

Dummy Title
By Dummies
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.

We found that countermeasures significantly impacted the spread of coronavirus and that there is a correlation between infection hotspots and subway lines as well as commute time.

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 20.06.2020. 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

Background

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.

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.

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.

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.

Daily turnstile entries



Data via MTA
toddschneider.com

Epidemic Modeling

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.

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** - These people are susceptible to getting the disease from someone infected.
- **Exposed** - These people are no longer susceptible to the disease, and do not infect others. After a latent period, they become infectious.
- **Infected** - These people will spread the disease to susceptible people. After a period of time they are removed by recovery, hospitalization, death.
- **Removed** - Sometimes known as resistant or recovered. We will do our modeling with the term 'removed'. These people are no longer spreading the disease.

Writer's note: just take the equations from someone's overleaf

$$\begin{aligned}dS/dt \\ ds/dt &= -\text{beta} * s(t) * i(t) \\ de/dt &= \text{beta} * s(t) * i(t) - \text{alpha} * e(t) \\ di/dt &= \text{alpha} * e(t) - \text{gamma} * i(t) \\ dr/dt &= \text{gamma} * i(t) \\ S + E + I + R &= 1\end{aligned}$$

Newer Compartmental Models

While we have done some research into 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.

Epidemics on Transportation Networks

Writer's Note: epidemics on social networks blah blah linear threshold epidemics on transportation networks a little different, passenger flow is one of the most important numbers

Agent Based Models

It's just a tool. like any model. Simulating agents in an environment. Key parameters are agent, model, and environment parameters. For example, we have the following list thus far:

```
(Agent)STATUS_SUSCEPTIBLE = 'Susceptible'
(Agent)STATUS_EXPOSED = 'Exposed'
(Agent)STATUS_INFECTED = 'Infected'
(Agent)STATUS_INFECTED_ASYMPTOMATIC = 'Asymptomatic'
(Agent)STATUS_RECOVERED = 'Recovered'
(Agent)TIME_TO_RECOVER = 10
(Agent)TIME_TO_INFECT = 3
(Model)RUN_SPAN = 60
(Model)TOTAL_POPULATION = 50000
```

The environment is new york subway mapped as a network.
All of this is still being discussed.

Transportation Networks and Epidemics

Airlines

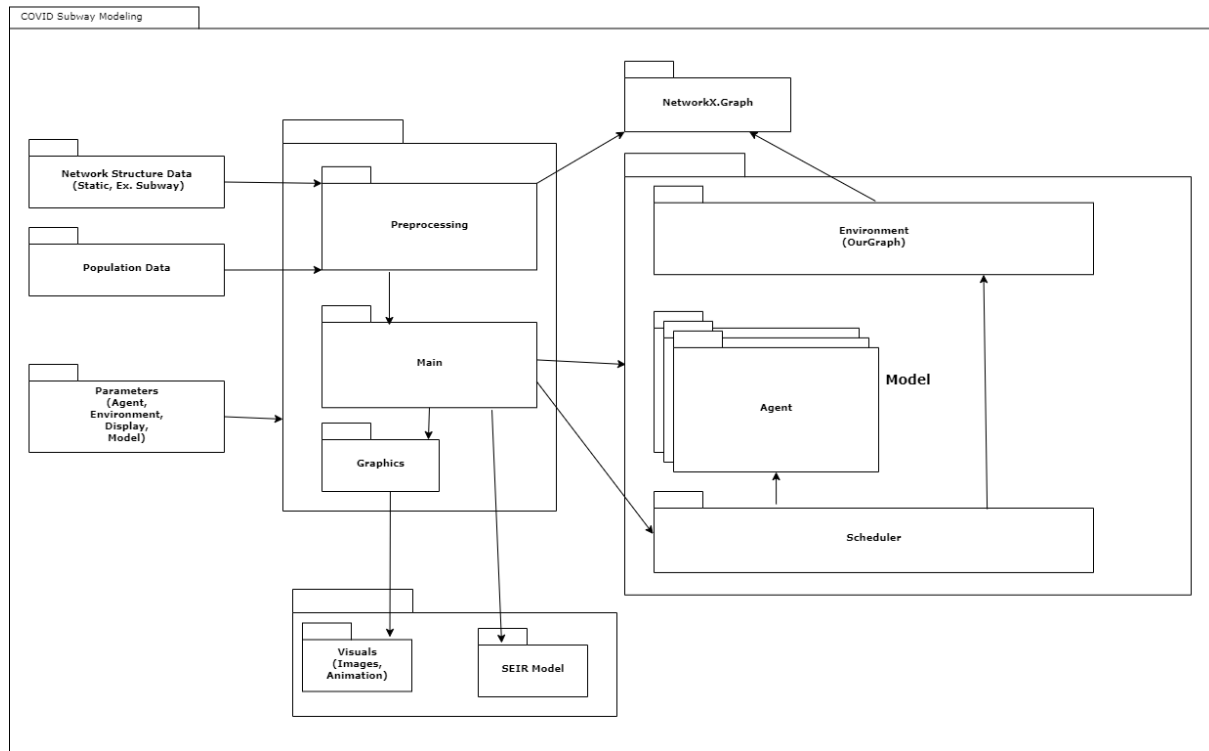
Subways and Subway Nomenclature

This guy is indispensable for figuring out how to parse some of this data:
https://en.wikipedia.org/wiki/New_York_City_Subway_nomenclature
The key is lines and routes

Methodology

MESA(Or our ABM)

Writer's note: I'm a bit too lazy to learn tikz-uml. I'll just make a regular uml and... screenshot it



NetworkX

MTA Station Data

Station, Complex, Line, Route

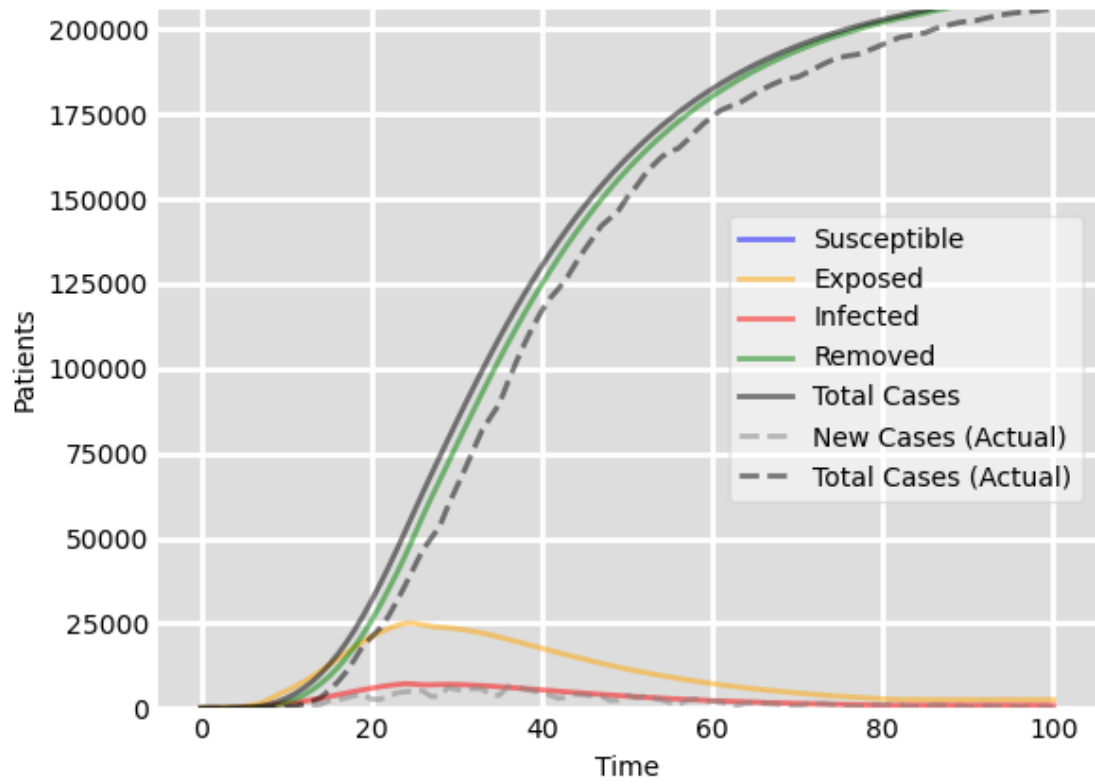
station 167... doubled

making edges. GTFS Data... no we do it ourselves

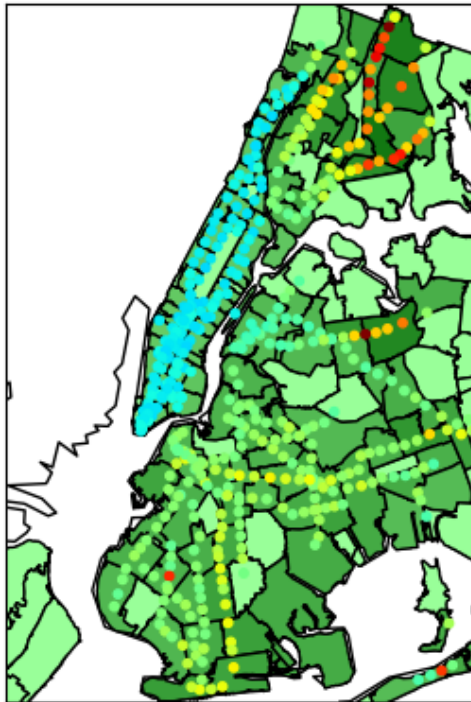
MTA Turnstile Data

discuss format. every 4 hours, every machine. aggregation.

Fitting the data



normalized_hotspot t=18



Writer's note: compare this to positive case coloring

Conclusion

Writer's Note: All Models are Wrong, Some are Useful

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://toddschneider.com/dashboards/nyc-subway-turnstile/>.
- [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.