Modeling the Effects of Transient Populations on Epidemics

Nidhi Parikh and **Sushrut Shirole** and **Samarth Swarup**

Network Dynamics and Simulation Science Lab
Virginia Bioinformatics Institute
Virginia Tech
Blacksburg, Virginia 24060, USA.

Abstract

A large number of transients visit big cities on any given day and they visit crowded areas and come in contact with many people. However, epidemiological studies have not paid much attention to the role of this subpopulation in disease spread. In the present work, we extend a synthetic population model of Washington DC metro area to include leisure and business travelers. This approach involves combining Census data, activity surveys, and geospatial data to build a detailed minute-by-minute simulation of population interaction.

We simulate a flu-like disease outbreak both with and without the transient population to evaluate the effect of the transients on outbreak size and peak day in terms of number of residents infected. Results show that there are significantly more infections when transients are considered. We also evaluate interventions like closing big museums and encouraging use of hand sanitizers at those museums. Surprisingly, closing museums does not result in a significant difference in the epidemic. However, we find that if the use of hand sanitizers reduces the infectivity and susceptibility to 80% or 60% of the original values, it is as effective as closing museums for a few days or entirely eliminating the effect of transients. If infectivity and susceptibility are reduced to 40% or 20%, it reduces the number of resident infections over the period of 120 days by 10% and 13%.

Introduction

Influenza, also referred as flu, is an infectious disease that spreads through air by coughs, sneezes or through contact with contaminated surfaces. The Centers for Disease Control and Prevention (CDC) estimated 24.7 million influenza cases annually based on 2003 population demographics (Molinari et al. 2007). It spreads as a seasonal flu but could potentially result into a pandemic. Modern trends like urbanization and increased travel provide further opportunities for the spread of infectious diseases. Governments must be prepared with the strategies to mitigate the spread. However, such contagious diseases spread through person to person contact over a social contact network. In areas like large cities, these networks can be very large and quite dense. This makes planning and responding for even a moderate disease outbreak quite challenging due to the difficulty of forecasting the epidemic spread, and finding and implementing the right interventions to contain the epidemic.

Computational approaches to epidemiology involve modeling the social contact network and exploring different outbreak scenarios and comparing various intervention strategies. However the social contact network cannot be obtained through traditional survey-based methods alone because of its size and complexity. At the Network Dynamics and Simulation Science Lab (NDSSL), detailed, high fidelity, social contact networks have been synthesized by combining data from multiple sources, including the US Census, geospatial data, the National Household Travel Survey, and others. Using these synthetic social contact networks, we conduct detailed and large-scale simulations of disease outbreaks using various high-performance computing softwares developed at NDSSL.

In the present work we explore the impact of transient populations—tourists and business travelers—on influenza-like illnesses (ILI). A large number of transients visit big cities through out the year, e.g., Washington DC has about 50000 transients including leisure travelers and business travelers visiting on any day. They usually visit high traffic areas in the city and come into contact with each other and also with area residents and hence are expected to play an important role in spreading disease. We extend the synthetic population for the Washington DC Metro Area to include a transient population consisting of leisure and business travelers. This population is constructed by combining data from Destination DC[1], the Smithsonian Institution and other geospatial data.

We simulate disease outbreaks both with and without the transient population to evaluate the effect of the transients on the size of the outbreak and the time at which the epidemic peaks using EpiSimdemics (Barrett et al. 2008), a high performance computing software, developed at NDSSL.

Related Work

Traditional epidemiological study has relied on differential equation based models (Hethcote 2000; Bailey 1975). They are useful to estimate if an epidemic will occur, when will

---

[1]http://washington.org

it occur, or peak infections. But this approach has some limitations. It assumes that a population is divided into a few sub-populations and there is uniform mixing within sub-populations. In reality, mixing depends upon demographics like age, social and economic background, travel patterns, etc. Neither these models nor random graph models like the Erdős-Rényi ($G_{n,p}$) model (Erdős and Rényi 1959) or the preferential attachment model (Barabási and Albert 1999) capture the complexity of human interactions which serves as a basis for disease transmission.

More recently, large-scale agent-based disaggregated models have been studied. They combine multiple datasets to represent individuals along with interactions among them to build a social contact network over which the disease spreads. Ferguson et al. (Ferguson et al. 2005; 2006) evaluated the spread of the $H5N1$ flu virus over a social contact network for Thailand and the spread of the flu in Great Britain and the United States respectively. However their approach does not consider much spatial detail. Germann et al. (2006) and Longini et al. (2005) evaluate the spread of the flu for the United States and Southeast Asia networks respectively. Their approach consists of hierarchical groups (household, household cluster, neighborhood, and entire community) and people have close and casual contacts within these groups with varying contact rates. However interactions in the real world are quite different from these templates.

At NDSSL, detailed *synthetic* social contact networks for the United States have been generated, originally for transportation modeling (Barrett et al. 2001), where nodes represent individuals and edges represent daily contact between them. These networks are then used in agent-based simulations to study process like disease spread (Barrett et al. 2010; Bisset and Marathe 2009; Eubank et al. 2004) or social contagion (Apolloni et al. 2009). The process involves five steps (Barrett et al. 2001):

1. *Generating Synthetic Population:* This step disaggregates the Census data using iterative proportional fitting to create a set of synthetic individuals grouped into households (Beckman, Baggerly, and McKay 1996). The generated synthetic population returns the same marginal distributions as given in the Census data, and also preserves anonymity of individuals because the generated synthetic individuals do not correspond exactly to any real individuals.

2. *Locating Households:* This step locates each synthetic household into a housing location using data from the US Census and street data from Navteq.

3. *Assigning Activities:* In this step, each individual is assigned a set of activities to perform during a day. Various activity templates are created by using the National Household and Travel Survey (NHTS) and the National Center for Education Statistics (NCES). Each synthetic household is matched to a survey a household based on its demographics and individuals in synthetic household are assigned the corresponding activities.

4. *Locating Activities:* An appropriate location (essentially a building) is chosen for each activity for each individual using a gravity model and Dun & Bradstreet location data.

5. *Sublocation modeling and constructing social contact network:* Each location is subdivided into sublocations (similar to rooms within a building). A person is assumed to come in contact with all people present at the same sublocation at the same time, which thus induces a social contact network.

Very few studies have been done to understand the effect of transient populations on epidemics. Ferguson et al. (2006) model air travel for the United States and Great Britain but they assume that tourists stay at hotels and do not travel within the city and hence the only place they come in contact with residents is at hotels. However, in big cities like Washington DC or New York, there are some popular tourism locations that many tourists visit and are highly crowded, e.g. the National Air and Space Museum at Washington DC has more than 80000 visits every day. Tourists come in contact with many people, including residents and other tourists at these places. We think such subpopulations serve as a constant reservoir of infected people and susceptible people (as there is a constant flux of new transients).

Most of the previous studies focus on pharmaceutical (vaccine, antiviral etc.) and non-pharmaceutical (based on social distancing, e.g. school closure, work closure etc.) interventions (Ferguson et al. 2006; Germann et al. 2006; Longini et al. 2005; Barrett et al. 2010; Bisset and Marathe 2009; Eubank et al. 2004). Since some tourist destinations attract very large numbers of visitors each day, they present a natural target for interventions. We evaluate two kinds of interventions: closing major tourist destinations (the top four museums at the Smithsonian Institution, in terms of numbers of daily visits), and a "hand sanitizer" intervention, where we assume that promoting sanitary habits at these locations, such as the use of hand sanitizers, can reduce the spread of the disease at these locations.

Next we describe the data and methodology used for generating the two transient subpopulations, after which we proceed to experiments.

Data Available

We generated an augmented synthetic population for the Washington DC Metro Area, which combines a previously generated resident population (the "base population" consisting of 4.13 million people) with a transient population consisting of tourists and business travelers. Demographic data about transients were obtained from Destination DC. We also used data from the the Smithsonian Institution about daily numbers of visits to various Smithsonian museums. Finally, we used data from Dun & Bradstreet to identify places that tourists visit, based on Standard Industrial Classification (SIC) codes. The data sets used for generating the synthetic population are listed in table 1. Since details about generating the base synthetic population are not novel to the present work and are described elsewhere (Barrett et al. 2001), we only describe in detail the methodology for generating the synthetic transient population.

The methodology for generating the transient population broadly follows that for generating the base population. We

Used for	data source
Base US population	American Community Survey
	National Center for Education Stat.
	National Household Travel Survey
	Navteq
	Dun & Bradstreet
Transient population (additional)	Destination DC
	Smithsonian visit counts

Table 1: Datasets used for population generation.

first use demographic data to represent transient individuals and transient parties (groups). Each transient party is placed in a hotel which serves as their home for the period of the visit. Each transient individual is then assigned activities to perform during the day like staying in the hotel, visiting museums and other tourist destinations (or work activities, for business travelers), going to restaurants, and various night life activities. Each activity is represented by the type of activity, the time each activity begins and ends, and the location for the activity. A location is chosen for each activity based on the type of activity using Dun & Bradstreet data.

Tourist Population

Generating Synthetic Tourists

The goal here is to combine various demographic distributions and represent synthetic tourist parties and individuals with demographics drawn from these distributions. According to data from Destination DC, about 50000 visitors visit Washington DC every day, 55% of these are leisure travelers and the rest are business travelers. For adult, overnight leisure travelers, they also provide distributions of age, household income, party size, marital status and if the household has children. Here, age and marital status are individual level demographics while household income and party size are household (or party) level demographics, and hence they need to be treated differently. These distributions are not independent of each other within a party i.e, a married couple is more likely to travel together and hence we cannot sample independently from the given distributions.

High Level View   Our approach is simple: we assume a small set of rules about party structure and then do sampling without replacement from the given demographic distributions (since we know the total number of individuals to be generated) in combination with these rules to generate the tourist population.

We start by generating first party member (called householder) by sampling age, marital status, income and party size independently from the corresponding distributions and then generate other party members in relation to the householder. For example, if a party member is married then with a certain probability his/her spouse will also be part of the party and the age difference between them is assumed to be within a certain range (± 5 years in this case). All party members should have same household/party level demographics, household income and party size. Whenever an

individual is assigned a demographic, the probabilities for selecting various categories of that demographic is adjusted to model sampling without replacement.

We also assign all individuals some other demographic variables as assigned in the synthetic base population, e.g., sex (at random and in accordance with marital status) and employee status record (*esr*) and occupation code (*socp*) (by finding an individual with the closest income from the synthetic population of Washington DC metro area and assigning corresponding *esr* and *socp* codes).

For the present study most of these demographic details are irrelevant because disease parameters are not chosen to vary with demographic. However, that could be done in future work, and the synthetic populations can also be used for other studies where the demographic details are important.

Assigning hotels

We identified hotels and lodging locations within I-495 loop area in Washington DC from Dun & Bradstreet (*D&B*) data. *D&B* is a commercial data set that gives information about business locations like longitude-latitude of buildings, number of employees (relative numbers), type of business etc.

Each tourist party is assumed to stay at a hotel, which serves as their home location for the duration of the visit. Taking into consideration that tourists prefer to stay near downtown and each hotel has a capacity proportional to number of employees there, a hotel location (i) is chosen from the available pool with probability proportional to $num\_employees(i) \times e^{\delta \times distance\_from\_white\_house}$.

Assigning Activities

Since we could not find any data about activity sequences for tourists, we assumed a template for it, as illustrated in figure 1. We assume that all individuals in a party travel together and hence have the same activity sequence and go to the same locations. However within a location (building) they may go to different sublocations (rooms). Each party's activity sequence contains information about the type of activity and the start time and duration. Location and sublocation are decided later.

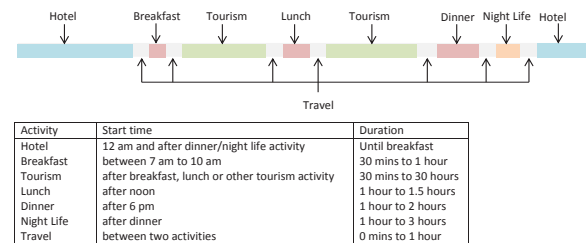| Activity	Start time	Duration
Hotel	12 am and after dinner/night life activity	Until breakfast
Breakfast	between 7 am to 10 am	30 mins to 1 hour
Tourism	after breakfast, lunch or other tourism activity	30 mins to 30 hours
Lunch	after noon	1 hour to 1.5 hours
Dinner	after 6 pm	1 hour to 2 hours
Night Life	after dinner	1 hour to 3 hours
Travel	between two activities	0 mins to 1 hour

Figure 1: Activity template for tourists.

Each party starts the day with a hotel activity. It is followed by breakfast which could happen at the same hotel (with 60% probability) or at some other location (with 40% probability). Each tourism activity shown in figure 1 is divided into one or more tourism activities with some travel

time between them. Each party goes for lunch after $12:00$ pm which is again followed by one or more tourism activities and then dinner. After dinner, with 50% probability they go back to the hotel directly and stay for the rest of the day. Otherwise they go for a night life activity and then back to their hotel. Each pair of activities is separated by travel time of 0 mins to one hour.

Locating Activities

We identified locations for tourism, eating, and night life activities from *D&B* data. Tourism activity locations include places like museums, art galleries, planetarium, historical societies, and botanical and zoological gardens. Eating activity locations include various restaurants and night life activity locations include bars and pubs, night clubs, and movie theatres.

Assuming that most of the transients to Washington DC visit museums which are around the National Mall and plan their trip around that area, we choose locations for all activities based on the distance from the National Mall. Each location has a capacity (again assumed to be proportional to the number of employees at that location according to *D&B*). Considering both of these factors, a location (i) is chosen from the available locations for a given activity type with probability proportional to $num\_employees(i) \times e^{\delta \times distance\_from\_national\_mall}$.

The Smithsonian Institution provides data about daily visit counts at various museums. To match the number of visits in our synthetic population at museum locations with these counts, we adjusted weights (number of employees) from *D&B* data for these museums. However, the number of transients is not sufficient to account for all the visits to some museum locations. For example, the National Air and Space Museum has about 80000 visits per day. So we adjusted the activities of some individuals in the base population and routed them to these locations to match the visit counts exactly. This also creates mixing between the transient and the base population, which is an important factor in the spread of disease.

Sublocation Modeling

An activity location typically corresponds to a building and sublocations correspond to rooms in the building. Sublocation modeling involves deciding which room a person visits and hence with whom he comes into contact.

All individuals in a party are assumed to meet each other at the hotel and hence are assigned same sublocation.

For other locations, we follow the assumption made in the creation of the base population (Barrett et al. 2001), that sublocations have a capacity of 25 people, and that each person, upon arriving at a location is assigned to a sublocation where he remains for the duration of his activity at that location.

In reality, people would come into contact with more than 25 people at major tourist venues like the National Air and Space Museum and the National Museum of Natural History. Also, inside museums, they do not stay at the same location during the entire period of their visit. They keep

moving from one exhibition to another. We therefore create a simple stochastic process modeling movement between sublocations at for the four biggest tourism locations - the National Air and Space Museum (NASM), the National Museum of Natural History (NMNH), the National Museum of American History (NMAH) and the National Art Gallery (NAG). For these four locations, we decided the number of sublocations by looking at their floor plans. While modeling visits to these locations, a person's visit is divided into the interval of 5 to 15 minutes and a person keeps moving to different sublocations (chosen at random) within the location.

Business Travelers

The process used is similar to the synthetic tourist population generation process.

Generating Synthetic Business Travelers

We could not find any demographic data for business travelers. The only information available is that about 45% of the transients are business travelers. We followed the same procedure as for generating tourists but with some assumptions. Each business traveler is assumed to be by himself and hence party size is assumed to be 1. Age is assumed to be between 18 to 70 years. Marital status is chosen from unmarried, married, and divorced/widowed with equal probability. The household income distribution is assumed to be Gaussian with peak and standard deviation equal to the maximum and average household income in Washington DC metro area respectively.

We assigned other demographic variables i.e., sex, *socp*, *esr* following the same process as for tourist population.

Assigning Hotels

Business travelers are assigned hotels the exact same way as tourists.

Assigning Activities

Here also, since we could not find any data about the activity sequences for business travelers, we assumed a template for it. The activity sequence created contains information about the type of activity and the start time and duration. Location and sublocation choice are described in the next subsections. The template for activities is as shown in figure 2.

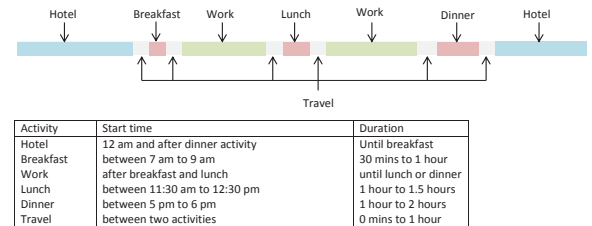| Activity	Start time	Duration
Hotel	12 am and after dinner activity	Until breakfast
Breakfast	between 7 am to 9 am	30 mins to 1 hour
Work	after breakfast and lunch	until lunch or dinner
Lunch	between 11:30 am to 12:30 pm	1 hour to 1.5 hours
Dinner	between 5 pm to 6 pm	1 hour to 2 hours
Travel	between two activities	0 mins to 1 hour

Figure 2: Activity template for business travelers.

Each business traveler starts the day at a hotel and hence with hotel activity. It is followed by breakfast which could

happen at the same hotel (with 60% probability) or at some other location. After breakfast, he leaves for work and stays there until lunch. After lunch, he goes back to work and stays there until dinner. After dinner, he goes back to the hotel and stays for the rest of the day. Here also, each pair of activities is separated by a travel time of 0 mins to one hour.

Locating Activities

The process used to assign activity locations is quite similar to that of tourists. Here we identified locations for work from the D&B data (these are also used as work locations for the base synthetic population) and eating (same as for tourists). Activity locations are assigned the same way as for tourists.

Sublocation Modeling

For sublocation modeling, we follow the same process as used for the base synthetic population.
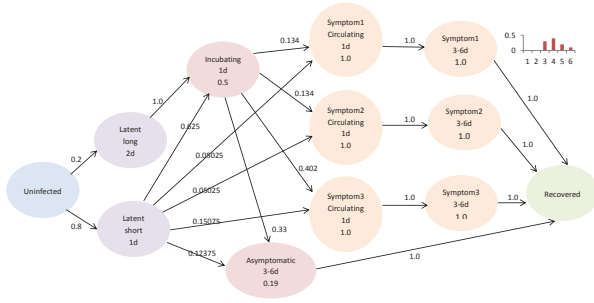
Experiments

Figure 3: Disease model.

We simulate a flu-like disease for Washington DC metro area using EpiSimdemics, an interaction based high performance computing simulation software for studying large scale epidemics (Bisset et al. 2009). To evaluate the effect of the transient population on disease spread, we simulate the disease spread for 120 days both with and without the transient population. The disease model, represented as a Finite State Machine (FSM), is as shown in Figure 3. Each node represents a state and transition probabilities are as shown on the edges. Each node label consists of the state name, number of days for which a person remains in this state, and the probability of him infecting others. The histogram in the upper right corner shows the probability of being in the given state versus the number of days for symptom1, symptom2, symptom3, and asymptomatic states.

We derive the epidemic curves for multiple cases, like considering only base population, base plus transient population, closing museums for a certain amount of time during the epidemic, and using more hand sanitizers in the four biggest museums. Each case is discussed in detail in following paragraphs. We run 50 simulations for each case. Each simulation takes one hour to 2.5 hours of computation. The total computational effort is on the order of 600 compute hours. Initial infections are the same for all cases. Only residents are initially infected so transients do not bring disease

to the city, they can get infected during their stay in the area and infect others and leave at the end of their trips (5 days in this case, which is the average trip length according to the data from the Destination DC). They are replaced by new incoming uninfected transients with exactly the same demographics.
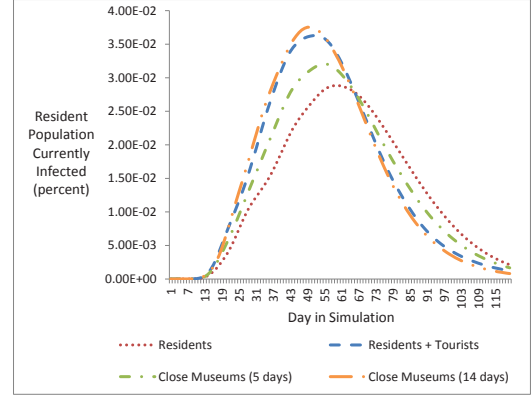

Figure 4: The fraction of residents currently infected vs. day in simulation (average over 50 iterations).
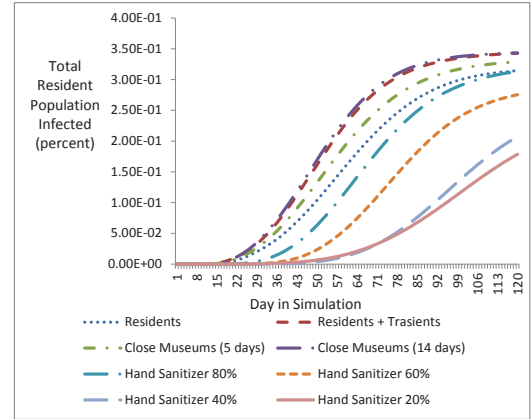

Figure 5: The fraction of residents infected cumulatively vs. day in simulation (average over 50 iterations).

As we are interested in the effect of transients on the number of residents being infected, figure 4 shows the average fraction of residents currently infected versus day in simulation and figure 5 shows the average of fraction of residents infected over 120 days versus day in simulation. The disease peaks about 7 days earlier and there are about 0.745% more resident infections at peak when the transients are considered. Over the period if 120 days, 2.85% more residents are infected. This is as expected as major tourist locations like the National Air and Space Museum (NASM), the National Museum of Natural History (NMNH), the National Museum of American History (NMAH), and the National Art Gallery (NAG) have about 40000 to 80000 visits (including visits from residents and transients) per day. Table 2

Location	Residents only	Residents + Transients	Museums closed for 5 days	Museums closed for 14 days	Hand sanitizer prob 80%	Hand sanitizer prob 60%	Hand sanitizer prob 40%	Hand sanitizer prob 20%
NASM	0.033413	0.067039	0.067036	0.066862	0.050779	0.038752	0.024469	0.002433
NMNH	0.028815	0.056805	0.056479	0.056117	0.042688	0.031852	0.015943	0.001745
NMAH	0.017859	0.031314	0.031088	0.031211	0.024012	0.016521	0.006404	0.000762
NAG	0.003242	0.007305	0.007267	0.007239	0.00403	0.001656	0.000459	0.000091

Table 2: The fraction of infections (residents + transients) at four major tourist locations (average over 50 iterations).

group	sample size	1	2	3	4
Hand sanitizer 80%	50	0.021482			
Hand sanitizer 60%	50	0.025398			
Hand sanitizer 40%	50		0.032125		
Resident only	50		0.034242	0.034242	
Hand sanitizer 20%	50			0.037873	0.037873
Museums closed (14 days)	50				0.041905
Museums closed (5 days)	50				0.042007
Residents + Transients	50				0.04202
Significance		0.093	0.794	0.154	0.060

Table 3: Tukey's HSD test to compare the fraction of residents currently infected at peak, for alpha = 0.05 and sample size = 50.

group	sample size	1	2	3	4
Museums closed (5 days)	50	50.46			
Museums closed (14 days)	50	51.58			
Residents + Transients	50	53.22	53.22		
Resident only	50	62.32	62.32	62.32	
Hand sanitizer 80%	50		66.46	66.46	
Hand sanitizer 60%	50			72.00	
Hand sanitizer 20%	50				88.48
Hand sanitizer 40%	50				88.52
Significance		0.135	0.061	0.367	1.0

Table 4: Tukey's HSD test to compare the day at which current resident infections peak, for alpha = 0.05 and sample size = 50.

group	sample size	1	2	3
Hand sanitizer 80%	50	0.178632		
Hand sanitizer 60%	50	0.207309		
Hand sanitizer 40%	50		0.275303	
Hand sanitizer 20%	50		0.313280	0.313280
Resident only	50		0.314311	0.314311
Residents + Transients	50			0.342675
Museums closed (14 days)	50			0.343243
Museums closed (5 days)	50			0.343936
Significance		0.350	0.057	0.265

Table 5: Tukey's HSD test to compare the fraction of residents currently infected over 120 days, for alpha = 0.05 and sample size = 50.

Figure 6: The fraction of residents currently infected vs. day in simulation (average over 50 iterations).
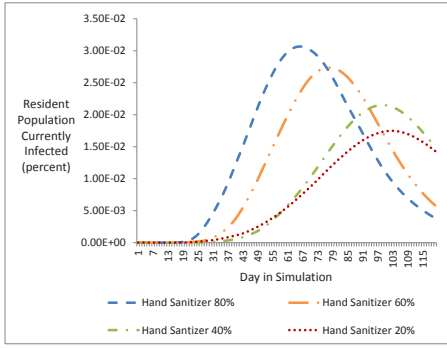

Figure 7: The fraction of residents currently infected vs. day in simulation (average over 50 iterations).

shows the average fraction of infections (residents+tourists) that happen over the period of 120 days at these four locations for both the cases. The number of infections increases by about two times when transients are considered. Also as the transients stay for a short period of time and at the end of their trips, new, uninfected, but susceptible transients replace them, there is almost constant number of susceptible and infected people at these locations, making them prominent sites for infection. Tables 3 and 4 compare the fraction of the resident population infected at peak and the day at which it peaks respectively, for different cases using Tukey's HSD test with significance level of 0.05. For the fraction of resident population infected at peak, the case when only residents are considered falls into a different group from the case when residents and transients both are considered. Hence, the fraction of resident population infected at peak is statistically significantly higher when transients are considered. Although on average the peak is about 9 days earlier, both cases fall into the same group and hence it is not statistically significant. Table 5 compares the fraction of the resident population being infected over the period of 120 days. They are also not significantly different from each other.

One way to reduce the infections is to apply social distancing measures like closing schools, work places etc., which reduces mixing and hence infections. Here, as we are interested in reducing the impact of transients, we apply interventions where we close the four big museums for a few days when the number of infections reaches a threshold. We assume that when museums are closed tourists go to other tourist locations and residents visiting these museums continue their other activities. We simulate two cases:

1. When the current number of infections (residents+transients) reaches 50000, we close these museums next day for 5 days. As shown Figure 4, the disease peak is delayed by 3 days and there are about 0.42% less infections at peak when museums are closed. Over the period if 120 days, about 1.42% less residents are infected. However they are not statistically different. The fraction of infections (as in table 2) that happen at the four big museums are slightly less.

2. When the current number of resident infections reaches 50000, we close these museums next day for 14 days.
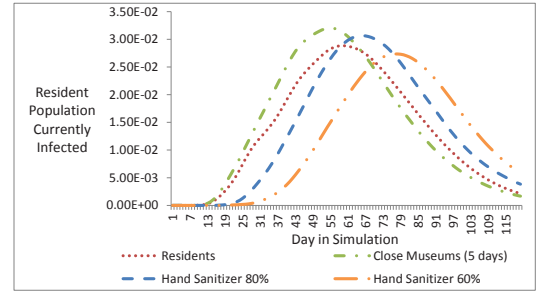
Though not statistically significant, the fraction of resident population infected at peak and the fraction of resident population being infected over the period of 120 days are surprisingly slightly higher than the case when museums are not closed. The number of infections also peak a couple of days earlier. This could be because in this case, we close museums the day after the number of current resident infections as opposed to the number of current infections (residents + transients) in the first case, reach 50000. This causes one day (on average) of delay in closing museums which may be too late to close as disease has already spread a lot. Another reason could be that when museums are closed, transients go to some other places and probably distribute to a larger area and hence spread disease more.

We also evaluate a scenario where people are encouraged to use hand sanitizers at the four big museums. As data about how much infectivity and susceptibility are reduced by application of hand sanitizers is unavailable, we did a series of experiments assuming that the use of hand sanitizers reduces infectivity and susceptibility to 20%, 40%, 60%, and 80% of the original values (effective only inside the four museums). We assume that 50% of the people going to these places use hand sanitizers. Results are as below:

1. If the use of hand sanitizer reduces the infectivity and the susceptibility to 80% of the original value, then as shown in figure 7 the peak is slightly less (though not significantly) than the case when museums are closed for 5 days and the peak is also delayed by 11 days which is statistically significant (Table 4). The total number of infections over 120 days are also significantly less.

2. If the use of hand sanitizer reduces the infectivity and the susceptibility to 60% of the original value, then the peak and the total number of infections over 120 days are slightly less (not significantly) than the case when transient population is not considered. The peak is also delayed by 18 days which is statistically significant.

3. If the use of hand sanitizer reduces the infectivity and the susceptibility to 40% of the original value, then the peak reduces by 0.73% from the base case. The fraction of infections over 120 days are reduced by 10%. The peak is also delayed by about 40 days. The results are significantly better than all the cases discussed so far.

4. If the use of hand sanitizer reduces the infectivity and the susceptibility to 20% of the original value, then the peak reduces by 1.14% from the base case. The fraction of infections over 120 days are reduced by 13.58%. The peak is also delayed by about 50 days. This is significantly better than all the other cases.

Also as expected, the use of hand sanitizer reduces the number of infections at the four big museums (Table 2).

Conclusion

In summary, including the transient population makes a significant difference in epidemic estimation. However, the most commonly recommended non-pharmaceutical intervention—social distancing—surprising does not show a statistically significant effect at reducing the outbreak. In this case, it seems that locations where a lot of mixing occurs are better thought of as presenting opportunities for reducing disease spread by promoting sanitary behavior such as the use of hand sanitizers. When we apply this intervention, depending on the efficacy of the behavior, we can make a significant difference to the peak and the cumulative number of infections, as well as the day of the peak.

There is much room for further work. We speculate that transients have an effect on the epidemic because they are a source of susceptible people that is constantly replenished. This can be examined analytically using a differential equation based model as well. The intervention policies that we have studied can be examined more systematically to determine optimal strategies. For example, it may turn out that there is a museum closure strategy that does produce a statistically significant difference.

Models like these can also be used for policy recommendations, for example to promote the use of hand sanitizers in museums. That in turn offers the opportunity to conduct field experiment to validate our model against actual epidemic and intervention data.

Acknowledgments

References

Apolloni, A.; Channakeshava, K.; Durbeck, L.; Khan, M.; Kuhlman, C.; Lewis, B.; and Swarup, S. 2009. A study of information diffusion over a realistic social network model. In *The IEEE International Symposium on Social Computing Applications*.

Bailey, N. 1975. *The Mathematical Theory of Infectious Diseases and its Applications*. London: Griffin.

Barabási, A. L., and Albert, R. 1999. Emergence of scaling in random networks. *Science* 286:509–512.

Barrett, C.; Beckman, R.; Berkbigler, K.; Bisset, K.; Bush, B.; Campbell, K.; Eubank, S.; Henson, K.; Hurford, J.; Kubicek, D.; Marathe, M.; Romero, P.; Smith, J.; Smith, L.; Speckman, P.; Stretz, P.; Thayer, G.; Eeckhout, E.; and Williams, M. D. 2001. TRANSIMS: Transportation analysis and simulation system. Technical Report LA-UR-00-1725, Los Alamos National Laboratory.

Barrett, C. L.; Bisset, K. R.; Eubank, S. G.; Feng, X.; and Marathe, M. V. 2008. EpiSimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks. In *Proc. SuperComputing Conference*.

Barrett, C. L.; Bisset, K. R.; Leidig, J. P.; Marathe, A.; and Marathe, M. V. 2010. An integrated modeling environment to study the coevolution of networks, individual behavior and epidemics. *AI Magazine* 31(1):75–87.

Beckman, R. J.; Baggerly, K. A.; and McKay, M. D. 1996. Creating synthetic baseline populations. *Transportation Research Part A: Policy and Practice* 30(6):415–429.

Bisset, K. R., and Marathe, M. V. 2009. A cyber environment to support pandemic planning and response. *DOE SCIDAC Magazine* 36–47.

Bisset, K. R.; Feng, X.; Marathe, M.; and Yardi, S. 2009. Modeling interaction between individuals, social networks and public policy to support public health epidemiology. In *Proc. Winter Simulation Conference*, 2020–2031.

Erdős, P., and Rényi, A. 1959. On random graphs. I. *Publ. Math. Debrecen* 6:290–297.

Eubank, S. G.; Guclu, H.; Kumar, V. S. A.; Marathe, M. V.; Srinivasan, A.; Toroczkai, Z.; and Wang, N. 2004. Modelling disease outbreaks in realistic urban social networks. *Nature* 429(6988):180–184.

Ferguson, N. M.; Cummings, D. A. T.; Cauchemez, S.; Fraser, C.; Riley, S.; Meeyai, A.; Iamsirithaworn, S.; and Burke, D. S. 2005. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature* 437(8):209–214.

Ferguson, N. M.; Cummings, D. A. T.; Fraser, C.; Cajka, J. C.; Cooley, P. C.; and Burke, D. S. 2006. Strategies for mitigating an influenza pandemic. *Nature* 442(7101):448–452.

Germann, T. C.; Kadau, K.; Longini, I. M.; and Macken, C. A. 2006. Mitigation strategies for pandemic influenza in the United States. *PNAS* 103(15):5935–5940.

Hethcote, H. W. 2000. The Mathematics of Infectious Diseases. *SIAM Review* 42(4):599–653.

Longini, I. M.; Nizam, A.; Xu, S.; Ungchusak, K.; Hanshaoworakul, W.; Cummings, D. A.; and Halloran, M. E. 2005. Containing pandemic influenza at the source. *Science* 309(5737):1083–1087.

Molinari, N.-A. M.; Ortega-Sanchez, I. R.; Messonnier, M. L.; Thompson, W. W.; Wortley, P. M.; Weintraub, E.; and Bridges, C. B. 2007. The annual impact of seasonal influenza in the US: Measuring disease burden and costs. *Vaccine* 25(27):5086 – 5096.