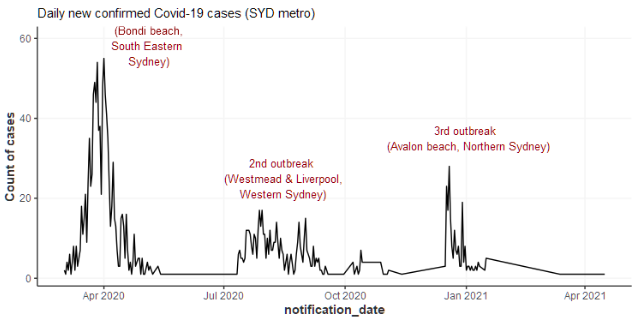
1. **Introduction**

Since the outbreak of Covid-19, governments around the world have relied on two key measures to curb local transmission of the disease are 1) contact tracing; 2) extensive testing. These two measures work well if a) the outbreak is discovered earlier; b) the number of confirmed cases is limited, and c) the presence of asymptomatic carriers is rare. However, in many parts of the world, it is extremely unlikely for all these three conditions to be met simultaneously. In most cases, local transmission cannot be contained without extreme measures such as extended periods of lockdown – which result in a heavy toll on the economic performance. Considering this, effective methods to select only a small subset of the geography or population to enter lockdown or perform group testing can be immensely helpful for the government to strike a balance between containing the virus and maintaining the economic output.

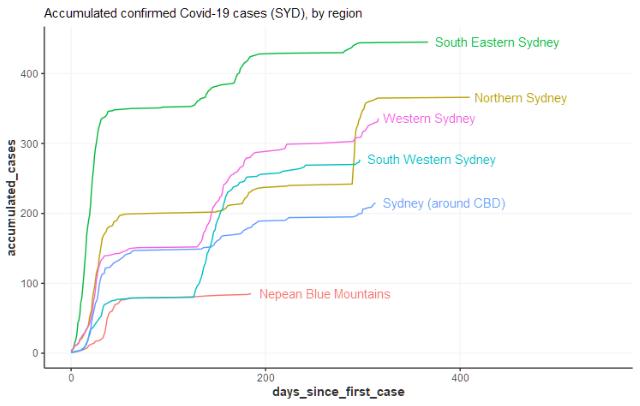
* 1. **Background**

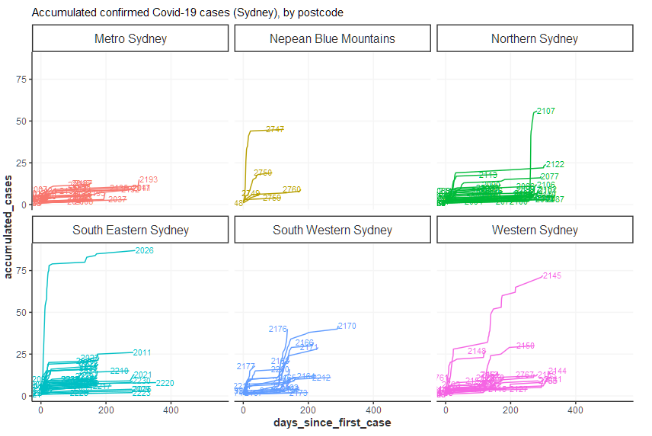
The first positive case for Covid-19 in Australia was reported on 25 January 2020 from incoming travellers from China. One month later, on 27th Feb, the Australian Prime minister Scott Morrison announced the Australian Health Sector Emergency Response Plan, 4 days before the first case of community transmission was reported in the state of NSW on 2nd March. Since then, Sydney, the capital city of NSW, has experienced three waves of the locally spread virus within its regional proximity – namely the “Bondi Beach Cluster”, the “Western Sydney Cluster” and the “Northern Beach Cluster”.

{Chart index} shows the daily count of new cases (locally acquired) in the metro Sydney area. One can clearly spot the three distinct waves from the three location in difference time. The magnitude of outbreak is relatively small compared to what happened in the rest of the world. This creates an opportunity to investigate the spatial diffusion of the virus in a relatively “controlled” environment, where the source of the virus during those periods is likely originated from the suburb that first recorded the spike in new cases.

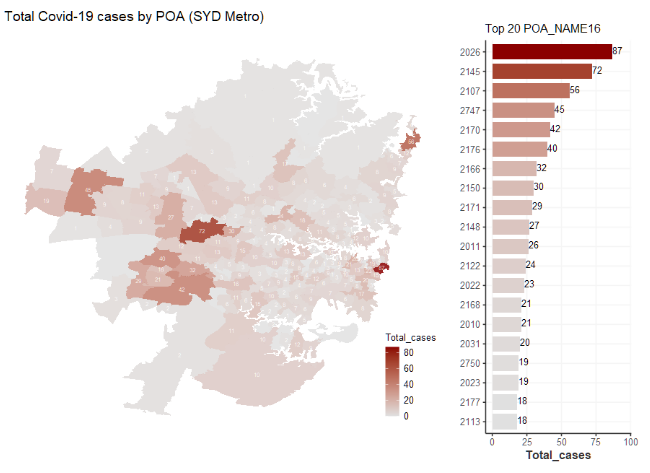


{Chart index} shows the accumulated cases by local health districts. The early spikes across all regions are driven by the South Eastern Sydney cluster (Bondi beach). The rise of cases in Western Sydney and South Western Sydney is more gradual which is also accompanied with rise in cases in other regions. The final cluster in Northern Sydney appears relatively isolated compared with the first two clusters.

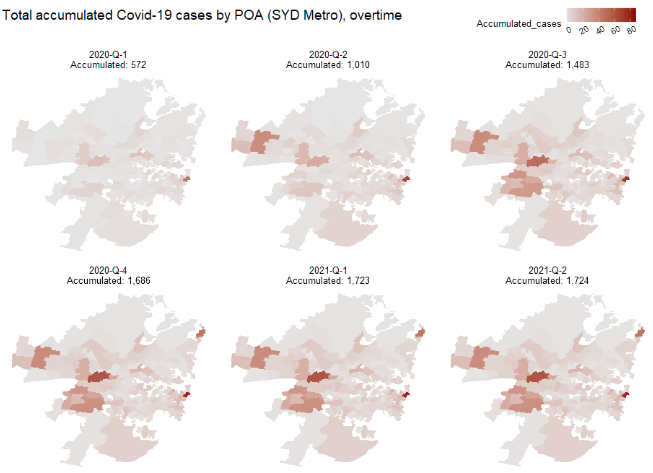




{Chart index} shows the accumulated cases in the map by postal area (closest shape file related to postcode which is what the Covid-19 case data is reported from). The locations of the three clusters are shown here. The Western Sydney cluster seems to have spread more widely than the other two beach clusters. It is conceivable that the flow of people from the beaches can be more scattered thus hard to predict. Consequently, there is a higher chance for our clustering algorithm to “predict” the spread of the virus in the Western Sydney area.

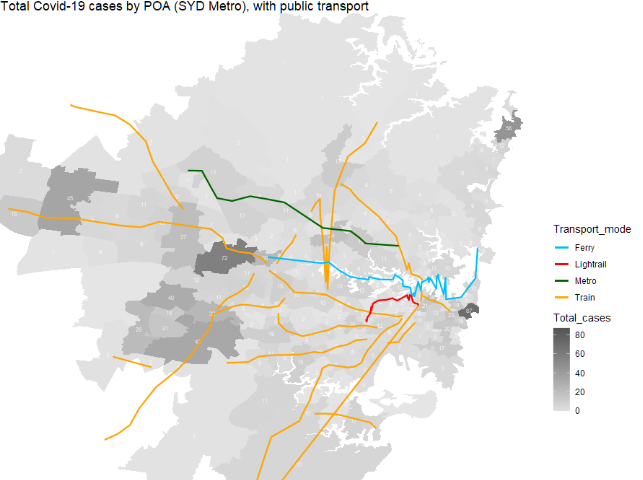


{Chart index} shows the accumulated cases over time. It appears that the Wester Sydney cluster could be originated from the further west area of Cambridge Park, before it spread out to the middle of the Western Sydney area which is close to a local transportation hub.

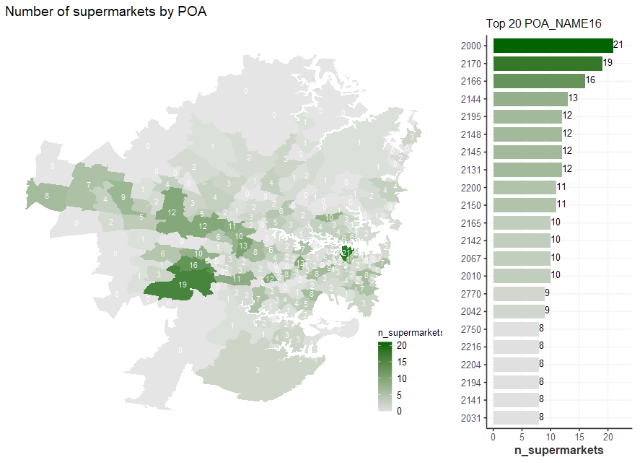


* 1. **Motivation of the gravity model**

{chart index} overlays the public transportation network on top of the choropleth (heatmap). It can be seen more clearly how the virus might have been transmitted along the train lines. This inspires the use of the Gravity Models to measure the level of “connectivity” between areas to predict the spread of the virus (see details in section 2.1).



The same pattern can be witnessed if we look at the number of supermarkets in each postal area.



* 1. **Motivation of the homophily model**

Need a map showing % of ethnic groups, etc., to demonstrate the existence of homophily in the geographic area.

* 1. **User story**

Example of lockdown around the world, how the lockdown areas were chosen and why our clustering method can be helpful.

* 1. **Literature review**

One paragraph on gravity model.

One paragraph on the homophily principle.

1. **Methodology**
   1. **The Gravity Model**

<https://oxfordre.com/economics/view/10.1093/acrefore/9780190625979.001.0001/acrefore-9780190625979-e-327#acrefore-9780190625979-e-327-bibItem-0048>

* 1. **The Homophily Principle**

<https://www.annualreviews.org/doi/abs/10.1146/annurev.soc.27.1.415?journalCode=soc>

* 1. **Geospatial Clustering**

<https://towardsdatascience.com/geospatial-clustering-kinds-and-uses-9aef7601f386>

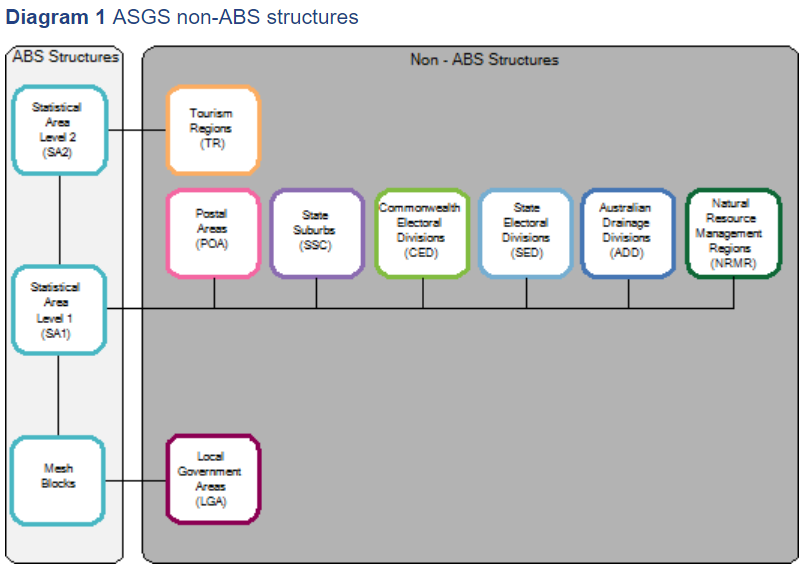
1. **Data**
   1. **Covid-19 confirmed cases SYD/AU**

<https://data.nsw.gov.au/data/dataset/nsw-covid-19-cases-by-location-and-likely-source-of-infection/resource/2776dbb8-f807-4fb2-b1ed-184a6fc2c8aa>

* 1. **Suburb demographic data SYD/AU**
  2. **SSC to POA mapping**

ABS data explanation

<https://www.abs.gov.au/websitedbs/censushome.nsf/home/factsheetsnas?opendocument&navpos=450>



Data source (mapping to meshblock) - <https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.003July%202016?OpenDocument>

1. **Model outcome**
2. **Model Evaluation**
3. **Conclusion**