SPATIOTEMPORAL MODELING OF DISEASE SPREAD THROUGH MICROMOBILITY SYSTEMS

by

Behnam Nikparvar

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Infrastructure and Environmental Systems

Charlotte

2021

Approved by:

Dr. Jean-Claude Thill

Dr. Eric Delmelle

Dr. Wenwu Tang

Dr. Dimitris Papanikolaou

Dr. Tiffany Gallicano

ABSTRACT

BEHNAM NIKPARVAR. SPATIOTEMPORAL MODELING OF DISEASE SPREAD THROUGH MICROMOBILITY SYSTEMS. (Under the direction of DR. JEAN-CLAUDE THILL)

New modes of public transportation such as micromobility are rapidly growing in urban areas. Bike sharing and e-scooter sharing, for example, have been advanced to solve the first/last mile problem, providing quick access to bus stops and train stations for their users. This efficiency, however, may come at the cost of transmitting disease since the surfaces on the bicycles or scooters are subject to germs and harmful pathogens when they are left in contaminated places or used by infectious individuals. This dissertation aims to understand various facets of the role of micromobility transportation in the spread of viral disease within dense urban areas. I propose a novel micro-level and spatially-explicit agent-based modeling framework to model the spread of viral infectious diseases through micromobility systems and a baseline population. I use this simulation framework to study the role of micromobility in the spread of viral disease in urban areas by breaking down the problem into three directions. First, I want to study how surfaces on the new micromobility transportation systems contribute to the emergence and dynamics of viral epidemics in urban areas. Second, I seek to find out how geographic space and time are organized concerning the risk of exposure to a viral disease out of using micromobility vehicles. Third, to inform decision-making in response to the spread of viral disease through micromobility systems, I examine what intervention methods and strategies, including random or systematic intervention, are more effective in controlling the spread of infectious diseases through micromobility vehicles. In order to test the proposed model, a case study is conducted in Cook County, Illinois, and uses the Chicago City public bikesharing system. Results show that the emergence of viral disease through micromobility transportation in Cook County is possible, but the overall impact of

the system on the disease dynamics in a worst-case scenario, especially with the current size of the system, is rather small. The proposed model, however, provides a better measure to evaluate the role of transportation in spread of disease compared to existing measures. The spatial pattern for the risk of exposure is higher in the central business district and in northern regions, where most of the shared bike transportation occurs. Moreover, the start day of exposure impacts the dynamics of the spread of disease through both micromobility and the baseline population. Finally, intervention success in a full-blown epidemic highly depends on human behavior, availability of disinfection equipment, and strategies to implement control methods. The proposed simulation framework can be used to assess the efficacy of interventions and make trade-offs between these factors when dealing with epidemics of the sort analyzed in this research.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my advisor Dr. Jean-Claude Thill for his invaluable academic and career advice, continuous support and mentorship, and patience during my Ph.D. study. Finishing this journey was not possible without his comments and encouragements that have been a source of inspiration for me. I am deeply grateful for this once-in-a-life opportunity of doing my Ph.D. research under his supervision.

Second, I would like to sincerely thank the committee members of my dissertation: Dr. Eric Delmelle, Dr. Wenwu Tang, Dr. Dimitris Papanikolaou, and Dr. Tiffany Gallicano, for their invaluable support, comments, and inputs to my Ph.D. research. Additionally, I would like to express my gratitude to other faculty members of the Geography and Earth Science Department and the Director and Associate Director of the Infrastructure and Environmental Systems Ph.D. program, Dr. Jay Wu and Dr. John Dimmer.

Third, I am thankful for the financial support from the Graduate Assistant Support Plan (GASP) at UNCC, and Graduate Research fundings from Minerva Research Initiative and Army Research Office, World Bank, Renaissance Computing Institute (RENCI), and Project Mosaic at UNCC.

Fourth, I would like to offer my special thanks to my colleagues and lifetime friends, including Dr. Daniel Yonto, Dr. Daidai Shen, Dr. Paul Jung, Faizeh Hatami, Md. Mokhlesur Rahman, Zurikanen Iddrisu, and many others for their friendship and support.

Finally, I must thank my real treasure in the world and my lifelong source of inspiration, my family, that finishing this journey was impossible without their encouragement: my dearest parents and sisters.

TABLE OF CONTENTS

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER 1: INTRODUCTION 1

1.1. Statement of Research 2

1.2. Structure of The Dissertation 6

CHAPTER 2: LITERATURE REVIEW 7

2.1. Transportation Systems and Epidemics 7

2.2. Micromobility Systems As a Novel Mode of Transportation 9

2.3. Transmission Modes 13

2.4. Epidemic Modeling 16

2.4.1. Homogeneous Mixing Models 16

2.4.2. Metapopulation Models 18

2.4.3. Contact Network Models 19

2.4.4. Agent-Based Models 24

2.5. Interventions and Control 28

CHAPTER 3: RESEARCH QUESTIONS 30

CHAPTER 4: RESEARCH DESIGN 34

4.1. Purpose of The Model 34

4.2. Agents, State Variables, And Scales 34

4.3. Process Overview And Scheduling 37

4.3.1. Relocation 37

4.3.2. Micromobility Ride 38

4.3.3. Transmission 39

4.3.4. Inactivation 43

4.3.5. Recovery 43

4.3.6. Disinfection 43

4.4. Design Concepts 45

4.5. Model Details 47

4.5.1. Human-Human (HH) Submodel 47

4.5.2. Micromobility Fomite-Human (MH) Submodel 50

4.5.3. Human-Micromobility Fomite (HM) Submodel 51

CHAPTER 5: A CASE STUDY: SPATIOTEMPORAL MODELING OF DISEASE SPREAD IN COOK COUNTY, ILLINOIS 52

5.1. Study Area 52

5.2. Modeling Framework 53

5.3. Data 54

5.3.1. Micromobility 54

5.3.2. Workplaces 55

5.3.3. Schools 56

5.4. Synthetic Population Module 56

5.5. Mobility Module 58

5.6. Disease Transmission Module 58

5.7. Programming And Software 59

5.8. Results 61

5.8.1. Micromobility Trips 61

5.8.2. Micromobility Stations 62

5.8.3. Synthetic Population 64

5.8.4. Sensitivity Analysis and Parametrization 68

5.8.5. Micromobility Impact on Viral Disease Emergence And Dynamics 69

5.8.6. Risk of Exposure to Viral Disease in Micromobility 76

5.8.7. Interventions and Control 83

CHAPTER 6: CONCLUSIONS 87

REFERENCES 93

APPENDIX A: GENERATING SYNTHETIC POPULATION WITH SOCIAL TIES 103

LIST OF TABLES

TABLE 4.1: Overview of processes and parameters 44

TABLE 5.1: Accuracy assessment of the synthetic population (based on percentage errors). 66

TABLE 5.2: Comparison of population infected in the presence and absence of micromobility. 73

LIST OF FIGURES

FIGURE 1.1: Micromobility and spread of viral disease in urban areas. 5

FIGURE 2.1: Homogeneous mixing compartment models. 16

FIGURE 4.1: Class diagram of agents in the model. 36

FIGURE 4.2: Synthetic baseline population. 37

FIGURE 4.3: Process overview (micromobility). 39

FIGURE 4.4: SIR-SC model. 41

FIGURE 5.1: Study area. 53

FIGURE 5.2: Modeling framework for Cook County. 54

FIGURE 5.3: Pseudo-code for the sequential version of the SIR-SC model. 60

FIGURE 5.4: Divvy shared bike system statistics during July 2018. 63

FIGURE 5.5: Origin-destination employment statistics. 64

FIGURE 5.6: Accuracy assessment of the synthetic population. 65

FIGURE 5.7: Sample census tract contact network. 67

FIGURE 5.8: Virus inactivation time impact on disease spread in Divvy system. 69

FIGURE 5.9: Emergence of outbreaks through micromobility. 71

FIGURE 5.10: Micromobility impact on disease dynamics 75

FIGURE 5.11: Spatial and temporal patterns of exposure risk to contaminated bicycles 80

FIGURE 5.12: Start day of exposure impacts the dynamics of the spread of disease 82

FIGURE 5.13: Random disinfection of bikes 83

FIGURE 5.14: Placing sanitizing stand at stations 86

FIGURE A.0: Sample census tracts. 106

CHAPTER 1: INTRODUCTION

New modes of transportation such as on-demand mobility services that usually involve shared use of vehicles instead of personal vehicles are becoming more pervasive due to creating more comfort and efficiency in the mobility of the population within urban areas. However, this efficiency may come at the cost of transmitting harmful pathogens and disease as people share space and surfaces in vehicles and infrastructure. For example, micromobility systems, such as bike and e-scooter sharing, have been advanced to solve the first/last mile problem, providing quick access to bus stops and train stations for its users. However, the surfaces on public bicycles or scooters are still subject to germs and harmful pathogens when they are left in contaminated places or used by infected individuals. In order to explore the role of these systems in the spread of infectious disease, a new way of studying the impact of transportation's sharing economy on the transmission of contagious diseases is critically needed.

To control infectious disease, we need to understand how the responsible pathogens and microorganisms spread throughout the environment [1] and how they may compromise the safety of transportation modes. The study of the spatiotemporal infectious disease patterns is usually limited to close contact and direct transmission between the hosts or transfer through common vehicles such as water, food, and air. Less attention has been placed on the role of inanimate surfaces –scientifically called fomites–, mainly because it is challenging to keep track of agents passing through surfaces. Also, it used to be assumed that the fomites are not a significant transmission route, but recent studies show that the role of these surfaces is not negligible in the spread of many types of infectious disease [2, 3]. While frequently touched surfaces are everywhere in transportation systems, they are the main points of contact

between susceptible and infected individuals in micromobility systems. The process of transmission, in this case, highly depends on the disease agent and individual behavior. Thus, to be effective, a modeling framework should account for these types of heterogeneities.

In addition, little attention has been paid to the spatial aspects of epidemics of this type in large and dynamic human contact networks, which is prevalent in urban areas [4]. While recent contact network models in mathematical epidemiology account for the structure of contacts among individuals, they hardly capture the spatial aspects of infectious disease transmission since the network perspective, in most cases, concerns the epidemic outbreak in a single population [5]. Thus, it is central to study these gaps by accounting for both spatial and temporal aspects [6].

Hence, this dissertation proposes to address the current shortcomings mentioned above with a new and original micro-level agent-based modeling framework for disease spread through micromobility in a dense urban environment. It involves an empirical simulation study of the Chicago public bikesharing system on how the velocity and spatial diffusion of infectious diseases are related to shared mobility vehicle surfaces.

1.1 Statement of Research

This study aims to model the spatiotemporal spread of a viral disease through surfaces on public micromobility vehicles within urban areas. In order to attain this goal, we need to look at the intersection between transportation and disease systems.

The Spanish flu of 1918-19 is the most widespread pandemic in modern history. It is estimated to have infected 30% of the world population and to have killed 50 to 100 million. The infection and morbidity rate of descendant pandemics, such as flu 1968, Middle East respiratory syndrome (MERS), and Severe Acute Respiratory Syndrome (SARS-CoV), have decreased thanks to progress in the prevention, diagnosis, surveillance, control, and treatment of infectious diseases[1]. Despite all this progress,

[1]Centers for Disease Control and Prevention (CDC):https://www.cdc.gov/flu/pandemic-

we are still vulnerable to the emergence and resurgence of infectious disease [7]. The emergence of the new Coronavirus (SARS-CoV2) in 2019-2021, which infected 251 million people and killed more than five million in almost all countries[2], is the most recent evidence that infectious diseases are still a serious threat for the foreseeable future.

The emergence, resurgence, and diffusion of infectious disease is a molecular or microbiological phenomenon and a matter of social, ecological, and geographical changes. This means that we need to determine disease causation and spread in social, geographical, and ecological contexts [8]. Stress factors of the modern era, such as urbanization, mobility, and transportation technology, as well as land-use and climate change, may contribute to the emergence and rapid diffusion of these communicable diseases [9].

Urban areas are at the forefront of disease outbreaks and spread initially because of the high population density. In the past hundred years, urbanization has resulted in the concentration of population in both developed and less developed regions in the world, changing from 30% back in 1950 to 56% in 2020, and projected to be 68% by 2050 [10]. Within urban areas, the population is heterogeneous in terms of distribution over space, contact network, and socioeconomic attributes, all of which can significantly affect the dynamics of epidemics.

The spatial distribution of population is a critical determinant from an epidemiological point of view. A single disease agent may cause different diffusion dynamics in two different cities only because the population contact network is different in those environments. This is the intuition behind network-based epidemiology, one of the most widely popular approaches in the past two decades [11]. However, very limited studies exist on the heterogeneity in behavior (e.g. mobility) using network

resources/reconstruction-1918-virus.html

[2]COVID-19 Map - Johns Hopkins Coronavirus Resource Center: https://coronavirus.jhu.edu/map.html

epidemiology [12].

Transportation systems can contribute to epidemics initially by introducing it to new regions. That means a larger population may be infected at the end of the epidemic [13]. The spread of many infectious diseases follows the pattern of international passenger flows, such as flights between gateway cities around the world [14]. It can be explained by the fact that the incubation period of many infectious diseases is longer than even long-distance travel owing to the increasing efficiency of transportation systems and to the resulting time-space convergence [15]. Thus, before symptoms of an infected person may be perceptible to health officials, the virus may already have spread to populations in other geographic regions [16]. For instance, a recent study of the spread of the early COVID-19 outbreak in 2019-2020 estimated that about 64% of exported cases from China to other countries were in the presymptomatic incubation period upon arrival [14]. At the same time, very limited empirical work exists that reveals the extent that transportation systems contribute to the spread of disease [17].

Transportation systems also have the potential to accelerate the diffusion of infectious diseases. For example, compared to the 1918 pandemic, the faster speed of spread of the influenza pandemic that originated in mainland China in 1957 is partially attributed to the availability of air and sea travel [18]. A systematic review of the role of transportation in the propagation of influenza and coronaviruses shows that global air transport plays an essential role in accelerating influenza. However, the degree to which transportation systems can accelerate epidemics of this type is not always easy to determine especially for ground-based transportation [17].

Public policies tend to lean toward efficient and accessible public transportation systems and infrastructures within urban areas. In this regard, micromobility, a relatively new transportation mode known for short-range trips, shares many of the fundamental characteristics of public transportation goods and is becoming more

pervasive in many cities worldwide. However, shared bicycle services are subject to pathogens and can become a vector to spread disease in the population (figure 1.1). The handlebars of public bikesharing systems, especially, are likely to be contaminated with disease-causing agents because they are in contact with the hands of different people with different standards of hygiene. Simultaneously, while bicycling is regaining popularity in many regions, including Europe, North America, and Asia, in some countries like China, the number of bikes is more than the carrying capacity of many cities, which raises serious concerns on safety and health. In order to fill these gaps, this study aims to model the spatiotemporal dynamics of contagious disease through micromobility systems within urban areas.

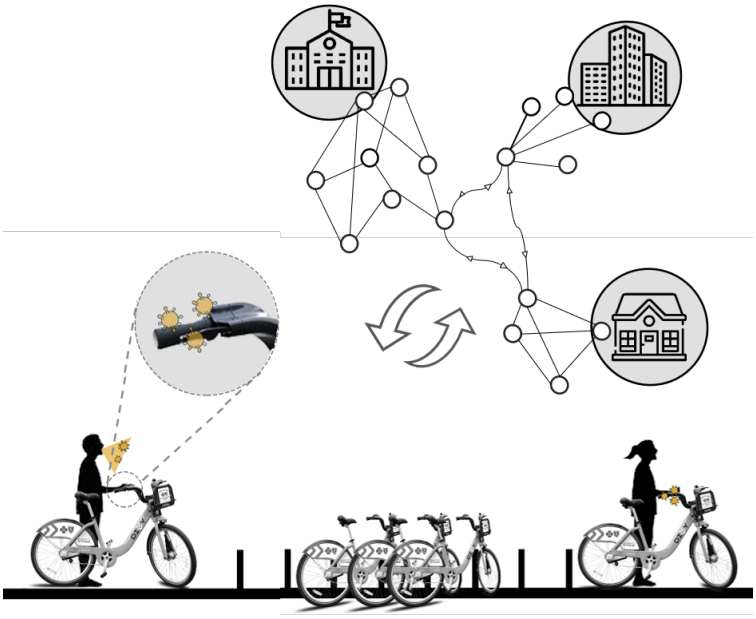


Figure 1.1: Micromobility and spread of viral disease in urban areas.

Intervention and control strategies related to micromobility and public transportation systems are also essential. When an epidemic occurs, partial or complete suspension of public transportation services is one of the early actions that local officials may take to stop or delay the spread of pathogens and diseases. However, other alternatives, such as disinfection and sterilizing environments in transportation systems, may be more effective since the service provided by public transportation remains

available[3].

Several studies have evaluated the use of control and intervention methods in transportation systems [19, 14]. For example, a recent study on the impact of individual hand hygiene at airports reveals a potential pandemic can be inhibited by 24% to 69% by increasing the engagement of travelers in washing hands [20].

Thus, we aim to empirically model potential intervention strategies pertaining to micromobility systems when infectious disease outbreaks happen. This can be informative both in terms of which methods to use and how to apply them. For example, the disinfection of a portion of bicycles may reduce the intensity of the disease agents on the surfaces and indeed decrease the possibility of transmission. Alternatively, placing hand sanitizing stands at stations and encouraging individuals to use them may reduce the possibility of spreading the pathogens.

1.2 Structure of The Dissertation

The dissertation is organized as follows. The next chapter reviews the literature in four sections. First, it reviews the theoretical and empirical literature on the intersection between transportation systems and epidemics. In the second section, it reviews the micromobility transportation systems. Third, it reviews the literature related to modes of transmission. Then, it reviews both the theoretical and empirical bodies of literature on epidemic modeling. It starts with the most popular and straightforward homogeneous mixing compartment models and continues with more complex metapopulation, contact network, and agent-based models. Finally, it reviews the intervention and control methods and strategies. Chapter 3 states the research questions and contributions of this research. Then, in chapter 4, we present a methodological framework and our proposed model's building blocks. It is followed by proposing a case study of a viral disease spread in Cook County in chapter 5. Conclusions are drawn in Chapter 6.

[3]Transportation Research Board, 2018 https://www.trb.org/Main/Blurbs/180591.aspx

CHAPTER 2: LITERATURE REVIEW

This section reviews the literature from multiple disciplines that cross-pollinate around transportation systems and the spread of infectious diseases in urban areas. It starts with a background and literature review on transportation systems and their relationship with epidemics, followed by the recent literature on micromobility systems. The next section reviews the theoretical underpinnings of contagious disease spread, including different agents, disease, and transmission modes. Then, we review the methodological aspect of the subject, which mostly covers the modeling of the dynamics of disease spread in space and time. Finally, we review the intervention and control strategies that can be envisioned.

2.1 Transportation Systems and Epidemics

Throughout history, transportation has contributed to the spread of vector-borne and non-vector borne infectious diseases locally, regionally, and globally. The interaction between infected and susceptible hosts and vectors usually creates a geographical pattern known as spatial diffusion of disease. The spatial diffusion of infectious disease follows a contagious, hierarchical, or network process [21]. In human to human infectious disease, these patterns emerge following the mobility of people along with transportation links. Contagious diffusion happens when the infection spreads outward from a source of infection to other locations according to proximity. For example, the spread of HIV in suburban areas is highly correlated with the number of workers who commute to urban areas [9, 22]. Human to human infectious diseases also diffuse hierarchically, starting from large cities and moving to medium and then smaller cities. Lastly, network diffusion is the spread through transportation and so-

cial networks. Most of the infectious diseases these days follow a hybrid pattern due to the complexity of transportation systems and of human mobility [9].

The geographic extent and the speed of many pandemics are attributed to the extent and speed to which humans could relocate by ground, water, and air transportation. Multiple reemergences of the plague around Europe during the 14-17 centuries and its later outbreaks that mainly happened in port cities involved modes of international transportation at the time [23]. The 1957 influenza pandemic originated in mainland China, spread around the world in only six months thanks to the availability of both air and sea travel, compared to the 1918 influenza pandemic, when passenger air travel was not yet available [18]. According to the World Health Organization (WHO) COVID-19 situation report, three months after the emergence of COVID-19, almost all countries worldwide reported infection cases.

The transportation systems and epidemics (or pandemics) exhibit a two-way relationship. Transportation can contribute to the spread of disease, and in reverse, epidemics can restrict operation or shut down transportation systems as soon as they occur. Transportation systems can contribute to the spread of diseases in three ways. First, by introducing it to new geographic regions via the interregional transfer of commodities, people, animals, etc. This contribution exposes a larger population to the agent that causes the disease, and as a result, a larger population may be impacted by the time the epidemic disappears. In other words, the transportation system affects the extent of the epidemic [13].

The second way is not by introducing the disease to a new region but through the speeding up or acceleration of diffusion in the areas that have already been affected by the disease. This happens by both increases in the volume of flows and decreases in travel time from one location to another [5]. The impact on the final epidemic curve is that the infected population reaches its maximum more rapidly. In these two ways, the infection of the new population happens in the destination.

[24] predicts the geographic spread of the SARS by considering global air traffic. Results show an agreement between the predictions and actual incidents. They show that the strong heterogeneity of the network causes a high degree of predictability. Multiple studies have shown the daily new cases of COVID-19 infections was associated with mobility records of three weeks before, indicating the significance of human mobility behavior in the spread of the disease [25, 26]. Third, for certain types of infectious diseases, such as gastrointestinal and respiratory diseases, the infection may happen within the transportation infrastructure or the vehicles where people have direct contact with one another and the environment [27]. In this regard, transportation systems can be considered as vectors [1]. Examples of these situations include global airports, public buses, and train stations. It is essential to distinguish this contribution from the other two introduced earlier because transportation infrastructure and vehicles act as intersection points of trips where people are in contact[28]. Furthermore, it can inform the choice of appropriate intervention methods. More investigation is required regarding the role of transportation systems, and transport hubs in the spread of infectious disease either in the form of pandemics or epidemics [29, 17].

2.2 Micromobility Systems As a Novel Mode of Transportation

Transportation systems operating on the principle of asset sharing can be categorized based on whether they share a vehicle, passenger ride, or a delivery ride. Micromobility systems are a subcategory of the first class and include sharing public bicycles, electric scooters, or other similar single-user vehicles [30]. Among micromobility systems, public bikesharing systems first appeared in Amsterdam in 1965 and became more popular in several other European countries during the 1990s. Be-

[1] The word vector is usually used for living organisms that can transmit infectious pathogens between humans, or from animals to humans. Our use of the word for transportation systems through out the document is mainly to make a distinction with other ways transportation may contribute to the spread of infectious disease.

fore 1998, all public bikesharing systems were coin based. The first computerized bike sharing system with 200 bicycles was developed in Rennes, France [31]. During the 2000s, many countries in other parts of the world started to open their public bikesharing systems. Still, they hardly became widely popular until the 2010s, when information technology facilitated their operation significantly. The number of docked bikesharing fleets experienced a worldwide increase from 139,000 to 2,300,000 between 2010-2016, with more than 80% deployed in P.R. China[2]. By the end of 2016, France, China, Belgium, and Taiwan are the top five docked bikesharing markets with 6.7, 6.2, 6.0, 5.0, and 4.8 bikes per 10,000 people [32].

Dockless bikes emerged in 2016, starting in China, and boomed to the other parts of the world. There is no exact market scale for dockless bikesharing. By the end of 2017, estimates of a fleet of 23 million bikes in China have been reported in the literature [32], with picks of 70 million rides a day, and 17 billion rides overall for dockless bicycles. In some countries like China, Denmark, and the Netherlands, cycling encompasses more than 25% of all trips, especially in large cities [13]. Public electric scooters started operation in 2015 from Europe and spread throughout the world [30].

Research has been conducted on the benefits of bikesharing systems on health against chronic disease [33], on bicycle rebalancing [34], demand [35], on factors impacting bicycle use [36], and on revealing spatiotemporal patterns of trips [37]. One way to look at a mode of transportation is to see how it affects the mobility of people. While micromobility systems are well known to serve the first or last mile of trips, they are also used for long-distance trips in urban areas [38]. The distribution of trip distance for users of public bicycles in China shows that 23% of the trips are less than 0.5 miles, 60% of the trips are between 0.5-2 miles, and 20% are more than 2 miles (See [32]).

[2]https://bikesharemap.com

These systems also impact the way people use other public transportation systems and their vehicles. According to a survey conducted by [30] on North American public bikesharing systems, more than 50% of the respondents reported a reduction in personal automobile usage. The increase in bus use in all cities has also been attributed to bikesharing improving access to/from bus stations. Also, the bikesharing systems of large cities tend to use buses less due to the cost benefits and speed of the bikesharing systems. Thus, bikesharing systems may work both as competitors and as complement to public transportation systems [39, 40].

Operations related to rebalancing and charging are the most critical challenges concerning micromobility systems. Research has been conducted on the costs, benefits and operational optimization of rebalancing [41] and to predict the demand to inform this logistical activity [42]. For example, [43] studied the individual destination preferences of bike users in Chicago using a multinomial logit model, where the destination station utility is affected by user attributes (e.g., age and gender), trip attributes (such as time of day) and destination attributes (e.g., distance from the origin station, bicycle infrastructure variables, and land use and built environment attributes). Their results show that higher capacity stations were more likely to be chosen as a destination. The network distance between origin and destination has negative impacts on selecting a station as a destination. The effects of the number and capacity of neighboring stations are opposite for casual and regular members. Members are more likely to use the higher number of stations with smaller capacities, while daily customers tend to use fewer stations with many docks. This information can inform the rebalancing of bicycles by operators.

While inventory management and rebalancing are still the main concerns in micromobility systems, equity of the systems, instead of equality, has been discussed in a few papers. It has frequently been reported in the literature that male, white, employed, and, compared to the average population, younger, more affluent, more

educated people have a higher tendency to use bikesharing systems [44]. Data from the New York City Bike-share system shows the distribution of bikesharing stations is uneven between low and high-poverty regions, having a higher share in more affluent areas [45]. Demand for bikesharing systems in disadvantaged communities is lower than in other areas, for instance, in Chicago [46].

Only limited research has been conducted to compare docked and dockless bicycles in the literature, which sometimes are not in agreement. [46] analyzed the performance of the two types of systems with regard to providing service to disadvantaged communities by considering bike availability, service areas, bike idling, rebalancing, and trip demand for these systems in San Francisco, CA. In terms of availability, dockless bikes are more readily available, and the bike idling time is higher on average for them in disadvantaged communities of San Francisco. These results are limited by assuming that trips are associated with local users, which is not necessarily true. In San Francisco, some tracts are both communities of concern zones and tourist areas. Tourists have been found to generate a large portion of bikeshare trips.

[47], however, compared the spatial access in these two types of systems based on bicycle data from 73 bikesharing systems and census level socioeconomic variables. Their result shows that dockless systems operate more equitably than docked systems by education but do not differ in spatial access by socioeconomic class.

The following are some of the challenges mentioned in the literature working with data from micromobility systems. While trip data are available from many micromobility systems, the user attributes are not, which makes it hard to study who are the users of the system and what their mobility behavior is. No information shows what the purpose of a trip is in micromobility records or whether it is part of a multimodal trip; not does it provide the origin and destination of a trip made by a user. While the user type is usually available, there is no information about the frequency of trips made by any individual. Within a geographic region, trips can be made by local or

non-local customers, which is not available in any of the current micromobility data. Surveys have been conducted in the case of micromobility systems to tackle some of these challenges. Still, in many cases, survey results are not in agreement, making it hard to generalize conclusions to other systems in other cities. Data from dockless bikes and electric scooters have more limited availability, making it hard to compare with station-based systems. Distinguishing real trips made by users from the relocation of bicycles made by operators is also not an easy task. Some methods to deal with are available by data mining methods such as looking at the charge level of the battery, the distance of two bike locations records, or the duration of a trip [46].

2.3 Transmission Modes

To control infectious disease, we need to understand how the responsible pathogens and microorganisms spread throughout the environment and between the hosts [1]. For a contagious disease to infect a person, the pathogen first needs to reach the host (transmission). It needs to grow or multiply in the host body (infection or exposed), which does not necessarily result in disease. In case that it causes illness, there is a period before the symptoms appear (incubation) which is typically in the range of a few hours (6-12) to several days (30-60). After a host becomes infectious, he/she may shed disease agents through cough, sneeze, vomit, etc. in the environment, and in so doing, contaminates the air, surfaces, water, or food. A secondary transmission happens when the pathogen enters the body of a healthy individual from the environment through the entrance membranes, and the process repeats [48].

The mode of transmission is an agent characteristic, and a specific agent may transmit through multiple routes. Since this research focuses on the transmission via surfaces on micromobility vehicles, we study the pathogens and the disease that are able to transmit through the fomite route. The fomite route is a common pathway for many respiratory and enteric diseases. However, for many of them, it may not be the main route of transmission. For example, for most respiratory disease-causing

agents, such as the influenza virus or coronavirus families (SARS, H1N1), the main transmission route is in fact through aerosols (close contact route). However, long-range airborne transmission or transmission through surfaces are also possible **??**. Some gastrointestinal disease-causing pathogens such as norovirus, on the other hand, transmit mainly through surfaces or food more efficiently, while they can also transmit through aerosol as a secondary route [49].

The study of the spatiotemporal patterns of gastrointestinal and respiratory disease is usually limited to close contact and direct transmission between the hosts (aerosol) or transfer through common vehicles such as water or food. However, less attention has been placed on the role of fomites. In fact, it used to be assumed that the fomites are not a significant route of transmission. Recent studies show that the role of these surfaces is actually not negligible in the spread of many types of pathogens such as rhinovirus and norovirus that are responsible for respiratory and gastrointestinal kinds of disease [2, 3]. Thus, we know that the surfaces are involved in transmitting certain types of infectious disease, but there is little data about how contamination propagates in the environment [50].

Several sources of contamination exist for the surfaces. One way is to divide the fomite contamination sources into direct and indirect ways. Contamination can be deposited on a surface directly from the aerosols in the air, droplets from a cough, or aerosolization of other body secretions such as vomit [51]. This is considered a host-fomite route. Contaminated hands, either after coughing on hands (host-hand-fomite) or touching another contaminated surface (fomite-hand-fomite), can also indirectly transmit pathogens on a new surface. [50] characterized the fomite-hand-fomite route to create a notion of surface contamination network. After a surface is contaminated, either through a direct or indirect route, an individual may touch and pick up a portion of contamination on his/her hand. The contaminated hand may touch mucus membranes of the host and cause infection, or it may touch another surface and

transmit contamination to that. This chain of hand and fomite touch creates a network of connected surfaces. The nodes of this network are fomites; they are linked if they have been touched with at least a hand. The authors found that the number of contaminated surfaces in an airplane grows logistically, and in a period of five to six hours, nearly all touchable surfaces are contaminated. This logistic growth shows that, in crowded environments, surfaces may become contaminated very quickly. Also, evidence exists that high-touch surfaces are more likely to be exposed to severe contamination and, as a result, play an essential role in the transmission of the contamination to other surfaces like human hands [52, 53].

A combination of pathogen and fomite related characteristics and human behavior regarding the above sources of contamination defines whether or not a fomite-mediated route can sustain the disease spread. Pathogen characteristics, such as environmental persistence, shedding, recovery rate, and dose-response, are among the main factors that influence disease transmission through surfaces [51]. For example, rhinovirus, norovirus, and influenza can all transmit through surfaces. However, the first two are more infectious through fomite routes [54].

Not all of the surfaces are considered as high risk concerning disease transmission. A large body of literature has attempted to characterize fomite properties. Fomites have been classified according to different transfer efficiencies and die-off rates, mainly divided into porous and non-porous [55]. The size of the surface, together with human touching behavior, has also been used to differentiate fomites concerning pathogen transmission [51]. Some of the surfaces are not accessible to human touch at all. Thus, they are not involved in the transmission while they can still become contaminated. However, some others are subject to rates of touch by hands ranging from low to high (frequently and non-frequently touched surfaces) [56]. [51] shows large and frequently touched surfaces have the highest potential to be involved in fomite transmission of influenza. They also show the direct route (droplet-surface) in the transmission of

contamination is more likely than the indirect route (droplet-hand-surface), except for small frequently touched surfaces such as doorknobs.

Surfaces and human behavior have also been characterized according to the venue where the contact may occur (e.g., home, school, office, public transportation) [54]. Fomites in different venues can be differentiated based on their physical (size, type, quantity) and human-related characteristics (touch frequency, the density of host, duration of time spent in a venue, and shedding rate concerning age). For example, the number of accessible surfaces, their touching frequency, and the shedding rate is high at schools compared to home or office.

2.4 Epidemic Modeling

We can categorize the individual level epidemiological models according to the degree to which they represent reality and their complexity. The main model strands can be divided into the compartment, network, and agent-based models, in order of increasing complexity [57].

2.4.1 Homogeneous Mixing Models

Homogeneous mixing models are the simplest dynamic models for spreading disease in the host population. The origins of the subject are attributed to the Swiss mathematician Bernoulli in the 18th century, while he was working on the smallpox [58]. Hamer (1906) and Soper (1929) [59, 60] established the foundation of such type of system view models, known as recurrent epidemic models. The model organizes people in different compartments according to their health status. In a simple model, each individual in the population is susceptible (S) or infected (I), or recovered (R) at any point in time, and their compartment will change over time based on different transition rates (figure 2.1).

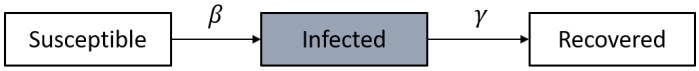


Figure 2.1: Homogeneous mixing compartment models.

In homogeneous mixing approaches, deterministic mathematical models are used to model temporal dynamics of the epidemic using the following differential equations:

$$d_s/d_t = -\beta i(t)s(t), \tag{2.1}$$

$$d_i/d_t = \beta i(t)s(t) - \gamma i(t), \tag{2.2}$$

$$d_r/d_t = \gamma i(t) \tag{2.3}$$

Where $s(t)$, $i(t)$, and $r(t)$ are the fractions of the population within each state. Transmission rate β characterizes the ease with which the agent is transmitted from person to person, and γ represents the rates with which individuals are removed from the infectious state. $i(t)s(t)$ is a representative of the contacts between individuals of the two compartments $s(t)$ and $i(t)$. This term makes a few critical assumptions about the population. It assumes that each person has an equal probability of contacting every other individual and that all individuals have the same number of contacts at each point in time [11]. These types of models are known as homogeneous mixing models. The assumption of homogeneous mixing contact among individuals may be not hold in reality, thus limiting the practical value of this approach to properly model some critical aspects of disease spread [57].

Homogeneous mixing models have also been used to model the spread of infectious disease through the environment, including inanimate surfaces. To provide a theoretical framework for examining the role of the environment in pathogen transmission of non-vector born disease, [61] suggested the so-called environmental infection transmission system (EITS). This model enables the researcher to characterize different fomite-mediated transmission routes by incorporating deposition and pick up of the pathogens in a homogeneous environment. This is realized by defining a surface contamination ratio, which measures the total pathogen deposited by an infectious individual during the contagious period. The environmental persistence ratio creates

a link between the pick-up and elimination of the pathogens. Using this conceptual framework, [51] characterized the impact of fomite route transmission for influenza. They estimated the basic reproductive rate based on human behavior and the two main characteristics of the surfaces.

2.4.2 Metapopulation Models

To account for heterogeneity in contacts, simple spatial models were initially used. These models assumed that contacts between susceptible and infectious individuals decline with the distance between them and create a heterogeneous mixing population. While explicitly spatial models are challenging to derive using mathematical equations 2.1, 2.2, and 2.3, a simplified representation of space where individuals are located at regular distances from one another and the spread of infection is isotropic is possible [18]. In this case, each spatial unit (e.g., county) has separate compartments sets, and transportation links connect these units. These links can be weighted by the flow of population between them, which then acts as a covariate. In the absence of flow data, more sophisticated spatial interaction models (e.g., gravity models) may be used to account for the mobility of the population between these spatial units per unit of time [62]. A gravity model uses the mass of the population in regions together with their distances to estimate the flow between areas [18].

These approaches are known as metapopulation models and have been among the main spatial modeling methods over the past two decades [19, 63]. The well-known global epidemic and mobility (GLEaM) model is based on a metapopulation approach [3]. In this model, the world is divided into regions defining subpopulations connected by transportation and mobility infrastructure. [64] compares GLEaM with an agent-based approach to model the spread of influenza-like disease between municipalities in Italy and shows the results of both models show agreement for epidemic spatial patterns in the accessible levels of granularity. The difference is in the peak time,

[3] http://www.gleamviz.org/

which is in the order of a few days. The extent of population infected at the end of the epidemic is higher for metapopulation models, which is expected based on the structure of within municipality populations. At the urban scale, [65] uses a dynamic metapopulation model informed by hourly subway transit data in Shanghai to estimate epidemic risks across different spatial and temporal scenarios. They found that time of day and the neighborhood of disease introduction significantly impact the initial epidemic growth. These results show that the metapopulation method is an effective way to model spatial aspects of disease spread.

In both homogeneous mixing and metapopulation models, ordinary differential equations are used to model the dynamics of transmission. Overall, these models are able to model the temporal dynamics of epidemics. When it comes to integrating systems together, in terms of replicability, these models are generalizable and easy to communicate. Homogeneous mixing models are less flexible when it comes to highly parameterized, disaggregated, and spatially explicit applications [62]. This is not necessarily a drawback and in applications where the aggregated behavior is of interest, it can smooth out fluctuations. Metapopulation models can account for heterogeneities between subpopulations, but they still make simplistic assumptions about the intra-subpopulation heterogeneities.

2.4.3 Contact Network Models

Network-based models are another way to incorporate human contact structure or mobility of populations. The nodes of the network are individuals or geographic locations. If the aim is to create a network of host contacts, it is called a contact network. As with the other individual-based methods, all individual humans within the population are modeled. These individuals are represented as nodes and connected by edges that represent their social ties.

The first question that needs to be answered when setting up a contact network model is how to create a network of contacts. Whether or not an edge exists be-

tween two individuals is a function of the mode in which the disease transmits in the population, and its weight is a function of disease transmissibility [11]. There are three ways to create contact networks: real, simulated, and semi-real networks. Using real contact networks is the best, but they are rarely available. Instead, simulated contact networks with network parameters that resemble the real networks are frequently used. The degree distribution of the individuals in the network is usually used to create a simulated network of contacts. Several well-known network structures have been used in the literature, according to the specific disease and the nature of the real contact networks. Random graphs, small-world networks, and scale-free networks have been used because they represent many natural and social phenomena [66]. Suppose the network is a random graph with a Poisson degree distribution. In that case, it is equivalent to the homogeneous mixing compartment models. The degree to which the real network departs from this structure determines whether the traditional compartment models are valid [67]. Small-world networks are characterized by high levels of local clustering and global connectivity. In contrast, scale-free networks are characterized by degree distributions that follow a power law, where a small number of nodes have a large number of contacts [67]. The problem with these networks is that they cannot model the spatial aspects of the epidemic, and real contact networks can still be very different from them.

As a third alternative, closer to the real contact networks, semi-real networks can be created using synthetic populations from census, travel behavior surveys, and other ancillary data sources [68]. While creating synthetic population networks is a recent addition to research practices, creating synthetic populations has been around for many years. These synthetic populations are usually categorized in micro-simulation models (MSM), which are closely related to agent based models (see section 2.4.3) [69]. The goal of a synthetic population is to simulate individuals in a way that the distribution and correlation of their attributes are similar to those in census and

the number of people with specific attributes matches the aggregated population data [70]. Synthetic population methods can be partitioned into sample-based and sample-free based [71]. Among sample based methods, synthetic reconstruction [72] and combinatorial optimization techniques [73] have frequently been used in the literature. The former use the iterative proportional fitting method [74] together with a sample of the target population to obtain the target joint-distributions. Examples of that are[4], TRANSIMS population synthesizer [75], and the synthetic population of Virginia Bioinformatics Institute Synthetic [4]. The combinatorial optimization methods create a population and modify it with the sample population until it meets a required constraint [76]. Using these synthetic populations, very large contact networks for purposes including disease spread have been created on national, regional, and urban scales in the past few years [77, 4, 76, 77]. A method developed by [76] is used in my research to create a large contact network as the baseline population.

Various types of disease, modes of transmission, and nodes may require different network types. For example, within a healthcare center, patients treated in different wards are typically not directly in contact. However, caregivers moving from ward to ward may transmit the infection from one ward to another. Bi-partite networks can be used to model contact networks of this type, where two different types of nodes exist in the network [78]. Sometimes the disease can only be transmitted in one direction. For example, many gastrointestinal diseases transmit mainly through surfaces. If person A (not necessarily infected) touches a surface with a contaminated hand and person B touches the same surface afterward, the contact AB is in one direction. It is not possible for person B to infect person A. These networks are usually called semi-directed networks. Each node has three degree distributions: undirected degree, directed indegree, and directed outdegree distributions. Epidemiological quantities for these types of networks are in [79]. [80] show that the probability that an outbreak

[4]https://www.rti.org/impact/rti-us-synthetic-household-population%E2%84%A2

happens and the fraction of the population that is infected can be different in semi-directed networks compared to undirected networks. The transmissibility may not be the same in both directions, making contact asymmetric. For instance, not all the population age groups are vulnerable to disease to the same degree.

The links in networks can also change over time. For instance, contact networks on weekdays and weekends are likely to be different [65], and they would also be distinctively different during holidays and other events of social significance. Most of the mathematical models on contact networks assume that the contacts are fixed during an outbreak. While static networks can capture diversity in the number of contacts and their intensity among individuals in the population, they miss the heterogeneity that results from the dynamicity of contacts of a single individual [81]. Whether or not modeling a dynamic network is needed can be informed by looking at the nature of contacts in the population. If the rate of change in contacts compared to the length of the outbreak is very slow, then a static network can probably capture the structure of contacts appropriately. However, suppose the rate of change in contacts is very high. The population structure is then similar to the homogeneous mixing setting, and maybe a mass action (i.e., homogeneous mixing) model is the best choice. Anywhere between these two cases is when a dynamic network structure can impact the dynamics of the epidemic [82].

Disease propagation in contact networks is modeled using different methods, one of which is bond percolation theory from statistical physics [11]. The spread of disease in network is similar to the homogeneous mixing populations except that a structured contact network replaces the Poisson distribution of contacts. Imagine a vertex is infected initially and remains infectious for a period of time. During this time, it is possible for the disease to transmit to each of the contacts of that individual. This process repeats for those secondary cases and, in this way, the disease percolates in the network. Bond percolation on networks describes this behavior of connected groups

of vertices in spreading disease. It can predict the number of vertices reached via disease transmission along the edges in the network, the size of the infected cluster, and outbreaks [67].

Contact network epidemic models, similar to percolation, can incorporate the heterogeneity of contacts to the model. However, even when the contacts in network are realistic, these models have important limitations. The realistic contact pattern is just one way to increase the fidelity of epidemic models. The spread of disease is different in various age groups. It changes depending on the type of contact (e.g., home, school, and public transportation), various behavior in population, and geographic and seasonality [57]. Models like bond percolation cannot account for these heterogeneities.

Also, while they are faster than other approaches lie agent-based models, they can still be computationally heavy when the contact network is very large and dynamic. Parallel algorithms such as EpiFast have been developed to reduce computation time for these types of networks [77].

While in most epidemic models, the population is divided into compartments, in some other models, population, attributes, and behaviors are aggregated. For example, EpiRank is a model that uses the commuting network of townships in Taiwan together with Markov Chains and a page-rank like algorithm to estimate disease risk for different townships. The model distinguishes inward and outward commuting loads for different daytime settings between townships. The movement is usually more toward city centers from satellite areas in the morning while the heavy load is reversed in the afternoon. Compared to compartment models, these models are less computationally expensive and more suitable to incorporate geographic aspects. They can also identify commutes made by different transportation modes and various environmental and socioeconomic factors [83]. These methods are useful, especially when the purpose is to capture spatial diffusion patterns. However, they cannot

provide detailed information about the epidemic and cannot model the impact of intervention methods.

2.4.4 Agent-Based Models

Agent-based models (ABM) are decentralized, dynamic, individual level (bottom-up) modeling approaches composed of autonomous and independent entities, called agents, that have a purpose, are capable of interaction with each other and with their environment, can make decisions and learn new things and indeed are capable of adapting to new conditions [84, 85]. ABM approaches are a subcategory of a larger class of models that are called individual-based models. Individual-based modeling started with Von Neumann's idea of an autonomous machine in the late 1950s (see [86]). It was followed by the Game of Life model of Conway (1970) [87]. The segregation model of Schelling (1971) [88] was among the first social individual-based models. Although the computational methods started in the 1950s, lack of data and of computation resources imposed severe restrictions in many ways, specifically on individual-based models, that impeded their progress. It is in the 1990s that these computationally intensive models found a chance to become more widespread. The word 'agent' was introduced by Holland and Miller (1991) [89] in the context of the economy. The Epstein and Axtell Sugarscape model (1996) was among the first large scale models that incorporated the geographic space into the agent-based models [90].

There is no agreed definition of agents, but most of the existing literature shares the view that agents are autonomous, heterogeneous, active entities. [91] has investigated the properties of agents that are frequently shared in the literature and their meaning for how agents operate. Agents are autonomous because they are not governed by a centralized control system and are free in setting their interaction and decision-making. Heterogeneity can feature in the agent population properties, in their relationship with other agents, in the environment, and in the rules of behavior. Being active has a more complex meaning. Agents can follow a specific goal

or be reactive to the environment and interact with other agents by having sensory properties. They may be mobile and move in the space, whether it is continuous, discrete (cell or patch-based), or topological/physical network. They can communicate and exchange information. More progressive active properties include bounded rationality, learning, and adaptation, as well as having memory. Bounded rationality comes from the fact that agents (especially humans) do not have infinitive analytical capabilities and do not behave with perfect logical analysis [92].

Since humans are probably the most complex agents in terms of behavior, a brief review of the methods to model this behavior can be informative to other social systems. [93] specifies some of the strategies to model human behavior in ABM. The first category of methods is mathematical models. One of the simplest ways is to randomly select a value for variables that represent the behavior of humans. However, this can end in completely wrong claims because human behavior is not random, although they may have long-range values. A better choice is to use threshold-based and simple IF-THEN rules. Alternatively, differential equations and dynamic modeling can be used, which links the rate of change to its previous states. The second category of methods to model human behavior is more abstract and conceptual. One example of that would be to analyze data on human decisions and create decision trees. Lastly, cognitive architectures like Soar [94] can model the intelligent agentâs behavior. Understanding and measuring human behavior and validation remains the main challenge for modeling human behavior [93].

Adaptation, learning, memory are among the most complex properties of agents. The notion of learning and memory opens space for a variety of inductive learning methods to be integrated with agent-based models. An example of approaches to the modeling of spatial memory and mind maps in ABMs is in [85] in the context of ecological individual-based models.

Previous work has focused on existing ABM [95, 96]. [96] highlighted seven chal-

lenges in developing agent-based modeling as follows: (1) Purpose of the model ranging from exploratory to predictive; (2) Theory and model; (3) Replication and experiment; (4) Verification calibration and validation; (5) Agent representation, aggregation and dynamics; (6) Operational modeling; and (7) Sharing and dissemination of the model.

A significant portion of the research agenda in ABM has been dedicated to addressing these various challenges. One of ABM's benefits is that we can simply create a wide range of scenarios, even for a simple model [97]. However, this can be challenging since we need to make sure that we develop appropriate research problems and scenarios. Models need to have a proper level of abstraction [98]. ABM is a relatively late arrival in modeling. There have been many other widely accepted models like deterministic mathematical (e.g., system dynamics) or statistical models that are usually easier to implement [99]. Thus, it is essential to make sure that ABM makes a difference, which is mainly realized by clearly specifying the purpose of the model and also thinking about the question of what the agents are, what they do, and how their decision may impact the system-level properties or decision of the other individuals? [99]. This can be even more important when there is an explicitly spatial aspect. [99] explore the answer to these questions in the purpose of the agents themselves. Certain model features are critical to attempt to answer the above questions. These features are the heterogeneity of the decision-making context, the impact of interactions, and the size of the system. Heterogeneity appears in characteristics of agents (e.g. gender) and their decision-making context (e.g. health status), which may further vary spatially. If, for example, agents are heterogeneous both in their characteristics and (spatial) decision making context, and their individual choice is likely to cause unexpected outcomes, then there is stronger evidence that using an ABM (as opposed to other types of modeling) in an explicit geographic area makes a difference in our understanding of the system or capturing the unexpected behavior.

Calibration and validation remain the main challenges in developing effective ABM. Calibration involves the fine-tuning of the parameters of the model using real-world data. Validation assesses how much the model results match with the real world by measuring the goodness of fit of the model [100, 91]. Verification is a process to ensure the model serves the goal for which it has been designed (See [101]). Numerous validation methods have been studied in the literature. Reviews of different validation methods can be found in [100, 102].

ABMs are very sensitive to initial conditions. That means small variations in interaction rules may significantly change outputs [98]. If the purpose of the ABM is prediction, this can be a fundamental challenge [103]. ABM is a stochastic model and different outputs are generated at each instance of the model. Thus, it is necessary to have several model runs to produce a statistically robust output [97] that is not dependent on a single specific set of stochastic parameters. In that past few years and with parallel and GPU computing progress, some studies have focused on modeling parallel ABM [104]. One of the complaints about many, if not most of the ABM works, especially those performed in spatially explicit environments and have output maps, is that the outcomes are usually compared or investigated only visually. Instead, spatial statistical approaches can be used to have a more in-depth and more robust assessment of the results.

Representation of the geographic space is also a significant challenge since multiple alternatives are available. Agents do not interact in a vacuum [105]. While they are not necessarily mobile, it should be recognized that they can be both in physical or social environments. The spatial extent of the model is said to have a significant impact on urban ABM models [106]. The use of highly granular resolutions has been reported to produce overly fragmented outputs in some studies [107]. However, a coarse granularity of the data relaxes spatial heterogeneity and spatial dynamics [108]. Aside from all these considerations, the fact is that different processes may

happen at different scales. The interaction between agents and their environment and other agents may occur in multiple scales [108]. Thus, it is essential to fine-tune the model to select the appropriate analysis scale or develop ABM in multiple scales [109].

A vast body of literature exists that has used agent-based modeling to model both vector-borne [110, 111] and non-vector borne infectious disease [112, 113]. However, spatially explicit agent-based models of respiratory and gastrointestinal disease are less frequent. [64] models the spread of influenza-like illness between municipalities in Italy using an agent-based model where a synthetic population is created based on census data. The mobility of individuals is modeled using a gravity model. Large-scale disease spread agent-based models are usually computationally expensive and need a significant amount of effort to implement. [92] suggest a global scale agent-based model (GSAM), a distributed platform, to model international scale disease transmission.

A few empirical studies have been conducted to reveal the role of fomites during real outbreaks using agent-based modeling. A SARS outbreak distribution in a hospital in Hong Kong represented a significant spatial pattern. [114] explored the role of fomites and long-range airborne contact routes to study this spatial pattern using multi-agent-based modeling. In the SARS coronavirus case, results show that the combined route transmission is very likely the correct transmission scenario, and the role of fomites is not negligible. A similar approach has been used in [115] to study the transmission of MERSâCoV during a hospital outbreak in the Republic of Korea.

2.5 Interventions and Control

Effective control of infectious diseases needs a quantitative comparison of different intervention strategies. Intervention success depends on the transmissibility of the disease and the contact network [116]. Accordingly, intervention methods can be categorized into three main approaches: transmission reduction, contact reduction,

and immunization [57]. Transmission reduction can be implemented by using barriers that reduce the spread of pathogens in the environment, such as wearing masks or disinfecting the hands and surfaces. Study shows both these approaches can affect the spread of respiratory and gastrointestinal disease [7]. A recent study on the impact of individual hand hygiene at airports reveals a potential pandemic can be inhibited by 24% to 69% by increasing the engagement of travelers in washing hands [20]. [116] shows that using face masks and general vaccination will only moderately affect the spread of mildly contagious disease. In contrast, quarantine and targeted vaccination can prevent the spread of a broader spectrum of disease.

Case identification and isolation, social distancing measures, and closures can reduce contact between individuals. The role of travel restrictions and control on international flights and hubs has been studied for influenza-like diseases [19, 14]. The social behavior of people usually impacts these intervention strategies. The individual willingness to follow the precautions is related to many factors such as the media or whether a member of their family or friends is already infected. Sometimes it is limited by the necessity of going to work or school [117].

Lastly, by either deploying antiviral agents or vaccination, the number of susceptible individuals decreases within the population. Both of these methods are labor-intensive and need a proper implementation strategy to manage resources. At the same time, both these methods have their uncertainties. On the one hand, the first approach is subject to false positives, unless adequate contact tracing and distribution capacities are available [24]. On the other hand, vaccination may not be effective after some lapse of time because the virus evolves to different variants or the antibodies are no longer exist in the host blood. The strategy with which the population receives the vaccination or antiviral is crucial for effective implementation.

CHAPTER 3: RESEARCH QUESTIONS

My main goal in this research is to better understand various facets of the role of micromobility transportation in the spread of viral disease within urban areas through spatially-explicit modeling. I make a distinction between the risk of exposure that occurs at the destination and exposure that occurs while an individual is using the micromobility vehicle by touching surfaces. In the former, the attention is on the part of the disease spread that results from the mobility of human individuals between an origin and a destination. This, in fact, is less affected by the mode of transportation in urban areas for two reasons. First, micromobility is only one of many options in an individual's modal choice set, which is replaced by other modes of transportation if micromobility is not available. In other words, the contact network of people is expected to remains the same whether or not the modal choice is micromobility. Second, the trip duration across various modes of transportation has no impact on the pace of contagious disease transmission, mainly due to short travel distances in urban areas.

However, in the latter case, the concentration is on the part of the disease spread where micromobility vehicle plays a role as a vector (similar to vector-borne diseases). This role is the center of interest to this research. The first question that I seek to answer in this regard is how surfaces on the new micromobility transportation systems contribute to the emergence and dynamics of viral epidemics in urban areas.

The second objective of this research is to study the spatiotemporal pattern of disease spread through micromobility systems. Thus, I am interested to understand how the interplay of geographic space and time with the use of micromobility services influences the risk of exposure to a viral illness. Both spatial and temporal profiles

of micromobility usage exhibit very different use patterns during different times and for various locations in urban areas.

The last objective of this research is related to intervention methods and control strategies to reduce or eliminate the risk of a viral disease spread through micromobility vehicles. For example, in the presence of an outbreak we can decrease the transmissibility of the disease agent by using disinfection methods. At the same time, strategies to implement these intervention methods are important. Thus, I identify which intervention methods and strategies are more effective in disrupting the spread of a viral disease through micromobility vehicles.

In summary, I seek to answer the following research questions:

1. How do surfaces on the new micromobility transportation systems contribute to the emergence and dynamics of viral epidemics in urban areas?

2. How are geographic space and time organized concerning the risk of exposure to a viral disease out of using micromobility vehicles?

3. What are intervention methods and strategies, including random or systematic intervention, more effective in controlling the spread of infectious diseases through micromobility vehicles?

Given the above research questions and the fact that this research is at the intersection of public transportation, environmental health, and urban modeling, I underscore the following contributions to each discipline.

While modeling the role of transportation and mobility in the spread of infectious diseases has been around for decades, modeling the role of different modes of transportation in the spread of the infectious disease has less been a focus of attention. This makes it less practical for intervention and control strategies, which are specific to each mode of transportation. Micromobility is a relatively new and fast-growing transportation mode that has induced change in mobility behavior in urban areas. Thus, it is less studied than other modes of transportation, specifically when it comes

to the spread of diseases. Besides, recent models of contagious disease spread in transportation systems usually assume homogeneity in population behavior while the spread of contagious diseases is highly dependent on the behavior of individuals [27]. This research contributes to the literature in public transportation by modeling the spread of disease in micromobility systems and then explicitly modeling the heterogeneity in human contacts and behavior. Last but not least, I model the transmission through micromobility vehicles by considering actual individual trips. The proposed model can be used to simulate scenarios as new micromobility data streams into the system.

From a modeling perspective, I propose a novel large-scale spatially explicit agent-based model to study the spread of a viral disease through micromobility and a large urban human contact network. The model has several attractive features. First, the model integrates actual individual micromobility transportation trips into epidemic modeling. This means it exploits actual mobility behavior of individual trips, and spatio-temporal movement patterns of the micromobility vehicles. Second, most epidemic modeling methods make simplistic assumptions about the heterogeneity of contacts, spatial distribution, and mobility of the baseline population. I use a synthetic human contact network generated from census data, which also considers individual social ties in the population. Third, the proposed SIR-SC epidemic model integrates the calculation of the spread of viruses through micromobility vehicles and through other processes in the baseline population. In epidemiology, it is common to divide the population into different compartments that represent health status of individuals. SIR represents the health status of human individuals (susceptible, infectious, and recovered), and SC represents the contamination status of micromobility vehicles (susceptible and contaminated). Thus, the SIR-SC models the complexity in the behavior of individuals in ways that are consistent with epidemiologic theories. From an epidemiological perspective, to my knowledge, this research is the first attempt to

model fomite-mediated disease transmission in explicit large urban scales. Previous studies only focus on a single or limited number of indoor spaces. This enables us to experiment with spatial intervention methods.

CHAPTER 4: RESEARCH DESIGN

This chapter builds the methodological framework for modeling the spatiotemporal spread of contagious disease through micromobility systems at an individual level using an agent-based model. I propose a hybrid modeling framework where agent-based modeling features prominently owing to its ability to model the behavior of individuals explicitly. The spread of contagious diseases highly depends on the behaviors of individual humans. These micro-level behaviors may result in outbreaks of disease at the system level. Agent-based modeling, which is a bottom-up approach, can model these behaviors. I describe the agent-based model using the overview, design, details protocol (ODD) [118].

4.1 Purpose of The Model

The model aims to simulate the spatial and temporal dynamics of a viral disease through surfaces on micromobility vehicles in urban areas. I propose a spatially explicit hybrid agent-based model in an urban scale. The current model focuses on micromobility transportation and on viral diseases that can transmit through surfaces (e.g., Norovirus and Rhinovirus) as surfaces are the only contact points between human individuals in micromobility systems. To this end, a large contact network is generated and used as baseline population. The model uses multiple discrete-time intervals for different processes. To calculate the disease transmission process in the micromobility system, a 1-day time-step is used.

4.2 Agents, State Variables, And Scales

Figure 4.1 shows the class diagram of the agents in the model. Human agents are independent and heterogeneous meaning they are characterized by age, gender,

household ID, work/school ID, and health status. They are also connected within a social network of family members, co-workers, or schoolmates. Such a social network is created in the process explained in section 5.4. and used as the baseline population in my model. Human agents are also active and mobile meaning they can move around either by if-then rules (to follow their schedule) or by picking up micromobility vehicles. They interact with other human individuals in the baseline population through proximity at school, home, and workplaces, and with micromobility vehicles. Micromobility vehicles are characterized by their contamination status and by their current location (latitude and longitude for dockless micromobility or station for dock-based systems). A unique property of the proposed model is that it is designed to use actual micromobility trips made by human individuals in urban areas. At the same time, the model can be adapted to include trips with other modes of transportation. Disease agents are not explicitly modeled in this research. In this model, a viral disease that can transmit through surfaces is characterized by the basic reproductive rate, inactivation time of a virus on surfaces, and recovery rate.

Static agents shape the other entities and geographic spaces in the environment. Schools are defined by their enrollment level and location. Households are characterized by the number of members, type, and location. Workplaces have number of employees and location. These geographic entities are considered as static point features. Areal geographic entities are divided into census units and micromobility service area units. The granularity and partitioning of the former geographic units are flexible and defined based on the desired level of complexity, constrained by population data available at that granularity level. The latter, however, is defined based on the service area of the micromobility stations (in station based systems).

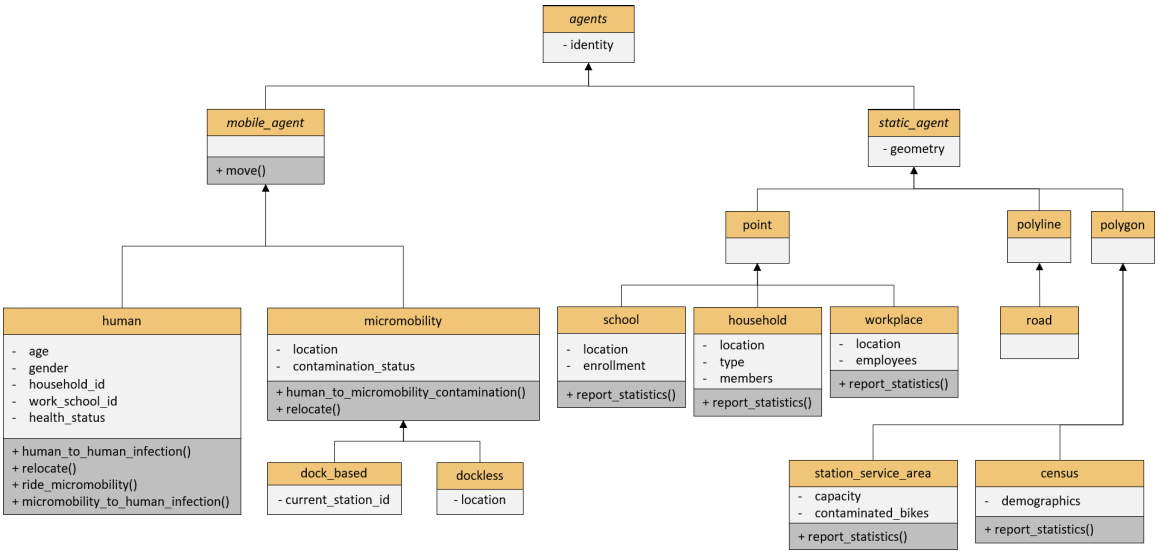


Figure 4.1: Class diagram of agents in the model.

Every human agent is simulated and located in a household. A human individual has a schedule and may go to either school or work during working/schooling hours or stay home for the entire day, representing preschool-age children, homemakers, people who work at home, retirees, and the elderly. People usually spend most of their time at home, workplace, or school, which increases the risk of infection in these small subpopulations, making it reasonable to distinguish these activities from others.

Such synthetic population with their daily schedule serves two important purposes in the model. First, it determines whether or not a connection exists between every pair of individuals in the population. That is, two individuals are connected if they are members of the same household, school, or workplace group (i.e., in the same proximity). Such a connection, if it exists, creates a potential route of transmission for viruses. At the population level, this would be a large individual contact network through which disease transmit. The schedule of human agents remains the same during workdays of the week, but it changes over weekends and holidays, meaning that they stay at home. That means two such contact networks are created both for weekdays and weekends (figure 4.2). These contact networks are created once and before the transmission process starts. The second purpose of a synthetic population

with daily schedule is to model the change in location of individual humans during the day due to a travel activity, which will be explained in section 4.3.1. where we describe this relocation process.

There are two discrete time-steps in the model. The first time-step is one hour and the second is a day. The former is used for considering relocation of the individuals. The latter is used to calculate the spread of disease in micromobility system and the baseline population at the last hour of each day.

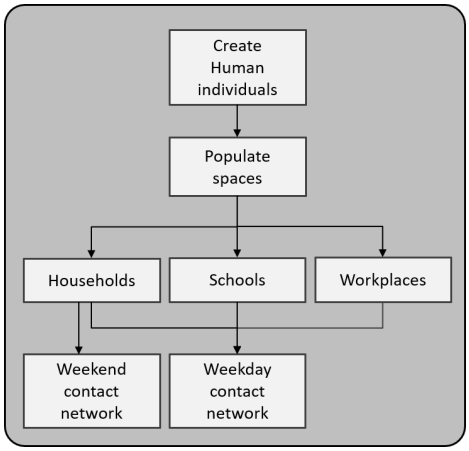


Figure 4.2: Synthetic baseline population.

4.3 Process Overview And Scheduling

The model comprises several processes, including relocation, micromobility ride, human-human transmission, human-micromobility vehicle transmission, micromobility vehicle-human transmission, inactivation, disinfection, and recovery. In what follows, I explain each process in more detail.

4.3.1 Relocation

Two mobile agents are present in the model: micromobility vehicles and human individuals. Micromobility vehicles relocate in the simulator to match actual trip records between start and end station service areas or start and end locations (for dockless systems).

As mentioned at the end of section 4.2, the second purpose of a synthetic population

with daily schedule is to model the relocation of individual humans during the day. The mode of transportation with this relocation is not important in itself. People relocate during a day based on their schedule and visit different micromobility service areas near their home, school, or workplace locations at what point they may decide to initiate a micromobility trip from that location. This means people from any part of the urban area can pick up a micromobility vehicle currently located in a specific zone as long as they are present in that zone. The distinctive objective of this research is to assess the implications of such decision of the spread of a contagious disease at the system level, spatially and temporally.

Virus loads may be exchanged between an individual hands and the micromobility surfaces and be transferred to other zones by the individual or the micromobility vehicle (section 4.3.3). One thing that is important about these two types relocation is that their purpose is not important for the model. Next I describe how human individuals in the model are assigned to ride micromobility vehicles.

4.3.2 Micromobility Ride

As mentioned earlier in section 4.2, a unique property of the proposed model is that it is designed to use actual micromobility trips made by human individuals in urban areas. These trips determine origin and destination zones of a micromobility vehicle associated with that trip. Based on the relocation process explained in the previous section, and the current location of human individuals and micromobility vehicles, individuals are assigned to trips that start from a specific zone in a stochastic way. In case that the demographic attributes of a rider is available in actual trip records, the process of assigning individual humans to micromobility vehicles is conducted by matching these attributes in a stochastic way (see section 4.4). So far, we know how human individuals and micromobility vehicles relocate in an urban area (figure 4.3). Next I describe how pathogens transmit and cause infections in micromobility systems and in the baseline population.

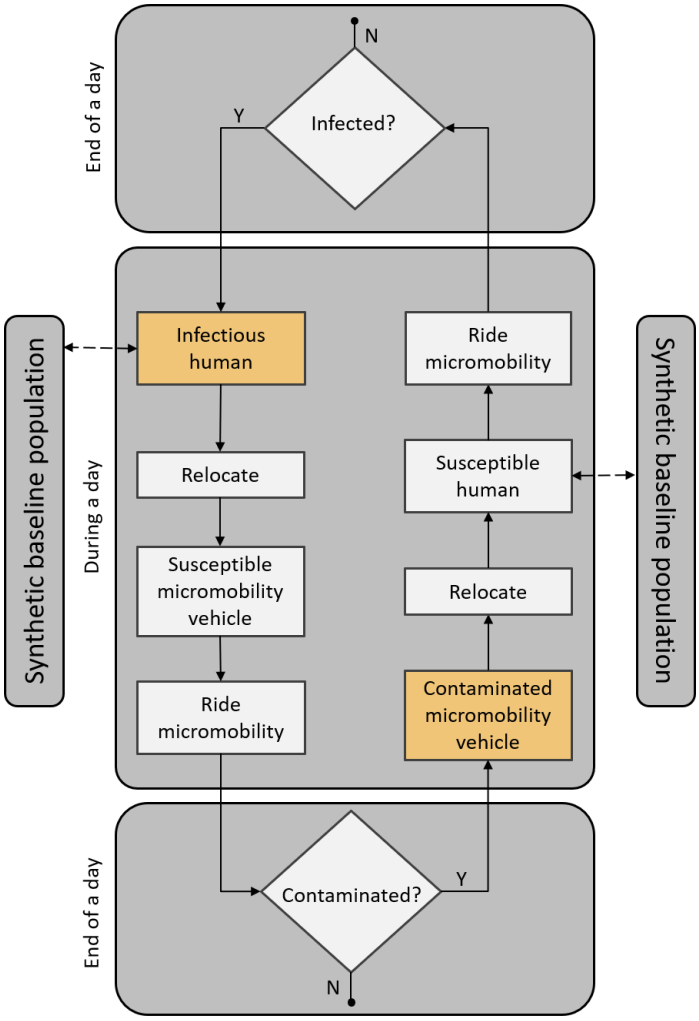


Figure 4.3: Process overview (micromobility).

4.3.3 Transmission

This research focuses on a viral disease transmission through fomites on micromobility systems and a baseline population. Within this context, fomites are a sufficient route of transmission, while they may not be the main pathway (see [54]). For example, surfaces are the main route of transmission in Norovirus outbreaks, while they are only a secondary route for rhinovirus transmission which sustains mainly through droplets route ([119]). Thus, I make a distinction between transmission in micromobility and transmission in the baseline population because in the latter transmission may occur through multiple pathways. Within micromobility systems, interaction ex-

ists between human individuals in an indirect way through surfaces on micromobility vehicles.

In section 4.2, I defined the potential route of transmission in the baseline population based on proximity in places where people spend significant amount of their time together. Thus, within the baseline population transmission occurs in the synthetic contact network where the number of existing contacts with infectious individuals determines one of the risk factors for a susceptible individual to be infected.

In this research, I assume the virus loads are deposited in the environment only by infectious human individuals (no spillover from other species to humans or vice versa occurs). I also assume the virus does not evolve into multiple variants, and the antibodies remain in the bloods of recovered individuals for the entire epidemic period.

Based on the above assumptions, I propose the SIR-SC transmission model depicted in figure 4.4. The host population of human individuals is divided into three compartments: susceptible (S^H), infectious (I^H), and recovered (R^H) (representing SIR in the naming of the model). The micromobility vehicles are divided into two compartments: susceptible (S^M) and contaminated (C^M) (representing SC in the naming of the model). With each iteration t, a susceptible human individual may transit to an infectious state with a certain probability P^{HH} or P^{MH} ($S^H \rightarrow I^H$). At the same time, infectious individuals progress to the recovered state after a certain number of days $1/\gamma$ ($I^H \rightarrow R^H$). I assume the immunity of individuals after recovery based on assumption made earlier in this section. For micromobility vehicles, a susceptible micromobility vehicle may become contaminated with a certain probability P^{HM} at the end of each day ($I^M \rightarrow C^M$), and a contaminated micromobility vehicle may transit to the susceptible state by either disinfection or inactivation ($C^M \rightarrow S^M$).

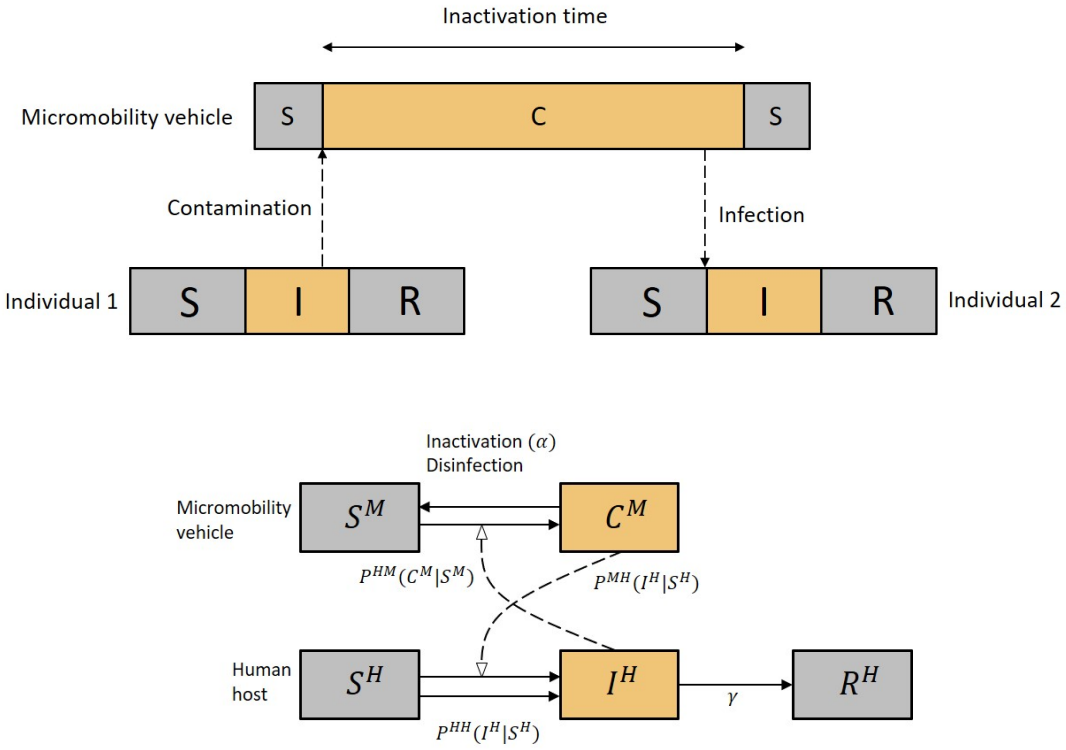


Figure 4.4: SIR-SC model.

According to SIR-SC model, multiple processes exist to model transmission and contamination in the population, namely human-human (HH), micromobility fomite-human (MH), human-micromobility fomite (HM), inactivation, recovery, and disinfection (Table 4.1). These are now discussed in turn.

4.3.3.1 Human-human (HH) transmission

This process models the spread of disease within the population when people spend time in the same proximity at home, school, or office (indoor environments). Infection through this process occurs with a certain probability $P^{HH}(I^H|S^H)_{i,t}$, which is defined based on the transmission rate of the virus and the number of infectious contacts for each individual. When people spend significant amount of time together in a certain environment, it is more likely for disease to spread in the population. We assume the same transmission rates for different venues, but one may calculate separate probabilities for different indoor environments and routes of transmission by acquiring

appropriate transmission rates.

4.3.3.2 Micromobility fomite-human (MH) transmission

As far as micromobility vehicles are concerned, fomites are the only possible pathway for virus to spread. When a human individual uses contaminated micromobility vehicles, infection occurs with a certain probability $P^{MH}(I^H|S^H)_{i,t}$, which is defined based on the transmission rate of the virus between fomites on micromobility vehicles and hands of human individuals, as well as the number of surfaces (number of micromobility vehicles) one interacts with during a day. In other words, at the end of each day, the number of contaminated micromobility vehicles and the rate of transmission from surfaces to hands define the risk of infection, which is used to calculate probability $P^{MH}(I^H|S^H)_{i,t}$. This means that people who use micromobility transportation frequently are more likely to be infected by contaminated bikes. Another parameter that impacts transmission of disease from micromobility vehicles to human is the ability of virus to survive on the micromobility surfaces. A virus must survive a sufficient amount of time on micromobility vehicles until the next human individual use the vehicle. I assume transmission through a micromobility vehicle occurs until the virus is inactivated or disinfected. Thus, a single contaminated micromobility vehicle can infect multiple human individuals.

4.3.3.3 Human-micromobility fomite (HM) transmission

This process models contamination of a micromobility vehicle when loads of virus are deposited on the vehicle surfaces by infectious individuals. HM is also referred to as contamination process and occurs with probability $P^{HM}(C^M|S^M)_{b,t}$, which depends on the transmissibility of the virus and the number of infectious individuals who use the vehicle during a day. Thus, if a bicycle is used frequently in a day, it is more likely to be contaminated.

4.3.4 Inactivation

Virus loads must survive a sufficient amount of time on micromobility surfaces until a new human host picks up the vehicle and become exposed to the virus. I define inactivation rate of the virus on micromobility surfaces as a constant value of α.

4.3.5 Recovery

The recovery rate of individuals is independent of the environment. Individuals are removed by transitioning from the infectious state to the recovered state at rate γ. I assume individuals are immune after infection for the duration of the epidemic. That means the virus does not evolve into multiple variants, and the antibodies remain in the blood of recovered individuals through the end of the epidemic.

4.3.6 Disinfection

Disinfection may occur in either of two ways in the model. The first one is by disinfecting bikes. Each day, a random number Q of micromobility vehicles are disinfected. The second one occurs by disinfection of human hands, which depends on the availability of a sanitizing stand at each station and the probability of sanitizing (P^{snt}) by individuals.

Table 4.1: Overview of processes and parameters

Process	Parameter	Description
Transmission		
	R_0	Basic reproductive rate
	β^{HH}, β^{MH}, β^{HM}	Transmission rate in human contact network, transmission rate from micromobility vehicle to human, and transmission rate from human to micromobility vehicle
	λ^{HH}, λ^{MH}, λ^{HM}	Human-human (HH), micromobility fomite-human (MH), and human-micromobility fomite (HM) risk of infection
	$P^{HH}(I^H\|S^H)_{i,t}$	Human-human (HH): probability of transition from $S^H \to I^H$ for human individual i at time t given contacts with other infectious humans
	$P^{MH}(I^H\|S^H)_{i,t}$	Micromobility fomite-human (MH): probability of transition from $S^H \to I^H$ for human individual i at time t given contacts with contaminated micromobility vehicles
	$P^{HM}(C^M\|S^M)_{b,t}$	Human-micromobility fomite (HM): probability of transition $S^M \to C^M$ for micromobility vehicle b at time t given contacts with infectious human individuals
	ρ	relative transmission factor
Recovery		
	γ	Recovery rate
Inactivation		
	α	Inactivation time ($C^M \to S^M$)

Table 4.1. (Continued) Overview of processes and parameters

Process	Parameter	Description
Disinfection		
	Q	Micromobility vehicles disinfected a day ($C^M \rightarrow S^M$)
	D	Fraction of stations with sanitizing stand
	P^{snt}	Probability sanitizing by individual humans

4.4 Design Concepts

The design concept in an agent based model provides a common framework to design and communicate the model by defining a few concepts including, interaction, emergence, sensing, stochasticity, observation, adaptation, fitness, and prediction [118]. Individual agents communicate through interaction with each other. Within micromobility systems, interaction exists between human individuals in an indirect way through surfaces on micromobility vehicles. In the baseline population, the interaction between human individuals is defined based on proximity in places where people spend a significant amount of time together (section 4.3.3). Emergence is a property of the system which arises from interaction among individual agents. Epidemics are the emergent phenomena attributed to the system of disease in this model, and emerge out of the above mentioned interactions between individuals. Individuals know their schedule during a day and week, as well as their school, home, and workplace locations and contacts. However, this information is imposed to the individuals and do not change during time.

Stochasticity may be defined for different behaviors or processes through pseudo-random numbers in an agent-based model [118]. Deciding on where and to what

degree include stochasticities is a challenging task in agent-based models. This decision highly depends on the purpose of the model and the questions that a model aims to answer. In my model, stochasticity exists in a number of processes and behaviors.

Referring back to section 4.3 where I define the processes, let me start with relocation and micromobility riding processes. On the one hand, relocation is not stochastic in the model. First, each individual follows a constant schedule during a day and week. Second, actual micromobility trips are used as input to the model. That means, the model always starts with the same set of origin-destination micromobility trips. However, it includes actual spatial and temporal patterns of the micromobility system usage in the urban area over time. Thus, one does not need to add stochasticity to the mobility unless the research questions call for that.

On the other hand, the micromobility riding process is stochastic in the model. Depending on their current location, individuals are assigned to micromobility trips in a stochastic way. Let us assume demographic attributes of riders exist in the actual trip records. Then, for a specific station at a specific time-step, trips are grouped based on existing age and gender values in the trip records. Accordingly, from the population currently present in the service area of the station (or zone in dockless systems), similar age-gender groups are extracted, and one of them is randomly assigned to the trip. If no individual with that attribute values is present in the service area (zone), a person from the entire population is selected with the sought age and gender via the same process.

The transmission process is stochastic and based on certain probabilities. Recovery and inactivation rates are assumed constant in the model. Disinfection of micromobility vehicles and sanitizing hands by individuals are also stochastic. Depending on the scenario, disinfection of micromobility vehicles and placing sanitizing stands at stations can be stochastic or systematic.

Finally, I observe the number of infectious and recovered individuals in micromo-

bility and baseline population, separately. Human host infections (either in micromobility or baseline population) can be reported based on their household, school, or workplace locations, or in the census geographic level as an aggregated variable. I also observe the number of contaminated micromobility vehicles as outputs of the agent based model. Micromobility vehicle contamination is reported based on the vehicle current docked stations or census units (for dockless micromobility systems). Individuals do not make strategic choice, follow explicit goals, or predict future. Thus, adaptation and fitness are not part of the model.

4.5 Model Details

Initialization of parameters in the model changes based on different scenarios. In the current model, a virus is roughly characterised by using different R_0 values. Also, actual micromobility data (including age and gender of users) are used as input to the ABM. Whenever data is not available for a parameter, we conduct sensitivity analysis to see how a range of values of the parameter may impact the system (e.g. β^{MH} and β^{HM} through sensitivity analysis of relative transmission factor ρ). No other environmental conditions are entered as inputs in the ABM. Next, I provide the details of transmission processes in the following submodels.

4.5.1 Human-Human (HH) Submodel

Let us start with modeling transmission of a disease within a homogeneous mixing population following [120]. Within a homogeneous mixing population, every individual has the same probability of contact with other individuals. Within this system, imagine an individual with an average of k contacts per unit of time. Only a fraction X/N of these contacts is formed of infectious individuals (X is the number of infectious individuals in the entire population and N is the number of possible contacts or population size). Thus, the number of contacts with infectious individuals during time interval ΔT is $k(X/N)\Delta T$. Now, let us define $w(0\leq w\leq 1)$ as the probability of

successful disease transmission following a single contact. The probability that the transmission does not occur is then $(1-w)$. Furthermore, by assuming independence of contacts, the probability that a susceptible individual escapes infection following all these contacts is:

$$1 - p = (1 - w)^{k(X/N)\Delta T} \tag{4.1}$$

And the probability that an individual is infected following any of these contacts is:

$$p = 1 - (1 - w)^{k(X/N)\Delta T} \tag{4.2}$$

However, probability p is the same for all individuals in the population and does not consider any heterogeneity in space, time, or attributes. Instead, let us consider G as a spatiotemporal graph of human contact networks with N nodes and l edges. Nodes represent human individuals (with their location) and edges represent their connections. The number of nodes remains constant from time step t to t+1 (the system is closed), but the number of edges may change over time (dynamic network). Using such a contact network, one can write a separate probability of infection for each individual i based on the number of contacts at time t. Compared to the equation 4.2, here one has the exact number of infectious contacts for each individual ($k/N = 1$) and the above probability can be rewritten as:

$$p_{i,t} = 1 - (1 - w)^{X_{i,t}\Delta T} \tag{4.3}$$

Where $p_{i,t}$ is the probability of infection for individual i following any of its current contacts with infectious individuals during time interval ΔT. $X_{i,t}$ is the number of infectious contacts for individual i at time t. Note that the probability depends on both time and location of the individual i (i is tied to a household, school, or

workplace location). Now, to create a connection between the probability in 4.3 and disease parameters, let us define $\beta = -log(1-w)$ following [120]. By substituting β for w in equation 4.3, I can rewrite the probability as:

$$p_{i,t} = (1-e)^{-\beta X_{i,t} \Delta T} \tag{4.4}$$

In this equation, $\beta X_{i,t}$ is called the force of infection, which measures the risk of acquiring infection by individual i at time t and is usually represented by $\lambda_{i,t}$. β is a parameter which defines how transmissible a virus is. In this research, I use one single β value for all indoor environments and approximate the value of that by using the basic reproductive and recovery rates. As explained in section 4.2, in this research, a viral disease is characterized by the basic reproductive rate R_0, inactivation time α of a virus on surfaces, and recovery rate γ. The β value in the baseline population is approximated from the following threshold equation, which is derived from a homogeneous mixing condition [120].

$$R_0 = \beta^{HH}/\gamma \tag{4.5}$$

In this equation, γ is easier to measure because it does not depend on the environment of transmission (1/ number of days to recover). Thus, it is usually known. The R_0 value is determined in either of two ways. One is through epidemiological surveillance at the very beginning of an outbreak in a population (independently of the above equation). Such value is determined by measuring the number of secondary cases that arise from a single infectious individual in a completely susceptible population. The R_0 value constantly changes during the epidemic and may even not be the same in two populations at a time due to many virus, environment, and population related factors. Because of this condition, determining the accurate value of R_0 with surveillance is not easy after the disease starts to spread. Thus, in the second method,

the value of R_0 is derived by evaluating the largest eigenvalue in next generation matrix of the infection dynamics in a disease-free equilibrium [64]. Either way, the range values of R_0 has been reported in the literature for different viruses. Given the fact that ranges of γ and R_0 are available in the literature, we can use them in equation 4.5 to have an approximation of β value for a viral disease in the baseline population. By testing scenarios with range values of these parameters, I can simulate the spread of different viral diseases with in the baseline population.

I rewrite equation 4.4 with more representative notations as follows:

$$P^{HH}(I^H|S^H)_{i,t} = (1-e)^{-\beta^{HH} X_{i,t} \Delta T} \tag{4.6}$$

Where $P^{HH}(I^H|S^H)_{i,t}$ is the probability of transition from $S^H \rightarrow I^H$ for human individual i at time t, given contacts with other infectious humans, and β^{HH} is a parameter that determines the transmissibility of the virus in human populations.

4.5.2 Micromobility Fomite-Human (MH) Submodel

Using the same process, I derive a probability for transmission of a virus from micromobility fomites to human body. Note that this time an individual human i rides the exact number of $Y_{i,t}$ micromobility vehicles during time step ΔT.

$$P^{MH}(I^H|S^H)_{i,t} = (1-e)^{-\beta^{MH} Y_{i,t} \Delta T} \tag{4.7}$$

Where $P^{MH}(I^H|S^H)_{i,t}$ is the probability of transition from $S^H \rightarrow I^H$ for human individual i at time t given interaction with contaminated micromobility vehicles, and β^{MH} is a parameter that determines transmissibility of the virus from micromobility vehicle to human individuals.

4.5.3 Human-Micromobility Fomite (HM) Submodel

I use a similar strategy to calculate the probability of contamination of a micromobility vehicle b when $Z_{b,t}$ infectious human individuals ride the vehicle during time step ΔT.

$$P^{HM}(C^M|S^M)_{b,t} = (1-e)^{-\beta^{HM} Z_{b,t} \Delta T} \tag{4.8}$$

Where $P^{HM}(C^M|S^M)_{b,t}$ is the probability of transition from $S^M \rightarrow C^M$ for micromobility vehicle b at time t, given interaction with infectious human individuals, and β^{HM} is a parameter that determines transmissibility of the virus from human individuals to micromobility vehicles.

In order to determine the values of β^{HM} and β^{MH}, I define the relative transmission factor ρ which specifies the transmission rates β^{HM} and β^{MH} as a factor of β^{HH}. For simplicity's sake, I also assume these two rates have the same value.

$$\beta^{HM} = \beta^{MH} = (1/\rho)\beta^{HH}, \tag{4.9}$$

The rate of transmissions β^{HM} (from hands of an infectious person to bicycle surfaces) and β^{MH} (from contaminated bicycle surfaces to hands) are smaller than β^{HH} because the virus may have multiple path ways in other venues and people spend more time in the same proximity. Thus, ρ has a value larger than one. I conduct sensitivity analysis for different values of ρ to experiment with different possible scenarios.

Using equations 4.6, 4.7, and 4.8, we can calculate the transmission of virus during each time step ΔT in the baseline population and micromobility.

CHAPTER 5: A CASE STUDY: SPATIOTEMPORAL MODELING OF DISEASE SPREAD IN COOK COUNTY, ILLINOIS

In this chapter, I conduct a case study simulating an epidemic in Cook County in the presence of micromobility as a transportation mode to demonstrate the use and effectiveness of the proposed agent-based modeling framework of disease spread.

5.1 Study Area

Cook County in the state of Illinois has a 2010 estimated population of 5,194,675, among which 2,746,388 live in the City of Chicago, part of the third largest metropolitan area in the United States[1]. It is highly connected to other large cities both nationally and internationally, making it more vulnerable to the emergence and spread of a disease. The city bikesharing system, Divvy, has been in place and growing since 2013, and several previous studies have been conducted on it. Thus, the Divvy bikesharing is used as the micromobility service, and bicycles are considered the micromobility vehicles in this case study. The trip data of Divvy are available for free and includes some demographic information about bicycle users. These properties make Cook County and Chicago a suitable region for this case study. While, the shared bike stations are distributed in the Chicago City, I model the baseline population for the entire Cook County (Figure 5.1).

[1]United States Census Bureau: QuickFacts.

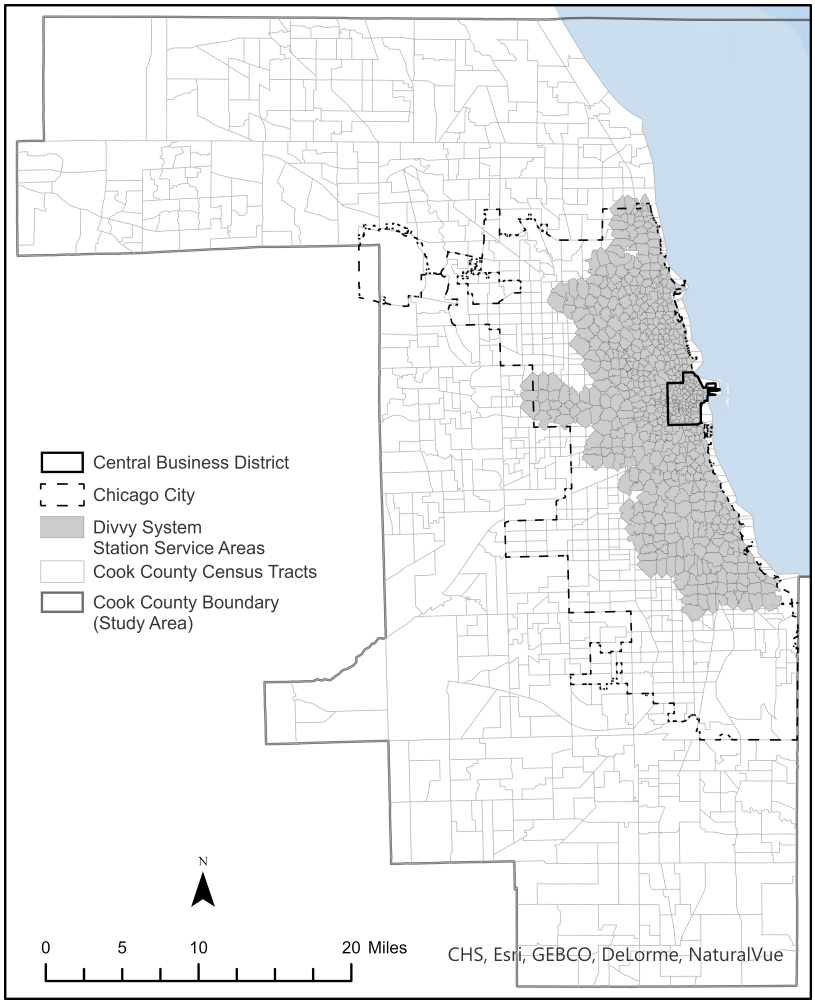


Figure 5.1: Study area.

5.2 Modeling Framework

Figure 5.2 represents a diagram of modules in the model. As explained in section 4.2, a network of individual contacts is needed to calculate the probability of infection for each individual in the baseline population (HH) at the end of each time step. I create a synthetic population based on census data and define household and work/school locations of the individuals using the origin-destination employment statistics (LODES), and school and educational institutions (ORNL education). Using this synthetic population, a human contact network is created which is used for the purpose of simulating the disease spread in the baseline population. Human in-

dividuals follow their daily schedule to go to work or school or stay at home. During the day, they may use Divvy public bikesharing as part of their trips. The human contact network is dynamic in the sense that, during weekends and holidays, individuals stay at home. Thus, only family contacts are present in the contact network during these days. Finally, the interaction among people in the indoor environments and with bicycles facilitates the spread of viruses and, as a result, disease, through population, which is modeled by the proposed SIR-SC model. In what follows, I explain the implementation of each module in more detail.

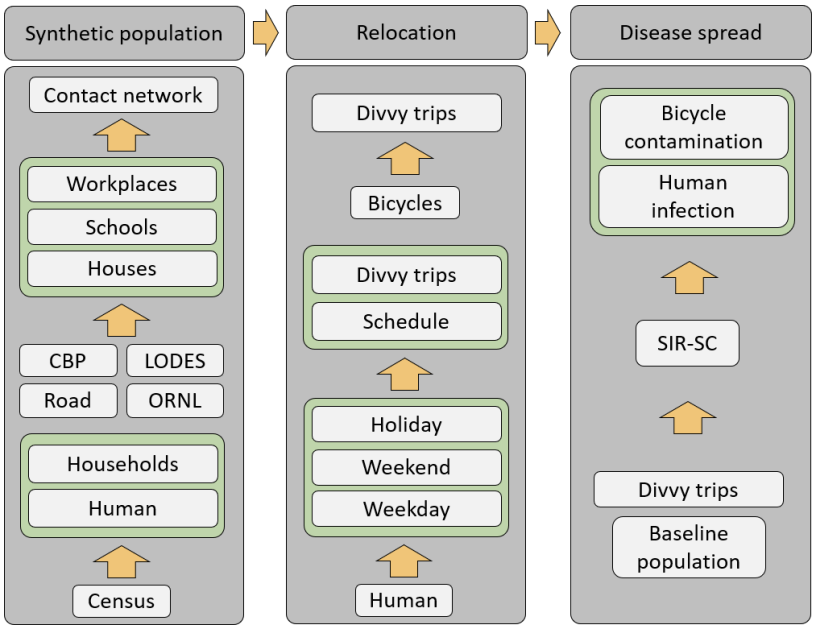


Figure 5.2: Modeling framework for Cook County.

5.3 Data

5.3.1 Micromobility

I use historical micromobility data from the Divvy bikesharing system in Chicago[2] for June and July 2018. Divvy bikesharing is a station-based system. This means that all the micromobility trips originate and finish at specific locations (stations). The data set has two main tables: bikesharing stations and trips. Each station has a

[2]https://www.divvybikes.com/system-data

name, a geographic location, and capacity. The trip table keeps a record of all trips. Every trip has a bike ID associated with it, a start and end date and time, start and end station, gender, and birth year. The urban space is divided into areal units based on the service areas defined according to the travel distance on the road network. These areas are used to assign trips to human individuals currently present in the service area and to represent the current location of bicycles.

5.3.2 Workplaces

I use two sets of data to assign workplaces to each individual in work-age. County Level Business Establishment Counts (CBP)[3] and Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES)[4]. These datasets are used initially to determine the number of businesses in each census tract. CBP provides subnational economic data by industry (6-digit NAICS industry codes) each year. It includes the number of establishments by different employment size at the county level. LODES is provided by the Census Bureau and contains the residence and workplace flow between census blocks. Two sets of origin and destination tables are available at the census block level. One includes jobs with both workplace and residence in the state, and the other has jobs with the workplace in the state and the residence outside of the state. The data include residence and census block codes, the total number of jobs, and the total number of jobs by age, income level, and type. LODES is also used to assign individuals to workplace locations (see appendix A for the details).

The employment data of LODES are from three different sources: Unemployment insurance (UI) wage records reported by employers and maintained by each state, the Office of Personnel Management (OPM), and the Quarterly Census for Employment and Wages (QCEW), which provides information on the firm structure and establish-

[3]https://www.census.gov/programs-surveys/cbp.html
[4]https://lehd.ces.census.gov/data/

ment locations. There are some data quality considerations concerning LODES data. For example, LODES does not include all workers covered by OPM. That means trips that are made by these individuals are missing in the origin-destination statistics and our mobility model underestimates the number of individuals who commute between origin-destinations. LODES data of 2018 is used for our study.

5.3.3 Schools

I use school and daycare locations from the United States Environmental Protection Agency (EPA), Office of Environmental Information[5]. The data include the grade and enrollment at each institution, which I use to assign school age individuals to school locations. I use the most recent update of the file, which is for 2015. Finally, the Cook County road network of 2018 from the Census Bureau TIGER/Line Shapefiles data set[6] is used to locate household and workplace locations.

5.4 Synthetic Population Module

I use the methodology proposed by [76] to create a baseline population for Cook County, which includes social ties among individuals based on their family, work, and school membership (a detailed description is available in Appendix A). The process includes four steps:

- Creating home, work, and school environments across geographic space,
- Generating and assigning individual agents to households,
- Assigning work and school locations to each individual,
- Creating a network of social contacts based on membership in family, school, and, workplace social ties.

Home and work locations are created along the road network in Cook County. For each census tract, the number of occupied houses are extracted from the census data

[5]https://geodata.epa.gov/arcgis/rest/services/OEI
[6]https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html

and located 50m apart or on top of each other on local roads. The number of work locations in each census tract is derived from the county-level business establishment counts after disaggregation by population size. Then, they are placed on secondary roads 20m apart or on intersections of the local roads. School locations are extracted from the $ORNL\_education$.

Eleven household types together with group quarters are identified in census data and are simulated. All Cook County individual residents are simulated using Census tables. These individuals are assigned to households based on their age-gender groups and a method called synthetic reconstruction (SR) [121]. SR assigns samples derived from joint distribution of relevant attributes (here, age and gender) to population based on an iterative proportional fitting (IPF) method [74]. Finally, households are assigned to the house locations within each census tract. Because the process of fitting population to households is stochastic and it is based on specific attributes, one needs to assess the accuracy of the fitting by aggregating population attributes and compare it with other actual census attributes. After placing individuals into households, I aggregate population to reconstruct four variables namely, the average family size, the average household size, the number of households with children under 18, and the number of households with individuals over 65 years old. Then, I compare these synthesized values with their actual value for the census tracts using the percentage error (E^{syn}) with the following equation:

$$E^{syn} = \frac{(Synthesized - Actual) * 100}{Actual} \tag{5.1}$$

Individuals of school age are assigned to the nearest school based on their age and school enrollment and those of work age are randomly assigned to the work locations based on the origin-destination employment statistics (LODES), either in their residential geographic unit or another geographic unit. I assume the work/school location of a single individual remains the same for the duration of the model execution.

Finally, a large scale social network is generated based on the contacts at residences, schools, and workplaces. We do not expect an individual to be connected to all other individuals at school, workplace, or household/group quarter. Previous research revealed small-world properties for human contact network of infectious disease [122]. When the number of individuals in one of these groups is larger than five, a small-world network is created [123]. Otherwise, a complete graph is created. I assume this contact network stays the same for the workdays of the week. For weekends and holidays, however, I assume the work and school connections are removed.

5.5 Mobility Module

I divide transportation modes into micromobility and other modes of transportation (including personal vehicles and walking). A human agent may use micromobility as part of his/her trip to the desired locations. The mobility of population, however, is calculated every hour to take into account commute of individuals to work/school. Every hour, trips originating from a specific station are randomly assigned to the individuals currently available in the service area based on the trip and individual attributes (riding). When no individual with the matching age and gender as a trip is available in the service area of a station, a individual with appropriate age and gender is assigned to the trip from the entire population via a random process. An individual may use any type of transportation to commute to work or school, including micromobility vehicles.

5.6 Disease Transmission Module

I use the SIR-SC model proposed in section 4.3.7 to calculate the transmission of disease within urban areas. As the simulation unfolds, individuals follow their activity schedule to go to work/school or stay home. If they are infected, they may transmit virus on the surfaces on micromobility vehicles or infect other individuals in the population.

In the SIR-SC model, time is discrete and the spread of the disease is calculated at the end of each day (24 hours). The contamination probability of a bicycle is calculated based on interactions with individuals during a day, and infection of an individual is calculated based on the interaction with bicycles and other individuals in the population all at once at the last hour of the day. That means new infections and contamination during a day have no impact on each other. If a bicycle is contaminated, the start time of contamination is stochastically assigned based on the distribution of the number of trips at weekdays and weekends for the next day. Then, the bicycle remains contaminated for the inactivation time (the amount of time a virus can survive on a surface) of the virus, or until it is disinfected. This approach allows us to experiment with virus inactivation times less than a day. Initialization parameters are varied based on our scenarios (sections 5.8.4-5.8.8).

5.7 Programming And Software

The synthetic population is implemented by adapting codes from [76] for Cook County. All other codes are in Python. Graph-tool[7], and Networkx[8] packages are used for the manipulation of graphs. The core data structures and algorithms in graph-tool are implemented in C++, making it comparable to a pure C/C++ library (both in memory usage and computation time). Thus, graph-tool is specially used for parallel processing tasks. Model and scenarios are implemented on Orion, a general use Slurm partition on a Redhat Linux based high-performance computing environment. Finally, ArcGIS Pro 2.8 is used for network analysis and some of the visualizations.

Pseudo-code for the sequential version of the SIR-SC model is available in figure 5.3.

[7]https://graph-tool.skewed.de/
[8]https://networkx.org/

Algorithm1: Sequential SIR-SC algorithm.

Input: $(G, F, A, \Lambda, \alpha, \gamma, \beta^{HH}, \beta^{MH}, \beta^{HM})$ where G is the human contact network, F is micromobility trips, A is intervention, Λ is initial conditions, α is inactivation rate, γ is recovery rate, and the last three variables are rate of transmission in control network, micromobility to human, and human to micromobility, respectively.

Output: C^M, I^{HM}, I^{HH} where C^M is the number of new contaminated micromobility vehicles per day, $I^{HM}$is the number of new infections in micromobility per day, and $I^{HH}$is the number of new cases in the control population per day.

Temporary variables: $health\ state\ and\ contaminattion\ state$ which are the current state of a human and micromobility vehicle, respectively.

initialization

for $t = 0$ **to** T **do**

  foreach $person\ v \in V$ **do**

    update his/her location according to day of week and time of day.

  foreach $trip\ f \in F(t)$ **do**

    randomly assign a nearby rider with the same attributes to $trip\ f$ and update F with rider id.

  if $t == last\ hour\ of\ a\ day$ **then**

    update edges of G according to day of the week.

    foreach $rider\ i \in F(day)$ **do**

      get interventions A and apply

      calculate probability of infection for $rider\ i$ using β^{MH} and number of contaminated micromobility vehicles used by $rider\ i$.

    foreach $micromobility\ vehicle\ b \in F(day)$ **do**

      get interventions A and apply

      calculate probability of contamination for $micromobility\ vehicle\ b$ using β^{HM} and number of infectious individuals who used b.

    foreach $infectious\ node\ u \in V$ **do**

      foreach $contact\ v\ of\ node\ u$ **do**

        if $v\ is\ susceptible$ **then**

          calculate probability of infection for $individual\ v$ using β^{HH} and number of infectious contacts of v.

    foreach $infectious\ node\ u \in V$ **do**

      if $individual\ u$ passed γ

        update $health\ state\ u$ to $recovered$

    $update\ C^M, I^{HM}, I^{HH}$

    update $health\ state$ of new infections in micromobility and contact network to $infectious$ at the same time

  foreach $contaminated\ micromobility\ vehicle\ b$ **do**

    if duration of contamination $b > \alpha$ **then**

      update $contaminattion\ state$ of b to $clean$

$output(C^M, I^{HM}, I^{HH})$

Figure 5.3: Pseudo-code for the sequential version of the SIR-SC model.

5.8 Results

5.8.1 Micromobility Trips

I processed shared bike data of Divvy system for June and July 2018. Publicly available Divvy data have already been processed by the data providers to remove trips taken by staff and trips with duration under 60 seconds. The latter are potentially false starts or re-dock by users. The farthest trip in the data set is 35 km, which has happened in 2.5 hours. I assume trips are not valid when their duration is larger than five hours (2,605 trips). Trips larger than five hours are less likely to be valid especially for the purpose of this research.

A large number of trips exist for the period of study with a distance of 0 (26,973 trips). I keep these records in the dataset for analysis since their duration is higher than 60 seconds and they still represent an interaction between a cyclist and a bike, which may lead to spread of disease agents. I also calculate the speed of trips. It is less likely that speed is faster than 25 km/h, and I remove those trips from the database as well (14 trips). 308,359 records do not include the age of the cyclist. I use the age distribution by gender from the rest of the database to impute age values for these records. I use the same strategy to impute new age values for records with age 80 and over (2,126 trips) because the values are not logically valid. Age and gender are customer reported in the database.

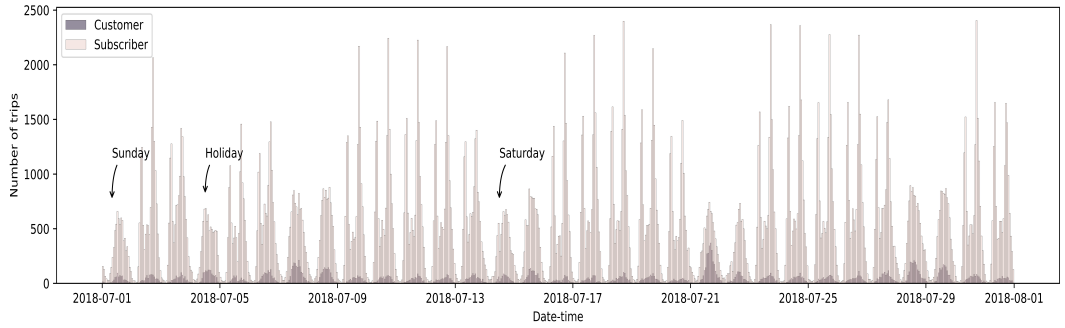
Figure 5.4 shows some of the statistics about the final data set after all corrections, with 1,209,808 trips for the period of study. 5,874 unique bikes exist within this data set. For visualization purposes, I demonstrate the hourly distribution of starting Divvy trips only for July 2018 in figure 5.4(a) and figure 5.4(b) Similar patterns are observed for weekends and holidays. Most of the shared bike users are subscribers to the system. There is no way to exactly determine the purpose of the trips from the data set. However, the hourly pattern of weekdays for subscribers implies that significant portion of these trips would be for commuting purposes. At the same

time, the similarity between hourly weekday patterns of trips by non-subscribers with weekend patterns (including subscribers and non-subscribers) may imply that most of these non-subscriber patterns are likely to be for non-commute purposes. Men use public bicycles more than women, but with the same hourly pattern.

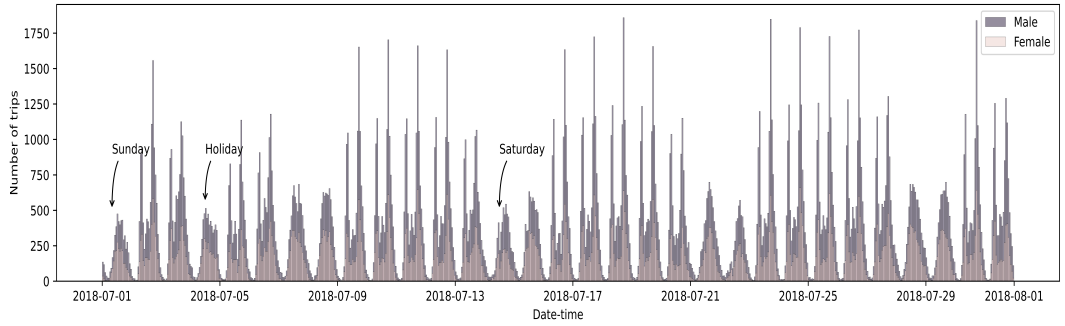
Figure 5.4.c shows the duration, length, and speed of the trips for the entire period of study. Average duration of trips by subscribers is less than customers (non-subscribers). Customers are more likely to use the bicycles for recreational purposes, which is expected to be longer than commuting trips by bicycles. The same inference can be made with regard to trip length. Also, there is no way to find out the types of trips (single or multimodal) or infer the origin and destination of an individualâs trip from the data set.

5.8.2 Micromobility Stations

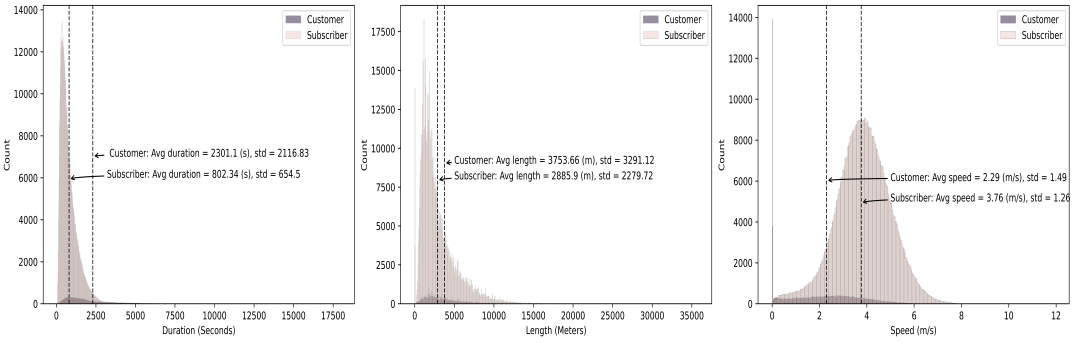
The station table contains 595 stations, among which 564 are either start station or end station in the trip data set during the period of study. If a station has no record during the time period of our study, I check if trips are available outside of that period within 2018. If the trips are available before and after, I keep the station in the micromobility stations table. Otherwise, I assume the station has been established after the period of study or stopped operation before the period of study. Then, these stations are removed from the station table. The Divvy system stations are rather systematically distributed across the city. The average Euclidean distance to the nearest station from each individual station is 586.38m with standard deviation of 244.26m. The service area of each station is calculated based on network distance and reported in figure 5.1.



(a)



(b)



(c)

Figure 5.4: Divvy shared bike system statistics during July 2018. (a) Hourly distribution of Divvy trips by type of users. (b) Hourly distribution of Divvy trips by gender. (c) Duration (left), length (center), and speed (right) of trips.

5.8.3 Synthetic Population

I use the 2010 decennial census to generate a synthetic population. Business establishment counts (CBP) data are at the county level. I disaggregated these business counts into the census tract level in proportion to the employees population size (see Appendix A for details). Alternatively, one can use other existing business points of interest (POIs) data sets such as Data Axle. These business locations were used as workplaces, which are placed along the secondary roads or intersections of local roads. I used the LODES origin-destination statistics after aggregation from blocks to census tracts to assign workplaces to each individual in work age. Figure 5.5 shows the origin and destination in- and out-flows for census tracts.

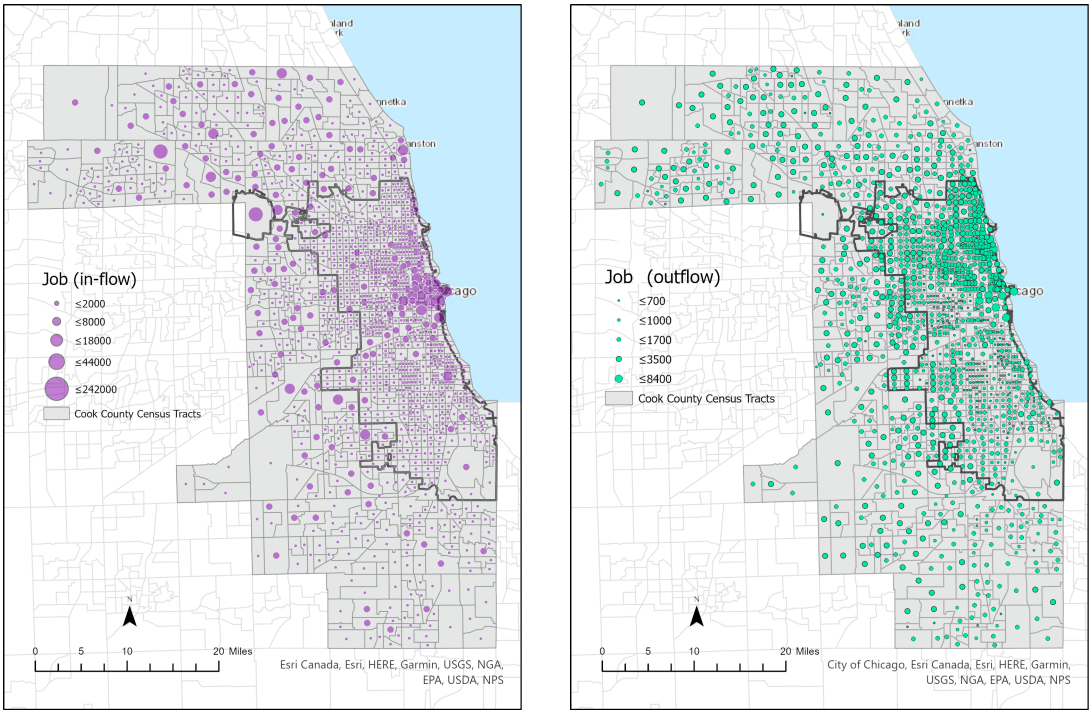
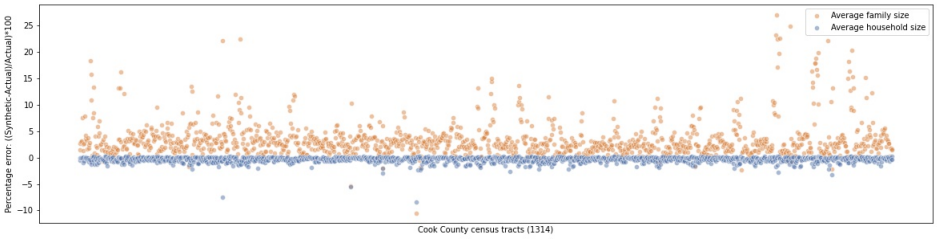


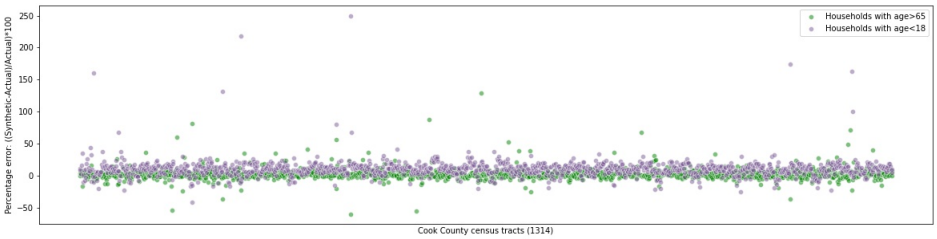
Figure 5.5: Origin-destination employment statistics.

Finally, the contact network was created by connecting individuals with their social ties at school, home, and work. The accuracy assessment of the synthetic population for four variables (average family size, average household size, number of households with children under 18, and number of households with individuals over 65 years old)

is shown in figure 5.6 by measuring percentage error (equation 5.1). The vertical axis shows the percentage errors of the synthesized attributes compared to the actual values for 1314 census tracts (1318 census tracts exist in 2010 Cook County census records, four of which have zero population). Each dot represents a census tract. Average family and household sizes are shows in figure 5.6(a) and the other two attributes are shown in 5.6(b). In general, the number of households with minors and seniors (b) show larger margins of error. A detailed description of the errors are available in table (5.1).



(a)



(b)

Figure 5.6: Accuracy assessment of the synthetic population. (a) Average family size and average household size. (b) Households with age>65 and households with age<18.

The percentage errors for the number of households with seniors ranges between -59.52 and 129.5, and the percentage errors for the number of households with minors ranges between -41.66 and 250. The larger standard deviation and outliers for these variables may be due to varying number of them in group quarters and households.

Table 5.1: Accuracy assessment of the synthetic population (based on percentage errors).

Statistic	Average family size	Average household size	Families with senior >65	Families with children <18
Count	1314	1314	1314	1314
Mean	3.40	-0.37	2.88	10.96
Std	3.38	0.58	9.58	15.35
min	-10.49	-8.35	-59.52	-41.66
Median	2.61	-0.20	1.93	10.70
Max	27.07	0.3	129.5	250

Two contact networks are created for weekdays (all edges) and weekends (household edges only), respectively. A sub-network of a weekday is represented in figure 5.7 for census tract 17031491400 for illustration purposes. Nodes are all individuals who live in this census tract plus the individuals that are direct work or school contacts of these residents from outside of this tract. The cone shaped light green edges in the large cluster (center), for example, are work contacts to locations outside of their tract). Six schools and daycare centers are in this census tract, which have created strong ties among households. Note how the households are creating a link between school and work contacts. A large number of connected components show households and group quarters that are not connected with other households in the census tract (see household types and group quarters in Appendix A). See how groups (households, work and school groups) with a small number of members (<=5) are all connected to each other. The workplace groups that are not all connected have larger than 5 members (those that are outside of the census tract are not shown for illustration purposes). At the same time groups with larger than five members are connected

using a small-world network mentioned earlier. 26 individuals are grouped in group quarters within this tract which stands out in the bottom-right of the figure. We assume all individuals that are categorized in group quarters are in a single group and assigned to a single place. Some individuals in single member households have no contacts at all, which is expected, given the assumptions of the model.

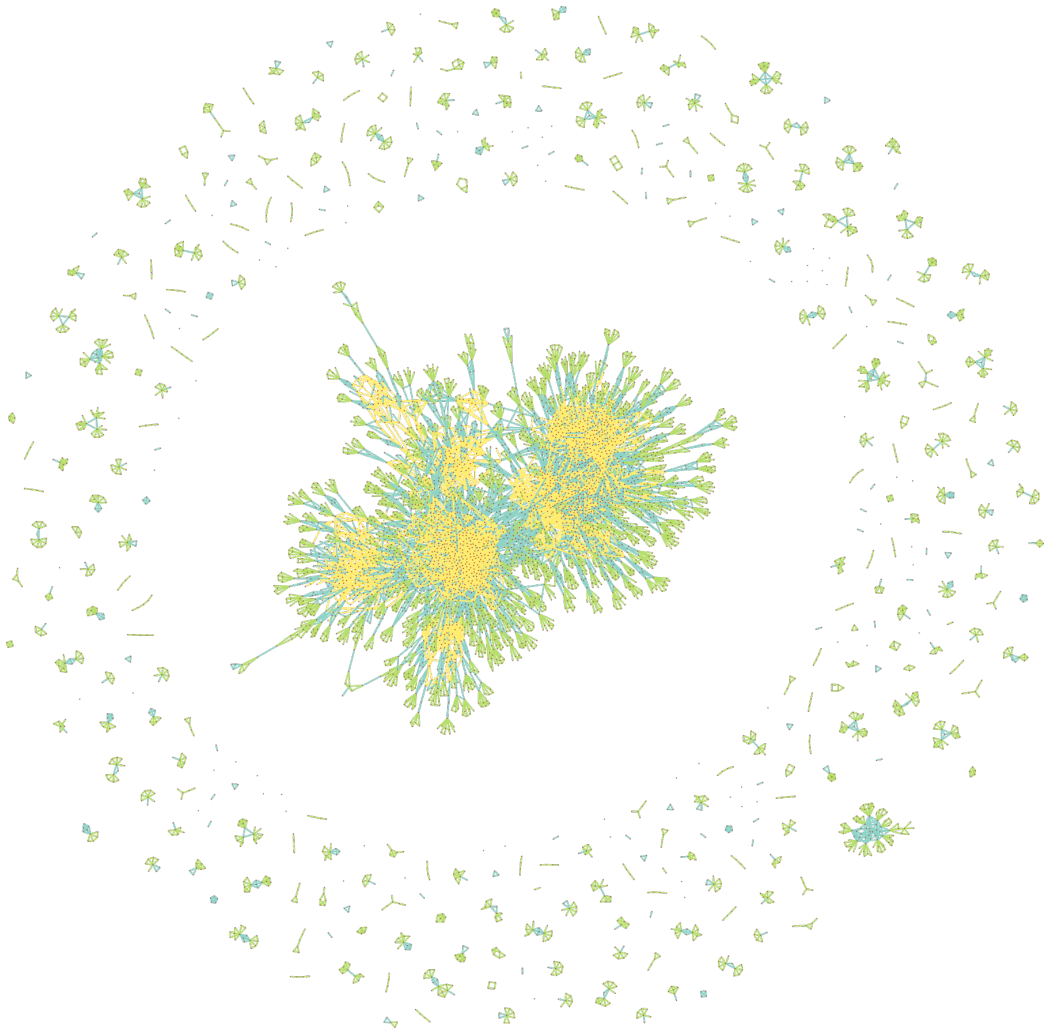


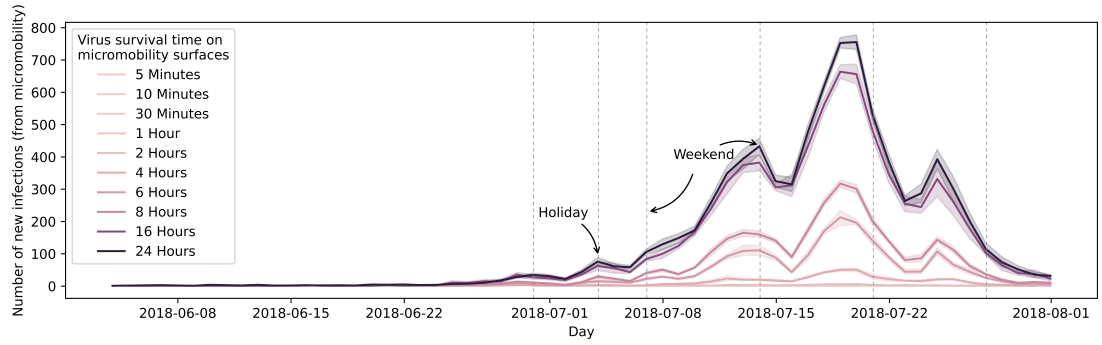
Figure 5.7: Census tract 17031491400 contact network. The vertices along the end of work edges (light green) are contacts outside of their census tract. 29% of the contacts are household, 25% are school, and 45% are work contacts (including the ones from outside).

5.8.4 Sensitivity Analysis and Parametrization

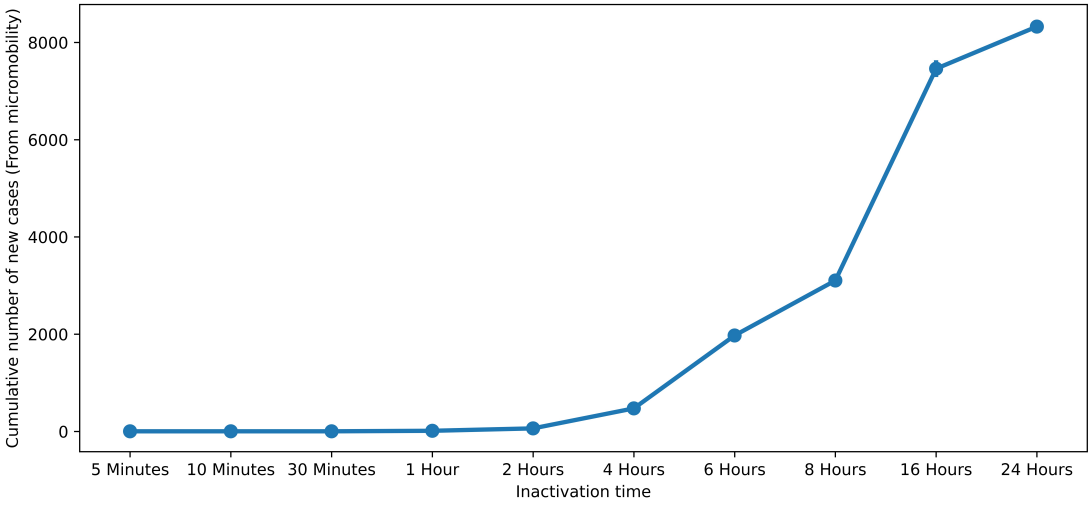
To study the sensitivity of the epidemic dynamics to the variations of the main parameters, I conduct sensitivity analysis. These parameters are virus inactivation time (α), number of bikes (B), basic reproductive rate (R_0), relative transmission factor ρ, probability of sanitizing by human individuals (P^{snt}), and fraction of stations with sanitizing stand (D). Depending on the experiment, I implement a different number of replications, ranging between 25-50. In general, a higher number of replications is recommended (at least 100). However, because I use the same trips (actual micromobility trips) during the study period and because most of the variables are not stochastic (random variables), the number of replications for the current study is still valid. All of the sensitivity analysis results are reported while answering research questions in the upcoming sections except for the virus inactivation time (α), which I present hereunder.

Virus inactivation time

It is expected that the dynamics of a viral disease epidemic in micromobility systems changes significantly by variations of virus inactivation time. Virus inactivation time is the amount of time a virus can survive on surfaces of micromobility vehicles. If the time period between two micromobility trips, made by a single vehicle, is more than the inactivation time of a virus, then there is no possible route of transmission between the individuals who ride that bicycle. Thus, in order to find out how much this parameter affects the disease spread in Divvy system, we conduct sensitivity analysis for this parameter by adjusting its values α =[5min, 10min, 30min, 1 hour, 2 hours, 4 hours, 6 hours, 8 hours, 16 hours, 24 hours]. Results are shows in Figure 5.8. These results show that when the inactivation time of a virus is short, even in full blown epidemic conditions, a very small number of individuals in the population are infected.



(a)



(b)

Figure 5.8: Virus inactivation time impact on disease spread in Divvy system. (a) The number of new infections per day in Divvy system. (b) The cumulative number of infected individuals for different inactivation times. Initial conditions: R_0=3, ρ=2, α=1-day, γ=0.2, I_0^H=100, I_0^M=0, rep=25.

5.8.5 Micromobility Impact on Viral Disease Emergence And Dynamics

In the first research question, I aim to find how surfaces on the new micromobility transportation systems contribute to the emergence and dynamics of viral epidemics in urban areas. In order to answer this question, I design two scenarios. The first scenario studies the emergence of a viral disease through micromobility in urban areas while the second scenario studies the dynamics of disease spread.

Emergence of viral disease through micromobility system

In the first experiment, I initialize the model with a single contaminated bicycle, which is located in a station in the central business district (CBD) of Chicago where the risk of infection is high due to high demand during the rush hours. Then, I count the number of times an outbreak occurs after 40 replications. I say an outbreak occurs when a person who has contracted the disease, out of using a contaminated bicycle, infects at least one other susceptible host in the population. Also, I conduct a multi-variable sensitivity analysis for R_0 (0.5, 1, 1.5, 3, 5) and ρ (2, 5, 10) to experiment with different types of viruses. In a deterministic homogeneous mixing condition, an outbreak can only occur with $R_0 > 1$. In network and agent based models, however, small outbreaks are possible for that condition because the probability of infection is calculated for each individual separately. Thus, I include R_0 values under one in the series of experiments.

The results are reported in Figure 5.9. Figure 5.9 shows the relationship between the basic reproductive rate (left) and the relative transmission factor (right) with the number of outbreaks. Outbreaks occur with R_0=1 only when ρ=2. As expected, the larger the basic reproductive rate, the larger the number of outbreaks. Both parameters show a quadratic relationship with the number of outbreaks. It is interesting to see that, even for small R_0 values (e.g., 1 and 1.5), outbreaks are likely to happen. Out of 40 replications, one outbreak has even occurred in a scenario with ρ=2 when the R_0=1.

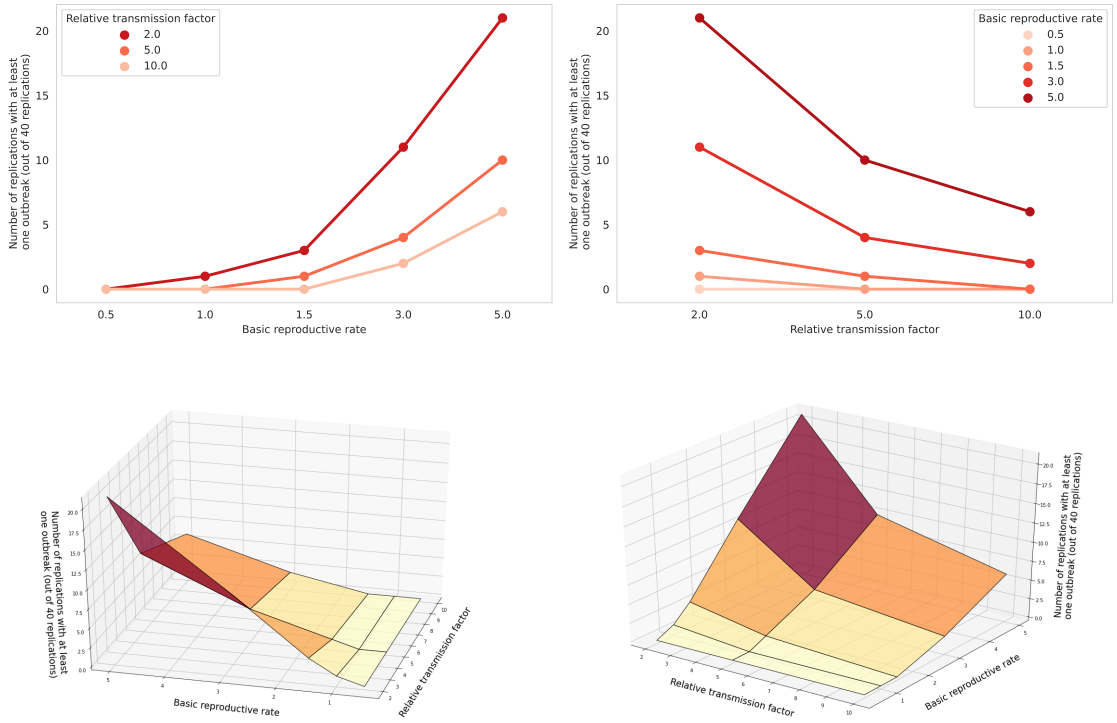


Figure 5.9: Emergence of outbreaks through micromobility. Initial conditions: R_0=(0.5, 1, 1.5, 3, 5), ρ=(2, 5, 10), α=1-day, γ=0.2, I_0^H=0, I_0^M=1, rep=40.

I need to highlight two important points here. First, We must be careful when interpreting these results. The basic reproductive rate that I use in the simulations as initial conditions provides a basic representation of a virus transmission ability. In other words, two viruses with the same R_0 may show quite different transmission behaviors based on their intrinsic properties (e.g., respiratory and gastrointestinal) and properties of the environment (frequency and type of the surfaces, temperature, humidity, type of venue). Also, I define β^{HM} and β^{MH} relative to the β^{HH} because I do not have their actual rates.

Second, the scenario I designed here is an extreme case, in which only one contaminated bicycle is placed in the City. In a city as large as Chicago and with the size of the Divvy system, it is likely there would be more than one contaminated bicycle at a time. Moreover, a positive feedback exists between contamination and infection processes, meaning new infections reinforce new bicycle contamination and vice versa.

Thus, the role of micromobility is not limited to the emergence of a disease and we need to study its role during the epidemic. Thus, I also design a second scenario where I study the impact of micromobility on dynamics of disease.

Micromobility impact on dynamics of viral disease spread

In order to study the impact of micromobility on dynamics of the viral disease, I design a worst case scenario with high basic reproductive rate (R_0=3) and low relative transmission factor (ρ=2). Going back to equation 4.9, the rate of transmission β^{HM} and β^{MH} are defined relative to the transmission rate β^{HH} using multiplication factor ρ. ρ has a value larger than one based on explanations provided in section 5.6. At the same time, the larger the value of ρ the lower the rate of transmission between individual hands and bicycle surfaces and vice versa. Thus, a value of 2 for parameter ρ corresponds to a case with high probability of virus transmission. That is why I call a condition with R_0=3, and ρ=2 a worst case scenario. Then, I measure the infected population P at the end of the epidemic and the amount of time it takes for a full blown epidemic to reach half its course T(mid). These two measures have been used to evaluate the impact of public transportation on epidemics [13]. Since my model has considered individual trips it is possible to directly measure the infections that occur in micromobility system. I implement the scenario with and without micromobility transportation. In this experiment, I also use a range of values for the number of available bicycles in the system to investigate the sensitivity of dynamics (micromobility impact) to the size of the micromobility system.

The Divvy data set has 5,874 unique bicycles during the period of the study. I use the current bicycles and trips in the data set to impute new bikes and trips. To create new bikes, I first calculate the number of times each bicycle in the system has been used during the period of study. The outcome is a distribution that shows the frequency of use for a bicycle in the system. Then, I randomly sample from this distribution to impute new bicycles and trips. New trips are imputed similar to

the trips associated with each sampled bicycle in the main data set. I design four experiments with 50%, 100%, 300%, and 500% increase in the number of bicycles, respectively.

Table 5.2: Comparison of population infected in the presence and absence of micromobility.

# bikes	# infected (micromobility)	Infected micromobility & baseline population (P)	ΔP	ΔP(%)
0	0	4,474,198	-	-
5,874	11,469	4,482,363	8,165	0.182
8,811	17,463	4,487,315	13,117	0.293
11,748	23,093	4,490,795	16,597	0.370
23,496	45,131	4,506,340	32,142	0.718
35,244	69,898	4,523,073	48,875	1.092

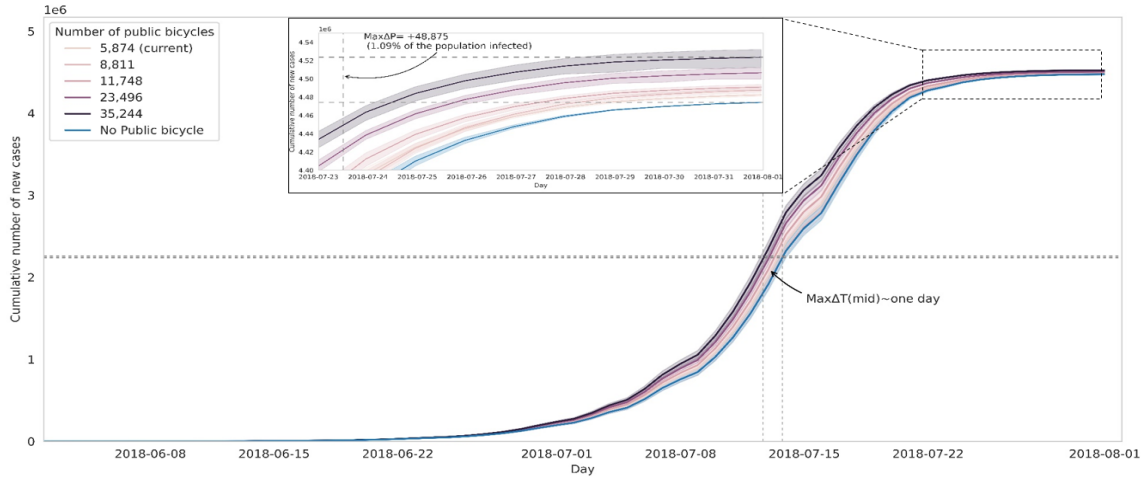
Figure 5.10 shows the cumulative number of new cases for all experiments in this scenario. The blue curve is the dynamics of new cases for the experiment without micromobility transportation, while the black curve represents the dynamics for the experiment with the largest number of bicycles and trips. Compared to the experiment without micromobility, the epidemic reaches its half course about one day earlier (ΔT(mid)=one day).

Table 5.2 shows the infected population and their percentage increase compared to the controlled experiment. The maximum increase in the infected population (ΔP) is 48,875 new cases which is about one percent increase in the extent of the overall infected population. For the Divvy system with current number of bicycles and trips, 8,165 (0.182%) new infections are added by the end of the epidemic. While these results show that the impact of micromobility in the spread of the disease is

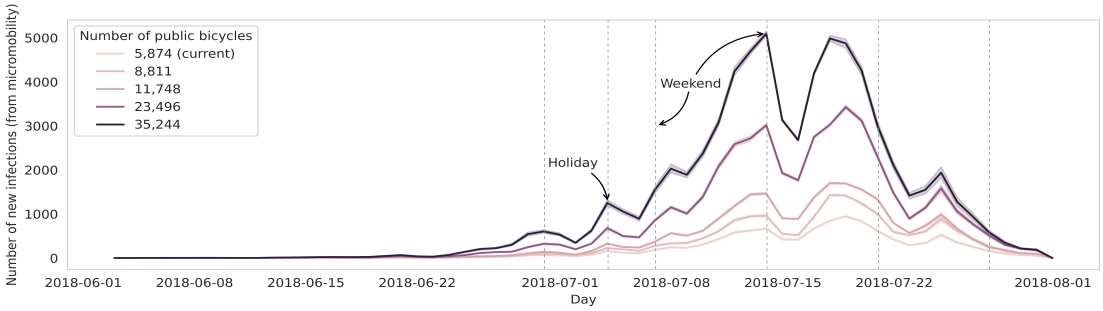
rather small, they are underestimated. Part of the population who would be infected in the baseline population when the micromobility is absent, may now be infected in micromobility. The second column in table 5.2 shows the number of infected individuals in the micromobility system at the end of the epidemic (also see figure 5.10(b)). Comparing these numbers with ΔP (column 4) reveals that the number of infected individuals are higher by 33-43% . This finding is important because it shows that ΔP is not sufficient to measure the role of public transportation systems and an individual based model such as the one presented here is needed to reveal the actual impact of transportation systems.

Moreover, one should notice that new cases from interaction with bicycles can cause secondary new infections in the population. Also, in the Cook County contact network, 332,610 individuals have no contact. They represent families in the census who are single individuals, are not working nor they are going to school. In the absence of micromobility, these individuals will not be infected at all since no connections exist between them and other individuals in the population. In the presence of micromobility, however, this part of population may be impacted by using contaminated bicycles.

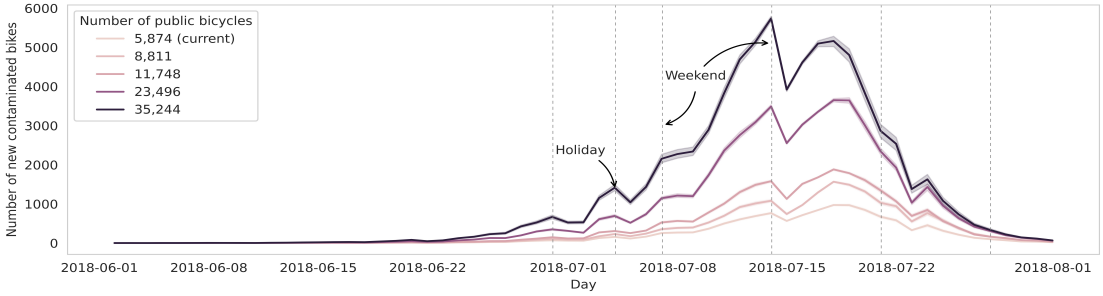
In conclusion, and based on our simulations, the emergence of viral disease through micromobility transportation in Cook County is possible, but the overall impact of the system on the disease dynamics in a worst case scenario, especially with the current size, is rather small (number of infections in micromobility= 1,469 or 0.26% of all infections, ΔT(mid)=less than an hour), but underestimated with the current measures by 43%.



(a)



(b)



(c)

Figure 5.10: (a) Micromobility impact on disease dynamics in terms of infected people P and the time it takes for epidemic to reach half its course T(mid). (b) Daily new infections in micromobility. (c) Daily new bike contamination. Initial conditions:R_0=3, ρ=2, α=1-day, γ=0.2, I_0^H=100, I_0^M=0, rep=25.

5.8.6 Risk of Exposure to Viral Disease in Micromobility

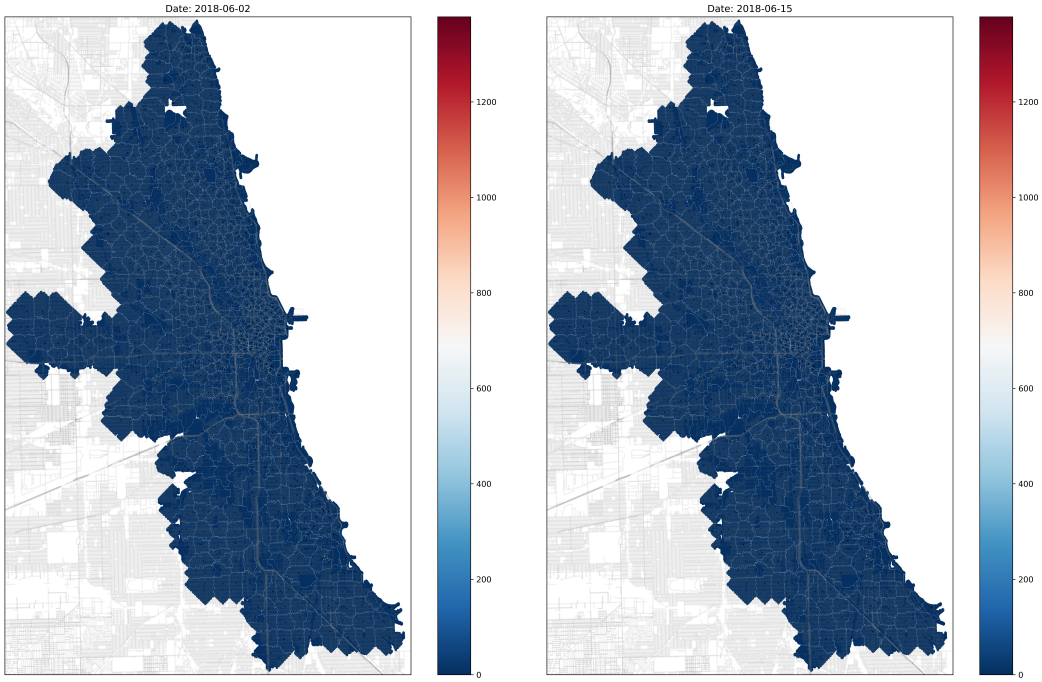
In the second question, I set out to find out how geographic space and time are organized concerning the risk of exposure to a viral disease out of using micromobility vehicles. I design two experiments to answer this question. In the first experiment, I study the spatial pattern of exposure risk to viral contamination in micromobility vehicles. In the second scenario, I aim to find out how the start time of the epidemic impacts the dynamics of the epidemic with respect to individuals who are infected by using bicycles.

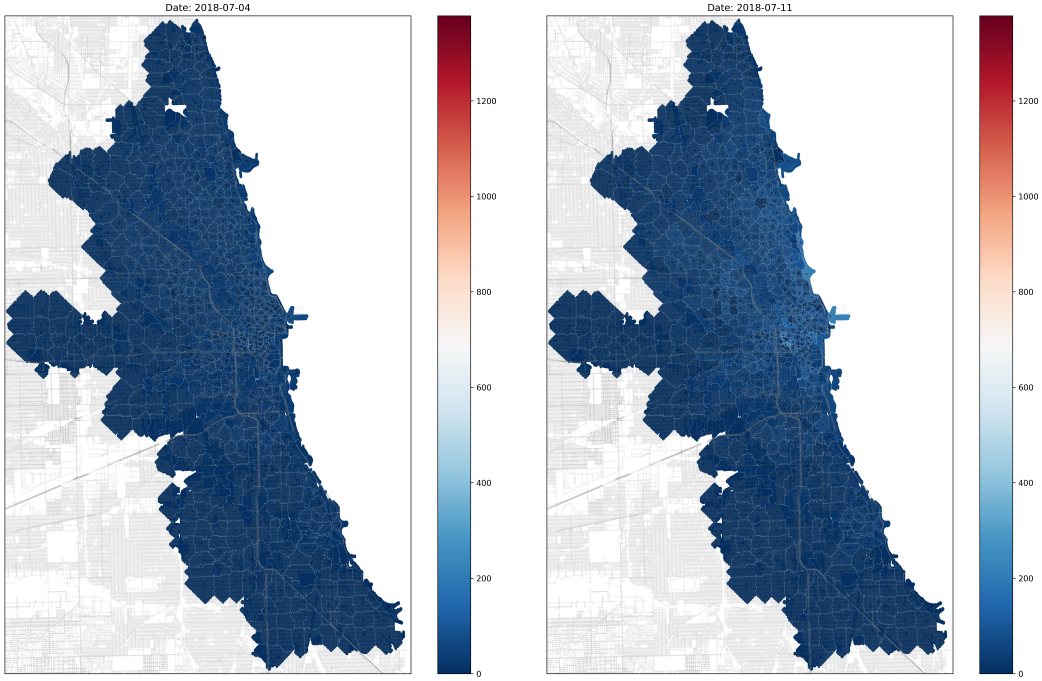
Risk of exposure in urban space

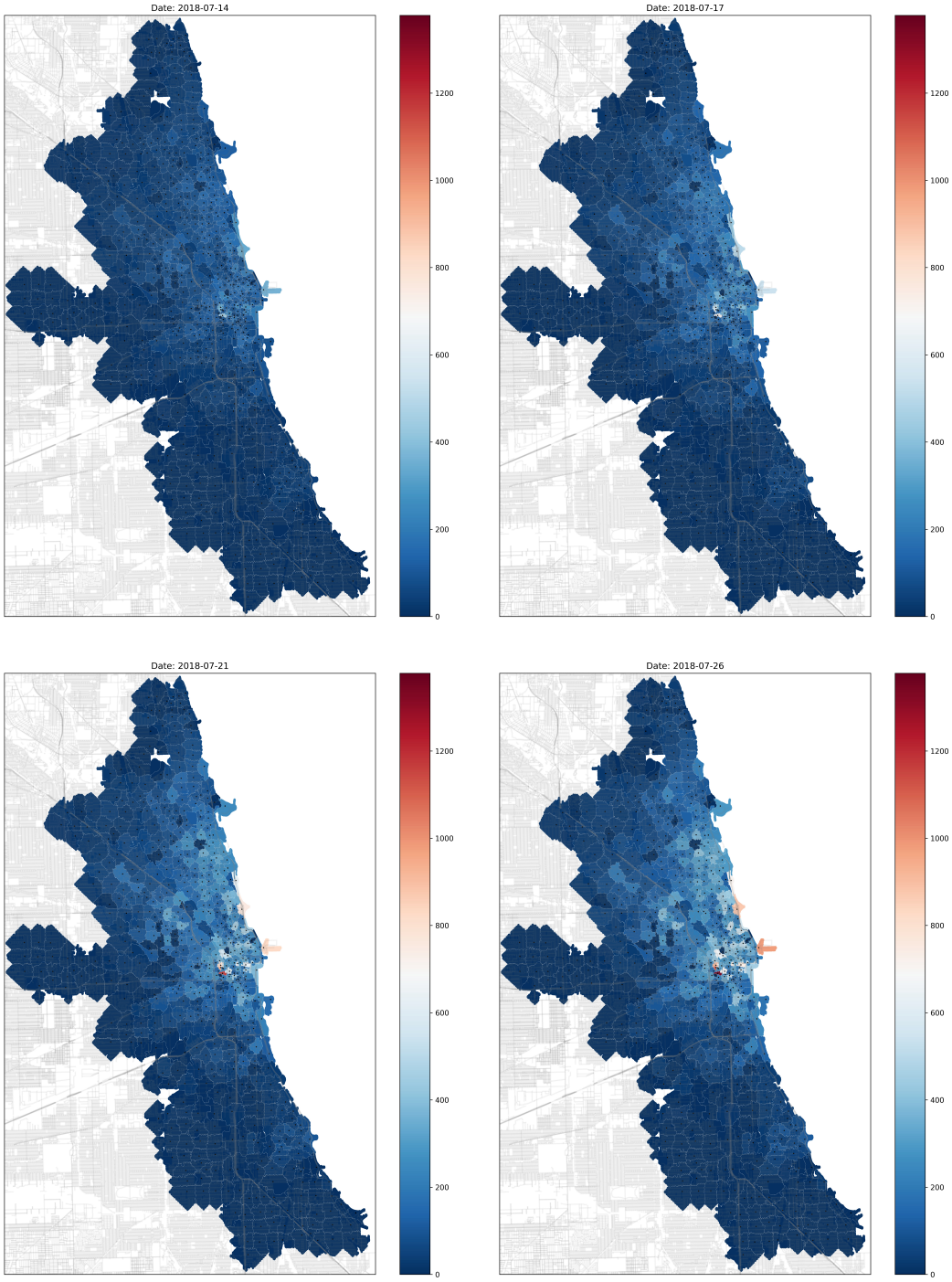
To study the spatial pattern of exposure, I track the average cumulative number of contaminated bikes docked to each station over the course of an epidemic (2 months) for 50 replications. The epidemic starts with 100 randomly selected infectious individuals, R0=3, ρ=2, γ=1/5, and survival time of one day. Then, I rank the stations based on the cumulative number of contaminated bikes. Figure 5.11 shows the spatial and temporal progress of the epidemic in terms of the cumulative number of contaminated bikes visiting each station during each time step. Zones are the service areas of the shared bike stations. For station based micromobility systems, the location of bicycles are recorded only when they are docked to a station. Thus, the color coded values are the number of contaminated bicycles.

At the end of the epidemic, a number of stations remain not impacted, meaning that no contaminated bicycles were docked to these stations during the epidemic. These stations have either a small number of trips (green service areas), or have no trips, which means they are out of service for the period of study (gray service areas). The Global Moran's I index for the spatial pattern of the last day of epidemic shows a highly spatially clustered pattern (0.56), as I expect stations close to each other to have similar exposure risks (last row, right).

The spatially local clusters are demonstrated on the last row (right) of figure 5.11. The spatial pattern on the last day of the epidemic (2018-07-31) shows that the risk of exposure is higher in the central business district and northern regions of CBD where most of the shared bike transportation occurs. Apart from stations that are out of service, it is interesting to see a number of stations with low risk (blue clusters) in the CBD, which is because not many bicycles are docked to these stations. I defined a neighborhood for each station based on the inverse distance method with a 1500 meter threshold (Divvy stations are 500m apart in average) to identify the local clusters. The threshold value can be defined based on the intervention application. For example, later in intervention scenarios I use the high-high clusters of this pattern to apply systematic interventions. One may use different thresholds based on the resources available for the intervention.


Date: 2018-06-02
Date: 2018-06-15
1200
1000
800
600
400
200
0


Date: 2018-07-04
Date: 2018-07-11
1200
1000
800
600
400
200
0


Date: 2018-07-14
Date: 2018-07-17
Date: 2018-07-21
Date: 2018-07-26
1200
1000
800
600
400
200
0

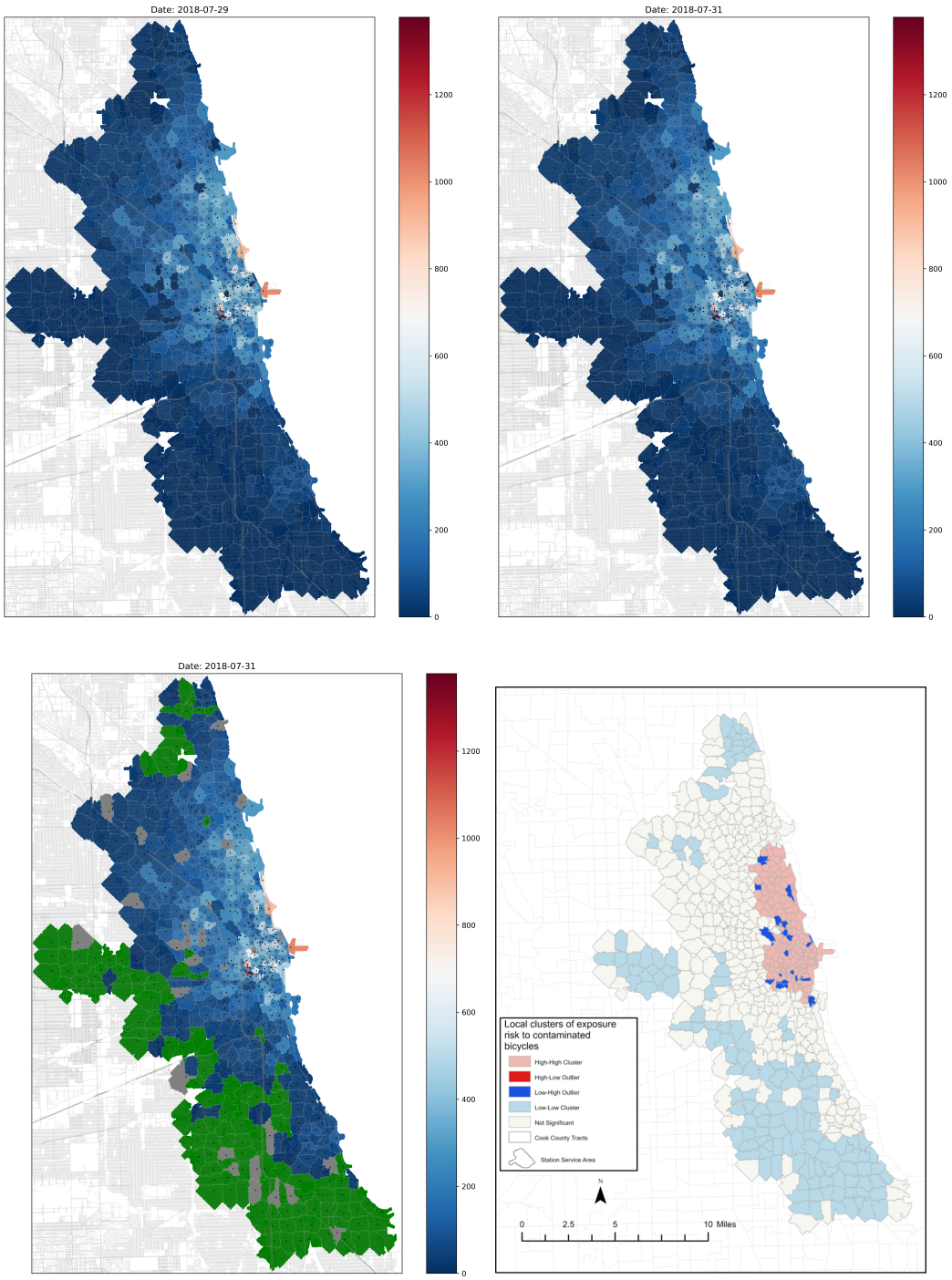


Figure 5.11: Spatial and temporal patterns of exposure risk to contaminated bicycles. The service areas in green and gray colors (last row) have no exposure to contaminated bicycles by the end of epidemic. Stations associated with green service areas have a small number of trips, while stations in gray areas are out of service for the period of study. Local clusters of high exposure areas are concentrated around downtown and the central business district (last row, right). Initial conditions: R_0=3, ρ=2, α=1-day, γ=0.2, I_0^H=100, I_0^M=0, rep=50.

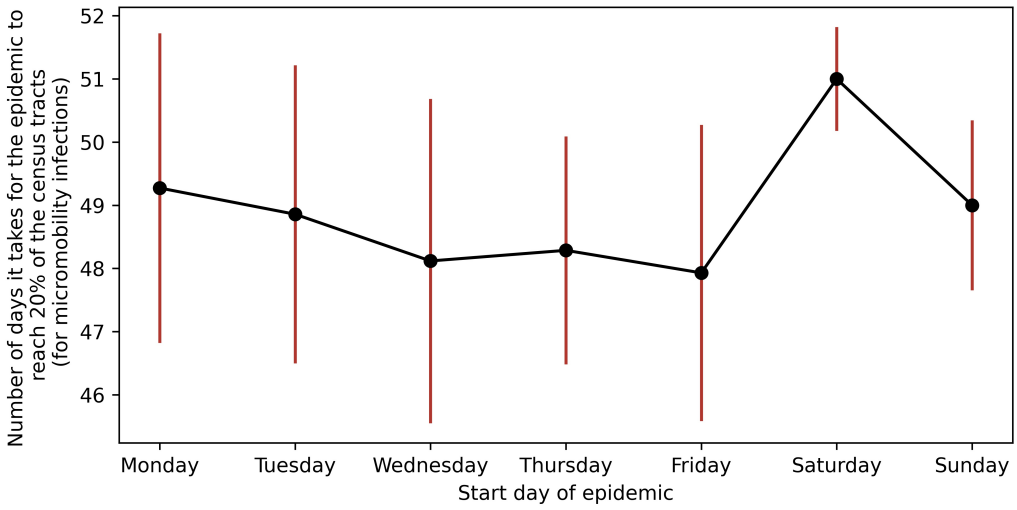
Start day of epidemic impacts dynamics of disease

To determine whether the start day has an impact on dynamics of disease spread in micromobility, I initialize the epidemic with 10 contaminated bikes on different days of a week. All these bicycles start a trip from a station in the central business district. Then, I record the time it takes until 20% of the census tracts in Cook County are reached by the disease through people who are infected using a micromobility vehicle.

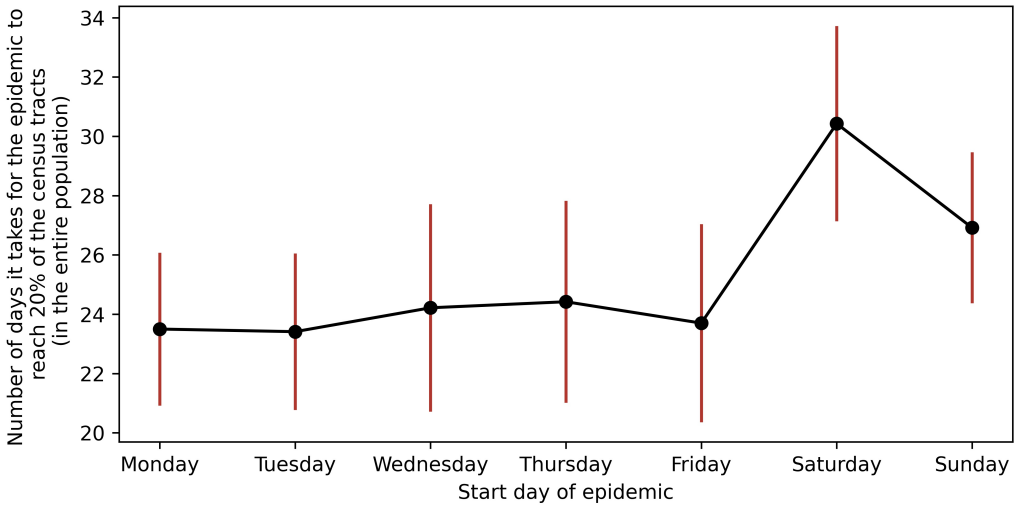
Results are shown in figure 5.12. According to these results, it takes about 48 days on average until 20% of the census tracts in the city are reached by the disease through people who are infected using a micromobility vehicle when the epidemic starts on Wednesday, Thursday, or Friday. This time is increased to 51 days for an epidemic that starts on Saturday (Figure 5.12(b)). That means if an epidemic starts on Saturday, it takes more time for it to reach 20% of the census tracts compared to the time to reach that same benchmark if it starts on Friday. There is a possible explanation for that based on both Divvy trips and baseline population relocation with respect to the schedule of individuals. During weekends, number of trips tend to be less than weekends. At the same time, only household contacts are present in the baseline population. During weekdays, however, number of trips tend to be larger compared to weekend. Moreover, the contacts in the baseline population include all household, workplace, and school contacts. Thus, in general during the weekdays, the disease has more opportunity to spread faster and reach the census tracts.

The same pattern of differences based on the starting day of the infection is observed if I monitor the time it takes for the epidemic to reach 20% of the census tracts (including infections outside of micromobility), but in a shorter period of time (Figure 5.12(b)). The slight difference between the two patterns for monday, tuesday and Wednesday is due to stochasticity in bikesharing system trips patterns. The baseline population is assumed to have the same weekday pattern for all days. These results confirm the importance of human mobility and social distancing in dynamics

of epidemics.



(a)



(b)

Figure 5.12: Start day of exposure impacts the dynamics of the spread of disease through both micromobility (a) and baseline population (b). Initial conditions: R_0=3, ρ=2, α=1-day, γ=0.2, I_0^H=0, I_0^M=10, rep=25.

5.8.7 Interventions and Control

Last but not least, with my last research question, I aim to find out how intervention methods and strategies, including random or systematic interventions, affect controlling the spread of infectious diseases through micromobility vehicles. Thus, I design two scenarios. In the first scenario, I apply a random daily disinfection of bicycles by maintenance staff. In the second scenario, I place hand sanitizing stands at micromobility stations either randomly and systematically across the city. Then, I record the impact on the number of people who are infected in the micromobility system by the end of the epidemic.

Disinfection of bicycles

For this scenario, I assume Q=5%, 15%, 25%, and 50% bicycles are randomly disinfected by maintenance staff each day in a worst case scenario. The epidemic starts with 100 infectious individuals, while $R_0 = 3$ and $\rho = 2$. Figure 5.13 shows the number of new infections from micromobility vehicles. Results show that, until a significant number of bicycles are cleaned (50%), the infection in the system remains high. Disinfection of a large number of bicycles is costly, and may not be practical for micromobility service providers.

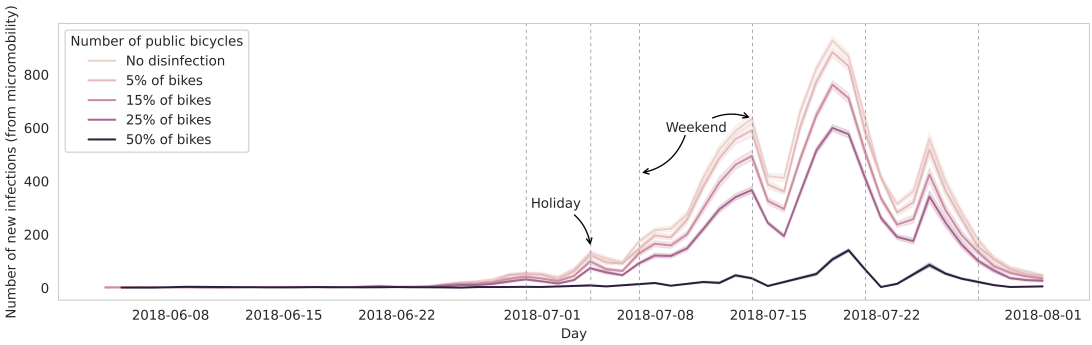


Figure 5.13: Random disinfection of bikes reduces the number of new infections in the micromobility system. Initial conditions: R_0=3, ρ=2, α=1-day, γ=0.2, I_0^H=100, I_0^M=0, rep=25.

Placing sanitizing stands at the stations

In this scenario, I assume that a fraction D of stations are randomly or systematically equipped with sanitizing stands nearby. Also, I assume people sanitize their hands with a probability P^{snt} after using a bicycle. In the experiment with random strategy, I place D=(0%, 25%, 50%, 75%, and 100%) sanitizing stands at stations randomly. When systematically placing sanitizing stands, I use the high-high clusters of the exposure risk in section 5.8.7 as the target stations (25.7% of the entire stations). Then, I place hand sanitizing at D=(0%, 25%, 50%, 75%, and 100%) of these stations, which corresponds to D=(0%, 6.4%, 12.9%, 19.3%, 25.7%) of all the stations. That means, in the second scenario, I do not apply any intervention on part of the stations, those with low level of risk. Then, for both experiments, I measure the cumulative number of people infected in the shared bike system at the end of a worst case epidemic.

The results are demonstrated in figure 5.14. Each color represents the experiment with a specific probability of sanitizing by cyclists. First, the relationship between the percentage of stations with sanitizing and the number of infections is quadratic in the case of systematic placement (first row, left) and linear in the case of random (first row, right) distribution. The reason is that in the first case stations are also ranked by risk of exposure. Second, the success of intervention highly depends on both human behavior and on the presence of sanitizing equipment. If 50% of the individuals use hand sanitizing, even when all of the stations have sanitizing stands, one can only reduce the number of infections by 50% (first row, right). If more people use hand sanitizing, however, the number of required sanitizing stands may reduce to 60% (when 75% of people sanitize their hands) or less that 50% (when all people sanitize their hands). Third, imagine the goal is to reduce infection in the system by 50% and assume that 75% of the individuals are likely to sanitize their hands after using a bicycle. Then, in the random experiment, about 65% of the stations must be

equipped with hand sanitizing stands (first row, right). In the systematic strategy, however, only 25.7% of the stations must be equipped with sanitizing stand (first row, right). Finally, by changing the distance threshold when identifying the local clusters, one can control the number of target stations for the second experiment. Other intervention scenarios are possible to test. For example, what if people wipe the handlebars of bicycles before cycling, or how many stations should stop operation to make sure no infection occurs in the system (flattening the curve).

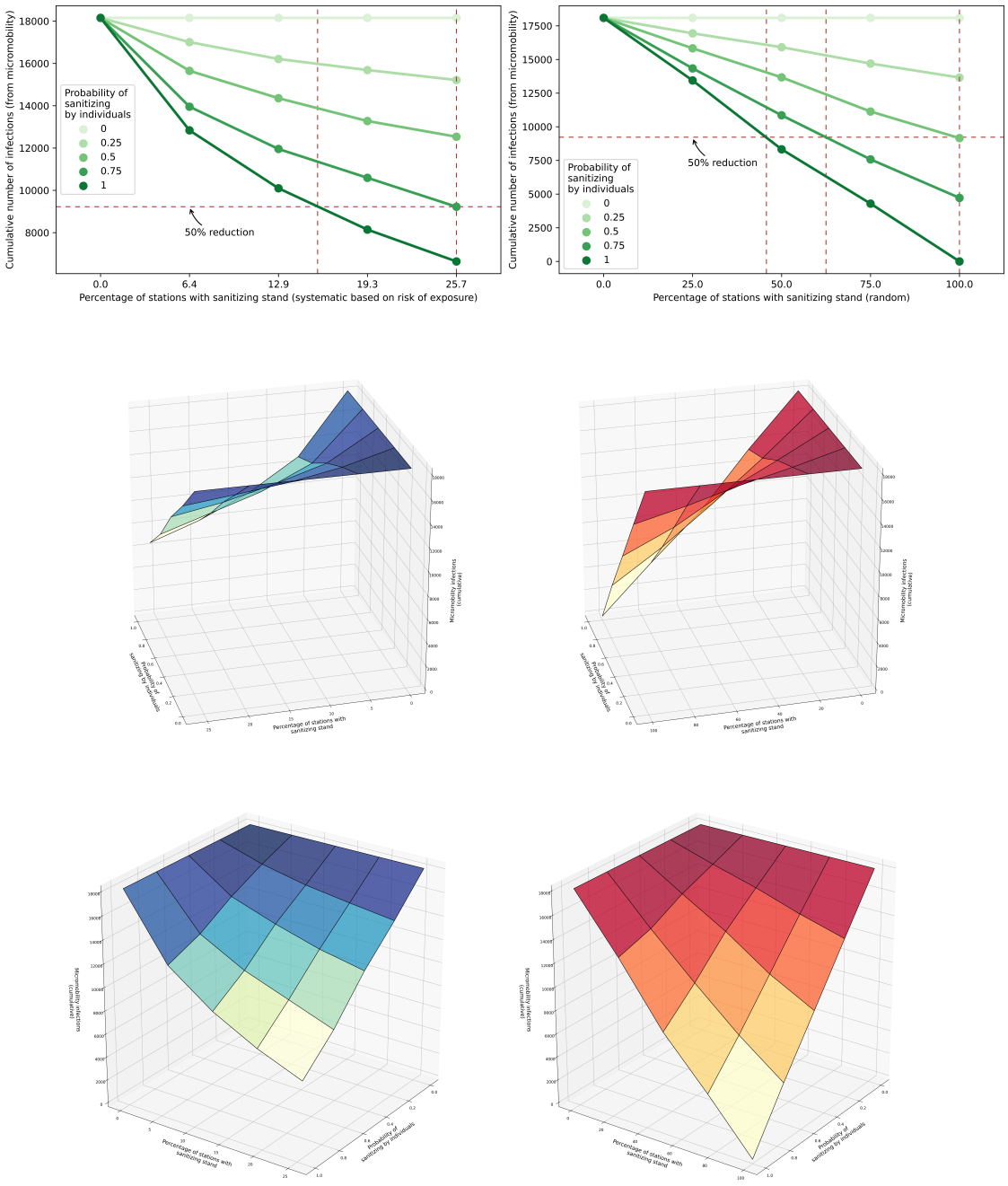


Figure 5.14: Placing sanitizing stand at stations by systematic (left) and random strategies (right). Initial conditions: R_0=5, ρ=2, α=1-day, γ=0.2, I_0^H=100, I_0^M=0, rep=25.

CHAPTER 6: CONCLUSIONS

The main idea of this research was grounded in the realization that the efficiency in mobility gained through novel methods of public transportation may come at the cost of transmitting harmful pathogens and disease as people share space and surfaces in vehicles and infrastructure. In recently popular micromobility systems, for example, the surfaces on the bicycles or scooters are subject to germs and harmful pathogens. The purpose of the current study was to better understand various facets of the potential role of micromobility transportation in the spread of viral disease within dense urban areas. To this end, I designed and implemented a novel spatially-explicit agent-based model of spread of infectious diseases and demonstrated its applicability in a case study of the Chicago Divvy bikesharing system.

In conclusion of the case study in Cook County, even in the extreme case where only a single contaminated bicycle is left in the County, an outbreak can emerge. With the current size of the Divvy system in Chicago, the impact of the system on overall infection rates is rather small (11,469 infections or 0.26% of all infections). However, ignoring the role of micromobility as a vector may underestimate its role in the spread of disease by 43%. The spatial pattern of exposure risk to viruses through micromobility follows the pattern of trips between origins and destinations (CBD and downtown in Chicago), and the start time of the epidemic can highly impact the dynamics of the disease. In terms of intervention methods, disinfection of bicycles is effective only when a large portion of them are cleaned regularly. When hand sanitizing stands are placed at stations, success highly depended on the behavior of individuals and the availability of the equipment. Finally, by systematically placing hand sanitizing stands at micromobility stations, one can save a large amount of

resources compared to the random placing of sanitizing stands at stations.

The contributions of this dissertation research fall in three areas, namely public transportation, disease transmission modeling, as well as policy and decision making.

From a public transportation perspective, this research is one of the first attempts to thoroughly study the spread of disease within public transportation vehicles and infrastructure by providing a platform to study the spread of disease in micromobility systems. In the first place, the model provides a more accurate measure of the role of public transportation in the spread of diseases by considering the vehicles as vectors. Secondly, it is designed based on origin-destination micromobility trips and can be adapted to other modes of transportation.

From a modeling perspective, I proposed a novel agent-based framework to study how micromobility systems contribute to the spread of disease in urban settings. The proposed model has multiple key components that distinguish it from existing models in the literature. First, it was central to this research to distinguish individual humans in the population with their spatial distribution across the urban environment, their heterogeneity in the number of contacts, their mobility in the environment, and their behavior while using the micromobility vehicles. I developed a baseline human contact network with social ties derived from their home, school, and workplace relationships to account for these considerations. Such a network was synthesized from census, public surveys, and empirical data for the entire population in the urban area. Then, two different networks were created for different times and days of a week.

Second, the model stands out by using historic individual micromobility trips, which allowed me to incorporate actual spatial and temporal dynamics (e.g., seasonality) of micromobility vehicles. The availability of such data was an essential contribution to the modeling approach. Historically, agent-based models have depended on simulating the behavior of individual agents. In the past few years and with wider data availability, however, the behavior can now be entered as an input to

ABMs. Third, I proposed the SIR-SC model, a novel epidemic model which leverages the previous two components and mimics vector-borne diseases modeling to calculate the spread of viruses through micromobility systems and a large urban human contact network for the baseline population.

Finally, from the intervention perspective, I created a platform that can be used to test a wide range of control and intervention scenarios.

At the same time, there are some limitations in the model. Since I use actual micromobility trips and a micro-level model with details about the spatial distribution of population and heterogeneity of contacts, the findings may be less generalizable to other micromobility systems.

Similar to other ABMs, evaluating the internal and external validity of the model is a challenge. I externally evaluated the synthetic population by comparing the synthesized values with the four other attributes. The large value of errors and outliers for some census tracts are associated with the group-quarters. More information about the age-gender structure within these groups is needed to improve the performance of the synthetic population. Also, because types of group quarters are not available in the census, they can cause issues related to contacts. For example, a group quarter in jail is supposed to have no contact with individuals outside, but the current model cannot identify those (see appendix A).

External evaluation of the spatial pattern of the exposure risk and the new infections in the population is rather not a simple task. Field sampling from surfaces on bicycles can be conducted to evaluate the density of pathogens on surfaces and compare the spatial pattern with the output of the current model. Other data sources such as master address files and business POIs can be used to determine house and workplace locations instead of placing them on roads by simple rules. When using LODES data, one must be aware that all origin-destination commuting patterns made by workers are not available in this data set.

At the same time, it is important to keep in mind that agent-based models are not necessarily the best way to model problems such as the one discussed in this research. Depending on the purpose of the model, one may find simpler modeling approaches more useful. A good rule of thumb is to start with the simplest methods and switch to more complex approaches as needed.

Modeling frameworks such as the one presented in this research are in the early stages. I suggest multiple interesting lines of research for the future of these types of models.

With respect to the human contact network, I did not model all possible social contacts for individuals. Friends and relatives and contacts made by trips with discretionary purposes (e.g., shopping, social events, restaurants, etc.) are not present in the current contact network. The implication for the model is the spread of disease within the population maybe highly underestimated. These contacts may be created using existing surveys or based on data from social media and recently available cellphone foot traffic and origin-destination.

Alternatively, one can substitute the synthetic contact networks with real contact networks (usually based on cell phone tracking) such as the ones that are used in contact tracing in some countries. In any case, it would be interesting to explore the further impact of the contact network properties by sensitivity analysis of the size of the network for different cohorts of family, school, work and other types of social contacts.

Models like this can be calibrated to current conditions in an epidemic and provide quick and short-term predictions, while they are highly capable of testing different scenarios and providing suggestions. Most of the current prediction models forecast new cases only based on the historic data. Historic data can be integrated in the environments like the one proposed here and improve predictions by considering behavior and other heterogeneities of population.

More complex properties and rules of behavior can be considered for human agents in the model. One can add other characteristics to agents such as ability to change their goals (e.g., schedule) or their behavior (e.g., adapting to seasonal changes by not using micromobility vehicles or wearing gloves during winters) in response to the events that occur during the epidemics.

Finally, environments such as the one presented here are critically needed to experiment with different what-if and intervention scenarios to smartly choose the best choices among alternatives. Let us consider an example with respect to what-if scenarios. What if one manipulates the number of trips to concentrate their load on specific geographic regions (e.g., due to a sports event) and see how the mobility pattern changes the dynamics of disease? In the absence of such a tool, it is not easy to find answers to these questions.

Also, there are some minor areas for future research. It is interesting to distinguish exposure risk in weekday and weekends because of the difference between trip patterns. During weekends trips are likely to be for recreational purposes, so they happen more in places such as parks. During weekdays, however, most trips are for the commuting purpose, so they may concentrate in places such as business districts. That means the high risk stations may have different patterns for weekdays and weekends.

Concerning the SIR-SC transmission model, one can flexibly distinguish rates of transmission in these groups based on the properties of these environments and calculate a separate risk of infection $\lambda_{i,t}$ (equation 4.4) for each venue. Moreover, one may consider contamination sources from the environment. While calculating the adjusted R0 was not needed for this research, it may be interesting to observe the variations of that during epidemic by considering a variation of first-generation matrix approach similar to [64].

I conducted a study on station-based micromobility. Particularly, it is interesting to see how these results may change when dockless bicycles are used. Also, this

research considered the age and gender of the individuals to assign the micromobility trips. However, it did not make a distinction of individual vulnerability against the viral disease. The implication for the results is that it may impact the final number of infections since the distribution of trips changes by gender-age groups of the micromobility users. It is also interesting to find out how the spatial pattern of the exposure risk is associated with other transportation infrastructure and land use such as subway and bus stations, parks, and different business sectors in future research.

REFERENCES

[1] D. A. Goldmann, "Transmission of viral respiratory infections in the home," *The Pediatric infectious disease journal*, vol. 19, no. 10, pp. S97–S102, 2000.

[2] C. Aitken and D. J. Jeffries, "Nosocomial spread of viral disease," *Clinical microbiology reviews*, vol. 14, no. 3, pp. 528–546, 2001.

[3] W. Seto, D. Tsang, R. Yung, T. Ching, T. Ng, M. Ho, L. Ho, J. Peiris, A. of Expert SARS group of Hospital Authority, *et al.*, "Effectiveness of precautions against droplets and contact in prevention of nosocomial transmission of severe acute respiratory syndrome (sars)," *The lancet*, vol. 361, no. 9368, pp. 1519–1520, 2003.

[4] S. Eubank, H. Guclu, V. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, "Modelling disease outbreaks in realistic urban social networks," *Nature*, vol. 429, no. 6988, pp. 180–184, 2004.

[5] Z. Wang, L. Wang, A. Szolnoki, and M. Perc, "Evolutionary games on multilayer networks: a colloquium," *The European physical journal B*, vol. 88, no. 5, pp. 1–15, 2015.

[6] C. L. Barrett, R. J. Beckman, M. Khan, V. A. Kumar, M. V. Marathe, P. E. Stretz, T. Dutta, and B. Lewis, "Generation and analysis of large synthetic social contact networks," in *Proceedings of the 2009 Winter Simulation Conference (WSC)*, pp. 1003–1014, IEEE, 2009.

[7] S. F. Bloomfield, A. E. Aiello, B. Cookson, C. O'Boyle, and E. L. Larson, "The effectiveness of hand hygiene procedures in reducing the risks of infections in home and community settings including handwashing and alcohol-based hand sanitizers," *American journal of infection control*, vol. 35, no. 10, pp. S27–S64, 2007.

[8] J. D. Mayer, "Geography, ecology and emerging infectious diseases," *Social science & medicine*, vol. 50, no. 7-8, pp. 937–952, 2000.

[9] E. K. Cromley and S. L. McLafferty, *GIS and public health*. Guilford Press, 2011.

[10] U. N. D. of Economic, *World urbanization prospects: The 2003 revision*, vol. 216. United Nations Publications, 2004.

[11] M. E. Newman, "Spread of epidemic disease on networks," *Physical review E*, vol. 66, no. 1, p. 016128, 2002.

[12] S. V. Scarpino, J. G. Scott, R. M. Eggo, B. Clements, N. B. Dimitrov, and L. A. Meyers, "Socioeconomic bias in influenza surveillance," *PLoS computational biology*, vol. 16, no. 7, p. e1007941, 2020.

[13] J.-P. Rodrigue, C. Comtois, and B. Slack, *The geography of transport systems*. Routledge, 2016.

[14] C. R. Wells, P. Sah, S. M. Moghadas, A. Pandey, A. Shoukat, Y. Wang, Z. Wang, L. A. Meyers, B. H. Singer, and A. P. Galvani, "Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak," *Proceedings of the National Academy of Sciences*, vol. 117, no. 13, pp. 7504–7509, 2020.

[15] D. G. Janelle, "Time–space convergence," in *Handbook of research methods and applications in spatially integrated social science*, Edward Elgar Publishing, 2014.

[16] S. H. Ali and R. Keil, "Contagious cities," *Geography Compass*, vol. 1, no. 5, pp. 1207–1226, 2007.

[17] A. Browne, S. St-Onge Ahmad, C. R. Beck, and J. S. Nguyen-Van-Tam, "The roles of transportation and transportation hubs in the propagation of influenza and coronaviruses: a systematic review," *Journal of travel medicine*, vol. 23, no. 1, p. tav002, 2016.

[18] R. Thomas, *Geomedical systems: intervention and control*. Routledge, 1992.

[19] J. M. Epstein, D. M. Goedecke, F. Yu, R. J. Morris, D. K. Wagener, and G. V. Bobashev, "Controlling pandemic flu: the value of international air travel restrictions," *PloS one*, vol. 2, no. 5, p. e401, 2007.

[20] C. Nicolaides, D. Avraam, L. Cueto-Felgueroso, M. C. González, and R. Juanes, "Hand-hygiene mitigation strategies against global disease spreading through the air transportation network," *Risk Analysis*, vol. 40, no. 4, pp. 723–740, 2020.

[21] A. D. Cliff, J. Ord, P. Haggett, G. Versey, *et al.*, *Spatial diffusion: an historical geography of epidemics in an island community*, vol. 14. CUP Archive, 1981.

[22] A. G. Ferguson and C. N. Morris, "Mapping transactional sex on the northern corridor highway in kenya," *Health & place*, vol. 13, no. 2, pp. 504–519, 2007.

[23] A. J. Tatem, D. J. Rogers, and S. I. Hay, "Global transport networks and infectious disease spread," *Advances in parasitology*, vol. 62, pp. 293–343, 2006.

[24] T. C. Germann, K. Kadau, I. M. Longini, and C. A. Macken, "Mitigation strategies for pandemic influenza in the united states," *Proceedings of the National Academy of Sciences*, vol. 103, no. 15, pp. 5935–5940, 2006.

[25] A. Cartenì, L. Di Francesco, and M. Martino, "How mobility habits influenced the spread of the covid-19 pandemic: Results from the italian case study," *Science of the Total Environment*, vol. 741, p. 140489, 2020.

[26] B. Nikparvar, M. Rahman, F. Hatami, J.-C. Thill, *et al.*, "Spatio-temporal prediction of the covid-19 pandemic in us counties: modeling with a deep lstm neural network," *Scientific Reports*, vol. 11, no. 1, pp. 1–12, 2021.

[27] F. Xu, C. C. McCluskey, and R. Cressman, "Spatial spread of an epidemic through public transportation systems with a hub," *Mathematical biosciences*, vol. 246, no. 1, pp. 164–175, 2013.

[28] R. E. Howland, N. R. Cowan, S. S. Wang, M. L. Moss, and S. Glied, "Public transportation and transmission of viral respiratory disease: Evidence from influenza deaths in 121 cities in the united states," *PloS one*, vol. 15, no. 12, p. e0242990, 2020.

[29] O. Mohr, M. Askar, S. Schink, T. Eckmanns, G. Krause, and G. Poggensee, "Evidence for airborne infectious disease transmission in public ground transport–a literature review," *Eurosurveillance*, vol. 17, no. 35, p. 20255, 2012.

[30] S. Shaheen and N. Chan, "Mobility and the sharing economy: Potential to facilitate the first-and last-mile public transit connections," *Built Environment*, vol. 42, no. 4, pp. 573–588, 2016.

[31] J. Larsen, *Bike-sharing programs hit the streets in over 500 cities worldwide*. Earth Policy Institute Washington, DC, 2013.

[32] T. Gu, I. Kim, and G. Currie, "To be or not to be dockless: Empirical analysis of dockless bikeshare development in china," *Transportation Research Part A: Policy and Practice*, vol. 119, pp. 122–147, 2019.

[33] C. Taddei, R. Gnesotto, S. Forni, G. Bonaccorsi, A. Vannucci, and G. Garofalo, "Cycling promotion and non-communicable disease prevention: health impact assessment and economic evaluation of cycling to work or school in florence," *PLoS One*, vol. 10, no. 4, p. e0125491, 2015.

[34] E. O'Mahony and D. Shmoys, "Data analysis and optimization for (citi) bike sharing," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, 2015.

[35] I. Frade and A. Ribeiro, "Bicycle sharing systems demand," *Procedia-Social and Behavioral Sciences*, vol. 111, pp. 518–527, 2014.

[36] D. Duran-Rodas and G. Wulfhorst, "Spatial factors influencing bike–sharing usage-a âmobility cultureâ approach," 2019.

[37] X. Zhou, "Understanding spatiotemporal patterns of biking behavior by analyzing massive bike sharing data in chicago," *PloS one*, vol. 10, no. 10, p. e0137922, 2015.

[38] S. A. Shaheen, E. W. Martin, A. P. Cohen, N. D. Chan, and M. Pogodzinski, "Public bikesharing in north america during a period of rapid expansion: Understanding business models, industry trends & user impacts, mti report 12-29," 2014.

[39] S. A. Shaheen, H. Zhang, E. Martin, and S. Guzman, "China's hangzhou public bicycle: understanding early adoption and behavioral response to bikesharing," *Transportation research record*, vol. 2247, no. 1, pp. 33–41, 2011.

[40] C. Zhang and H. Zhou, "The study of coopetition between public bus and bike sharing based on environmental protection," in *E3S Web of Conferences*, vol. 136, p. 04015, EDP Sciences, 2019.

[41] C. M. De Chardon, G. Caruso, and I. Thomas, "Bike-share rebalancing strategies, patterns, and purpose," *Journal of transport geography*, vol. 55, pp. 22–39, 2016.

[42] J. Zhang, X. Pan, M. Li, and S. Y. Philip, "Bicycle-sharing system analysis and trip prediction," in *2016 17th IEEE international conference on mobile data management (MDM)*, vol. 1, pp. 174–179, IEEE, 2016.

[43] A. Faghih-Imani and N. Eluru, "Analysing bicycle-sharing system user destination choice preferences: Chicagoâs divvy system," *Journal of transport geography*, vol. 44, pp. 53–64, 2015.

[44] M. Ricci, "Bike sharing: a review of evidence on impacts and processes of implementation and operation. research in transportation business and management 15," 2015.

[45] M. A. Babagoli, T. K. Kaufman, P. Noyes, and P. E. Sheffield, "Exploring the health and spatial equity implications of the new york city bike share system," *Journal of transport & health*, vol. 13, pp. 200–209, 2019.

[46] X. Qian, M. Jaller, and D. Niemeier, "Enhancing equitable service level: Which can address better, dockless or dock-based bikeshare systems?," *Journal of Transport Geography*, vol. 86, p. 102784, 2020.

[47] S. Couch and H. K. Smalley, "Encouraging equitable bikeshare: Implications of docked and dockless models for spatial equity," *arXiv preprint arXiv:1906.00129*, 2019.

[48] I. L. Pepper, C. P. Gerba, T. J. Gentry, and R. M. Maier, *Environmental microbiology*. Academic press, 2011.

[49] R. A. Canales, K. A. Reynolds, A. M. Wilson, S. L. Fankem, M. H. Weir, J. B. Rose, S. Abd-Elmaksoud, and C. P. Gerba, "Modeling the role of fomites in a norovirus outbreak," *Journal of occupational and environmental hygiene*, vol. 16, no. 1, pp. 16–26, 2019.

[50] H. Lei, Y. Li, S. Xiao, X. Yang, C. Lin, S. L. Norris, D. Wei, Z. Hu, and S. Ji, "Logistic growth of a surface contamination network and its role in disease spread," *Scientific reports*, vol. 7, no. 1, pp. 1–10, 2017.

[51] K. Zhao, Q. Wu, X. Shen, Y. Xuan, and X. Fu, "I undervalue you but i need you: the dissociation of attitude and memory toward in-group members," *PLoS One*, vol. 7, no. 3, p. e32932, 2012.

[52] S. Xiao, J. Tang, D. Hui, H. Lei, H. Yu, and Y. Li, "Probable transmission routes of the influenza virus in a nosocomial outbreak," *Epidemiology & Infection*, vol. 146, no. 9, pp. 1114–1122, 2018.

[53] A. Kramer and O. Assadian, "Survival of microorganisms on inanimate surfaces," in *Use of Biocidal Surfaces for Reduction of Healthcare Acquired Infections*, pp. 7–26, Springer, 2014.

[54] A. N. Kraay, M. A. Hayashi, N. Hernandez-Ceron, I. H. Spicknall, M. C. Eisenberg, R. Meza, and J. N. Eisenberg, "Fomite-mediated transmission as a sufficient pathway: a comparative analysis across three viral pathogens," *BMC infectious diseases*, vol. 18, no. 1, pp. 1–13, 2018.

[55] M. P. Atkinson and L. M. Wein, "Quantifying the routes of transmission for pandemic influenza," *Bulletin of mathematical biology*, vol. 70, no. 3, pp. 820–867, 2008.

[56] N. Zhang, Y. Li, and H. Huang, "Surface touch and its network growth in a graduate student office," *Indoor Air*, vol. 28, no. 6, pp. 963–972, 2018.

[57] N. B. Dimitrov and L. A. Meyers, "Mathematical approaches to infectious disease prediction and control," in *Risk and optimization in an uncertain world*, pp. 1–25, INFORMS, 2010.

[58] D. Bernoulli, "Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir," *Histoire de l'Acad., Roy. Sci.(Paris) avec Mem*, pp. 1–45, 1760.

[59] W. H. Hamer, *Epidemic disease in England: the evidence of variability and of persistency of type*. Bedford Press, 1906.

[60] H. E. Soper, "The interpretation of periodicity in disease prevalence," *Journal of the Royal Statistical Society*, vol. 92, no. 1, pp. 34–73, 1929.

[61] S. Li, J. N. Eisenberg, I. H. Spicknall, and J. S. Koopman, "Dynamics and control of infections transmitted from person to person through the environment," *American journal of epidemiology*, vol. 170, no. 2, pp. 257–265, 2009.

[62] M. Batty, "A generic framework for computational spatial modelling," in *Agent-based models of geographical systems*, pp. 19–50, Springer, 2012.

[63] R. F. Grais, J. H. Ellis, and G. E. Glass, "Assessing the impact of airline travel on the geographic spread of pandemic influenza," *European journal of epidemiology*, vol. 18, no. 11, pp. 1065–1072, 2003.

[64] M. Ajelli, B. Gonçalves, D. Balcan, V. Colizza, H. Hu, J. J. Ramasco, S. Merler, and A. Vespignani, "Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models," *BMC infectious diseases*, vol. 10, no. 1, pp. 1–13, 2010.

[65] Z. Du, S. J. Fox, P. Holme, J. Liu, A. P. Galvani, and L. A. Meyers, "Periodicity in movement patterns shapes epidemic risk in urban environments," *arXiv preprint arXiv:1809.05203*, 2018.

[66] H. Rahmandad and J. Sterman, "Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models," *Management science*, vol. 54, no. 5, pp. 998–1014, 2008.

[67] L. Meyers, "Contact network epidemiology: Bond percolation applied to infectious disease prediction and control," *Bulletin of the American Mathematical Society*, vol. 44, no. 1, pp. 63–86, 2007.

[68] K. Harland, A. Heppenstall, D. Smith, and M. H. Birkin, "Creating realistic synthetic populations at varying spatial scales: A comparative critique of population synthesis techniques," *Journal of Artificial Societies and Social Simulation*, vol. 15, no. 1, 2012.

[69] M. Birkin and B. Wu, "A review of microsimulation and hybrid agent-based approaches," *Agent-based models of geographical systems*, pp. 51–68, 2012.

[70] K. Müller and K. W. Axhausen, "Population synthesis for microsimulation: State of the art," *Arbeitsberichte Verkehrs-und Raumplanung*, vol. 638, 2010.

[71] M. Lenormand and G. Deffuant, "Generating a synthetic population of individuals in households: Sample-free vs sample-based methods," *arXiv preprint arXiv:1208.6403*, 2012.

[72] R. J. Beckman, K. A. Baggerly, and M. D. McKay, "Creating synthetic baseline populations," *Transportation Research Part A: Policy and Practice*, vol. 30, no. 6, pp. 415–429, 1996.

[73] D. Voas and P. Williamson, "An evaluation of the combinatorial optimisation approach to the creation of synthetic microdata," *International Journal of Population Geography*, vol. 6, no. 5, pp. 349–366, 2000.

[74] W. E. Deming and F. F. Stephan, "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known," *The Annals of Mathematical Statistics*, vol. 11, no. 4, pp. 427–444, 1940.

[75] A. Hobeika, "Transims fundamentals: Chapter 3: Population synthesizer," *US Department of Transportation*, 2005.

[76] A. Burger, T. Oz, A. Crooks, and W. G. Kennedy, "Generation of realistic megacity populations and social networks for agent-based modeling," in *Proceedings of the 2017 International Conference of The Computational Social Science Society of the Americas*, pp. 1–7, 2017.

[77] K. R. Bisset, J. Chen, X. Feng, V. A. Kumar, and M. V. Marathe, "Epifast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems," in *Proceedings of the 23rd international conference on Supercomputing*, pp. 430–439, 2009.

[78] L. A. Meyers, M. Newman, M. Martin, and S. Schrag, "Applying network theory to epidemics: control measures for mycoplasma pneumoniae outbreaks," *Emerging infectious diseases*, vol. 9, no. 2, p. 204, 2003.

[79] M. L. Ciofi degli Atti, S. Merler, C. Rizzo, M. Ajelli, M. Massari, P. Manfredi, C. Furlanello, G. Scalia Tomba, and M. Iannelli, "Mitigation measures for pandemic influenza in italy: an individual based model considering different scenarios," *PloS one*, vol. 3, no. 3, p. e1790, 2008.

[80] L. A. Meyers, M. Newman, and B. Pourbohloul, "Predicting epidemics on directed contact networks," *Journal of theoretical biology*, vol. 240, no. 3, pp. 400–418, 2006.

[81] J. Enright and R. R. Kao, "Epidemics on dynamic networks," *Epidemics*, vol. 24, pp. 88–97, 2018.

[82] E. Volz and L. A. Meyers, "Susceptible–infected–recovered epidemics in dynamic contact networks," *Proceedings of the Royal Society B: Biological Sciences*, vol. 274, no. 1628, pp. 2925–2934, 2007.

[83] C.-Y. Huang, T.-H. Wen, Y.-H. Fu, Y.-S. Tsai, *et al.*, "Epirank: Modeling bidirectional disease spread in asymmetric commuting networks," *Scientific reports*, vol. 9, no. 1, pp. 1–15, 2019.

[84] S. J. Alam and A. Geller, "Networks in agent-based social simulation," in *Agent-based models of geographical systems*, pp. 199–216, Springer, 2012.

[85] D. A. Bennett and W. Tang, "Modelling adaptive, spatially aware, and mobile agents: Elk migration in yellowstone," *International Journal of Geographical Information Science*, vol. 20, no. 9, pp. 1039–1066, 2006.

[86] J. Von Neumann, "The general and logical theory of automata," in *Systems Research for Behavioral Sciencesystems Research*, pp. 97–107, Routledge, 2017.

[87] J. Conway *et al.*, "The game of life," *Scientific American*, vol. 223, no. 4, p. 4, 1970.

[88] T. C. Schelling, "Dynamic models of segregation," *Journal of mathematical sociology*, vol. 1, no. 2, pp. 143–186, 1971.

[89] J. H. Holland and J. H. Miller, "Artificial adaptive agents in economic theory," *The American economic review*, vol. 81, no. 2, pp. 365–370, 1991.

[90] R. Axtell, R. Axelrod, J. M. Epstein, and M. D. Cohen, "Aligning simulation models: A case study and results," *Computational & mathematical organization theory*, vol. 1, no. 2, pp. 123–141, 1996.

[91] A. T. Crooks and A. J. Heppenstall, "Introduction to agent-based modelling," in *Agent-based models of geographical systems*, pp. 85–105, Springer, 2012.

[92] D. C. Parker, S. M. Manson, M. A. Janssen, M. J. Hoffmann, and P. Deadman, "Multi-agent systems for the simulation of land-use and land-cover change: a review," *Annals of the association of American Geographers*, vol. 93, no. 2, pp. 314–337, 2003.

[93] W. G. Kennedy, "Modelling human behaviour in agent-based models," in *Agent-based models of geographical systems*, pp. 167–179, Springer, 2012.

[94] J. E. Laird, A. Newell, and P. S. Rosenbloom, "Soar: An architecture for general intelligence," *Artificial intelligence*, vol. 33, no. 1, pp. 1–64, 1987.

[95] R. Axelrod, *The complexity of cooperation*. Princeton university press, 1997.

[96] A. Crooks, C. Castle, and M. Batty, "Key challenges in agent-based modelling for geo-spatial simulation," *Computers, Environment and Urban Systems*, vol. 32, no. 6, pp. 417–430, 2008.

[97] M. Abdou, L. Hamill, and N. Gilbert, "Designing and building an agent-based model," in *Agent-based models of geographical systems*, pp. 141–165, Springer, 2012.

[98] H. Couclelis, "Modeling frameworks, paradigms, and approaches," *Geographic information systems and environmental modeling. Upper Saddle River, NJ: Prentice Hall*, pp. 36–50, 2002.

[99] D. OâSullivan, J. Millington, G. Perry, and J. Wainwright, "Agent-based models–because theyâre worth it?," in *Agent-based models of geographical systems*, pp. 109–123, Springer, 2012.

[100] T. Berger, M. Goodchild, M. A. Janssen, S. M. Manson, R. Najlis, and D. Parker, "Methodological considerations for agent-based modeling of land-use and land-cover change," in *Agent-based models of land-use and land cover change: Report and review of an International Workshop Irvine, California, USA October*, pp. 4–7, 2001.

[101] L. See *et al.*, "Calibration and validation of agent-based models of land cover change," in *Agent-based models of geographical systems*, pp. 181–197, Springer, 2012.

[102] F. Klügl, "A validation methodology for agent-based simulations," in *Proceedings of the 2008 ACM symposium on Applied computing*, pp. 39–43, 2008.

[103] M. Batty and P. M. Torrens, "Modelling and prediction in a complex world," *Futures*, vol. 37, no. 7, pp. 745–766, 2005.

[104] Z. Gong, W. Tang, D. A. Bennett, and J.-C. Thill, "Parallel agent-based simulation of individual-level spatial interactions within a multicore computing environment," *International Journal of Geographical Information Science*, vol. 27, no. 6, pp. 1152–1170, 2013.

[105] J. M. Epstein, "Agent-based computational models and generative social science," *Complexity*, vol. 4, no. 5, pp. 41–60, 1999.

[106] S. Saura and J. Martinez-Millan, "Sensitivity of landscape pattern metrics to map spatial extent," *Photogrammetric engineering and remote sensing*, vol. 67, no. 9, pp. 1027–1036, 2001.

[107] A. Ménard and D. J. Marceau, "Exploration of spatial scale sensitivity in geographic cellular automata," *Environment and Planning B: Planning and Design*, vol. 32, no. 5, pp. 693–714, 2005.

[108] M. Batty, *Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals*. The MIT press, 2007.

[109] B. M. Ardestani, D. O'Sullivan, and P. Davis, "A multi-scaled agent-based model of residential segregation applied to a real metropolitan area," *Computers, Environment and Urban Systems*, vol. 69, pp. 1–16, 2018.

[110] C. J. Dommar, R. Lowe, M. Robinson, and X. Rodó, "An agent-based model driven by tropical rainfall to understand the spatio-temporal heterogeneity of a chikungunya outbreak," *Acta tropica*, vol. 129, pp. 61–73, 2014.

[111] C. Linard, N. Ponçon, D. Fontenille, and E. F. Lambin, "A multi-agent simulation to assess the risk of malaria re-emergence in southern france," *Ecological Modelling*, vol. 220, no. 2, pp. 160–174, 2009.

[112] A. T. Crooks and A. B. Hailegiorgis, "An agent-based modeling approach applied to the spread of cholera," *Environmental Modelling & Software*, vol. 62, pp. 164–177, 2014.

[113] L. Perez and S. Dragicevic, "An agent-based approach for modeling dynamics of contagious disease spread," *International journal of health geographics*, vol. 8, no. 1, pp. 1–17, 2009.

[114] S. Xiao, J. W. Tang, and Y. Li, "Airborne or fomite transmission for norovirus? a case study revisited," *International journal of environmental research and public health*, vol. 14, no. 12, p. 1571, 2017.

[115] S. Xiao, Y. Li, M. Sung, J. Wei, and Z. Yang, "A study of the probable transmission routes of mers-cov during the first hospital outbreak in the republic of korea," *Indoor air*, vol. 28, no. 1, pp. 51–63, 2018.

[116] B. Pourbohloul, L. A. Meyers, D. M. Skowronski, M. Krajden, D. M. Patrick, and R. C. Brunham, "Modeling control strategies of respiratory pathogens," *Emerging infectious diseases*, vol. 11, no. 8, p. 1249, 2005.

[117] J. L. Herrera-Diestra and L. A. Meyers, "Local risk perception enhances epidemic control," *PloS one*, vol. 14, no. 12, p. e0225576, 2019.

[118] V. Grimm, U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S. K. Heinz, G. Huse, *et al.*, "A standard protocol for describing individual-based and agent-based models," *Ecological modelling*, vol. 198, no. 1-2, pp. 115–126, 2006.

[119] A. L'Huillier, C. Tapparel, L. Turin, P. Boquete-Suter, Y. Thomas, and L. Kaiser, "Survival of rhinoviruses on human fingers," *Clinical Microbiology and Infection*, vol. 21, no. 4, pp. 381–385, 2015.

[120] M. J. Keeling and P. Rohani, *Modeling infectious diseases in humans and animals*. Princeton university press, 2011.

[121] K. Chapuis and P. Taillandier, "A brief review of synthetic population generation practices in agent-based social simulation," in *submitted to SSC2019, Social Simulation Conference*, 2019.

[122] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones, "A high-resolution human contact network for infectious disease transmission," *Proceedings of the National Academy of Sciences*, vol. 107, no. 51, pp. 22020–22025, 2010.

[123] R. I. Dunbar, "The social brain hypothesis," *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, vol. 6, no. 5, pp. 178–190, 1998.

APPENDIX A: GENERATING SYNTHETIC POPULATION WITH SOCIAL TIES

I used the methodology proposed by [76] to generate a realistic large scale synthetic baseline population with social ties using the census data, work and school locations and commuting patterns. Here, I describe in detail the process of generating the synthetic population with social ties for Cook County, IL.

The synthetic population is created in four steps:

1. Creating home, work, and school environments across geographic space,
2. Generating and assigning individual agents to households,
3. Assigning work and school locations to each individual,
4. Creating a network of social contacts based on membership in family, school, and, workplace social ties.

A.1 Creating the Geographic Space

A.1.1 Household locations

House locations are created along the road network in Cook County[1]. The number of occupied houses are extracted from census data for each census tract (number of occupied houses equals the number of households in census). Then, they are located 50m apart on local roads (MTFCC = S1400)[2]. When the population density is high they are placed on top of each other.

A.1.2 Workplace locations

Two sources of data are used to calculate the number of businesses within each tract. County level business establishment counts (CBP) and Origin-Destination Employ-

[1] https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html

[2] see technical documentation of the TIGER data set, Appendix E, MAF/TIGERFeature Class Code (MTFCC) Definitions athttps://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html

ment Statistics (LODES). The CBP database is available through the US Census Bureau County Business Patterns survey (CBP)[3]. CBP provides annual subnational economic data by industry, and includes the number of business establishments, employment during the week of March 12, first quarter payroll, and annual payroll. The number of business establishments are extracted from CBP table, but it is in county level. Thus I need to downscale the data into census tract level. Origin-Destination Employment Statistics (LODES) data from Longitudinal Employer-Household Dynamics (LEHD) program[4] is used for this purpose. LEHD provides the residence and workplace flow among blocks for the years between 2002-2019 (for this case study I use data of 2018). The data includes all jobs covered under state unemployment insurance law and most civilian federal employment. However, it does not cover the self-employment, military, the US Postal Service, and informal employment[5]. Jobs totals are associated with home and work Census Blocks. First the jobs totals are aggregated into the census tract level. Then, I count the number of jobs within each census tract and the number of jobs for the entire county. Using this new information together with the number of businesses within each county ($nworkplaces_{tract}$), one can calculate the number of workplaces within each census tract using the following equation:

$$nworkplaces_{tract} = nworkplaces_{county} * \frac{njobs_{tract}}{njobs_{county}} \tag{A.1}$$

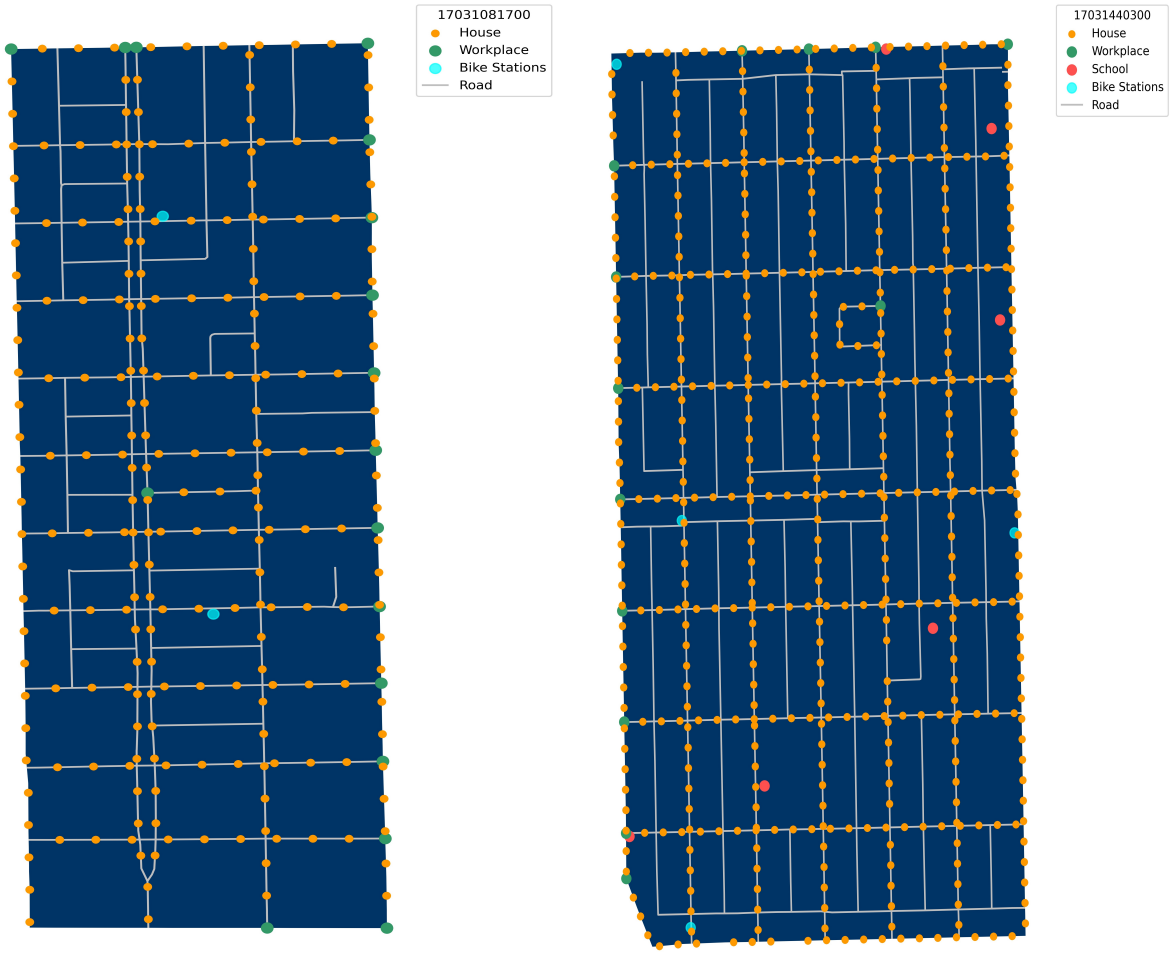
Finally, Workplaces are placed 20m apart on secondary roads or on intersection of local roads (MTFCC = S1200, S1400).

[3]https://www.census.gov/programs-surveys/cbp.html
[4]https://lehd.ces.census.gov/
[5]https://ideas.repec.org/p/cen/wpaper/14-38.html

A.1.3 School locations

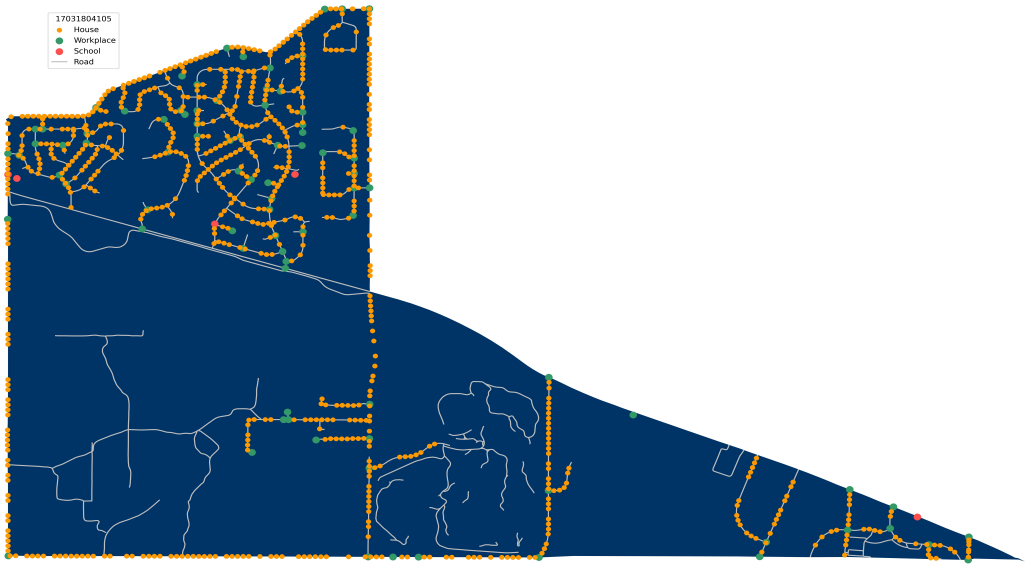
School locations are extracted from the United States Environmental Protection Agency (EPA), Office of Environmental Information[6]. The data set is called ORNL_Education, which is collected by Oak Ridge National Laboratory and includes colleges and universities, supplemental colleges, private schools, public schools and day care centers. Location, enrollment, and grade of Private schools, public schools, and daycare homes were extracted and all used as school locations in the model.

Figure A.0 shows some example census tracts with simulated house, workplace, and school, as well as actual public bicycle stations in Cook County, IL.



(a)

[6](https://geodata.epa.gov/arcgis/rest/services/OEI/ORNL$_{E}ducation/MapServer$)



(b)

(c)

Figure A.0: Sample census tracts.

A.2 Generating and Assigning Individual Agents to Households

In order to create synthetic population, one needs to simulated every individual and then place them in the households. I use two sets of information from the decennial census tables to generate individuals and assign them to households. The first one is age and sex groups and the second one is the household types. Within census tables individuals are divided into two sex types each with 18 age groups. Thus, within each census tract the number of individuals for 36 sex-age groups are available. For each census tract, individuals within these sex-age categories are generated. Then, we need to populate the households with these individuals.

Within census tables, population is divided into household and group quarters. Two types of each group exist in the tables. Households are divided into family and non-family households and group quarters are divided into institutionalized and non-institutionalized groups. Family households include, husband and wife with and without children under age 18, male householder with and without children under age 18, and female householder with and without children under age 18 (1-6). Non-family households are divided into non-family groups (e.g. friends, unmarried couple), male younger and older than 65, and female younger and older than 65 (6-11). These 11 household types together with group quarters (both group quarter types are considered together as category 12) are generated for the Cook County and randomly assigned to household locations in section A.1.1. For simplicity, I call all these categories as households.

Using the first 11 household types, and a few other constraints imposed by American Community Survey (ACS) family arrangement data[7], individuals are assigned to households. First, the head of a household is assigned to each household based on household type constraints, where the sex and age of the head is stochastically derived from the ACS Families and Living Arrangements tables 2018 age and sex

[7] https://www.census.gov/data/tables/2018/demo/families/cps-2018.html

distribution of householders (nationally). Then, for household types with married couples, a simple rule is defined to add the wife or husband of a householder ($-4 <$ husband age $-$ wife age < 9)[76].

For the households with children under 18, the number of children is stocahstically defined by distributions extracted from ACS Families and Living Arrangements 2018. Group quarters are considered as one single population within each county and are placed to the same location. I did not distinguish different group quarters from each other. One may extract them using ACS tables related to group quarters (Tables B26101, B26201). To illustrate the importance of group quarters in creating synthetic population, let us consider an example. The largest group quarter in Cook County is located in census tract 17031843500 (Cook County Department of Corrections) with 10,273 individuals belong to this group out of 11,309 residents. However, we do not expect individuals in a jail to have connections with outside. There is no way in the model to account for that unless we know the types of a group quarter.

A.3 Assigning Work and School Locations to Each Individual

Individuals in work-age are randomly assigned to a working location within a tract using the LEHD origin-destination statistics. Employee size of the establishments is considered lognormal within each tract[8]. Schools are assigned based on grade and enrollment levels. School-age agents are assigned to the nearest school within 30 km distance to their home location if a spot is available for their age. Otherwise, they are assigned to the school with the lowest enrollment.

A.4 Creating a network of social contacts

Based on their connections in household, school, and workplace, individuals are linked to each other. When the size of population within one of these places is higher than 5, a small-world network is created.

[8]https://www.aeaweb.org/articles?id=10.1257/aer.97.5.1639