COVID-19
CORONAVIRUS DISEASE 2019

Public Opinion Reflection

Guidance for People

Reference for the Government

Methods

Micro & macro perspectives

1 Communication Networks

We can explore the connection between words and the semantic structure through social networks.

2 Word2Vec

We use Word2Vec similarity to different COVID 19 keywords to analyze the sentiment of the tweets.

3 Doc2Vec

We use the most similar words and word relations to explore people's concerns in different aspects of influence brought by COVID19.

4 Classification

We will analyze whether these tweets present obvious sentiment attributes in three different stages.

5 Sentiment Analysis

This part aims to find a perceptron algorithm among three ones with the optimal hyperparameters which can predict the emotional tendency of a review.
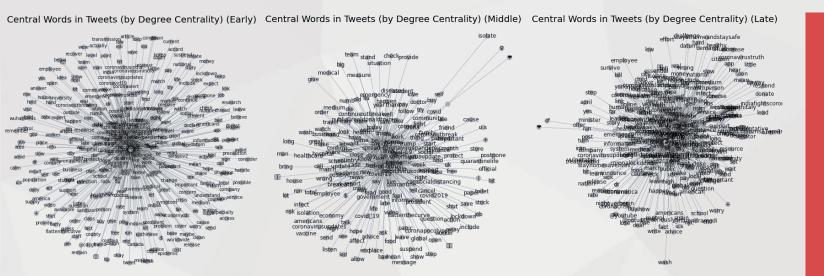
6 Gensim Model and Dynamic Modelling

We want to explore the topics covered in each stage of the tweets and also the changes in the topics over time. Two models will be used in this part: the Gensim model and the dynamic model.

Communication Networks

Central-word networks, "covid19" –related networks and ego-networks of "virus"

Central Words in Tweets (by Degree Centrality) (Early)

Central Words in Tweets (by Degree Centrality) (Middle)

Central Words in Tweets (by Degree Centrality) (Late)

Central-word networks

No matter what stage of the tweet, the nouns related to COVID 19 are the core words in the social network. In addition, China and Italy, two regions with severe epidemics, are also at the very core.
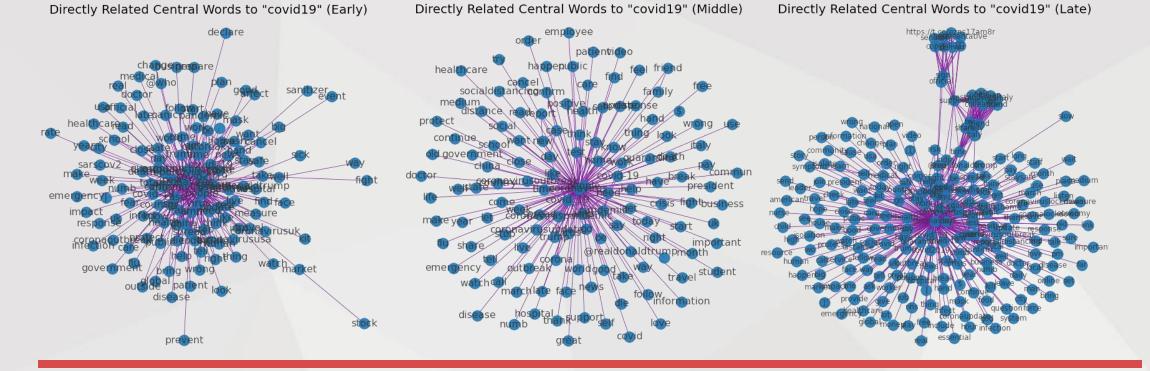
DEDUCTION 1

People pay attention to the epidemic itself and hope to learn about the relevant information of the epidemic in the disease-stricken regions and draw on their epidemic prevention experience.

DEDUCTION 2

People hope that the government will play a greater role in this public health crisis.

Directly Related Central Words to "covid19" (Early)

Directly Related Central Words to "covid19" (Middle)

Directly Related Central Words to "covid19" (Late)

Directly Related Central Words to "covid19"

In the early stage, the words related to prevention of COVID 19 are mostly concentrated on the material level, such as sanitizer, mask, market, emergency and so on.
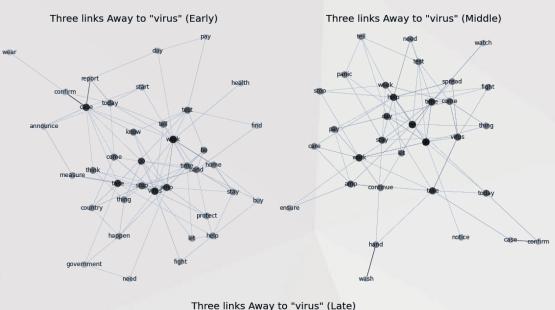
In the middle and late stages, the words 'socialdistancing', 'lockdown', 'community', 'stay', etc. appeared, and the President's Twitter account was also directly related to 'covid19'.

Three links Away to "virus" (Early)

Three links Away to "virus" (Middle)

Three links Away to "virus" (Late)

Three Links Away to "virus"

Ego-networks of "virus"

- With the proliferation of COVID 19, social conflicts have increased, and more social conflicts need to be resolved through legal issues.

- Local governments have perfected relevant management regulations and included some events involving COVID 19 in the management of laws and regulations.

- Responding to the COVID 19 crisis requires the joint efforts of the entire social group. Different individuals are connected and influence each other.

Table 1: Top 10 Words in Terms of Degree Centrality Related to 'Virus' in Early Tweets

One Link	Two Links	Three Links
protect	work	case
know	stop	amp
take	amp	work
come	take	hand
happen	hand	stop
go	case	today
think	go	take
left	come	time
tell	thing	test
stop	time	go

Table 2: Top 10 Words in Terms of Degree Centrality Related to 'Virus' in Middle Tweets

One Link	Two Links	Three Links
take	take	time
spread	come	help
come	virus	work
go	stay	be
stay	time	take
fight	help	day
	day	come
	week	virus
	spread	stay
	fight	week

Table 3: Top 10 Words in Terms of Degree Centrality Related to 'Virus' in Late Tweets

One Link	Two Links	Three Links
follow	stay	take
say	virus	time
save	be	stay
take	go	work
come	save	amp
go	stop	case
think	come	virus
break	work	be
try	time	go
stay	amp	save

The most relevant words to 'virus' do not change too much over time and the number of links

Top 10 Words in Terms of Degree Centrality Related to 'Virus'

Embedding

Word2Vec, Doc2Vec and Projection

Table 4: Most Similar Words of the 'coronavirus' + 'unemployment' - 'quarantine'

Early		Middle		Late	
Most similar words	Scores	Most similar words	Scores	Most similar words	Scores
die	0.999	economic	0.824	global	0.808
away	0.999	global	0.813	humanitarian	0.781
fear	0.999	declared	0.810	funds	0.769
m	0.999	economy	0.810	current	0.766
italy	0.999	funds	0.806	combined	0.765
cancelled	0.999	drtedros	0.796	financial	0.760

Word2Vec

Embedding

Word2Vec, Doc2Vec and Projection

Table 4: Most Similar Words of the 'coronavirus' + 'unemployment' - 'quarantine'

Early		Middle		Late	
Most similar words	Scores	Most similar words	Scores	Most similar words	Scores
die	0.999	economic	0.824	global	0.808
away	0.999	global	0.813	humanitarian	0.781
fear	0.999	declared	0.810	funds	0.769
m	0.999	economy	0.810	current	0.766
italy	0.999	funds	0.806	combined	0.765
cancelled	0.999	drtedros	0.796	financial	0.760

Word2Vec

Embedding

Word2Vec, Doc2Vec and Projection

Table 5: Most Similar Words of the 'coronavirus' + 'die' - 'quarantine'

Most similar words	Scores
prc (People's Republic of China)	0.732
spread	0.701
conceal	0.692
capitalist	0.689
erase	0.683
mistake	0.677
choose	0.675
incompetents	0.675
Isis (Islamic State of Iraq and Syria)	0.667
ccp (Communist Party of China)	0.667

Doc2Vec

Embedding
Word2Vec, Doc2Vec and Projection

Table 6: Most Similar Words of the 'coronavirus' + 'unemployment' - 'quarantine'
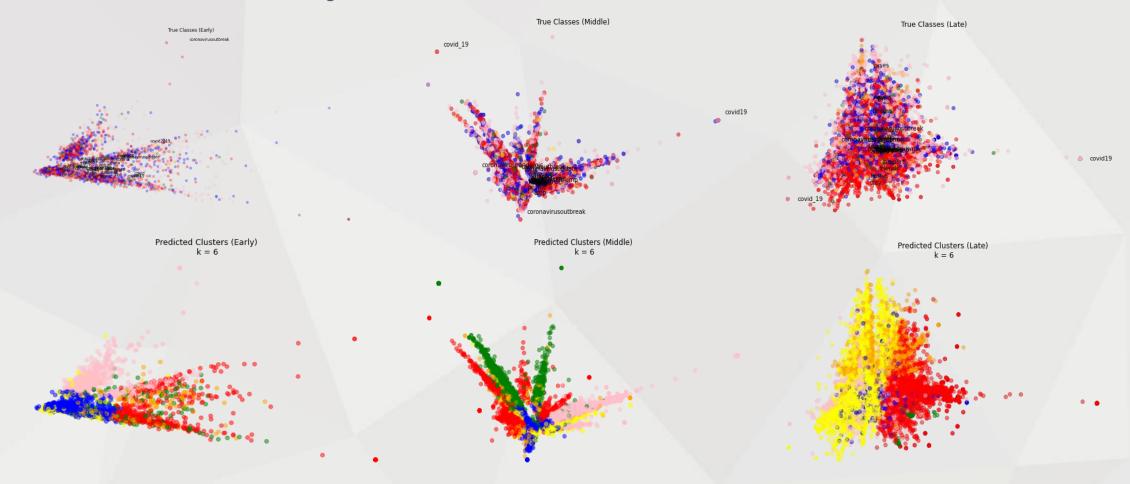
Most similar words	*Scores*
crippled	0.753
repo (repurchase agreement)	0.745
Davidndii	0.737
imf	0.736
worsens	0.735
monies	0.729
moderna	0.725
bil	0.721
investments	0.719
abrupt	0.713

Doc2Vec

Embedding

Word2Vec, Doc2Vec and Projection

Figure 12 : Projection results of two dimensions

Projection

Classification
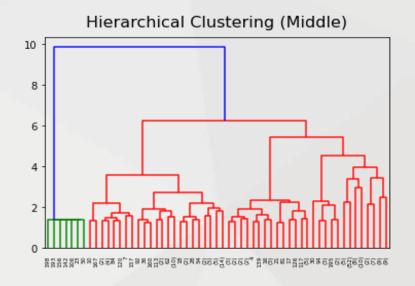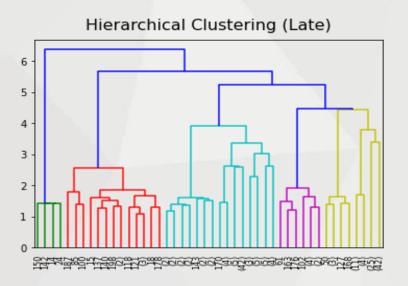
K-means(++) & Hierarchical Clustering

True Clusters and Predicted Clusters of Tweets of Different Stages

Classification

K-means(++) & Hierarchical Clustering

Hierarchical Clustering of Tweets of Different Stages

Sentiment Analysis

Perceptron algorithm, average perceptron algorithm and Pegasos algorithm

Positive (Early)

1
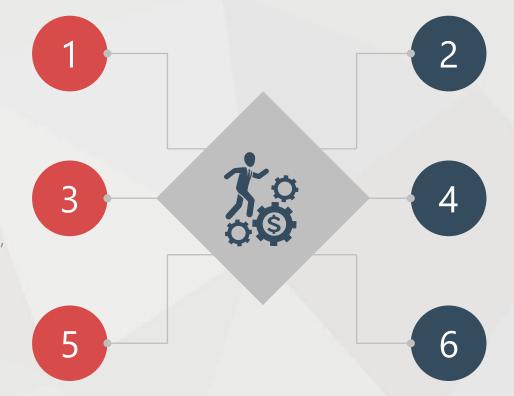
'positive', 'best', 'safe', 'sure', 'good', 'help', 'hand', 'chance', 'Please', 'protect'

Negative (Early)

2

'sick', 'stop', 'die', 'emergency', 'fear', 'crisis', 'deaths', 'dying', 'infected', 'isolation'

Positive (Middle)

3

'best', 'safe', 'positive', 'support', 'hope', 'important', 'good', 'Please', 'save', 'care'

Negative (Middle)

4

'crisis', 'death', 'avoid', 'fighting', 'alone', 'kill', 'crisis.', 'worst', 'stop', 'infected'

Positive (Late)

5

'great', 'positive', 'best', 'love', 'hand', 'Thank', 'hope', 'please', 'free', 'safe.'

Negative (Late)

6

'crisis', 'crisis.', 'death', 'infected', 'critical', 'stop', 'wrong', 'ass', 'fuck', 'isolation'

Gensim Model and Dynamic Modelling
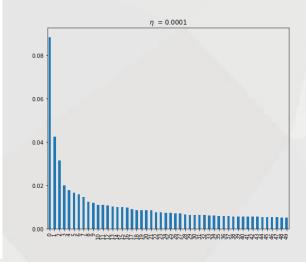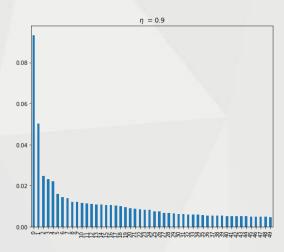
Independent model VS time-series model

Table 16: Gensim Topic Model for Early Tweets
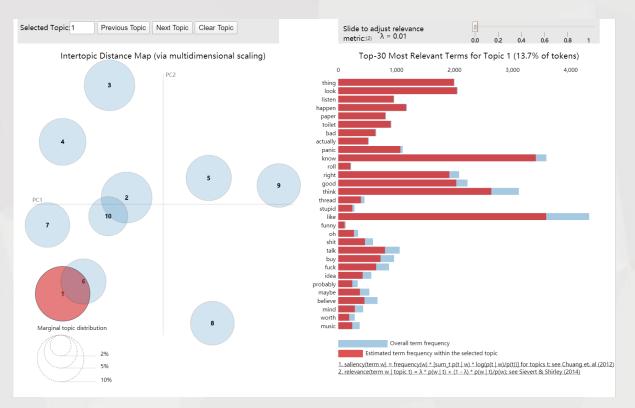
Topic_0	Topic_1	Topic_2	Topic_3	Topic_4	Topic_5	Topic_6	Topic_7	Topic_8	Topic_9
think	case	coronavirus	coronavirusoutbreak	coronavirus	trump	people	coronavirus	travel	coronavirus
coronavirus	coronavirusoutbreak	hand	coronavirus	coronavirusoutbreak	lot	case	coronavirusoutbreak	new	coronavirusoutbreak
health	coronavirus	wash	covid2019	close	say	say	corona	coronavirus	covid_19
time	new	people	case	kill	like	tell	covid_19	case	pandemic
people	total	amp	italy	test	thing	need	coronavirusupdate	state	amp
amp	like	coronavirusoutbreak	covid—19	wake	response	covid	cancel	coverage
public	bring	come	india	virus	coronavirus	work	covid2019	city	outbreak
coronavirusoutbreak	report	break	coronavirusinindia	country	way	quarantine	people	test	news
organization	virus	sick	care	school	say	coronavirus	virus	spread	virus
fast	question	home	govt	emergency	look	infection	case	people	people
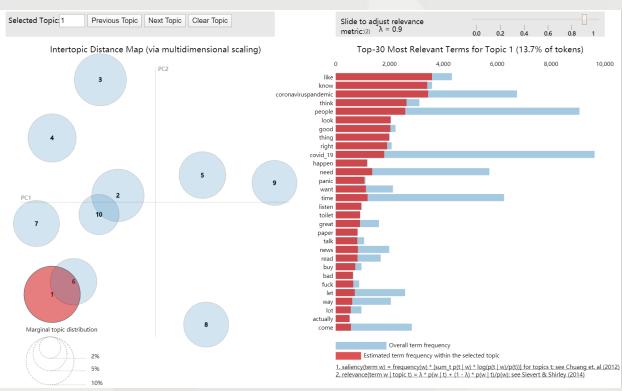
Gensin Model for Early Tweets with η changing

Gensim Model and Dynamic Modelling

Independent model VS time-series model

Topic Model for Early Tweets with λ changing

Discussion

Advantages and disadvantages of big data

Limitations

Since we only used 1.25% of English Tweets, we are worried about the generalization of the results to the population (here referring to people in a target society)

Strengths

We make the use of big data's advantages including big, always-on, and nonreactive, and research results are consistent with our common sense

Conclusion

How people react & what we can learn

WordCloud for Tweets in Early March

WordCloud for Tweets in Middle March

WordCloud for Tweets in Late March

Overall, people are most concerned about safety and health.

People do not understand and accept news.

Citizens and the government began to take strict measures to deal with COVID 19.

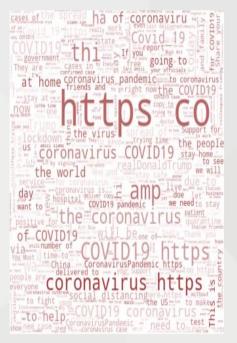People pay more and more attention on employment and economy.

Conclusion

How people react & what we can learn

WordCloud for Tweets in Early March

WordCloud for Tweets in Middle March

WordCloud for Tweets in Late March

Overall, people are most concerned about safety and health.

People do not understand and accept news.

Citizens and the government began to take strict measures to deal with COVID 19.

People pay more and more attention on employment and economy.