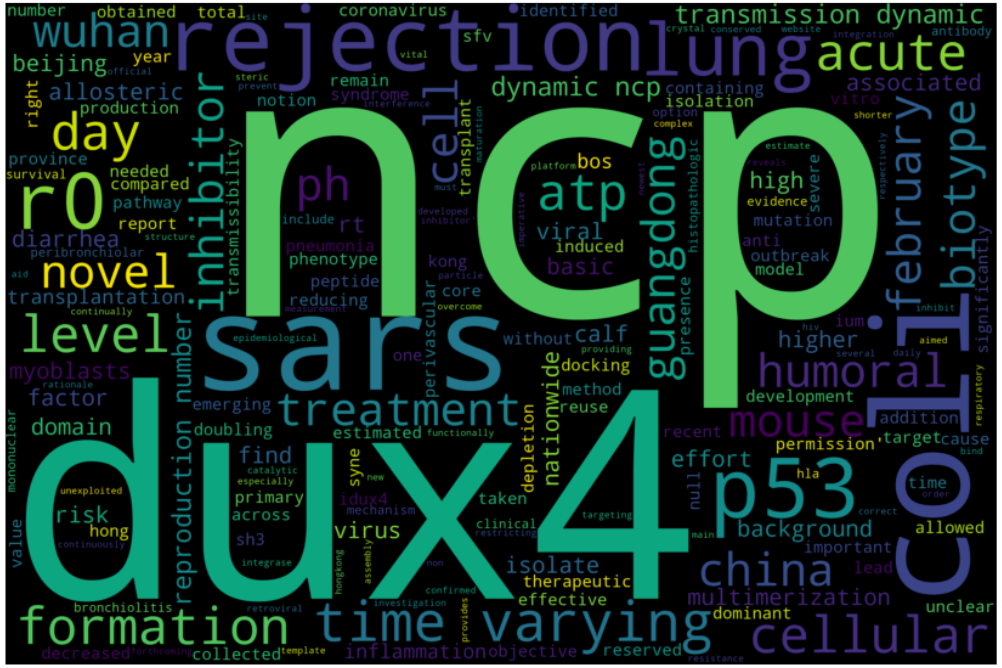
COVID-19 Open Research Dataset Challenge

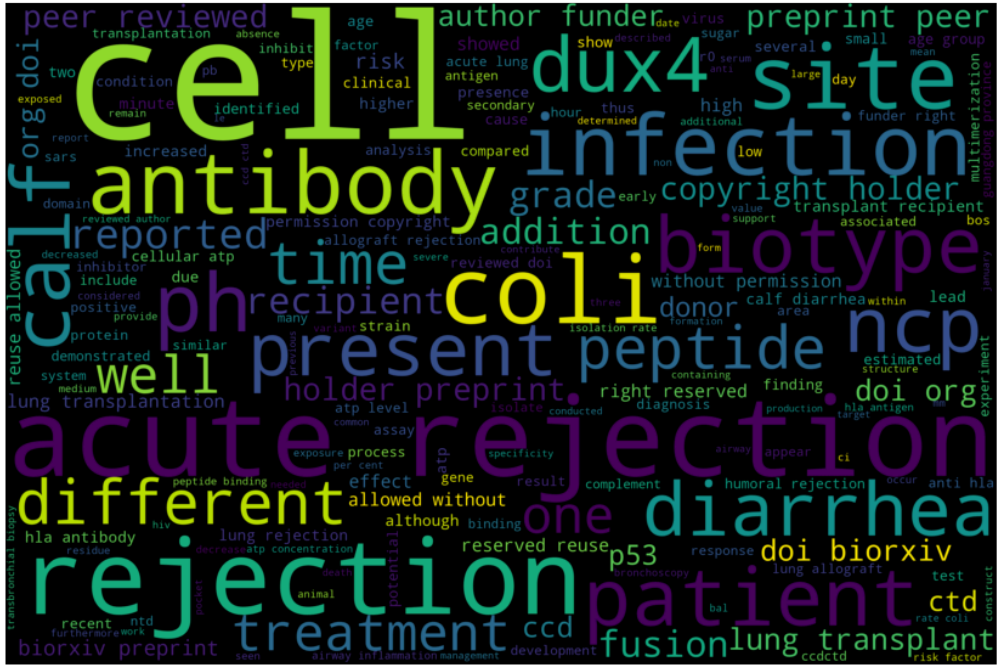
Tianyi Yu

April 9, 2020

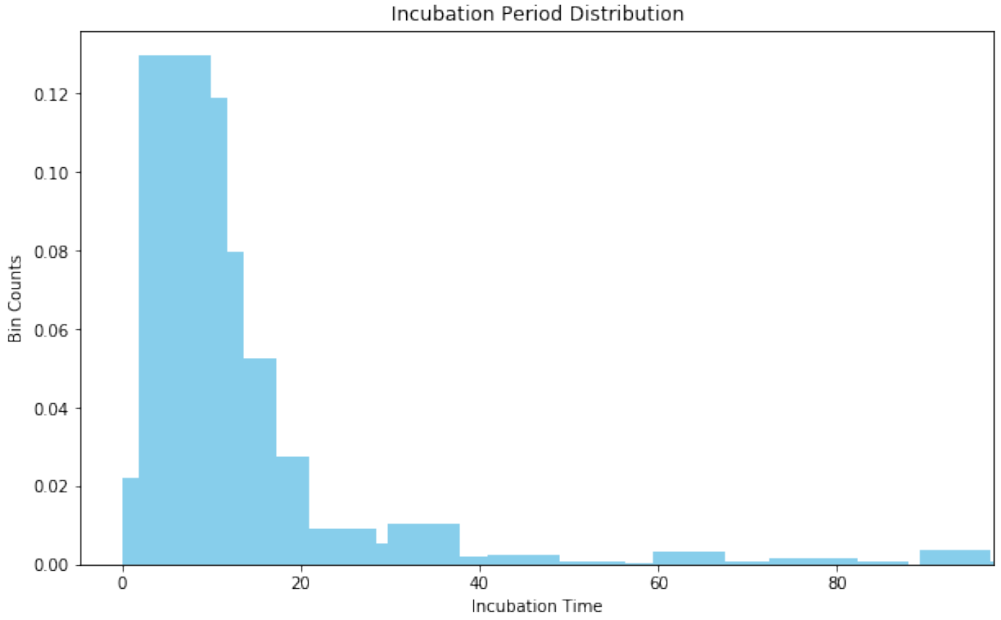
Keywords of Research Papers Abstract

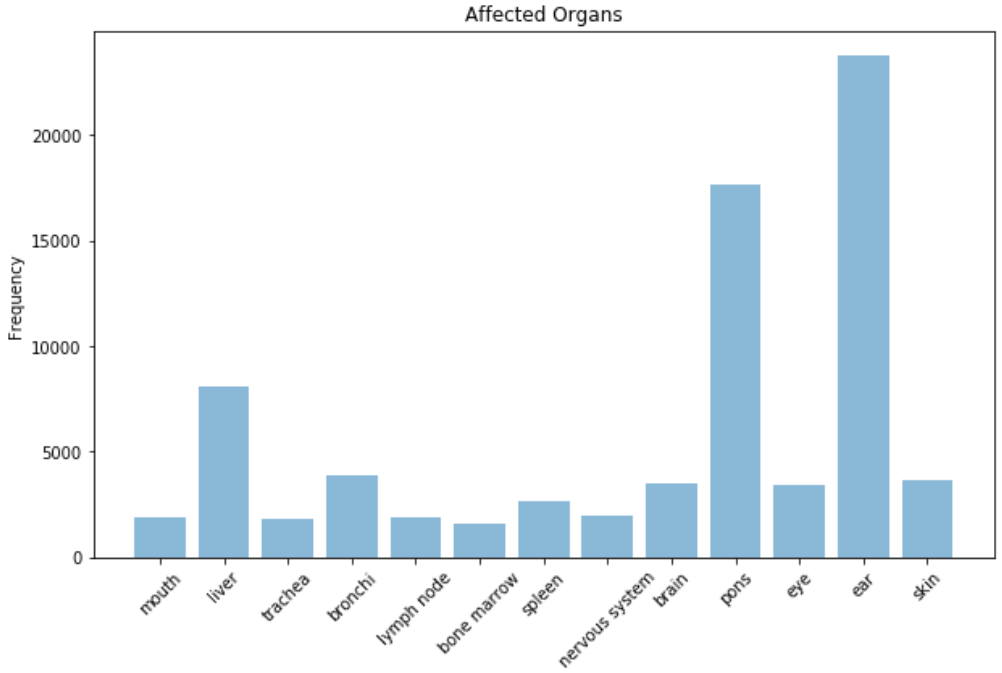


Keywords of Research Papers Full Text



Word Cloud of Abstract and Full Text



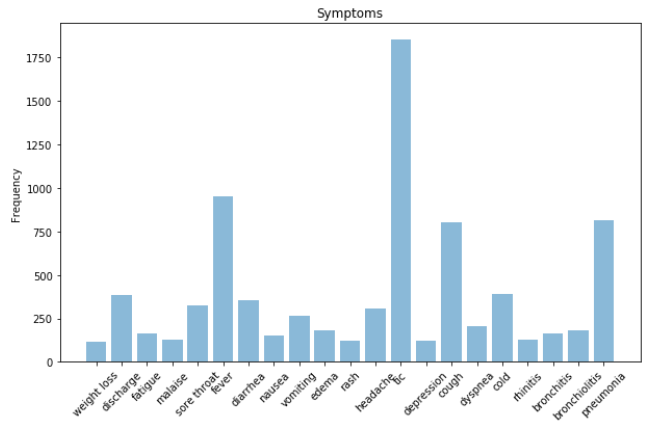


EDA: Incubation Period and Affected Organs

Transmission Analysis Using Latent Dirichlet Allocation

```
Topics found via LDA:

Topic #0:
dog rabies canine domestic lineage wild awareness interval wildlife appropri

Topic #1:
hrv chikv text full pubmed free rhinovirus infant closure coli

Topic #2:
patient care hospital sars respiratory infection hcws worker healthcare cont

Topic #3:
model disease epidemic outbreak number health infectious network contact pop

Topic #4:
recognized antimicrobial precaution acid delay lung isolates canine pcr path

Topic #5:
airborne air aerosol ventilation droplet particle room indoor building flow

Topic #6:
rsv tb ferret household nosocomial infant loss closure acid pcr

Topic #7:
virus human cell viral host bat infection protein specie animal

Topic #8:
mers cov camel middle east saudi arabia syndrome coronavirus respiratory

Topic #9:
h5n1 poultry avian bird market live interspecies pathogen wild chicken
```

Symptoms Analysis Using Apriori Algorithm



	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(fever)	(discharge)	0.509657	0.205472	0.143240	0.281053	1.367838	0.038520	1.105127
1	(discharge)	(fever)	0.205472	0.509657	0.143240	0.697128	1.367838	0.038520	1.618978
2	(discharge)	(tic)	0.205472	0.995172	0.204936	0.997389	1.002228	0.000456	1.849249
3	(tic)	(discharge)	0.995172	0.205472	0.204936	0.205930	1.002228	0.000456	1.000577
4	(discharge)	(cough)	0.205472	0.431330	0.140558	0.684073	1.585961	0.051932	1.800004

	support	itemsets
0	0.205472	(discharge)
1	0.174356	(sore throat)
2	0.509657	(fever)
3	0.189378	(diarrhea)
4	0.143777	(vomiting)
5	0.165236	(headache)
6	0.995172	(tic)
7	0.431330	(cough)
8	0.111588	(dyspnea)
9	0.210300	(cold)
10	0.437768	(pneumonia)

Insights about Policy and Guidance to Tackle the Outbreak

- The incubation period varies from 0 day to several months. Most of the researchers conclude the incubation period to be between 0 and 20 days. The most popular deduction would be between 0 and 10 days. It is suggested that people should self-isolate for two weeks after getting close to possible sources of infection to see if any symptoms appear and to avoid infecting others.
- Coronavirus involve a number of organs of our body. The most related ones are liver, bronchi, spleen, brain, pons, eye, ear and skin. Mild cases may not experience great challenges for their respiratory system, however, severe cases can be in danger as many organs would be attacked. It is suggested that the government should gather medical resources and be prepared for the outbreak. Doctors and nurses should be familiarized with special practices treating the patients and dealing with the symptoms.
- The coronavirus is highly infectious. The keywords retrieved by the LDA model reveal that the virus is transmitted by aerosols in the air and thus could transmit by contacting one another. Results also show that both human and animal can be infected. Wild animals such as bats are also hosts of the virus. In this case, it is suggested that the government should inform the public about the ways of transmission and methods to avoid infection, including social distancing, avoiding contacting with others, washing hands often, wearing facial masks and so on.
- The results from Apriori Algorithm show that tic, fever, cough, penumonia, cold, diarrhea, discharge, sore throat, headache, vomiting are the common symptoms of coronavirus. The association rule results show that discharge often appears with tic, fever or cough. Similarly, it is suggested that the government should inform the public about the common symptoms of coronavirus and the co-occurance of which of the symptoms is likely to be a coronavirus case and when they should go to the hospital or call the police.