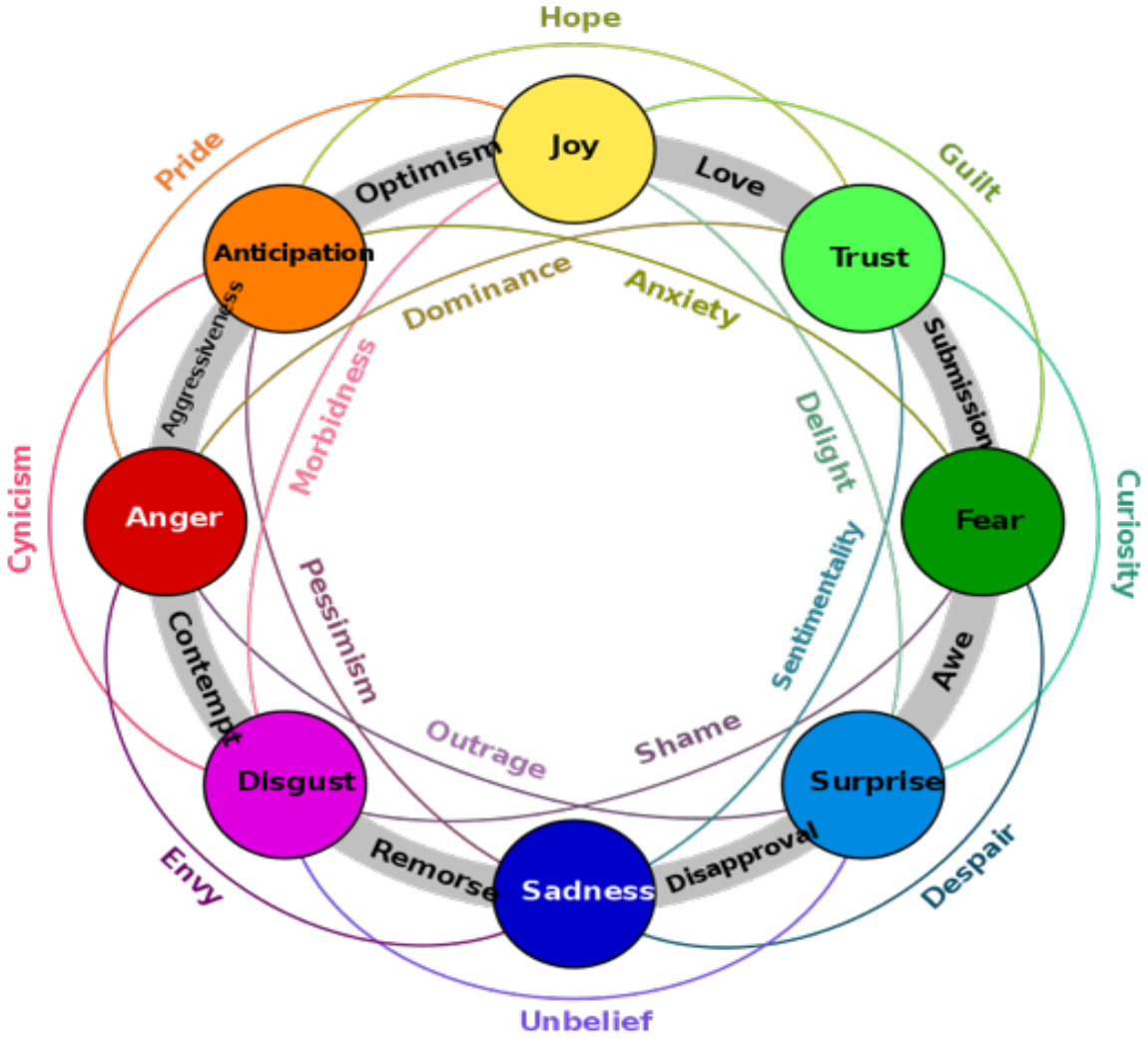
Emotion	Maximum expressed topic	Least expressed topic
Anger	Lockdown	Transmission
Anticipation	Economy	Transmission
Disgust	Outbreak/Pneumonia	Transmission
Fear	Origin	Transmission
Joy	Parenting	Transmission
Sadness	Lockdown	Origin
Surprise	Global concern	Transmission
Trust	Stay at home	Transmission

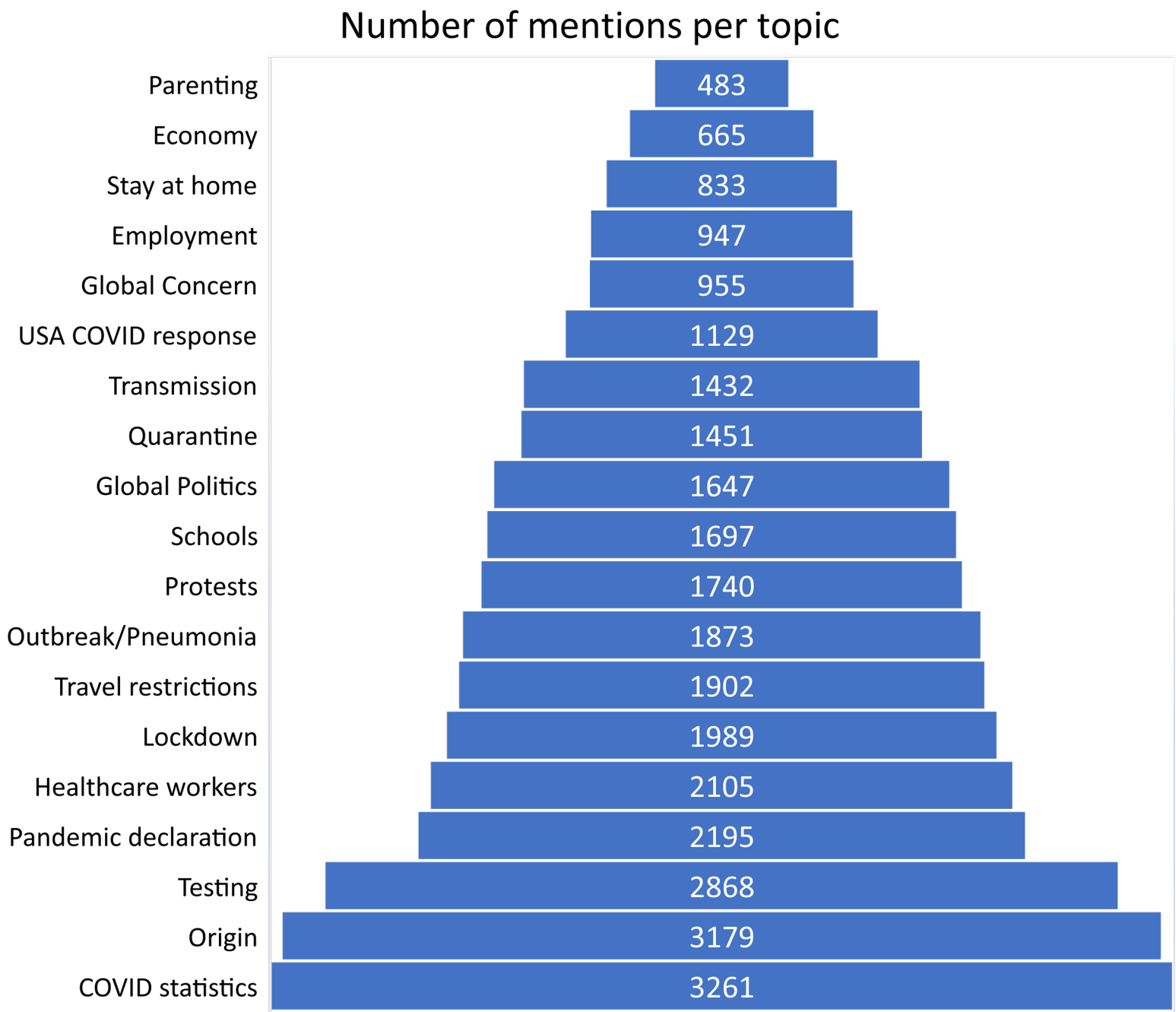
Topic	Highest emotion	Lowest emotion
Outbreak/Pneumonia	Fear	Surprise
Origin	Fear	Sadness
Travel restrictions	Fear	Surprise
Healthcare workers	Fear	Joy
Schools	Fear	Joy
Stay at home	Fear	Joy
Employment	Fear	Surprise
Parenting	Fear	Joy
Quarantine	Fear	Joy
Testing	Fear	Joy
Protests	Fear	Joy
Lockdown	Fear	Joy
Economy	Fear	Joy
Global concern	Fear	Joy
Global politics	Fear	Joy
USA COVID response	Fear	Joy
Pandemic declaration	Fear	Surprise
Transmission	Fear	Surprise
COVID statistics	Fear	Joy

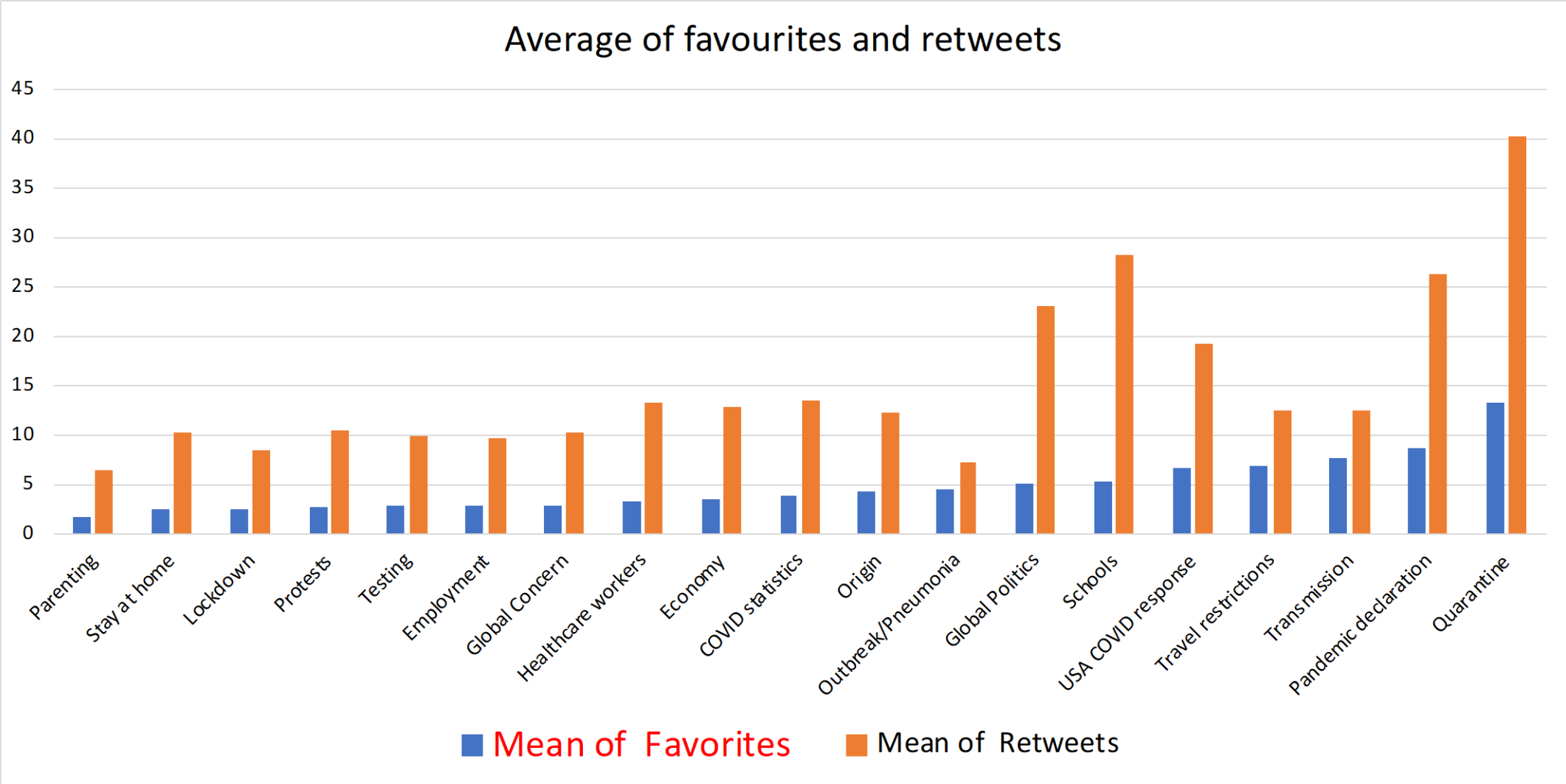

Emotion distribution of topics
COVID statistics
Transmission
Pandemic declaration
USA COVID response
Global Politics
Global Concern
Economy
Lockdown
Protests
Testing
Quarantine
Parenting
Employment
Stay at home
Schools
Healthcare workers
Travel restrictions
Origin
Outbreak/Pneumonia
0
2000
4000
6000
8000
10000
12000
14000
16000
18000
20000
Anger
Anticipation
Disgust
Fear
Joy
Sadness
Surprise
Trust

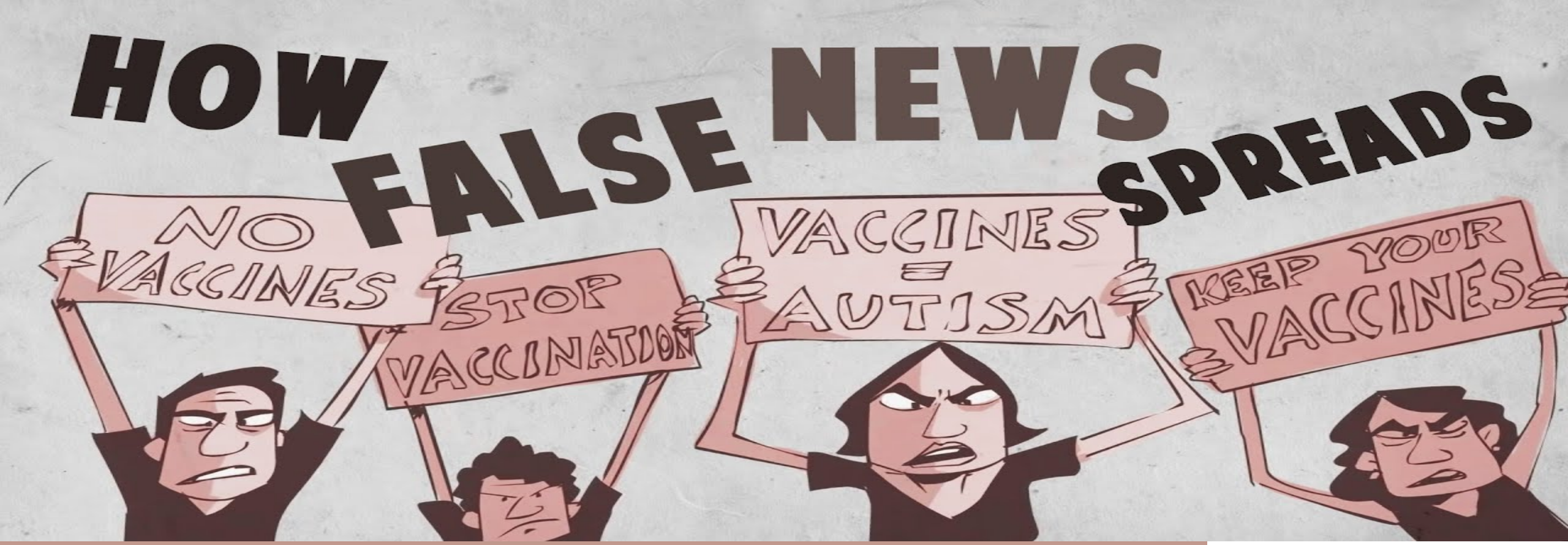
Plutchik wheel of dyads

Disgust is most expressed for Outbreak/Pneumonia which when linked with elements within it explains the ultimately leading xenophobia developed towards Chinese(real-world scenario).




Number of mentions per topic
Parenting
483
Economy
665
Stay at home
833
Employment
947
Global Concern
955
USA COVID response
1129
Transmission
1432
Quarantine
1451
Global Politics
1647
Schools
1697
Protests
1740
Outbreak/Pneumonia
1873
Travel restrictions
1902
Lockdown
1989
Healthcare workers
2105
Pandemic declaration
2195
Testing
2868
Origin
3179
COVID statistics
3261


Average of favourites and retweets
45
40
35
30
25
20
15
10
5
0
Parenting
Stay at home
Lockdown
Protests
Testing
Employment
Global Concern
Healthcare workers
Economy
COVID statistics
Origin
Outbreak/Pneumonia
Global Politics
Schools
USA COVID response
Travel restrictions
Transmission
Pandemic declaration
Quarantine
Mean of Favorites
Mean of Retweets



Misinformation?

- False or inaccurate information , especially that which is deliberately intended to deceive.
- Fighting with misinformation is even tougher than dealing with COVID-19
- Public engagement with misinformation related posts are raising day by day.
- Leads to confusion, anxiety, politicizing the situation, etc.

Social media

- Most powerful millennial tool
- Definition from Oxford: "*Websites and applications that enable users to* ***create and share content*** *or to participate in social networking.*"

IS THIS CONTENT CREDIBLE?

Credible sources

- CDC (Centers for Disease Control and Prevention)
- WHO (World Health Organization)
- NIH (National Institute of Health)
- State Departments of Health, etc.

How to combat misinformation?

- Educating the people
  - How to identify
  - How to prevent spread
  - How to report
- Promoting reliable/ credible sources

References

- WHO Coronavirus Disease (COVID-19) Dashboard | WHO Coronavirus Disease (COVID-19) Dashboard. (n.d.). Retrieved October 11, 2020, from https://covid19.who.int/
- Social Media | Definition of Social Media by Oxford Dictionary on Lexico.com also meaning of Social Media. (n.d.). Retrieved October 11, 2020, from https://www.lexico.com/en/definition/social_media
- Plutchik's color wheel of emotions, b) dyads of emotions | Download Scientific Diagram. (n.d.). Retrieved October 11, 2020, from https://www.researchgate.net/figure/a-Plutchiks-color-wheel-of-emotions-b-dyads-of-emotions_fig3_308972170
- Chakraborty, I., & Maity, P. (2020). COVID-19 outbreak: Migration, effects on society, global environment and prevention. *The Science of the Total Environment*, *728*, 138882. doi:10.1016/j.scitotenv.2020.138882
- Mian, A., & Khan, S. (2020). Coronavirus: the spread of misinformation. *BMC Medicine*, *18*(1), 89. doi:10.1186/s12916-020-01556-3

White paper report on HealthDX (Global health data exchange) – PROJECT 2

Chilukuri, Susmitha

Supervisor: Mostafa, Javed

Advisor: Gibson, Amelia

Introduction

- COVID-19 pandemic made us realize the importance of global health data exchange and how far we are from achieving it.
- Global health exchange/Interoperability is essential for handling the global health issues in an effective way.
- The progress in the space of inter-health data exchange is slowed by the lack of an aggregated knowledge base that is fed in real-time by a community of practice.

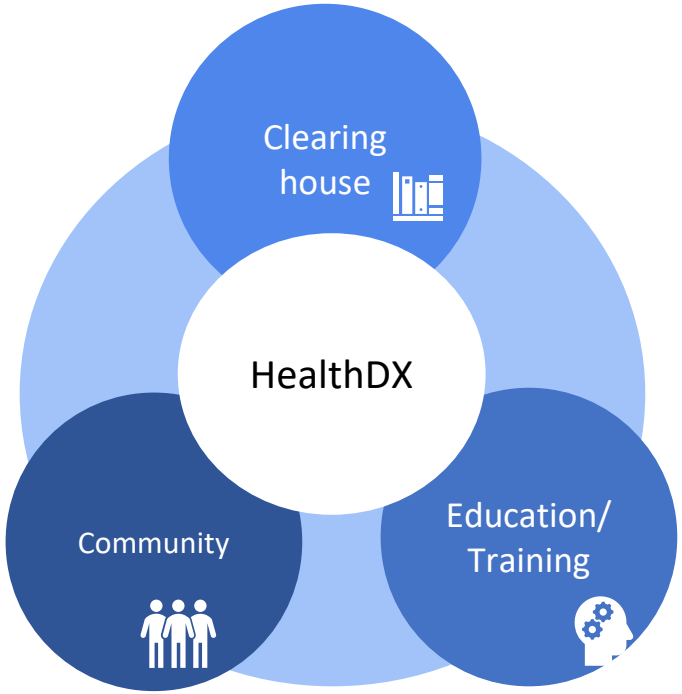
Aim/Objective

The main aim of this paper is to emphasize the importance of global health data exchange with the establishment of HealthDX platform and systematic review of the pillars/layers it stand by.

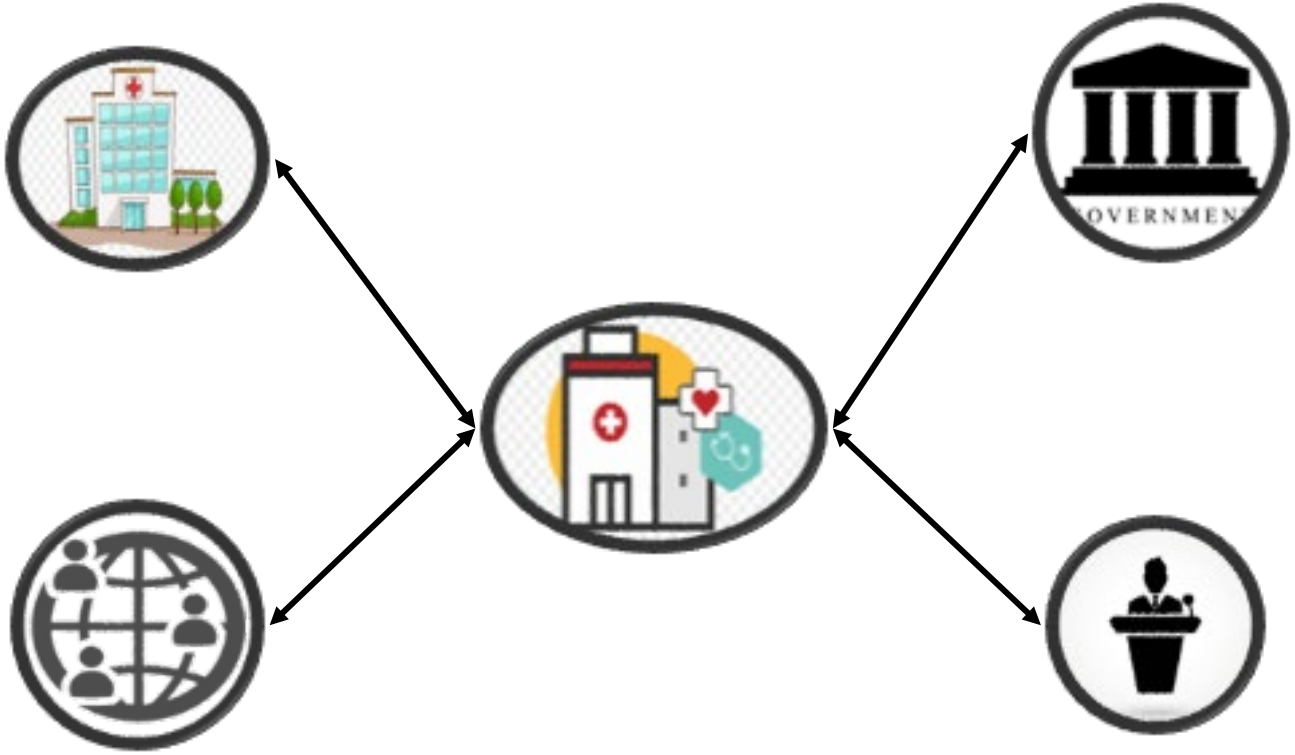
Method: Systematic literature review to report the whole process of creating a platform like HealthDX and what key features it is built on and makes it unique.

HealthDX

- Multifaceted platform that brings stakeholders who are passionate about improving healthcare by using technology and capacity in an innovative way, focusing on health data exchange.
- 3 layers
- Focus is inter- organizational exchange that has a major impact on decision making at organizational, district, state, national and international levels.



Inter – organizational Exchange



Website

- https://healthdx.unc.edu/
- Using this as a platform for partnering with potential organizations and hosting a global health conference/webinar in March 2021.

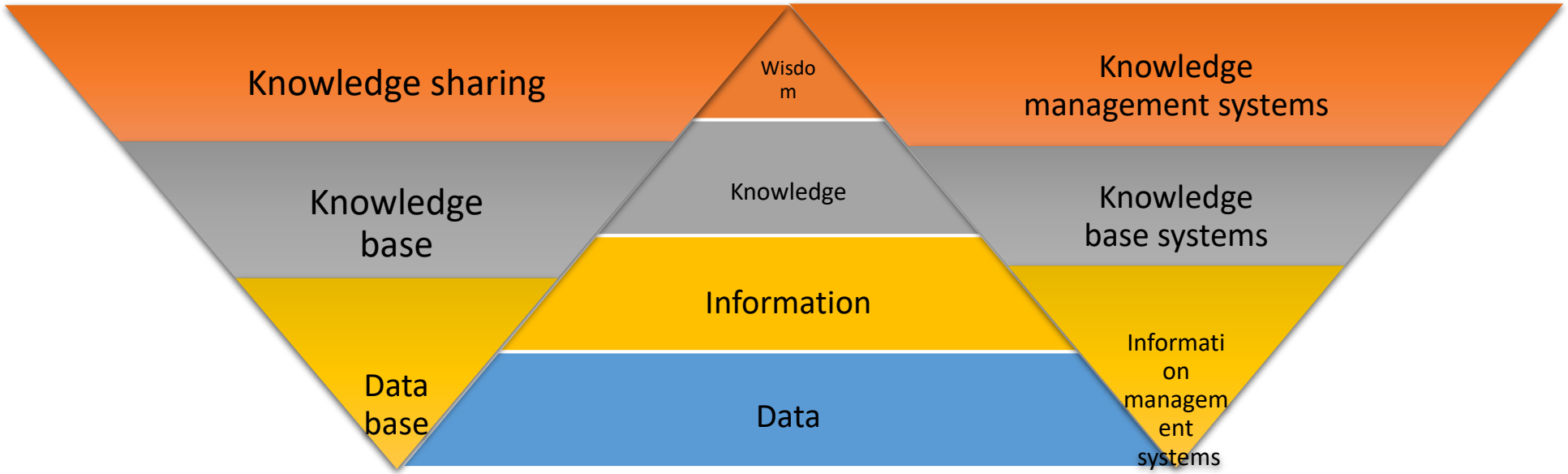
Knowledge base (KB)

- Collection of information about a particular subject
- Changes with context
- 6 types
  - Internal KB
  - Hosted KB
  - Self-Hosted KB
  - Customer KB
  - Open-source KB software
  - External KB

Database vs knowledge base

Data base	Knowledge base
Organized collection of data	Collection of information on a particular subject
Typically represents facts in their basic form	Fully developed and ready to be applied
Supports storage and retrieval of data	Supports rapid search, retrieval and reuse of information
Designed mostly for concrete (Surface thinking) knowledge	Designed for abstract (Thinking in depth) knowledge
Collection of data representing facts	Information at higher level of abstraction

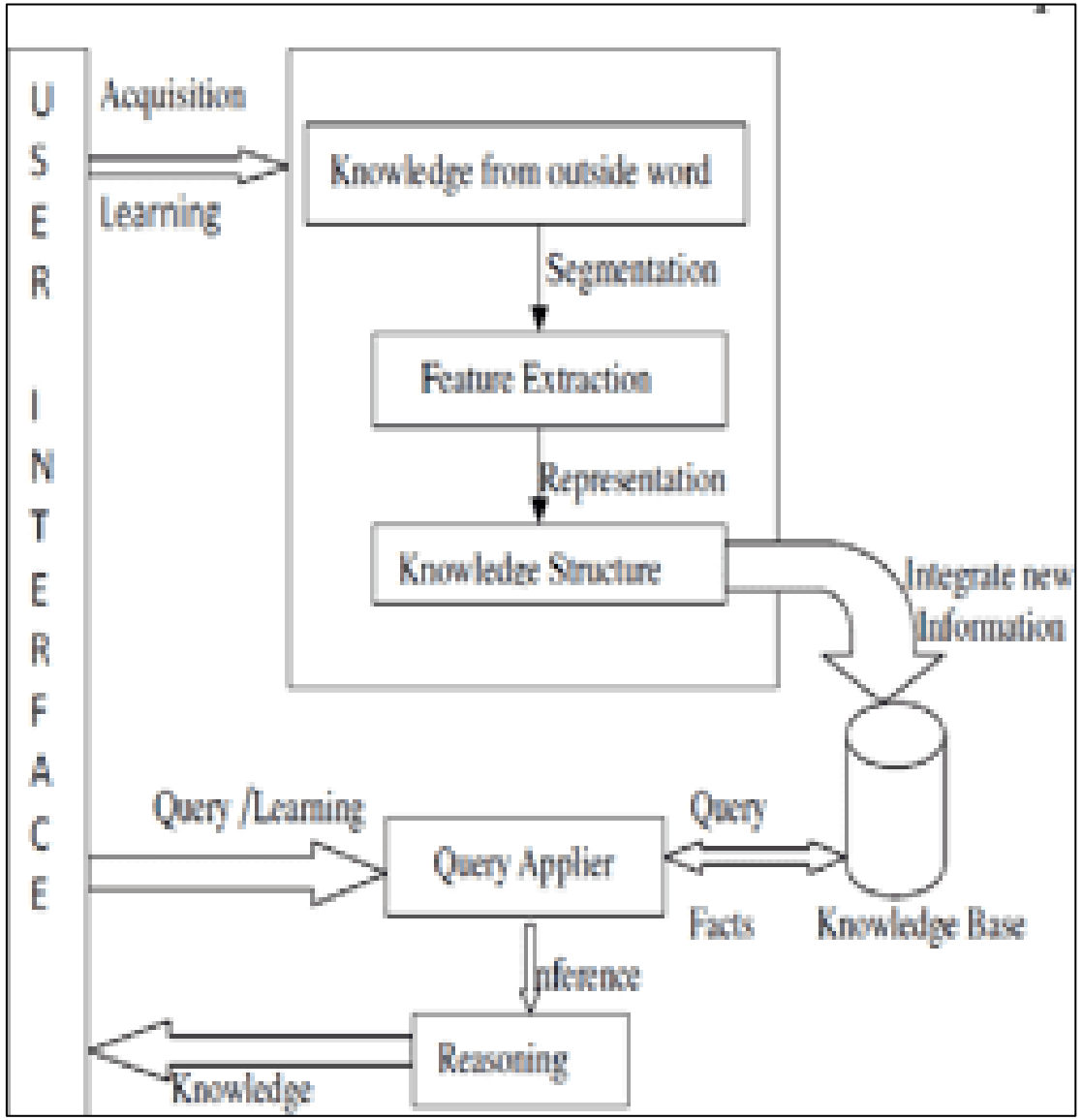
DIKW Pyramid



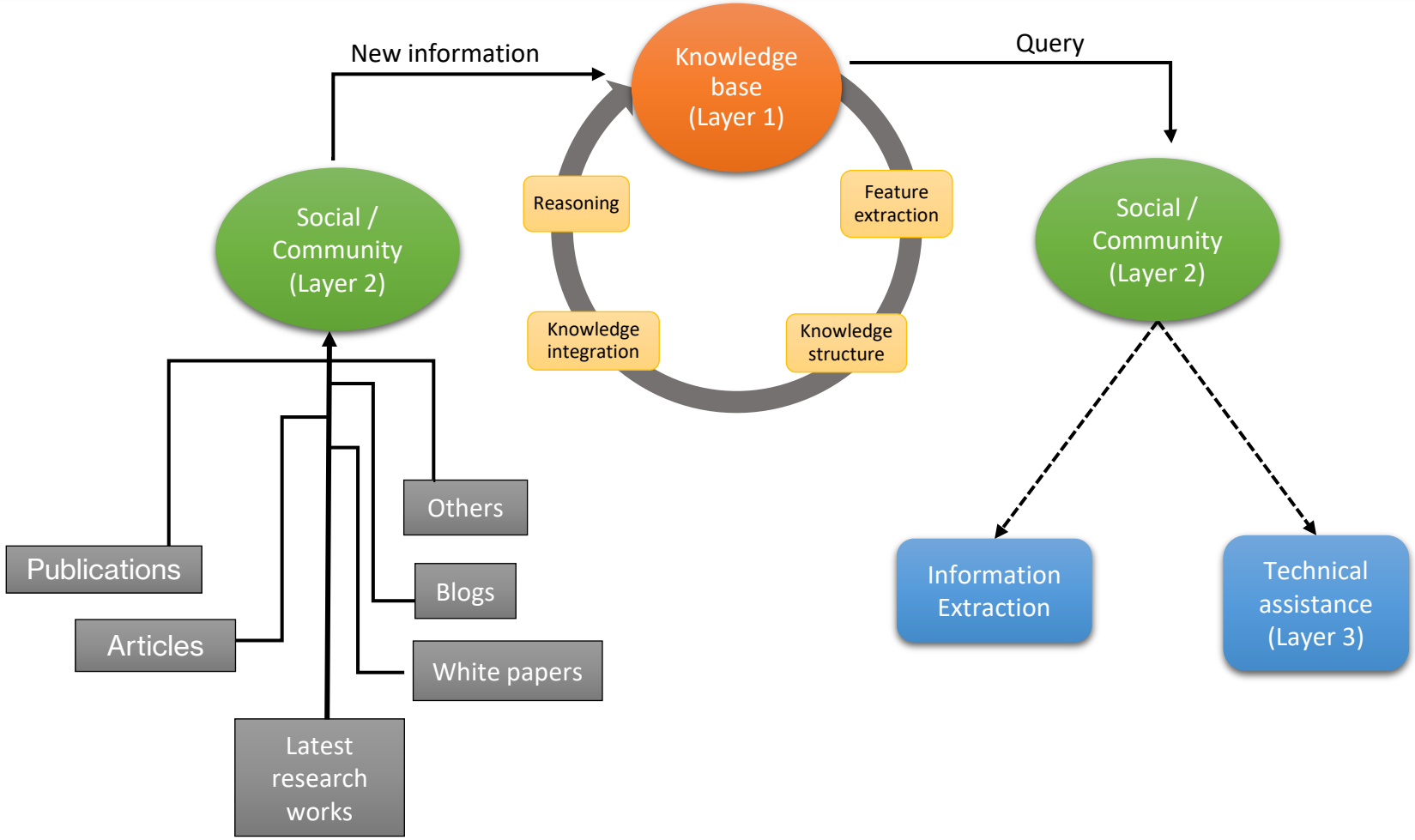
Characteristics of KB

- Comprehensive
- Accessible
- Intelligent
- Search
- Accurate

Structure/ architecture of KB



Working model of HealthDX



Thank you!

Questions?