Using a Dynamic Graph to Represent COVID-19 Virus Spread and Population Movements

Filipp Shelobolin
Carnegie Mellon University

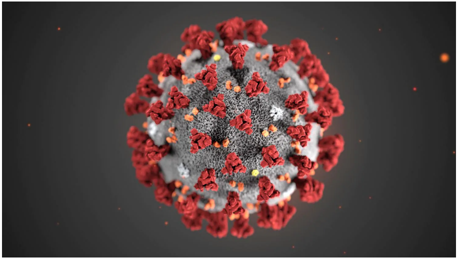
Advisor: Dr. Larry Holder
Washington State University

This material is based upon work supported by the National Science Foundation Research Experiences for Undergraduates Program under Grant No. 1757632.

August 3, 2020

Problem

- Covid-19 is highly contagious and is spreading fast.
- Most current models do not integreate both number of infection and population movements.
- Understanding patterns in global spread is essential for identifying effective policies and other methods for disruption.



Approach

- Create a dynamic graph that connects countries, states, counties (nodes) by their population movement (edges) over time.
- Develop a model that explains Covid-19 spread as a function of node attributes and edge attributes over time.
- Create a visualization app.
- Develop pattern discovery and anomaly detection methods for the graph.

Evaluation and Results: Data

- Data was combined from over 10 sources and daily updates automated.
- Over 25 node attributes describing Covid-19 cases, local mobility, geographical and meterological features, and policy decisions, mostly daily.
- Multiple sources of population movement approximations.

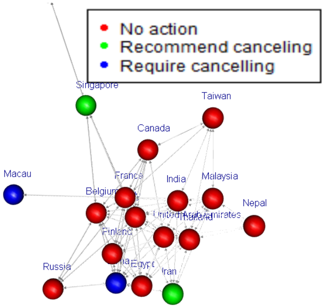
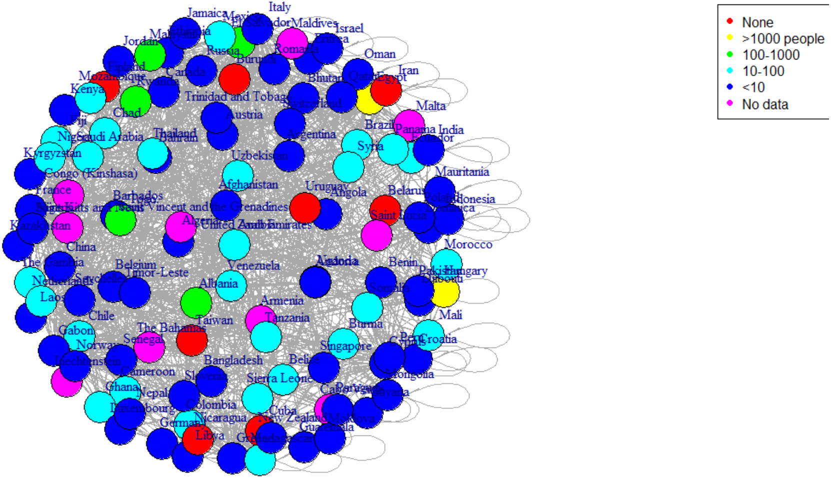


Figure: Public events policies and major population flow on 2020-02-15

Evaluation and Results: Data

Restrictions on gatherings and significant population movement on 2020-04-10



Sources

- Johns Hopkins COVID-19 data repository
- US Census Bureau International Population Database
- Google geocoding
- Weathersource.com
- Google mobility database
- Oxford policy database
- US T100 Airflow database
- Mao airflow predictions
- Google country data
- FCC FIPS codes encoding
- US Census Geocoder

Evaluation and Results: Model

- Goal: predict infection rate as function of features.
- Features we are most interested in:
  - $\mathcal{I}_{in}(C) = \sum_{\text{country} \neq C} \frac{\text{flow[country}, C] \cdot \text{infection\_rate[country]}}{\text{population}[C]}$
  - $\mathcal{I}_{out}(C) = \sum_{\text{country} \neq C} \frac{\text{flow}[C, \text{country}] \cdot \text{infection\_rate}[C]}{\text{population[C]}}$
- Define $S = \frac{\#\ \text{susceptible}}{\text{population}}$, $I = \frac{\#\ \text{infected}}{\text{population}}$, then our model is:

$$\frac{\partial S}{\partial t} = S(\beta_1 \mathcal{I}_{in} + \beta_2 I)$$

- $0 \leq \beta_1 \mathcal{I}_{in} + \beta_2 I \leq 1$, so fit with beta regression.
- Very high train and test error, bad predictions.
- Model is too simple, but shows effects of $\mathcal{I}_{in}$.

- Features: map, simulation of closing additional airports, simple graphs, variable list.
- Allows for running simulation and viewing results.
- In the future: better pairing with model to convey more information, and a better model.

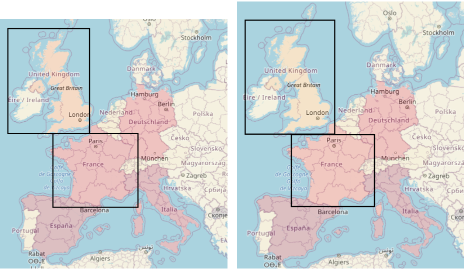
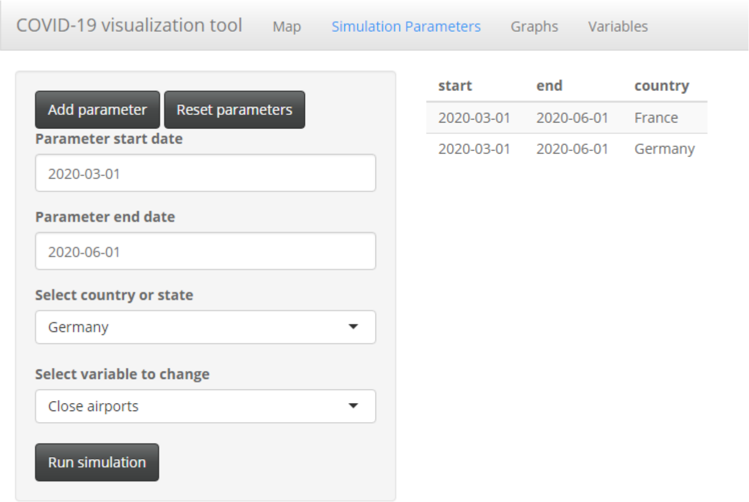


Figure: Coronavirus cases (left) vs. simulated Coronavirus cases (right) on May 1st after 1-week closure of airports. Note how United Kingdom has fewer cases due to high passenger flow from France.

Evaluation and Results: Visualization



Evaluation and Results: Visualization

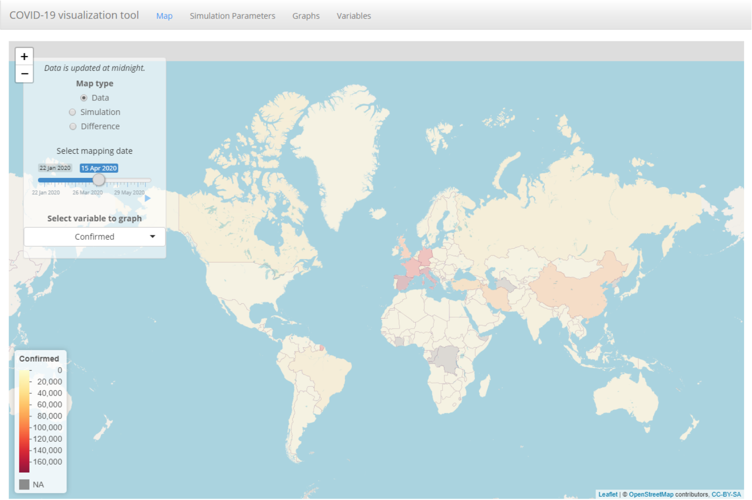
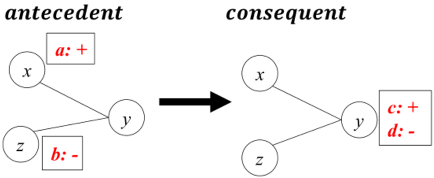


Figure: A map of Coronavirus cases on April 15th 2020 generated using the visualization tool. Variable and mapping data can be chosen in the top left panel. Key in the bottom left.

Evaluation and Results: Pattern Detection

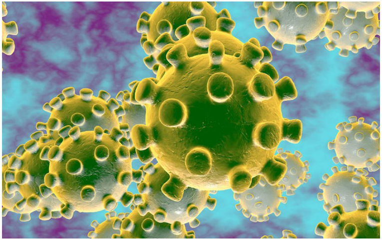
- Working on implementing Attribute Evolution Rule Miner (AER-Miner)[1], a new method that mines patterns from dynamic graphs.
- Will be implemented into visualization.



[1] Fournier-Viger, P., He, G., Lin, J. C.-W., Gomes, H. M. (2020). Mining Attribute Evolution Rules in Dynamic Attributed Graphs. Proc. 22nd Intern. Conf. on Data Warehousing and Knowledge Discovery (DAWAK 2020), Springer.

Conclusions

- Population movement matters (likely, but still need to do a deeper analysis of the extent methodology is accurate). Close airports early!
- Our dynamic graph representation effectively shows trends in attributes, but is very computationally expensive.
- Next steps: AER-miner to show more complicated patterns.



Thank you for your time

Questions?