Epidemics on Transportation Networks

By Ho Lum Cheung, Dimas Muntesinos, Frank Acquaye, Elie Wanko

27 JUN 2020

Abstract

Public transportation plays a vital role connecting people with each other. International trade and tourism is reliant on commercial aviation. And the most successful cities all have buses, trams, or subways connecting workers to their workplaces. However, the speed and connectivity of these transportation networks leave us vulnerable to disease propagation.

In this paper, we introduce a general agent-based framework for modeling disease spread on transportation networks. We use it to build models of major transportation networks, and take a special look at the New York City subway to see how well it can predict infection rate and infection hotspots during the 2020 COVID-19 epidemic.

Predicating disease spread on subway lines, ridership at stations, commute time, and city-wide countermeasures, we were able to fit our model very closely to empirical data.

Competencies Briefing (for other NS teams)

Purpose: Brief others on competencies so we can discuss things
Transportation Networks
specifically NYC subway, world airline network, Madrid trains, Moscow metro. mapping, passenger flow
python, networkx, mesa
we're working with these techs.
Writer's Note: Obviously to be removed before submission
This report was last pushed to OverLeaf on 21.06.2020. You will find our latest work at the link below: https://github.com/cheung-ho-lum/NS_Epidemics_ABM_Approach/blob/master/Report/ABM_NYC_Subway.pdf

1 Background

1.1 Epidemics and COVID-19

Coronavirus disease 2019 (COVID-19) is a disease caused by the SARS-CoV-2 coronavirus. Since being identified in December 2019, it has been labelled by the WHO as a pandemic, and spread around the world. Epidemics such as the coronavirus have been a subject of research for centuries, and is of special interest to those working in public health. Recent waves of new research came in 2002 (SARS), 2009 (H1N1), and 2014 (Ebola). However, in these prior epidemics, researchers did not have access to as much data as we have currently. In recent years, the state of data science research tools have also greatly improved, allowing researchers to answer questions in novel ways, but following

the scientific method. In this paper, we will try to respect the expertise of public health researchers, medical professionals, and the general public by being conservative when making conclusions.

1.2 Mathematical Modeling of Epidemics

There are two basic types of modeling for epidemics. Statistical and Mechanistic. While statistical modeling has typically given more accurate forecasts for a well-known situation [citation needed], mechanistic models such as the SIR and SEIR compartmental models help to explain why phenomena such as epidemics spread the way they do, and what impact various policy decisions will have on the result.

1.3 SEIR Model

As we will be using the SEIR model in our research, we will cover it briefly here. The SEIR compartmental model is a mathematical modeling of infectious diseases where a closed population of people move successively from compartments suspected to exposed to infected and to removed. We briefly provide an explanation of each compartment and leave the equations below.

- **Susceptible (S)** - These people are susceptible to getting the disease from someone infected.

- **Exposed (E)** - These people are no longer susceptible to the disease, and do not infect others. After a latent period, they become infectious.

- **Infected (I)** - These people will spread the disease to susceptible people. After a period of time they are removed by recovery, hospitalization, death.

- **Removed (R)** - Sometimes known as resistant or recovered. We will do our modeling with the term 'removed'. These people are no longer spreading the disease.

- **Contact Rate (β)** - Rate at which infected people infect susceptible people.

- **Latent Rate (α)** - Rate at which exposed people become infected.

- **Removal Rate (γ)** - Rate at which infected people become removed.

$$S(t) + E(t) + I(t) + R(t) = N$$

$$s(t) = S(t)/N, e(t) = E(t)/N, i(t) = I(t)/N, r(t) = R(t)/N$$

$$ds/dt = -\beta * s(t) * i(t)$$

$$de/dt = \beta * s(t) * i(t) - \alpha * e(t)$$

$$di/dt = \alpha * e(t) - \gamma * i(t)$$

$$dr/dt = \gamma * i(t)$$

The system of equations described above can be numerically solved given β, α, γ and initial values $S(0), E(0), I(0), R(0)$. And if we have values for S,E,I,R at certain times, we can fit β, α, γ to better define the disease's epidemic characteristics and predict its future course.

1.4   Other Compartmental Models

While we have done some research into simpler and more advanced models and are interested in cases such as super-spreaders, we believe the SEIR model to be sufficient for the modeling done in this paper. Basic SIR is insufficient because public health officials often make policy decisions based on positive case numbers. For example, an official may decide to impose strict isolation only after 100 positive cases. But by the time there are 100 cases of 'infected' people, there may be 1000 exposed people who will meaningfully impact epidemic statistics.

1.5   Agent Based Models

An agent based model(ABM) is a computational model used to simulate the effects of individual agents on a system. Some famous prior uses include Conway's Game of Life and Schelling's Segregation Model[**citations needed**]. ABMs typically consist of one or several types of agents with certain agent behavior interacting with an environment. For example, in Schelling's Segregation Model, the agents have a race and a tolerance level (of other races). If the agent finds the surrounding environment intolerable, they will independently move away.
ABMs offer a number of benefits over traditional mathematical models. Complex systems which cannot be easily solved mathematically can easily be simulated. And policy makers can help policy makers make decisions when mathematical results are not available and real world experiments are impractical.
For example, one popular use case of ABMs is in city urban planning to simulate modern traffic flow. Given empirical traffic flow data, it is virtually impossible to calculate the optimal traffic light configuration for a city and extremely expensive to run multiple

configuration experiments. However, ABM software with various vehicles as agents and the city traffic grid as the environment can easily find a satisfactory configuration **citation needed**.

1.6 Epidemics on Transportation Networks

Epidemics on transportation networks have been modeled in many different ways depending on the needs of the researcher. The most important differences are usually the type of transportation network, and the duration of interest, and a recurrent issue is modeling passenger flow. We briefly cite some prior research below:

- London subway guys

- Singapore bus guys

- Holistic model guys

- Tutu.ru guys

- NYC subway guys

In this paper, we focus on the NYC Subway at the start of the outbreak, but show a few other models to highlight common characteristics of transportation systems, explain general behavior of diseases on networks, and show the robustness of our framework for incorporating different models.

1.7 World Airline Network

1.8 Madrid Suburban Trains

1.9 Moscow Metro

1.10 COVID-19 and the NYC Subway

On April 24th, 2020, a researcher at MIT released a working paper finding that "The Subways Seeded the Massive Coronavirus Epidemic in New York City"[**citation needed**]. The paper received a lot of attention from the media. There is, at least, significant correlation between subway lines and severity of infection. But the factors involved are unknown and disputed. For example, high infection rate could be rated to dense housing, and housing is denser around subway stations[**citation needed**].

From this paper, and other prior research into subways, we bring into our own work two major points about disease spread on subways:

- Infection on a subway network depends much more strongly on the line than the geographical distance or network shortest path between stations.

- Infection on a subway network depends on the average commute time of commuters using the station.

An another important piece of information is the subway ridership. The NYC Metropolitan Transit Authority (MTA) publicly releases turnstile data. Below we show a figure demonstrating the gradual shutdown of NYC between March 9th and March 23rd.
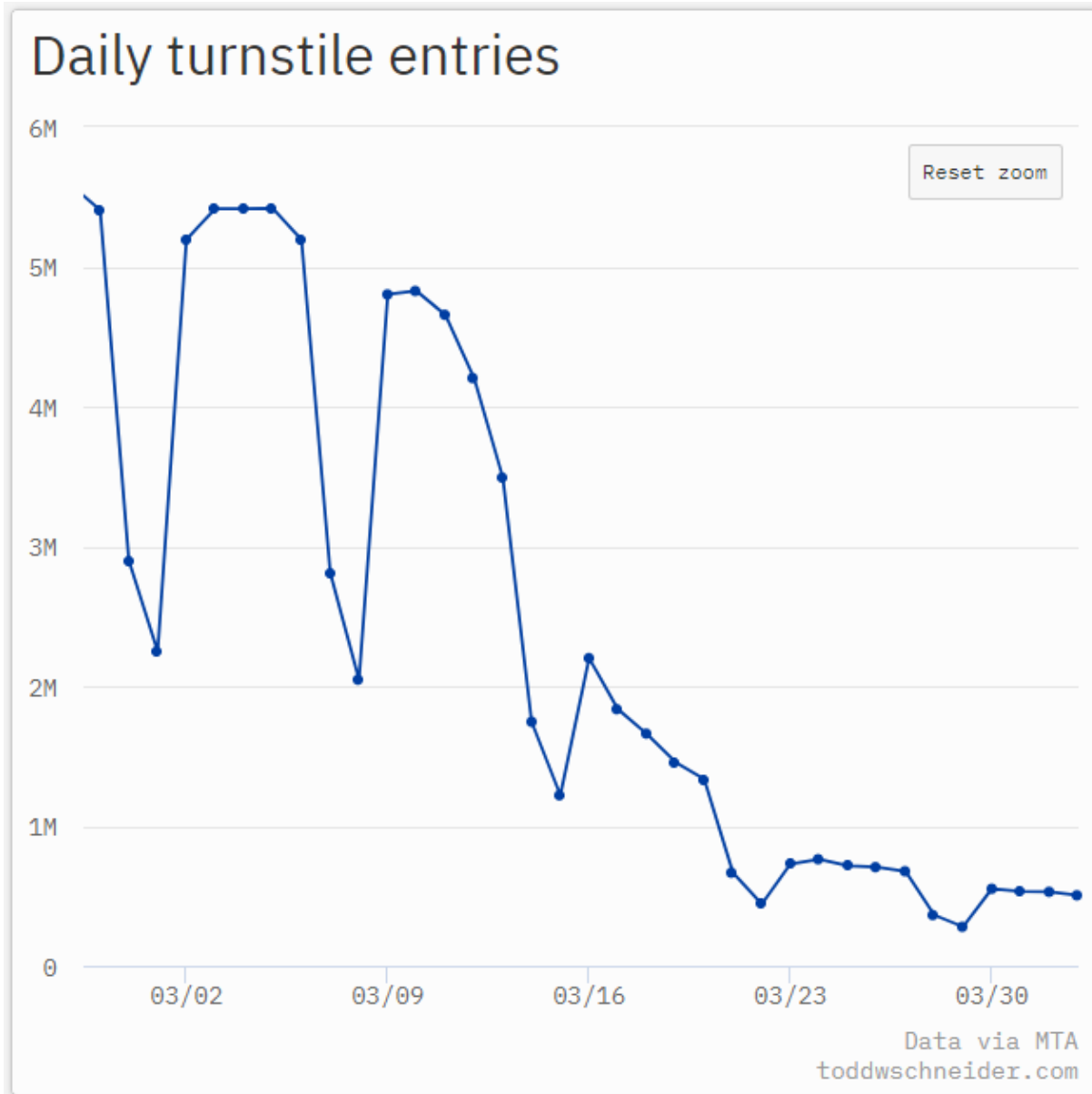
1.11 COVID-19 in NYC

In addition to general background knowledge about COVID-19, it would behoove the reader to know about the early spread of the disease in New York. Analysis of viral RNA in patients at the Mount Sinai Health System[**citation**] has lead researchers to conclude that the virus first came into the community through "multiple, independent but isolated introductions" from Europe and elsewhere in the USA. Below we have also given an approximate timeline of some of the most relevant events as of June 1st, 2020. We would also like to note that as forensic researchers begin examining the data, some significant

new dates or corrections may arise:

- February 25 - First positive test in NYC from a 39 year-old female healthcare worker flying back from Iran.

- March 3 - First confirmed P2P spread in NYC.

- March 9 - Mayor holds press conference and notes that there have been 16 confirmed cases.

- March 12 - Mayor declares a local state of emergency.

- March 15 - Schools officially close.

- March 22 - State order (PAUSE) to shelter in place comes into effect.

1.12 Subway Nomenclature and Other Definitions

- Station - Passengers enter subway stations in order to ride the subway to an exit station.

- Complex - Multiple stations can reside in one station complex.

- Turnstiles - Barriers at the entrance and exit of stations which record people entering and exiting.

- Line - The train tracks on which services and routes run.

- Service/Route - Trains follow specific routes between stations based on a timetable.

- Borough - A geographical region. NYC has 5 boroughs.

- MODZCTA - Modified Zip Code Tabulation Areas. They are very similar to zip (postal) codes, but are used due to postal codes sometimes not uniquely identifying geographical regions.

2 Data Sources and Preprocessing

2.1 MTA Station Data

This dataset has basic subway station information. It has stations, their longitude, latitude, associated complex (if any), line, route, and borough. We can find the MODZCTA from the longtitude and latitude data.

In this dataset, station 167 is listed twice, so we took care to combine the route data. We also split service 'S' into 3 different services as it represents 3 different shuttle services. We also split services on the 'A' line, into 3 different services depending on the destination.

2.2 MTA Turnstile Data

The MTA also publishes turnstile data with turnstiles reporting every 4 hours. Prior researchers have analyzed this data down to the most specific granularity**citation needed**. However, we only need a basic picture of the daily in and out flow of each station, so we take a look at aggregate passenger flow of each station between March 1 and March 21 (inclusive).

2.3 NYC COVID-19 Data

Since March 26, the NYC Health Department has been releasing and updating some COVID-19 on Github. Some data is incomplete or unavailable due to technical or privacy issues. For example, detailed case, death, and recovery numbers by MODZCTA only became available on May 18, 2020. However, a rudimentary record of positive tests for COVID-19 by modzcta has been available since April 1, 2020.

3 Methodology

3.1 MESA

To implement our ideas, we chose to use MESA, an ABM framework written in Python **citation needed**. There are many other ABM frameworks, but many of them are oriented towards beginners and lacking features to organize the code necessary for complex behavior. Others are not suitable for modeling a complex environment (network). MESA provides a simple framework with basic Agent, Model, and Schedule classes from which we can specify the necessary behavior. In addition, our group already has a working knowledge of modeling networks using Python's NetworkX, which will be necessary for specifying the environment of our model.
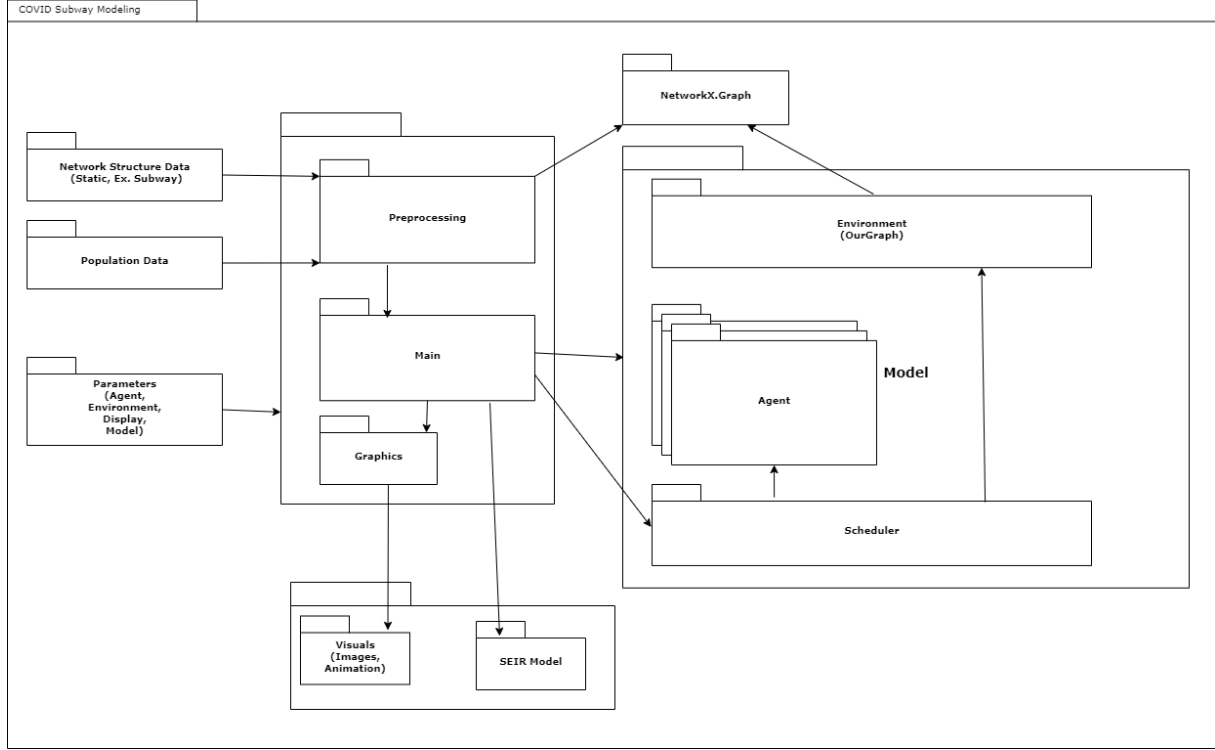
3.2 Modeling Framework

Before focusing on the NYC subways, we built a general framework for modeling various transportation problems. This allowed us to see specifically how subways (and subway commuters) differ from other types of transportation networks. It also allows us to extend our initial research into other networks and environments. In the appendix, the reader will find some other models we have built and the key point behind the modeling. Below is a UML diagram of our framework.

3.3 SubwayModel

...models NYC subway

3.3.1 SubwayAgent

represents population using the station and surrounding areas

NetworkX.Graph

Network Structure Data
(Static, Ex. Subway)

Preprocessing

Population Data

Environment
(OurGraph)

Parameters
(Agent,
Environment,
Display,
Model)

Main

Model

Agent

Graphics

Scheduler

Visuals
(Images,
Animation)

SEIR Model

3.3.2 SubwayGraph

represents the NYC subway. Nodes = stations, Edges = lines

3.4 NetworkX

used to track the graph

3.5 Important Parameters and Hyper-parameters

Hyper-parameters are capitalized and regular parameters are italicized. All parameters could be made into hyper-parameters, but we did not want to create unnecessarily complicated tuning.

3.5.1 Basic Epidemic Characteristics

- DEFAULT_BETA - The default β of the modeled disease in the SEIR model. It can be affected by countermeasures. Based on research, we set this to 1.75 **citation needed**.

- DEFAULT_GAMMA - The default γ of the modeled disease in the SEIR model. It can be affected by countermeasures. Based on research, we set this to 0.50.

- DEFAULT_ALPHA - The default α of the modeled disease in the SEIR model. It can be affected by countermeasures. Based on research, we set this to 0.20.

3.5.2 Countermeasures

- ISOLATION_COUNTERMEASURE - Models a government order to stay isolated after a certain number of people are infected (5000).

- RECOMMENDATION_COUNTERMEASURE - Models a government recommendation to be safe after a certain number of people are infected (500).

- AWARENESS_COUNTERMEASURE - Models increasing public awareness after a certain number of people are infected (5000). As time passes since the public first becomes aware of a problem, the infection and removal rates decrease due to self-initiative.

3.5.3 Location-specific Characteristics

- *'countermeasure defiance'* -

- *'exposure'* -

- *'commute time'* -

3.6 Algorithm

Below is a simplified algorithm of our model.

**Algorithm 1** Simulation of Disease Spread on Subways

---
1: **for** $i = 1; i < TIMESPAN; i + +$ **do**
2:     Check conditions (i, number of infected) to see if we should deploy COUNTER-MEASURES
3:     **for** Station in SubwayModel.Environment.Nodes **do**
4:         Calculate 'Local Exposure' from locally infected (and commute time).
5:         Calculate 'Route Exposure' from infected on the same route.
6:         Calculate 'General Exposure' due to city-wide infected.
7:         Update 'Exposure' At Station based on above conditions
8:     **end for**
9:     **for** Agent in SubwayNetwork.Agents **do**
10:         Get 'Exposure' At Location
11:         Get City-wide COUNTERMEASURES
12:         Get Percentage of commuters
13:         Calculate SEIR beta and gamma based on conditions
14:         Update SEIR numbers
15:     **end for**
16: **end for**

---

4 Fitting

4.1 Fitting to SEIR

We first fit overall SEIR numbers to NYC case, death, and recovery numbers. We only sought to fit the total cases to $I(t) + R(t)$ and adjusted the countermeasure properties to do so. Since we were at complete liberty to adjust these numbers, it is unsurprising that we created a near-perfect fit.
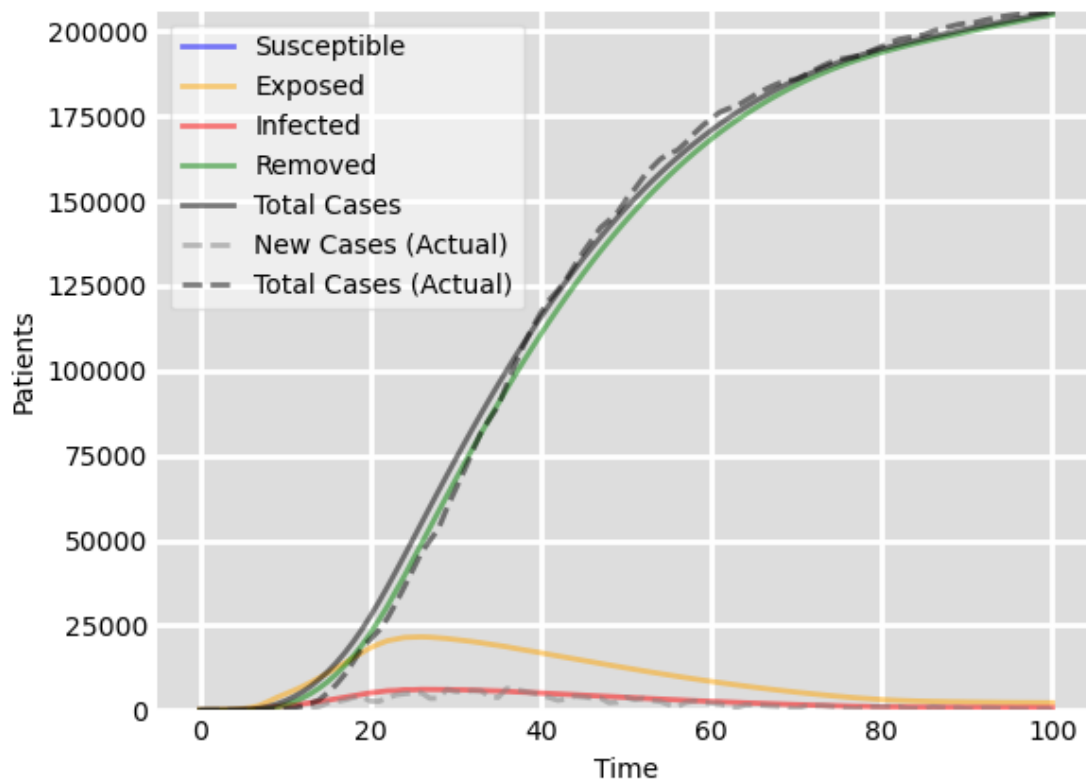
*MAPE for t=30+ is 0.0145*
We used the following hyper-parameters:

- RECOMMENDATION* - When 500 people are infected, divide β and γ by **1.5**

- ISOLATION** - When 5000 people are infected, divide β by **4**, and γ by **2**

- AWARENESS When 500 people are infected, β is modified by a negative exponential function down to a minimum of 25% of the original effect.

*Note that recommendations and isolation can be 'defied' if there are not enough infected people locally.
** Either the adjustment for recommendation or for isolation will take place.

4.2 Fitting to MODZCTA

Next we fit our data to MODZCTA. Due to the lack of numbers, there is no direct fitting. Also, it is difficult to visualize a timelapse in a report. Nevertheless, with some exceptions due to lack of subway coverage or missing data, our data fits actual infection hotspots very well.
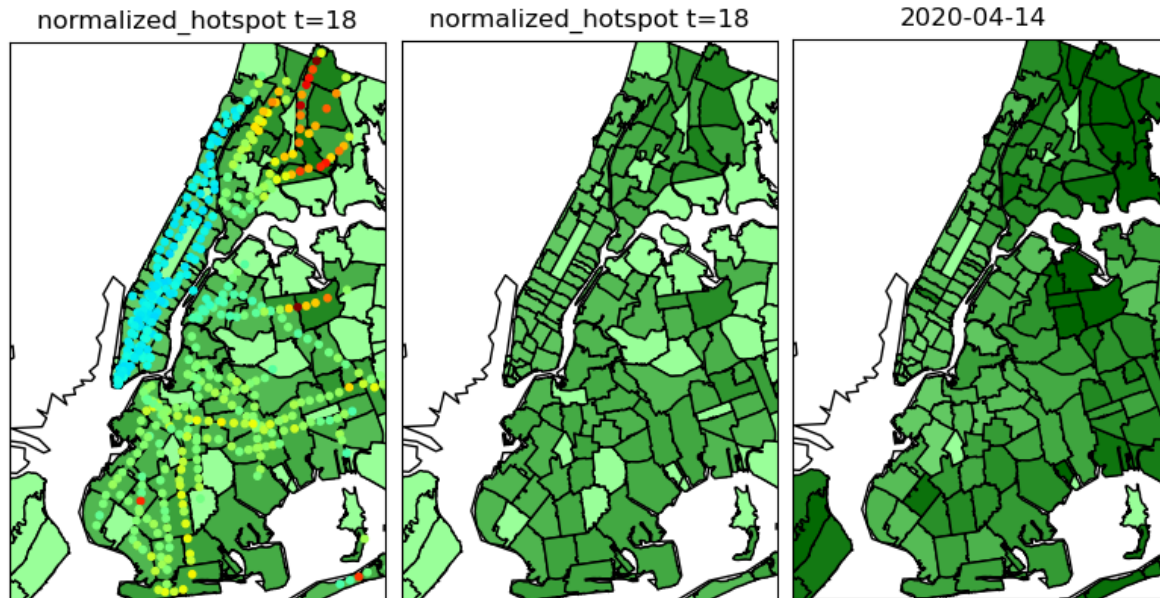
Scaling still TBD. Left is infected per capita (max = 0.1 percent) and Right is **total** *cases per capita (max = 0.2 percent).*

Discussion

Measure actual difference in geographical fitting Countermeasures shenanigans Fitting on a different network. London, Singapore, Tokyo, or Moscow Blah blah...

Conclusion

Writer's Note: All Models are Wrong, Some are Useful *...Predicating disease spread on subway lines, ridership at stations, commute time, and city-wide countermeasures, we were able to fit our model very closely to empirical data....*

References

[1] Mta station data. http://web.mta.info/developers/data/nyct/subway/Stations.csv.

[2] Mta turnstile data. http://web.mta.info/developers/turnstile.html.

[3] New york city subway usage. https://toddwschneider.com/dashboards/nyc-subway-turnstiles/.

[4] Albert-László Barabási and Márton Pósfai. *Network science.* Cambridge University Press, 2017.

[5] Philip Cooley, Shawn Brown, James Cajka, Bernadette Chasteen, Laxminarayana Ganapathi, John Grefenstette, Craig R. Hollingsworth, Bruce Y. Lee, Burton Levine, William D. Wheaton, and et al. The role of subway travel in an influenza epidemic: A new york city simulation. *Journal of Urban Health*, 88(5):982–995, Sep 2011.

[6] Ana S Gonzalez-Reiche, Matthew M Hernandez, Mitchell Sullivan, Brianne Ciferri, Hala Alshammary, Ajay Obla, Shelcie Fabre, Giulio Kleiner, Jose Polanco, Zenab Khan, and et al. Introductions and early spread of sars-cov-2 in the new york city area. Nov 2020.

[7] Jeffrey E. Harris. The subways seeded the massive coronavirus epidemic in new york city. http://web.mit.edu/jeffrey/harris/HarrisJE_WP2_COVID19_NYC_24-Apr-2020.pdf.

[8] M. Laskowski, B. C. P. Demianyk, J. Witt, S. N. Mukhi, M. R. Friesen, and R. D. Mcleod. Agent-based modeling of the spread of influenza-like illness in an emergency department: A simulation study. *IEEE Transactions on Information Technology in Biomedicine*, 15(6):877–889, 2011.

[9] Faïçal Ndaïrou, Iván Area, Juan J Nieto, and Delfim F M Torres. Mathematical modeling of covid-19 transmission dynamics with a case study of wuhan, Apr 2020.

5 Appendix

Here be all the models that we ended up not using