

Realistic Modeling and Simulation of Influenza Transmission Over an Urban Community

Jiechen Chen

Sponsor: Dr. Gino Biondini

Department of Mathematics, State University of New York, Buffalo, NY 14260

March 20, 2015

Abstract

Infectious diseases that are spread through human contact can progress very rapidly in a population. One of the key factors in the spreading of contagion, and a main concern in attempting to stop the spread of illness, is the particular configuration of links among individuals in local communities within the larger population. This study uses a detailed individual-based, three-partite model comprising about 245,000 individuals located in an urban area in the Northeastern United States. Interactions among individuals are divided into family, workplace and pastime (service places, shopping, etc.), each occurring during a separate time period (daytime, pastime, and nighttime). Thus, the network allows one to model the spatial and temporal heterogeneity in the transmission of communicable diseases and to capture the differences between various individuals' vulnerability to infection. We performed Monte-Carlo simulations of the spreading of influenza through this network. Simulation results correspond well to the reported epidemic information. Results also demonstrate a temporal and population threshold which if exceeded, result in the long-term spread of infection. We expect that the findings will offer a valuable platform to devise spatially and temporally oriented control and intervention strategies for communicable diseases.

1 Introduction

Disease progression has traditionally been studied with the SIR model, a model which examines the change in susceptible, infected, and recovered populations. A lesser known way of tracking and predicting the movement of communicable diseases, which we will look at, is the individual-based model. In our model, we zero in on individuals in a population by examining each person's day-to-day routines. We track who each person is in contact with throughout each day and define the level of contact that any two individuals have with each other based on their reasons for being present at that location. This type of individual-based modeling allows for a better assessment of individual dynamics in a population while the SIR model is ideal for considering population dynamics as a whole. With individual-based models, there are more parameters to consider [9].

Individual-based modeling allows us to take into account certain parameters that would not have been considered in the traditional SIR model. This includes: length of latent period (when an individual is infected with the disease but cannot yet infect others), length of infectious period (when an individual can infect others), infection rate (the probability for a susceptible individual to be infected with the disease upon contact), and number of connections or contacts an individual has [4].

Merler and Ajelli [11] performed an individual-based study emphasizing the importance of population heterogeneity in Europe. In individual-based models, the population is assumed to be heterogeneous and disease transmission likelihood is different from person to person. Like in our study, they take into consideration the fact that timing and human mobility play a large role in the course of an epidemic. Lukens et al. [10] also used an agent-based model, but their study focused on differences between individuals' symptoms and infections over time. While their study looks at the population dynamics and how it relates to each individual's infection response, our model assumes a predetermined infection response among all individuals of the same age group who acquire the disease. We are more concerned with the infection progression dynamics, rather than the biological details of the infections and the symptoms it produces.

Individual-based modeling has also been done on a more global scale, but this is not comparable to an urban population which is much smaller in size. In larger geographic areas, there is a focus on air travel, border control, and differences between urban and non-urban regions within the whole region. Ferguson et al. [7] produced large-scale epidemic simulations by examining influenza outbreaks in Great Britain and the U.S. while Balcan et al. [2] compared the distribution of disease that occurs on various scales of commuting - from local small-scale commutes to longer airtime travel. Charaudeau et al. studied the spread of influenza in France [5]. Their study, like ours, also used census data, but included over 34 million people in the dataset while ours used only a quarter of a million. With a smaller dataset, we are able to take a more detailed look at the network interactions and how types of individual clusterings within an urban environment may impact the spread of disease, thus providing a structure for modeling infection dynamics for a specific, but popular, subset of locations. Charaudeau et al. provide a model that is more applicable for looking at entire nations and larger populations.

Previous mathematical models have also been successful in helping to create public policies in response to infectious diseases. In 2001 and 2003, modeling was used to monitor the spread of the foot and mouth disease in the United Kingdom and the spread of the severe acute respiratory syndrome, respectively [8]. Future research may further aid in the development of policy to prevent epidemics.

2 Model

2.1 Network

Unlike many models which are population-based, this network is individual-based and thus accounts for the heterogeneity and mobility of its members. The network is comprised of three populations (daytime, pastime, and nighttime) and two scales (local and societal). The three populations allow for three different models to occur throughout each day. There is a daytime population network, a pastime population network, and a nighttime population network. These different networks are a result of the mobility and heterogeneity of the individuals. As these individuals travel from work to home at the end of the daytime period, their network changes locally as they go from interactions with coworkers to interactions with family members. A person cycles through multiple different networks and therefore has different vulnerabilities to infection based on contact and on his or her own age group. This is all accounted for in this individual-based modeling system, which is an extension and further exploration of the data-verified framework designed by Bian et al. [3].

In our model, we use census data to estimate a network structure for each of the three time periods in a day. We do not directly use actual mobility data. Vazquez-Prokopec et al. studied an urban city in Peru with similar considerations, but instead of using census data, they used Global Positioning System (GPS) data-loggers to track movements of individuals [12]. This provided them with the actual locations for each

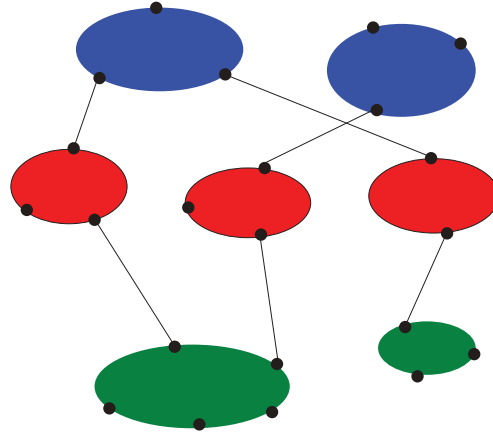


Figure 1: Schematic representation of the tripartite network. Red: homes; Blue: workplaces; Green: service places. Individuals are connected through home contacts, work contacts, and service contacts.

person for each period of each day. Their study specifically focused on urban environments which are resource-poor. From data-loggers, Vazquez-Prokopec et al. found that only up to 38% of individuals who were tracked showed predictable courses of movement through their days, thus revealing that resource-poor locations have less predictable temporal and spatial movements in comparison to more developed areas. This teaches us that an aspect of modeling disease progression that researchers need to take into consideration is the reliability of our prior. Our model assumes that we can predict daily movements because our model setting is not a resource-poor region.

We first examine the way that the daily routines of individuals are structured. This network is temporally three-partite; each day is broken down into three time periods: daytime, pastime and nighttime. Daytime is defined to be 9 a.m. to 5 p.m.; pastime is 5 p.m. to 12 a.m.; and nighttime is 12 a.m. to 9 a.m. This division is effective in modeling the realistic nature of most humans whose typical workday is in the 'daytime' or 'pastime' time frames, whose leisure activities are in the 'daytime' and 'pastime' time frame, and whose sleep and home interactions occur in the 'nighttime' time frame. Eubank et al. performed a very similar study to ours [6]. Their model was also individual-based and considered the spatial and temporal qualities of members of an urban community, but a noticeable difference is that Eubank's model employs second-by-second data while we break our 24-hour days into three periods. This parameter dissimilarity causes a major difference in the functionality of the two models. Eubank's model seems like an ideal fit when very detailed information is available. In our model, we generalize more and therefore provide a summary of disease progression dynamics. Our model runs very little risk of overfitting to data and therefore has a low bias. In other words, our model can be a middle ground between the SIR model (which is very general) and Eubank's model (which is very detailed).

The daytime network is comprised of people at one of three different types of locations. A person can be at home, at a non-service work location, or at a service work location (as an employee or a customer). Any other individual who is at the same location is considered a potential contact for the person within the person's daytime network. Figure 1 gives a pictorial representation of the way the network is comprised. Pastime is the next stage in a person's day. Once again, the three options for a person's location are: home, non-service work location, or service work location. Many individuals travel to a different location during the pastime. Those who were at work during the daytime may have returned home for the pastime period. Those who were at home during the daytime may have gone out to service places as customers during the

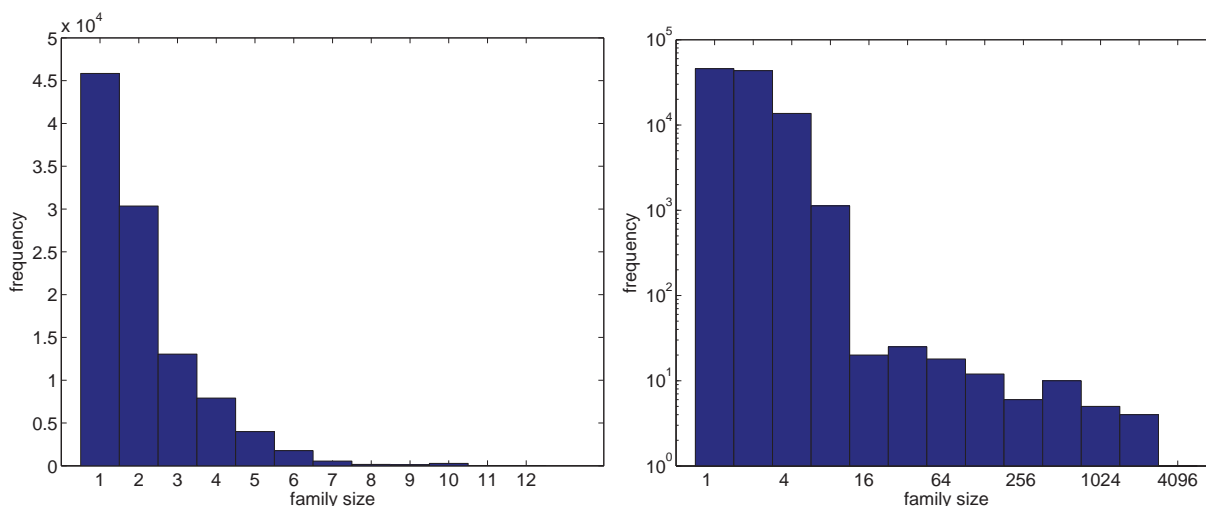


Figure 2: Size of family communities plotted on a linear scale (left) and double logarithmic scale (right). The linear scale shows the frequency of family which are of size 12 or less. Larger "families" are a result of community living arrangements rather than the typical definition of a related family unit.

pastime. The network changes as we transition to a different time.

Once nighttime occurs, each individual returns home. The nighttime population network is defined simply as the network of familial contacts. This is the single stage in which all members of the network are uniformly at their home locations. Once nighttime is over, it is daytime again. This cycle repeats for each day of the simulation.

This total population network has 245,809 members, divided mainly into cliques based on family communities and workplace communities. Every member of a family or workplace has a chance of interacting with every other member of the same family or workplace. There are a total of 104,137 different families, averaging 2.4 members per family. This number misrepresents the typical family size though as 44.0% of the families are singleton families (families with only one individual). Figure 2 displays histograms of family sizes. When considering interactions, we only allow a person to interact with a maximum of ten other individuals. This does not significantly restrict the network since only about 0.1% of families have more than 11 individuals. The largest family has 3,202 members; so individuals in this family have a much wider network of potential contacts, though each member still interacts with only ten of them.

There are a total of 37,283 unique weekday workplaces. 72.7% of the total population leave their homes to go to work during the daytime in 34,491 unique workplaces and 4.4% of the population leave their homes to go to work during the pastime in 5,765 unique workplaces. Daytime and pastime workplaces are not mutually exclusive to each other as there exists an overlap of work locations during the daytime and pastime. In total, 77.1% of the population regularly leave home to go to work locations. During the daytime, the average workplace size is 5.2 members while the largest workplace has 596 employees. Figure 3 displays a histogram of workplace sizes during the daytime. During the pastime, the average workplace size is 1.9 members while the largest workplace has 28 employees. Figure 4 displays a histogram of workplace sizes during the pastime. Of those who do leave their homes to go to work, 33.9% work at a location considered a service place. Service places are defined as location where employees have the possibility of interacting with customers, thus expanding their network of potential contacts. These contacts are very weak though

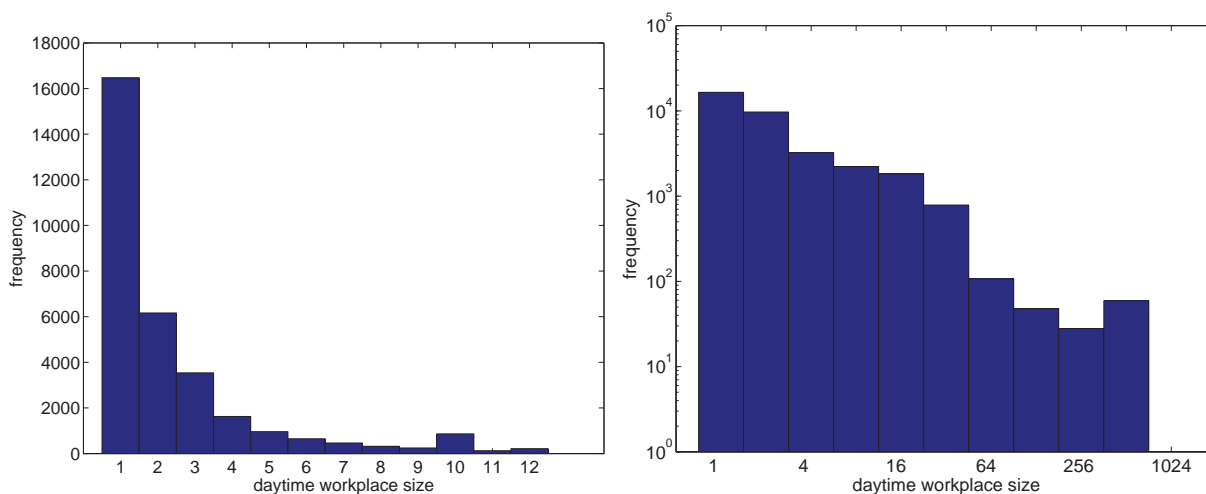


Figure 3: Size of workplace communities during the daytime plotted on a linear scale (left) and double logarithmic scale (right). The linear scale shows the frequency of daytime workplaces which are of size 12 or less.

compared to the type of contacts between members of the same family and coworkers in the same workplace.

While there are 19,753,981 edges in the home network, there are only 7,822,256 edges in the daytime work network and 13,072 edges in the pastime work network. Since there are almost triple the number of edges in the home network than in both of the work networks combined, we can hypothesize that, even before running simulations, the home network is more indicative of the spread of disease than the workplace networks.

Along with home and workplace interactions, individuals also have service interactions when they visit service places as customers. On weekdays, there are about 12,000 people who frequent service places as customers during the daytime and 43,000 people who frequent service places as customers during the pastime.

On weekends, the set of potential contacts for each individuals changes. Instead of going to their regularly listed locations of employment, some individuals stay at home while others have weekend jobs which are located at service places. On weekends, about 54,000 people visit service places during the daytime and about 8,000 people visit service places during the pastime as customers.

In summary, the network is constantly changing and each person is in contact with different people throughout different time periods of each day. The core of the network occurs through contacts within home communities though; there are more opportunities for infections to spread among those who reside in the same location.

2.2 Disease transmission

Before we examine the spread of illness, we classify each individual to be one of four main statuses: susceptible, latent, infectious, or recovered. A susceptible person is anyone who is able to get the illness. Once a person gets the illness, the individual is classified to be latent. The latency period, when a person is sick but cannot yet infect others, is two days. Once the latent period is over, a person is infectious. Adults are infectious for five days; children are infectious for eight days; and seniors are infectious for five days. Fol-

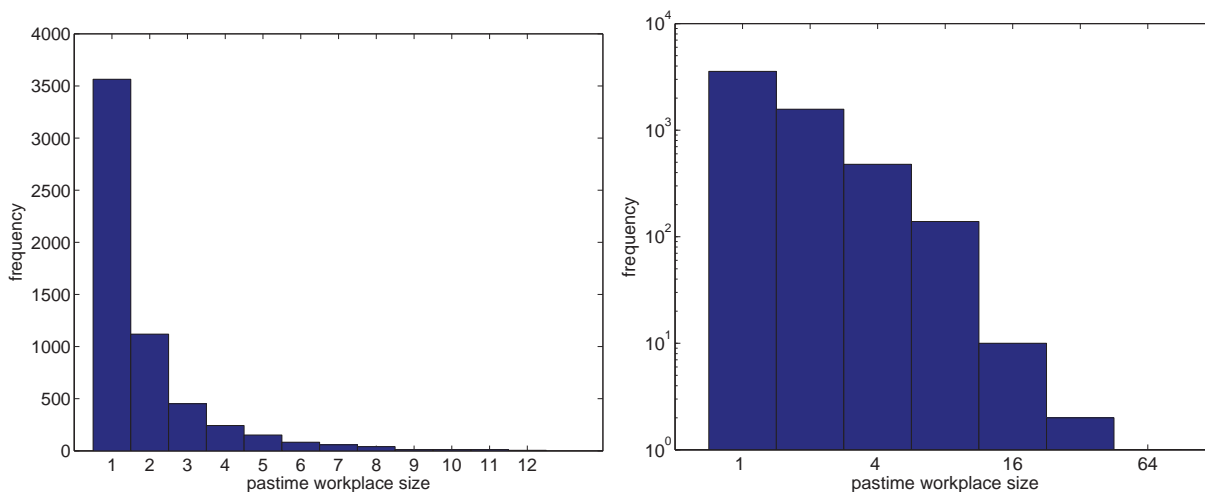


Figure 4: Size of workplace communities during the pastime plotted on a linear scale (left) and logarithmic scale (right). The linear scale shows the frequency of pastime workplaces which are of size 12 or less.

lowing the infectious period, a person moves to the recovered category and can no longer give or receive the infection. People in this category are immune for the rest of the simulation.

In order for the disease to be passed from one person to another, the two people have to be considered contacts for that period of that day. There are two types of contacts: regular contacts and casual contacts. Regular contacts are those which last a long time and will be more effective in allowing the disease to spread from one person to another. Casual contacts are less serious and have one tenth of the probability of passing the disease from one person to another. In our model, a regular contact is given a contact rate of 1 while a casual contact is given a contact rate of 0.1.

Not all contacts will result in a successful transmission of disease. Different age groups have different infection rates or likelihoods that a contact between two individuals will result in a susceptible person getting the infection from the infectious one. Adults have an infection rate of 3% while children and seniors have infection rates of 5%.

With such restrictions imposed on disease transmission, a successful transmittance of infection from one individual to another has a very low possibility. To obtain the actual probability of person A getting the disease from person B, we multiply the infection rate from person A's age group by the contact rate that is between person A and person B. For example, say that person A is a child visiting a store where person B is an employee at that store. Say that person B is infectious. Then the likelihood of person A getting sick from person B is $0.1 * 0.05 = 0.005$. Person A will have a 0.5% chance of getting infected since there is a 0.1 contact rate between the two individuals and a 0.05 infection rate for person A since he or she is a child.

Once a person has the disease for five days (including the two days in the latent period), symptoms of illness appear. Once symptoms appear, it is assumed that all children, all seniors, and 32.6% of adults do not leave their homes until they are recovered. Those who do not leave their homes are considered withdrawn. Once withdrawn, these individuals can no longer spread illness to coworkers, customers, or anyone else who doesn't live in the same home as them.

To summarize, we describe a single day in our model. We start the model on a Monday morning during the daytime. During this time, we have four scenarios that occur. In each scenario, we remember to limit

the number of people that a person may be in contact with to ten. So in the case that a person is at a location with more than ten other people, the simulation will randomly choose ten people to be that person's contacts for that specific time period of that day.

Scenario 1 consists of individuals who go to work at non-service locations during this time period. Since workplaces are considered locations where there is constant close contact, interactions in this scenario are considered regular contacts.

Scenario 2 consists of individuals who go to work at service locations during this time period. Instead of only being in contact with coworkers, these individuals can also interact with customers. With the ten-contact limit, the ten contacts will be randomly chosen from a heterogeneous mixture of coworkers and customers who are at the same location. Contacts between an employee and customer are not as serious, so these types of interactions are classified as casual contacts. Interactions between service workers are still considered regular contacts.

Scenario 3 consists of individuals who are not employed during this time period and go to service locations as customers. They interact with employees and other customers at the service location that they are visiting. Thus, their contacts are also randomly selected from a mixture of customers and employees. All contacts here are casual.

Scenario 4 is the remaining group of individuals who do not leave their homes during this time. They only interact with other members of their families who are also home. All contacts here are regular contacts since there is a lot of interaction between members of the same household. This group consists of both people who are unemployed during this time period and those who are withdrawn because disease symptoms have appeared.

For each scenario, individuals who are infectious can pass infection on to others who are in the same location. Once these scenarios have taken place ubiquitously in the daytime network, the four scenarios are repeated for the pastime network. After this is completed, we move on to the nighttime period, when every individual is considered to be in scenario 4. All children, adults, and seniors are at home and are solely interacting with members of their families (people residing in the same home).

Once nighttime is complete, the same three-partite day is simulated for Tuesday, Wednesday, Thursday, and Friday. On Saturday and Sunday, people do not work at their regular workplaces. Special data concerning service locations indicate that some individuals hold service jobs outside of their regular employment on the weekends. Therefore, we apply scenarios 2 - 4 to the weekend since the only people who are working are those who work at service locations. Similar to the weekday case, there are daytime and pastime periods when individuals leave their homes and interact with others who are at service places. There are still some people who stay at home and therefore scenario 4 takes place during both the daytime and pastime. During the nighttime, the weekends are just like the weekdays: all individuals are at home and only interact with family members. The simulations are allowed to run for a maximum of 1000 days.

3 Implementation

We first used shell scripting to get our files into a format that we could easily extract information from by manipulating files which were generated from census data and creating contact networks. Then we simulated disease transmission by creating algorithms in MATLAB¹ with the criteria described in the previous section.

There are three groups of files (totaling over 300 MB in size) that we received from Bian et al. which we adapted and combined to summarize data for the model: home/individual data files, work data files,

¹MATLAB code can be found at <https://github.com/jcc1125/SIURO.git>

and service data files. Any information from these files had an associated ID number which indicated the pertinent individual. This ID number was consistent through all of the data.

The home/individual files provided information such as age group, gender, and family index for each individual in the network. From this data, we were able to create the home matrix by matching members of the network who lived in the same home and are classified as part of the same family. The work data files were in a similar format to the home data files. Each individual has a row with their work locations and details such as what time of the day the person worked and the commute time. The set of service data files was the largest. For each day of the month, there was a list of the all service locations and individuals who were at each location (and delineated by the individual's reason for his or her presence as an employee or a customer). In order to efficiently gather the potential contacts a person has at home, work, and service locations, we store each type of contact in different matrices and list formations.

A sparse matrix is used to represent the adjacency matrix of family connections. For each entry e_{ij} in the matrix, e_{ij} is 1 if individual i and individual j reside in the same home. e_{ij} is 0 otherwise. The same representation is used for the daytime and pastime work adjacency matrices. Sparse matrices are used because they present a more efficient way of storing data. Since the average number of contacts each individual has at each location is under ten, a full network of 245,809 by 245,809 is unnecessary since a sparse network can capture our relatively small number of non-zero entries. The work matrices (one for daytime work connections and one for pastime work connections) are stored in a similar way with adjacency matrices.

The service data is stored in a different way. Therefore, we use a different method to store service contacts. Adjacency matrices are not ideal for this situation since 56 different adjacency matrices would be needed (there are 28 days in the service data cycle and each day has two time periods in which people can be at service locations). To retain the most data, we use a 245,809 by 56 sparse matrix where each row is the index of an individual in the network and each column is a time period (28 days * 2 periods per day). The entries in this matrix are location indices of the service places. When running simulations, we simply grab the location of each person and find others who also have the same location index for that time period. This saves us computational time and space since we are only using one matrix.

Once all of the initial data files are summarized in formats that we can easily pull information from, we use MATLAB to model the progression of disease using the criteria described in the previous section.

4 Simulations and results

We use the framework designed by Bian et al. to run many simulations [3]. In order to attempt to produce reproducible results, each type of simulation is run 500 times. A random employee is selected as the starting source on day 1 of the epidemic. This means that this employee is the only person who initially has the illness and all other infections are a result of the spread of infection from this source individual. We progress the disease according to the rules stated in the previous section. We also test whether the addition of service data is necessary for the model at all. Since service contacts have a contact rate that is only 10% of the regular contact rate, we hypothesize that the removal of such data is a way to make the model cleaner, yet have minor effects on the network simulations as a whole.

Figure 5 shows an overview of the number of new sick cases on each day of the simulation. There is no evidence to indicate that the addition of service data has any effect on the number of newly sick cases on each day of the simulation. Both graphs in Figure 5 display similar trends in newly appearing cases.

Figure 6 zooms in on the propagation of the disease over the first 200 days. The graph on the right in Figure 6 is used to verify the results with the simulations performed by Bian since it does take into account

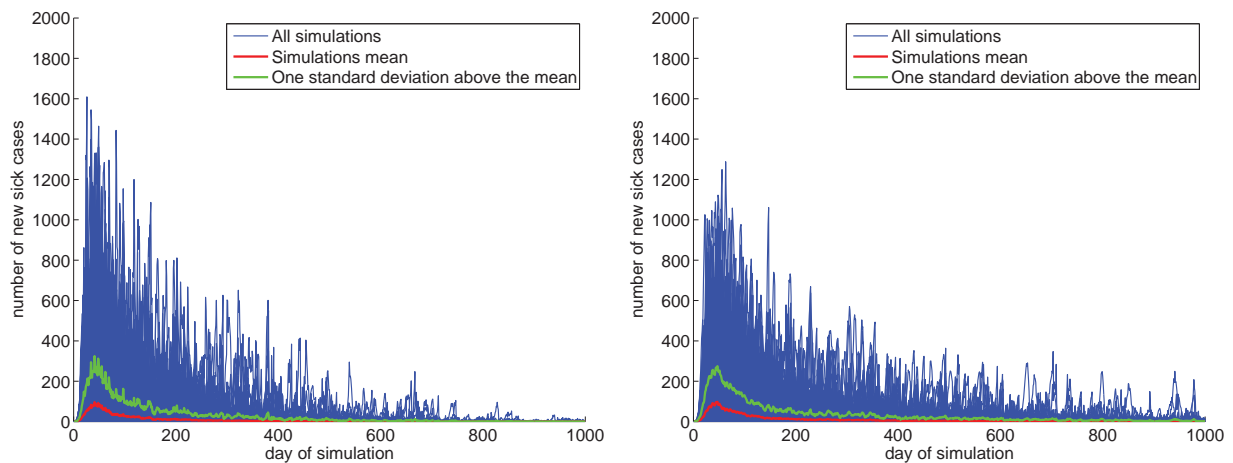


Figure 5: Number of new sick cases on each day of simulation. 500 runs were performed without the service data (left) and 500 runs were performed with the service data included (right).

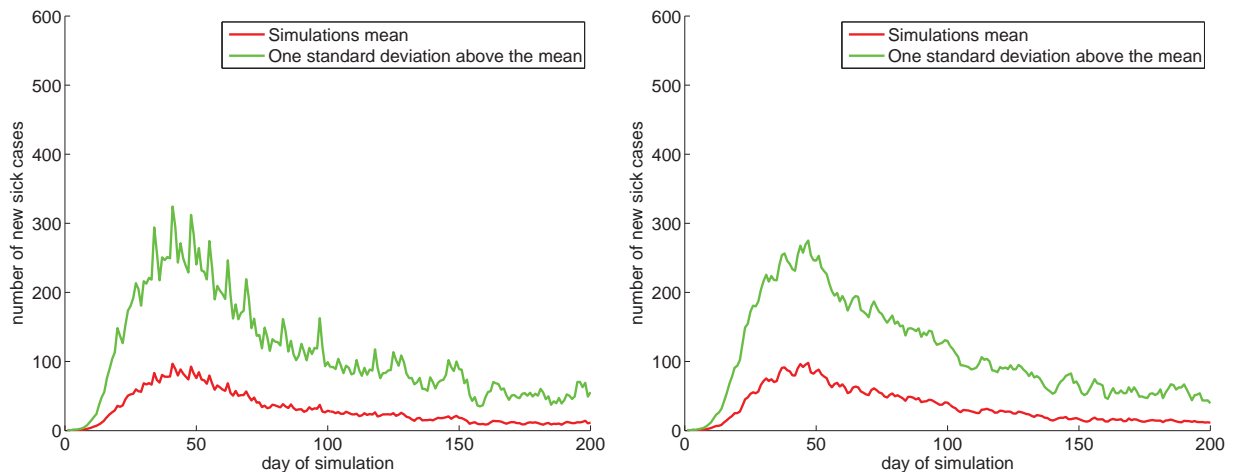


Figure 6: Number of new sick cases on each day of simulation; day 1 – 200. 500 runs without service data (left) and 500 runs with service data included (right).

service data like Bian’s model did. A comparison of the graphs indicate a similar curve for the mean, but our results produce a wider range of results for the standard deviation. Another difference is that our curve peaks around day 50 while Bian’s curve peaks around day 85. Our peaks show similar overall patterns though as both graphs demonstrate small dips throughout the simulations, a trend that occurs as a result of the different routines that occur on the weekends. This is because that during the weekends, people do not go to their regular places of employment and thus only have regular contact with family members and casual contacts with others at service places.

Along with the progression of disease, we also look at summary statistics of the simulations as a whole. The top left and bottom left graphs of Figure 7 show that out of the 500 runs of the Monte-Carlos simulations with only home and workplace data involved, a little more than half of the total runs affected less than 64 individuals. This affected subgroup is a very insignificant fraction of the entire population. If the disease

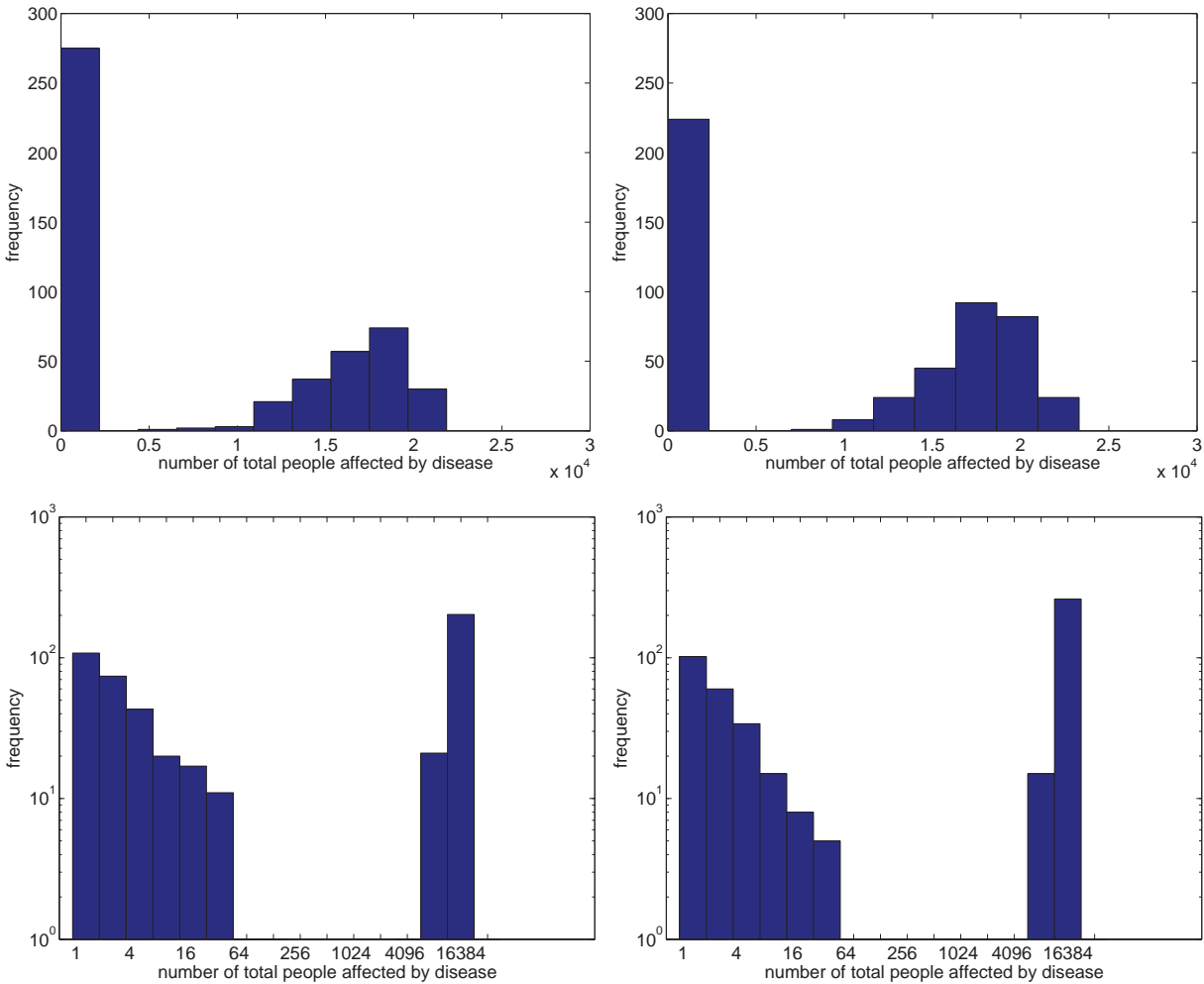


Figure 7: Total number of people affected by the disease. 500 runs without service data plotted in linear scale (top left) and double logarithmic scale (bottom left). 500 runs with service data plotted in linear scale (top right) and double logarithmic scale (bottom right).

does manage to affect more than 64 individuals, then it will likely affect more than 10,000 individuals, but still less than 25,000. The significant gap between 64 individuals and 10,000 individuals shows that there are two main options for the route of progression - either a very local network of individuals are affected (perhaps people who are only within a couple of degrees of separation from the source) or the disease manages to spread to about 5% of the entire network.

In comparison, the top right and bottom right graphs of Figure 7 show the total number of people affected by the disease when service interactions are taken into account. There is a similar bimodal distribution of the frequency of the number of people affected. In this case, a little fewer than half of the runs show the disease affecting less than 64 people. There is a slightly greater frequency of runs in which the disease affects about 20,000 individuals. We can see this in Figure 7 as the double logarithmic graph with service data shows the second peak at a frequency of about 150 runs while the double logarithmic graph without service data has a second peak at a frequency of about 100 runs. In summary, there is a greater chance that the disease leaves

simulation	infections at home	infections at workplace	infections at service places
without service data	90.40%	9.60%	n/a
with service data	83.91%	15.94%	0.15%

Table 1: Location of disease transmission - total of all runs

simulation	infections at home	infections at workplace	infections at service places
without service data	77.76%	22.24%	n/a
with service data	76.80%	22.96%	0.24%

Table 2: Location of disease transmission - averaged for each run

the local network of the source if service interactions are taken into account.

Infections can occur at a person's home, at a person's workplace, and at service places a person visits. There are at least two methods to calculate the relative size of infections that occur at each location: one by looking at the total percentage of people affected at each location across all 500 runs of the simulation and another by looking at the percentage of people infected at each location in each simulation. The first method uses:

$$\alpha_H = \frac{\sum_{i=1}^n H_i}{T}, \quad \alpha_W = \frac{\sum_{i=1}^n W_i}{T}, \quad \alpha_S = \frac{\sum_{i=1}^n S_i}{T}, \quad (1)$$

where α_H , α_W , and α_S are the percentage of infections that occur at home, work, and service locations respectively, n is the number of runs in the simulation, $T = \sum_{i=1}^n H_i + \sum_{i=1}^n W_i + \sum_{i=1}^n S_i$ (the total number of infections among all runs) and H_i , W_i , and S_i are the number of people who are infected at home, work, and a service location, respectively, during the i -th run of the simulation.

Alternatively, we can average the fraction of infections occurring in each individual run:

$$\beta_H = \frac{\sum_{i=1}^n \frac{H_i}{T_i}}{n}, \quad \beta_W = \frac{\sum_{i=1}^n \frac{W_i}{T_i}}{n}, \quad \beta_S = \frac{\sum_{i=1}^n \frac{S_i}{T_i}}{n}, \quad (2)$$

where β_H , β_W , and β_S are the analogues of α_H , α_W , and α_S computed with the alternative method, and $T_i = H_i + W_i + S_i$ is the total number of infections for the i -th run.

Table 1 shows the relative percentage of infections which occur at home, work, and service place from all infections over all 500 runs. Table 2, on the other hand, shows the averaged percentages of each run for all 500 runs (Note: only runs in which the disease progresses past the initial source(s) are taken into account). The discrepancy between the data in Table 1 and Table 2 suggests that the runs with more total infections occurring overall will have a higher percentage of infections occurring at home. This means that as the the disease progresses further on, it will have a higher chance of spreading at home.

We also observe a relationship between the total number of people who are affected by the disease and the total number of days that the disease lasts. Figure 8 shows that for both the simulations with and without service data, there is a positive and prominent correlation between the number of people affected by the disease and the length of outbreak. This allows us to predict how many people in total will be affected by the disease for each day of the simulation.

Figure 9 shows the frequency of outbreak duration over the 500 runs. The disease rarely progresses to 1000 days when service data is not included, as we can see in both the linear and double log graphs in

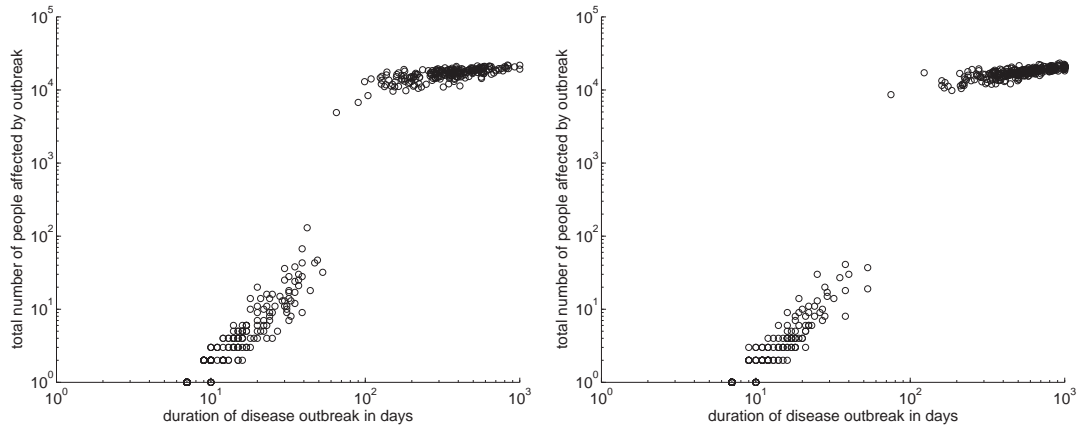


Figure 8: Number of people affected vs. Length of outbreak - scatterplots of 500 runs of each simulation without service data (left) and with service data (right).

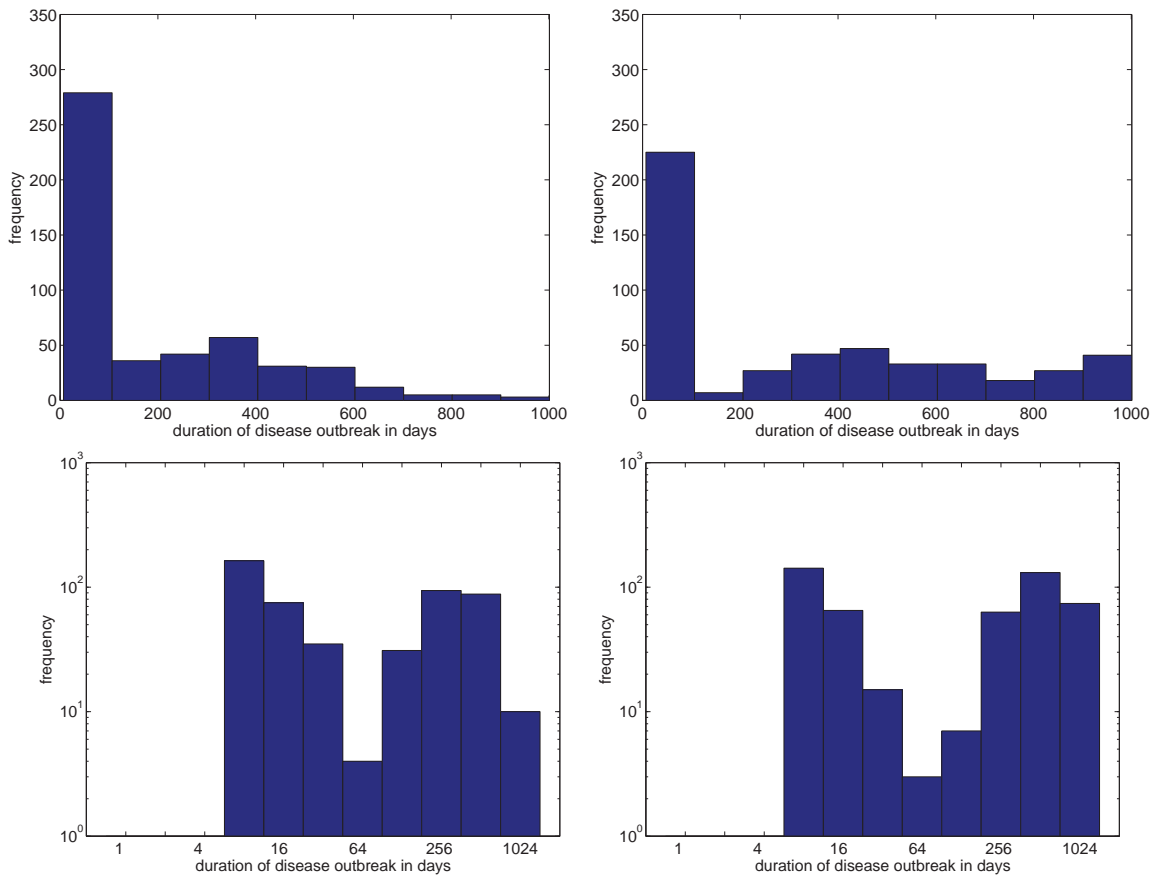


Figure 9: Length of days of disease outbreak. 500 runs without service data plotted on a linear scale (top left) and double logarithmic scale (bottom left). 500 runs with service data plotted on a linear scale (top right) and double logarithmic scale (bottom right).

Figure 9. In fact, the bimodal distribution of the loglog plot shows that about half of the runs end before day 64.

On the other hand, the linear and double logarithmic graphs for the runs with service data (Figure 9) show that when service interactions are taken into consideration, there are a number of simulations which progress to 1000 days. Without our cap of 1000 days, the outbreak may continue to progress. With the inclusion of service interactions, the linear graph (Figure 9 bottom right) shows a slower decline in the frequency of outbreak.

For both the simulations with and without service data, the loglog plots show a bimodal distributions. This suggests that there is a slight "hump" to get over around day 64 (the dip in the graphs). This is similar to the shape of the graphs in Figure 7. Both patterns may be used as benchmarks in the monitoring of disease progression. If the disease progresses past 64 days, then we may consider the disease to have a high potential of posing a longtime threat. Analogously, the disease could be considered a potential threat to a large portion of the population if more than 64 people are initially affected.

5 Discussion and further work

By carefully analyzing the spread of disease in our network, we can devise methods to monitor and prevent the progression of disease. Akin to general epidemiological models, this model has a "threshold" that results in a formidable outbreak when it is exceeded [1]. If this threshold is not met, the disease shows evidence of dying out as seen in the runs where outbreak duration is less than 64 days. Therefore, in future simulations of our data, one can use 64 days or 64 individuals as a benchmark in assessing the seriousness of the epidemic. When such a benchmark has been surpassed, there is a need to implement more dire quarantine or recovery measures. (Of course, such an approach would be hard to realize in practice.)

Also, a proposed method for stopping the spread of infection can focus on monitoring interactions at workplaces early in the epidemic and monitoring interactions at homes later in the epidemic. Specifically, two policies that have been examined are: prophylaxis using antiviral drugs and quarantine where those who are in the same household as someone who is sick are asked to stay home. Ferguson et al. [7] examined different ways of containing the outbreak in a nation-wide scale, but many of the concepts can be adapted to our urban community model. This method acknowledges not only the potential risk to others that those who are sick pose, but also the risk to others that those who are in the same family as those who are sick pose.

Now that we have modeled the population network and the spread of illness from a single source, we are considering various parameter changes in our simulation. We will look at what happens when there are multiple sources at the beginning of the epidemic and what happens when new sources are added to the model daily. In addition, we will explore methods of preventing the spread of illness through targeted vaccination strategies and other potential methods.

References

- [1] M. E. ALEXANDER, Mathematical modeling of infectious diseases: methods and some results, Fuzzy Information, 2004. Processing NAFIPS '04. IEEE Annual Meeting of the, 2(2004), pp.675-679.
- [2] D. BALCAN, V. COLIZZA, B. GONCALVES, H. HU, J. RAMASCO, and A. VESPIGNANI, Multiscale mobility networks and the spatial spreading of infectious diseases, Proceedings of the National Academy of Sciences of the United States of America, 106(2009), pp. 21484-21489.

- [3] L. BIAN, Y. HUANG, L. MAO, E. LIM, G. LEE, Y. YANG, M. COHEN, and D. WILSON, Modeling Individual Vulnerability to Communicable Diseases: A Framework and Design, *Annals of the Association of American Geographers*, 102(2012), pp.1016-1025.
- [4] L. BIAN and D. LIEBNER, A Network Model for Dispersion of Communicable Diseases, *Transactions in GIS*, 11(2007), pp.155-173.
- [5] S. CHARAUDEAU, K. PAKDAMAN, and P. BOELLE, Commuter Mobility and the Spread of Infectious Diseases: Application to Influenza in France, *PLoS ONE*, 9(2004).
- [6] S. EUBANK, H. GUCLU, V.S.A. KUMAR, M. V. MARATHE, A. SRINIVASAN, Z. TOROCZKAI, and N. WANG, Modelling disease outbreaks in realistic urban social networks, *Nature*, 429(2004), pp.180-184.
- [7] N. FERGUSON, D. CUMMINGS, C. FRASER, J. CAJKA, P. COOLEY, and D. BURKE, Strategies for Mitigating an Influenza Pandemic, *Nature*, 442(2006), pp.448-452.
- [8] T. D. HOLLINGSWORTH, Controlling Infection Disease Outbreaks: Lessons from Mathematical Modeling, *Journal of Public Health Policy*, 30(2009), pp.328-341.
- [9] M. J. KEELING and L. DANON, Mathematical modeling of infectious diseases, *Br Med Bull*, 92(2009), pp.33-42.
- [10] S. LUKENS, J. DEPASSE, R. ROSENFELD, E. GHEDIN, E. MOCHAN, S. BROWN, J. GREFFENSTETTE, D. BURKE, D. SWIGON, and G. CLERMONT, A large-scale immuno-epidemiological simulation of influenza A epidemics. *BMC Public Health* 14(2014).
- [11] S. MERLER and M. AJELLI, The role of population heterogeneity and human mobility in the spread of pandemic influenza, *Proceedings of the Royal Society B -Biological Sciences*, 277(2010), pp. 557-565.
- [12] G. VAZQUEZ-PROKOPEC, D. BISANZIO, S. STODDARD, V. PAZ-SOLDAN, A. MORRISON, J. ELDER, J. RAMIREZ-PAREDES, E. HALSEY, T. KOCHER, T. SCOTT and U. KITRON, Using GPS Technology to Quantify Human Mobility, Dynamic Contacts and Infectious Disease Dynamics in a Resource-Poor Urban Environment, *PLoS ONE*, 8(2013).