Micromobility Systems and Disease Transmission:

A Spatially Explicit Agent-Based Approach

Behnam Nikparvar[1,2*], Jean-Claude Thill[3,4]

[1]The William States Lee College of Engineering, University of North Carolina at Charlotte, 9201 University City Blvd, NC 28223, USA
[2]Department of Biology, Center for Infectious Disease Dynamics, Penn State University, University Park, PA, United States (e-mail: bzn5190@psu.edu)
[3]Department of Geography and Earth Sciences, University of North Carolina at Charlotte, 9201 University City Blvd, NC 28223, USA
[4]School of Data Science, University of North Carolina at Charlotte, 9201 University City Blvd, NC 28223, USA

Abstract

We 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. We use this simulation framework to study the role of micromobility in the spread of viral disease in urban areas. 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 shows higher risk 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. The proposed simulation framework can be used to assess the efficacy of interventions and make trade-offs between competing scenarios when dealing with epidemics of the sort analyzed in this research.

1. Introduction

New modes of public[1] and shared transportation such as micromobility are rapidly growing in popularity in urban areas (Lazarus et al., 2020; Shaheen et al., 2021). Bike sharing and e-scooter sharing, for example, have been advanced to solve the first/last mile problem, providing swift access to public transit stops for their users (Shaheen & Chan, 2016). Nonetheless, the expedience afforded by these modes of transportation may pose a risk of pathogen transmission, as the surfaces of bicycles and scooters are susceptible to contamination when they are left in contaminated places or used by infectious individuals. This research proposes an agent-based modeling framework to study the spatial and temporal dynamics of a viral disease through surfaces on dockless or dockbased micromobility vehicles (MVs) in urban areas.

[1] By public transportation, we mean transportation available for use by the general public. This does not imply vehicles are owned and/or operated by a public agency.

Transportation systems can contribute to the spread of diseases by introducing pathogens to new geographic areas (Epstein et al., 2007), by accelerating their diffusion (Cartenì et al., 2021; J. Li et al., 2021), or by becoming a disease vector (Goscé & Johansson, 2018; Rodrigue et al., 2016). In many respects, this mirrors the transmission of vector-borne diseases by living vectors such as mosquitoes (Dommar et al., 2014). The latter role manifests itself within moving vehicles or in facilities where individuals share space with others, thus leading to new contacts that may facilitate disease transmission (Borjigin et al., 2023; Xu et al., 2013). Despite the potential impact of transportation systems on disease epidemics, it is challenging to estimate or model this effect. Existing models tend to overlook the role of transportation systems as a disease vector or fail to distinguish it from the other two roles. This makes it difficult to develop effective intervention strategies specific to public and shared transportation systems (Browne et al., 2016; Fumagalli et al., 2021; Mohr et al., 2012).

Quantifying the impact of movement and of spatial relationships in the spread of human-to-human infectious diseases is a well-established and evolving area of research (Baker et al., 2022; Nita Bharti, 2021), yet scant attention has been directed to public transportation implications. This study focuses on viral respiratory and gastrointestinal diseases that can transmit through surfaces and close contact among human individuals who share environment while using transportation systems. Dynamic epidemic models, including metapopulation, contact network, and agent-based models, constitute the main strands in epidemic modeling that can account for movement and space.

Metapopulation models are spatial models designed to model the dynamics of a disease among subpopulations; they are mostly used for larger geographic regions where multiple population centers are connected through transportation links (Ajelli et al., 2010; Balcan et al., 2009; Du et al., 2018). They can account for some heterogeneities between subpopulations, but they still make a number of crude assumptions about intra-subpopulation heterogeneities (Riley et al., 2015). Moreover, the examination of individual passenger trajectories is critical in the analysis of viral transmission in public and shared transportation systems, whereas metapopulation models can only depict the collective movement between subpopulations (Tirachini, 2020).

Contact network models address the latter issue by modeling all individuals in the population as nodes of a network and linking them through edges that represent their social and physical connections along which disease transmission can be enabled (Eubank et al., 2004; Meyers, 2007; Newman, 2002). The existence of an edge between two individuals is determined by the route of disease transmission through the population, and its weight is determined by the transmissibility of the disease (Newman, 2002). As individuals move through space, new contact links are formed between them.

A passenger encounter network, commonly extracted from tap-in/tap-out smart carts, has been used to simulate the spread of a disease through urban public transit systems (Bóta et al., 2017; Mo et al., 2021). Passenger encounters refer to individuals who have at any point shared the same vehicle within a public transportation system (Sun et al., 2013). It is assumed that passengers who share a vehicle are in close enough proximity to potentially spread infections. However, this method is limited to accounting for direct encounters within transportation

vehicles, while disease agents may persist in the environment (e.g., on surfaces and in the air) and potentially infect a constantly changing passenger population. Additionally, transportation vehicles are not explicitly modeled while their distribution is important to develop spatially heterogeneous interventions. A common feature of many novel shared mobility systems is that they are not confined to operating on specific routes or between specific fixed stations.

The transmission of viruses in transportation vehicles is subject to a multitude of factors, including trip duration, individual behavior during travel (e.g., mask-wearing, social distancing, population density), demographic attributes of travelers, and virus survival time in the environmental setting. Agent-based models (ABMs), another strand of individual-based modeling, are well suited to account for these qualities. They have extensively been used in the study of vector-borne (Dommar et al., 2014; Linard et al., 2009) and non-vector-borne infectious diseases (Ajelli et al., 2010; Crooks & Hailegiorgis, 2014; Eubank et al., 2004; Germann et al., 2006; Kerr et al., 2021; Pastor-Satorras et al., 2014; Rahmandad & Sterman, 2008; Truszkowska et al., 2021) where the movement and the spatial distribution of humans and vectors are taken explicitly into account. However, with a few recent exceptions (Borjigin et al., 2023; Nicolaides et al., 2020), the potential role of transportation as a vector is not broached, nor are the contributions of specific transportation modes and their public transportation implications.

Given the state of the extant literature, this study proposes to address the current shortcomings mentioned above with a hybrid, micro-level, and spatially explicit agent-based modeling framework for disease spread through micromobility in an urban environment. The spread of disease in micromobility systems is influenced by the spatiotemporal distribution of both the vehicles and the human population. Like vectors, MVs can transport pathogens between various locations within the urban environment. Unlike living vectors, however, the prevalence and distribution of MVs are precisely known. The spatiotemporal dynamics of the human population, including their locations at home, school, workplace, and other places, determine access to MVs and the potential for disease transmission through contact. Thus, we create a large and realistic synthetic population and a human contact network to account for social and physical contacts across a large urban region. This contact network is dynamically updated by individuals who ride explicitly modeled MVs. Then, we propose a novel model motivated by vector-borne disease models in epidemiology to study the role of micromobility in spread of disease using individual trips from a micromobility system (Dommar et al., 2014). The study is implemented in Cook County, IL and various results of this implementation are presented to demonstrate the feasibility and scientific insights to be gained with this new modeling framework for the understanding of disease transmission through micromobility systems and for evaluating the spatiotemporal dynamic of such transmission.

There are three methods to create large-scale human contact networks: real, simulated, and semi-real networks. Simulated networks such as small worlds and scale-free networks are commonly used because real networks are rarely available for large populations. The degree distribution of individuals is used to create a simulated network. Small-world networks have high local clustering and global connectivity, while scale-free networks have a small number of nodes with many contacts (Meyers, 2007). While these networks showcase heterogeneity in contacts and are

attractive theoretically, they cannot model the spatial aspects of an epidemic, and real contact networks can differ significantly from them. Semi-real networks can be created using synthetic populations from census, travel behavior surveys, and other data sources (Harland et al., 2012). Synthetic population methods include sample-based and sample-free methods (Beckman et al., 1996; Lenormand & Deffuant, 2012; Voas & Williamson, 2000), and they are used to create very large contact networks for disease spread on national, regional, and urban scales. Examples of large synthetic populations include the TRANSIMS population synthesizer (Hobeika, 2005) and the synthetic population developed by (Eubank et al., 2004) at the Biocomplexity Institute of Virginia Tech (formerly the Virginia Bioinformatics Institute). A method originally developed by (Burger et al., 2017) is used in this research to create a large baseline contact network.

As frequently touched surfaces (fomites) are ubiquitous in transportation operations and services, they are the main points of contact between susceptible and infected individuals on MVs. The process of transmission through fomites is highly dependent on the disease agent and on individual behavior. The study of disease dynamics is traditionally limited to close contact and direct transmission between hosts or transfer through common vehicles such as water, food, and air (Cromley & McLafferty, 2011). Fomites have received relatively little attention as it is difficult to track the agents passing through surfaces. Moreover, it was previously assumed that fomites do not play a significant role in transmission, but recent studies have shown that these surfaces are far from negligible in the spread of many types of infectious diseases (Aitken & Jeffries, 2001; Kraay et al., 2018; Seto et al., 2003). Recently, epidemic models have been developed for the transmission of viruses through surfaces in indoor environments (Kraay et al., 2018; Lei et al., 2017; Li et al., 2009; Xiao et al., 2018).

The usage of public and shared transportation in urban areas is largely influenced by the perceived levels of comfort and safety, particularly during an ongoing epidemic (Chen et al., 2022; Przybylowski et al., 2021). Modeling frameworks such as the one presented in this research create a testbed to experiment with prevention and control strategies that ensure safety and comfort of the passengers.

2. Modeling framework

2.1. Overview

During epidemics, the role of public transportation systems becomes crucial in understanding two key aspects. First, it is important to determine the locations where disease transmission takes place. This can occur at the destination, as depicted in Figure 1.a and b, or within the transportation vehicles or infrastructure where people share the ambient environment and create new direct or indirect contacts that may be sufficient to contract the disease (Figure 1.c). Such understanding helps in guiding the selection of appropriate intervention strategies. Second, it is essential to assess how transportation systems influence the diffusion and dynamics of the disease. They can contribute to the spread of disease by introducing it to new geographic regions via the interregional transfer of commodities, people, animals, etc. (Figure 1.a). 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 (Rodrigue et al., 2016). The other way is through the speeding up or acceleration of diffusion in the areas that have already been affected by the disease (Figure 1.b). This happens by both increases in the volume of flows and decreases in travel time from one location to another. The impact on the final epidemic curve is that the infected population reaches its maximum more rapidly (Rodrigue et al., 2016). When transportation infrastructure is a vector, it can impact dynamics of epidemic in similar ways, either by accelerating the diffusion or exposing individuals who are rather isolated from infected populations outside of using public transportation.

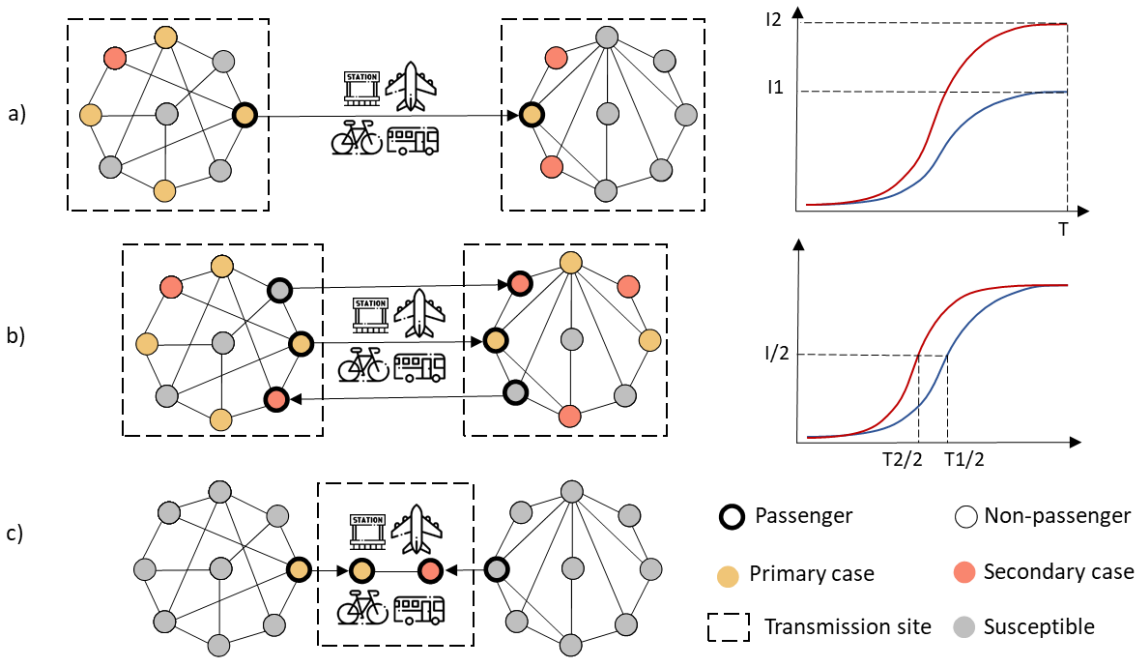


Figure 1. Contribution of transportation systems to the spread of diseases by introducing pathogens to new geographic areas (a), by accelerating their diffusion (b), or by becoming a disease vector (c).

This study focuses on a single mode of transportation, namely micromobility systems. To this end, we introduce an agent-based approach that explicitly models disease transmission through the use of MVs (Figure 1.c). Infectious individuals can potentially spread viruses by using MVs and contaminating the surfaces of these vehicles. If the viruses manage to survive on these surfaces for a sufficient duration until other individuals come into contact with the contaminated vehicle, there is a risk of disease transmission occurring. The number of contaminated MVs one uses during a day and the survival time of a virus on surfaces comprise the risk of contracting disease through a micromobility system. Disease transmission within transportation vehicles and infrastructure is independent of the purpose of a trip and of the type of trip destination. Thus, individuals may use MVs as part of their commuting trips or for other discretionary activities.

In this study, we do not explicitly model how transportation systems facilitate the spreading of disease at destinations (Figure 1.a and b). However, we still consider the movement of individuals between their main activity locations. By moving around in the city and relocating

between different main activity locations (school, work, or home) during a day, people make contacts based on proximity to others that might create sufficient pathways to transmit diseases. These contacts can be represented as edges of a spatial graph where nodes are individuals tied to different locations. The number of contacts an individual makes within different venues implies the risk of contracting the disease. This spatial contact network serves as a baseline population, as it will be referred to from now on.

Our model exclusively deals with 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.

2.2. Disease transmission

This research centers on viral disease and distinguish transmission through fomites on micromobility systems and through proximity in a baseline population. Within this context, fomites are a sufficient route of transmission, but they may not be the main pathway (Kraay et al., 2018). For example, while surfaces are the main route of transmission in Norovirus outbreaks, they are only a secondary route for rhinovirus transmission, which is sustained mainly through the droplets route (L'Huillier et al., 2015). Thus, we make a distinction between transmission through micromobility and transmission in the baseline population because, in the latter, transmission may occur through multiple pathways. Within micromobility systems, interaction between human individuals is possible only in an indirect way, namely through surfaces on MVs.

Based on the above assumptions, we propose the SIR-SC transmission model depicted in figure 2.a. 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). MVs are divided into two compartments: susceptible (S^M) and contaminated (C^M) (representing SC in the naming of the model).

With each time iteration t, a susceptible human individual may transit to an infectious state ($S^H \rightarrow I^H$) with a certain probability P^{HH}, for transmissions that occur in the baseline population, or with probability P^{MH}, for transmission that occurs on MVs. At the same time, infectious individuals progress to the recovered state ($I^H \rightarrow R^H$) after a certain number of days $1/\gamma$ where γ is recovery rate. We assume the immunity of individuals upon recovery. For MVs, a susceptible MV may become contaminated ($I^M \rightarrow C^M$) with a certain probability P^{HM} at the end of each day, and a contaminated MV may transition to the susceptible state by either disinfection or inactivation ($C^M \rightarrow S^M$).

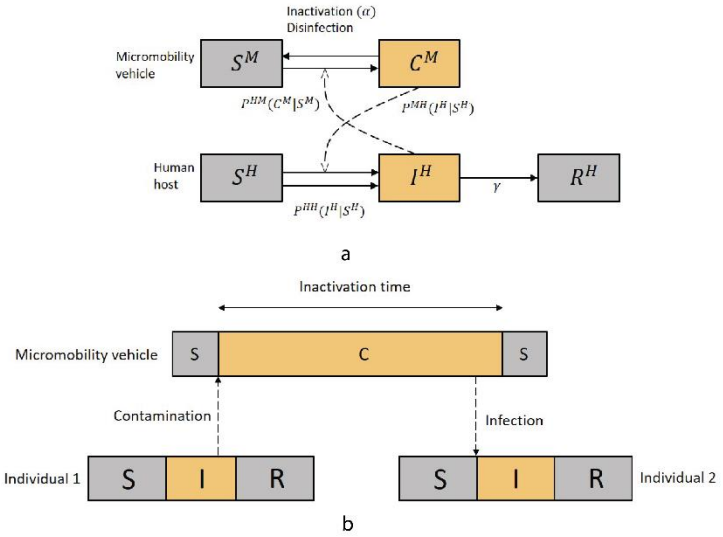


Figure 2. Disease transmission. a) SIR-SC model. b) Transmission in MVs.

2.2.1. Human-human (HH) transmission

This process models the spread of disease within the baseline population when people spend time in the same proximity to each other 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 a significant amount of time together in a certain environment, it is more likely for disease to spread in the population.

We model the transmission of disease between two hosts following Keeling and Rohani (2011) by adopting their homogeneous model for a spatial host contact network G with N nodes and l edges. Nodes represent human individuals with their locations 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 depending on day of the week. Using such a contact network, one can write a separate probability of infection for each individual based on the number of contacts at time t. Within this graph, imagine an individual i that has X contacts with infectious individuals at time 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:

$$1 - p_{i,t} = (1 - w)^{X_{i,t}}. \text{ (Eq. 1)}$$

The probability that an individual is infected following any of these contacts is then:

$$p_{i,t} = 1 - (1 - w)^{X_{i,t}} \text{ (Eq. 2)}$$

Where $p_{i,t}$ is the probability of infection for individual i following any of its current contacts with infectious individuals 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). Depending on the transmissibility of the virus (β), an infectious individual can infect a number of susceptible contacts. Let us define $\beta = -\log(1-w)$ following (Keeling & Rohani, 2011). By substituting β^{HH} for w in equation Eq. 2, we can rewrite the probability as:

$$p_{i,t} = 1 - (1-e)^{-\beta X_{i,t}} \text{ (Eq. 3)}$$

In this equation, $\beta X_{i,t}$, also called the force of infection, measures the risk of acquiring infection by individual i at time t; β is a parameter that defines how transmissible a virus is. The current model assumes the same transmission rate β for various venues, but one may calculate separate probabilities for different indoor environments and routes of transmission by acquiring appropriate transmission rates. We can rewrite Eq. 3 with more representative notations as follows:

$$P^{HH}(I^H|S^H)_{i,t} = 1 - (1-e)^{-\beta^{HH} X_{i,t}} \text{ (Eq. 4)}$$

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 baseline populations. We calculate the value of that by using the basic reproductive R_0 and recovery rates γ using equation $\beta = R_0\gamma$. In this equation, γ is easier to measure because it does not depend on the environment of transmission (1/ number of days to recover). The R_0 value is determined 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 range of values for R_0 is known for many viruses.

2.2.2 Micromobility fomite-human (MH) transmission

As far as MVs are concerned, fomites are the only possible pathway for viruses to spread. When a human individual uses contaminated MVs, 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 MVs and hands of human individuals, as well as the number of contaminated surfaces (number of MVs) one interacts with during a day. In other words, at the end of each day, the number of contaminated MVs and the rate of transmission from surfaces to hands define the risk of infection, which is used to calculate the above probability. This means that people who use micromobility transportation frequently are more likely to be infected by contaminated vehicles. Here, we are assuming that the transmission of viruses to hands is always followed by entry into the human body and subsequent infection. Another parameter that impacts transmission of disease from MVs to human is the ability of the virus to survive on the micromobility surfaces (Figure 1.b). A virus must survive long enough on MVs for the next human individual using the vehicle to be exposed. We assume transmission through an MV occurs until the virus is naturally inactivated or disinfected through intervention. Thus, a single contaminated MV can infect multiple human individuals.

Using the same process as in section 2.2.1, we 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 $Y_{i,t}$ contaminated MVs during time step t.

$$P^{MH}(I^H|S^H)_{i,t} = 1 - (1 - e)^{-\beta^{MH}Y_{i,t}} \text{ (Eq. 5)}$$

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 MVs, and β^{MH} is a parameter that determines transmissibility of the virus from MV surfaces to human body.

2.2.3. Human-micromobility fomite (HM) transmission

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

We use a similar strategy as in the previous two sections to calculate the probability of contamination of a micromobility vehicle b when $Z_{b,t}$ infectious human individuals ride the vehicle during time step t.

$$P^{HM}(I^M|S^M)_{b,t} = 1 - (1 - e)^{-\beta^{HM}Z_{b,t}} \text{ (Eq. 6)}$$

Where $P^{HM}(I^M|S^M)_{b,t}$ 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 body to MV surfaces, namely the shedding rate. Here, we are assuming that after the shedding of the virus, it consistently contaminates hands first and then subsequently vehicle surfaces.

Field surveys are needed to determine the values of β^{HM} and β^{MH}. In this study we 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, we also assume β^{HM} and β^{MH} have the same value.

$$\beta^{HM} = \beta^{MH} = (1/\rho)\beta^{HH} \text{ (Eq. 7)}$$

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

Using equations 4, 5, and 6 we calculate the transmission of virus during each time step t in the baseline population and in the micromobility system.

2.2.4. Inactivation

Virus must survive long enough on micromobility surfaces until a new host picks up the vehicle and becomes infected with the virus. We represent the inactivation rate of the virus on micromobility surfaces with α.

2.2.5. Recovery

The recovery rate of individuals is independent of the environment. Individuals transition from the infectious state to the recovered state at rate γ. For the sake of simplification, we assume individuals are immune after infection for the duration of the epidemic. This 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.

2.2.6. Disinfection

Disinfection may occur in either of two ways in the model. The first one is by disinfecting bikes. This is handled by assuming that, each day, a random number Q of MVs 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 1. Overview of processes and parameters

Process	Parameter	Description
Transmission	R_0	Basic reproductive rate
	$\beta^{HH}, \beta^{MH}, \beta^{HM}$	Transmission rate in human contact network, transmission rate from MV to human, and transmission rate from human to MV
	$P^{HH}(I^H \mid S^H)_{i,t}$	Human-human (HH), micromobility fomite-human (MH), and human-micromobility fomite (HM) risk of infection
	$P^{MH}(I^H \mid S^H)_{i,t}$	Human-human (HH): probability of transition from $S^H \rightarrow I^H$ for human individual i at time t given contacts with other infectious humans
	$P^{HM}(C^M \mid S^M)_{b,t}$	Micromobility fomite-human (MH): probability of transition from $S^M \rightarrow C^M$ for MV b at time t given contacts with other infectious human individuals
	ρ	Relative transmission factor
Recovery	γ	Recovery rate
Inactivation	α	Inactivation time
Social contacts	$k = 31$[2] $(school\ \&\ households)$ $k = 9$[3] $(workplace)$	Threshold for the average number of close contacts per person at school and households (Salathé et al., 2010) as well

[2] See appendix section S.6.5

[3] Same as previous.

as workplaces (Zhang et al., 2019).

3. Agent-based model

A spatially explicit agent-based model is designed to simulate the spatial and temporal dynamics of a viral disease in an urban area, when a transmission pathway is through surfaces on MVs following the framework presented in section 2. The proposed approach consists of three modules, as depicted in Figure 3. The first module generates the synthetic baseline population and the network of contacts among human agents using census data and auxiliary data sets about roads, school locations, and commuting patterns, and sets up the micromobility system. The second module is the relocation module, which simulates the movement of individuals and MVs within the urban area. The third module is the disease transmission module, which models the spread of the disease within both MVs and the baseline population. The model uses multiple discrete-time intervals for different processes. Relocations of human agents and of MVs occur every hour, and disease transmission is calculated at the end of each day. In this section, we discuss the structure of agents, geographic spaces, and modules in the model.

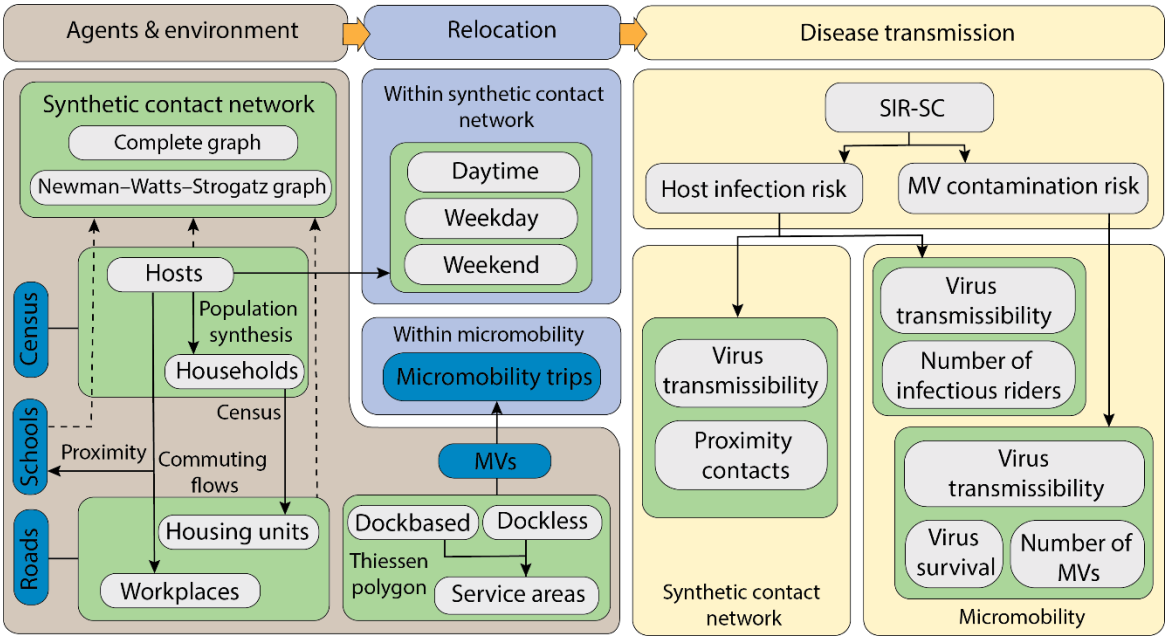


Figure 3: Proposed approach of the agent-based model with three modules: Synthetic contact network, relocation, and disease transmission.

3.1. Agents and environment

We start with the synthetic contact network module. Given the customary lack of microlevel data on urban populations, the gender, age, and household type information in each census tract is used to simulate and assign host agents to households using a Gaussian distribution. This is done using tract level census data and a sample free population synthesis method, as proposed by (Burger et al., 2017). Host agents are independent and heterogeneous, meaning they are characterized by age, gender, household ID, workplace/school ID, and health status, being one of

susceptible, infectious, or recovered (Figure S.4). Each host individual has an activity and travel schedule and may go to either school (age<18) or work (age>18) 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.

The model does not require microlevel data on residence and workplace locations. Instead, we extract the number of occupied housing units within each census unit (specifically, US census tracts) and generate them at least 50m apart along local road networks. Then, we randomly populate houses with households within each census tract. The number of workplaces countywide is extracted from census business establishment counts and downscaled to tracts in proportion to the population size. Then, they are generated at least 20m apart along the secondary or local road networks. We randomly assign work-age agents to a workplace location within a census tract based on the origin-destination statistics. School locations with their grade and enrollment level are extracted from publicly available data sets. School-age individuals are assigned to the nearest school based on grade and enrollment levels.

A large-scale social network is generated based on social ties 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 (Salathé et al., 2010). When the number of individuals in households (including group quarters), schools, or workplaces are larger than the average number of contacts k for an individual in these groups, a Newman–Watts–Strogatz small-world graph (Newman & Watts, 1999) is created indicating each individual makes contacts with k closest members in that group. Otherwise, a complete contact network is created for the group.

MVs are characterized by their contamination status, being susceptible or contaminated, and by their current location (latitude and longitude for dockless micromobility or station for dockbased systems). Micromobility systems also have service areas defined based on travel distance on the road network, which can be determined around micromobility stations or MVs depending on type of system (dockbased or dockless). A unique property of the proposed model is that it is designed to make use of real micromobility trips made by human individuals in urban areas, when they are available. At the same time, the model can be adapted to include trips with other modes of transportation if such data happen to be available.

Disease agents are not explicitly modeled, and 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.

3.2. Relocation

Two types of mobile agents are present in the model: MVs and hosts (Relocation module in Figure 3). MVs relocate between start and end station service areas or start and end locations (for dockless systems) based on their real trip records. The purpose of micromobility trips is not central to our model and a micromobility trip can be the only segment of a trip to work, home, or

any other discretionary activity, or it can be a partial segment of a larger trip to one of these places.

The synthetic human contact network described above determines the proximity contacts for each individual, therefore defining the risk of infection. However, it also serves a second purpose in the model, which is defining the current spatial distribution of population within micromobility service areas. Depending on their daytime schedule and their current location (at home, school, or work), hosts may initiate a micromobility trip based on the availability of MVs at the closest station/location. During weekends, individuals stay at home indicating they are in the micromobility services areas near their housing units. Human individuals are assigned to actual historic micromobility trips based on their current location, and demographic attributes in a stochastic way (Figure 4). As mentioned earlier, the purpose of micromobility trips is not important for the model. Thus, people may use MVs as part of their commuting trips or for other discretionary purposes.

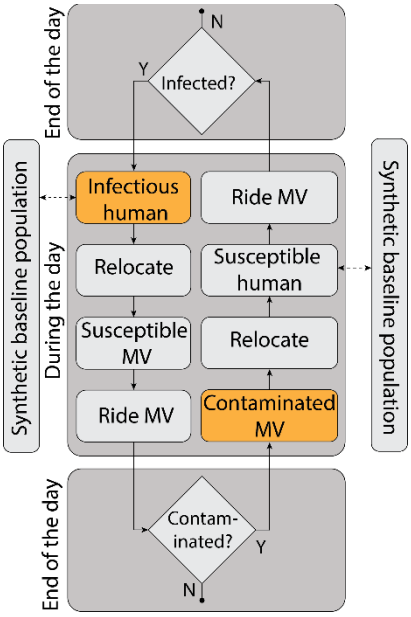


Figure 4: Relocation process

3.3. Disease transmission

We use the SIR-SC model proposed in section 2.2 to calculate the transmission of disease within urban areas. As the simulation unfolds, individual agents follow their activity schedule to go to work/school or stay home. If they are infected, they may transmit viruses on the surfaces on MVs or infect other susceptible individuals in their social networks. Conversely, if they are susceptible, they could ride contaminated MVs and contract the disease, or acquire the disease from infectious individuals in their proximity.

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 an MV is calculated based on interactions with individuals during a day and transmissibility of the virus. Constrained by the survival of

virus on surfaces of the MVs, infection of an individual is calculated based on the interaction with MVs and other individuals in the population all at once at the last hour of the day.

4. Cook County case study

Cook County, located in the state of Illinois, has a population of over 5.1 million people, with 2.7 million residing in the City of Chicago. Being part of the third largest metropolitan area in the United States and well-connected to other major cities nationally and internationally, the region is susceptible to the emergence and spread of diseases. In this case study, we focus on the Divvy bikesharing system, which has been operational in Chicago since 2013 and has been the subject of previous research. The Divvy system, with bicycles as the MVs (MVs), is utilized for this study due to the availability of trip data, including demographic information of users. Although the shared bike stations are primarily located in the City of Chicago, we model the baseline population for the entire Cook County to capture a broader representation (see figure S2)

4.1. Synthetic contact network

We use the methodology proposed by (Burger et al., 2017) 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 section S6 of the supplementary materials). The process includes four steps:

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

After placing individuals into households, we 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, we compare these synthesized values with their actual value for the census tracts using the percentage error (E^{syn}) with the following equation:

$$E^{syn} = \frac{Synthesized - Census}{Census} \times 100 \quad \text{(Eq. 8)}$$

4.2. Divvy bikesharing system

In this study, we utilized historical micromobility data from the Divvy bikesharing system spanning the months of June and July 2018. Divvy bikesharing system is categorized as a station-based system, meaning that all micromobility trips originate and end at specific locations. The data set encompasses two primary tables, namely, bikesharing stations and trips. Each station is characterized by its geographic location, name, and capacity. The trip table logs all trips, including the associated bike ID, start and end date and time, start and end station, gender, and birth year. The urban space is partitioned into areal units based on service areas that are established according to the travel distance on the road network. These areas are used in the

model to assign trips to human individuals currently present in the service area and to represent the current location of bicycles.

The population mobility is calculated on an hourly basis to consider the daily commute of individuals to work/school. At each hour, trips originating from a specific station are randomly assigned to individuals present within the service area based on the individual and trip attributes. In instances where no individuals with matching age and gender to the trip are accessible within the service area, an individual with the appropriate age and gender is assigned to the trip via a random selection process from the entire population.

4.3. Programming and software

The synthetic contact network is implemented by adapting codes from (Burger et al., 2017) for Cook County. All codes are in Python. Graph-tool[4], and Networkx[5] packages were used for the manipulation and computations on graphs. 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 road network analysis and some of the visualizations. A pseudo-code for the sequential version of the SIR-SC model is available in figure S1 and codes are available to access in a Git repository online. The parallelized implementation of the model will be presented elsewhere.

5. Results and discussion

5.1. Human contact network

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 S6 by measuring the percentage error (equation 8). 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 shown in figure S6(a), and the other two attributes are shown in figure S6(b). In general, the number of households with minors and seniors S6(b) show larger margins of error.

A detailed description of the errors is available in table 2. 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 is due to varying number of them in group quarters and households. Overall, margins of error are small and deemed acceptable for the intended purposes of our simulation model.

[4] https://graph-tool.skewed.de/
[5] https://networkx.org/

Table 2. Accuracy assessment of the synthetic population (based on percentage errors)

Statistic	Avg family size	Avg household size	Families with seniors >65	Families with children < 18
Count	1314	1314	1314	1314
Avg	3.4	-0.37	2.88	10.96
Std	3.38	0.58	9.58	15.35
Min	-10.49	-8.35	-59.52	-41.66
Med	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. Table 3 shows the statistics for these two contact networks. The clustering coefficient of the contact network is 0.4 which is significantly higher than a random graph of the same size and degree distribution. Note that household contacts include individuals who are living in group quarters such as dormitories, nursing homes, or prisons. These results are based on average number of close contacts per person $k = 31$ for school and households, and $k = 9$ for workplaces (see section S 6.5 for more details).

Table 3. Human contact network descriptive statistics

		# Nodes	Household & group quarter contacts	Work contacts	School contacts	Total contacts
Count	Weekday	5,170,895	7,971,408 (19%)	9,781,860 (24%)	23,760,993 (57%)	41,514,261
	Weekend	5,170,895	7,971,408 (100%)	0	0	7,971,408
Avg			3.1	3.8	10.0	16.1
Min			0	0	0	0
Med			2	0	0	11
Max			55	21	56	90

5.2. Micromobility

We processed shared bike data of Divvy system for June and July 2018. After corrections (see section S3 in supplemental materials), 564 active stations, 1,209,808 trips with 5,874 bicycles are recorded for the period of study. 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 m with standard deviation of 244 m. The farthest trip in the data set is 35 km, which has happened in 2.5 hours. Men use public bicycles more than women, but with the same hourly pattern. Similar patterns are observed for weekends and holidays. Figure S3 shows the duration, length, and speed of the trips for the entire period of study.

5.3. Sensitivity analysis and scenarios

To study the sensitivity of the epidemic dynamics to the variations of the main parameters of our model, we conduct sensitivity analysis. The parameters we focus on are virus inactivation time (α), number of bikes B, basic reproductive rate R_0, and relative transmission factor ρ. Depending on the experiment, we implement a different number of replications, ranging between 25-50. In general, a higher number of replications is recommended (at least 100). However, because we use the same trips (actual micromobility trips) during the study period and because most of the variables are not stochastic (random variables), a smaller 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 we present hereunder. We also, measure the sensitivity of the model to the starting location of the epidemic and find the Divvy stations more at risk of contamination.

5.3.1. 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 MVs. If the time period between two consecutive micromobility trips on the same MV 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 a sensitivity analysis for this parameter by adjusting its values $\alpha = [10\ min, 30\ min, 1\ hour, 2\ hours, 4\ hours, 6\ hours,\ 16\ hours]$. Results are shown in Figure 4. 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. The inactivation time needs to be over 6 hours for more than 100 new infections to be logged.

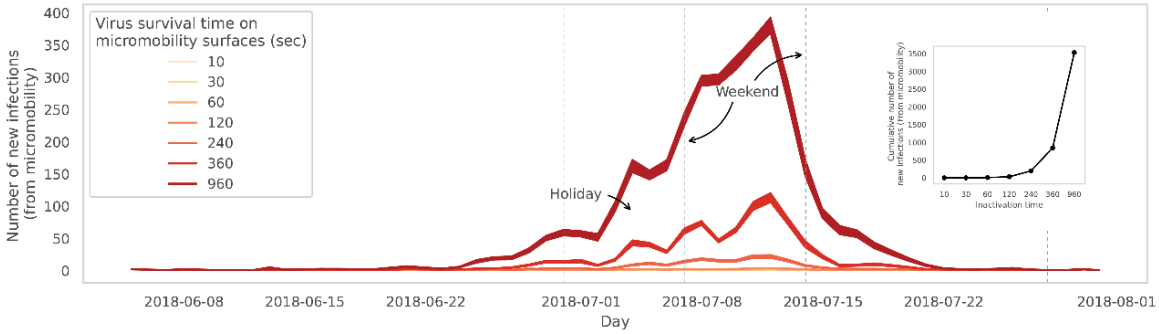


Figure 5. Virus inactivation time impact on disease spread in Divvy system. The number of new infections per day in Divvy system and the cumulative number of infected individuals for different inactivation times (small panel). Initial conditions: $R_0 = 2.5, \rho = 2, \alpha = [10\ min, 30\ min, 1\ hours, 2\ hours, 4\ hours, 6\ hours, 16\ hours], \gamma = 0.2, I_0^H = 100, I_0^M = 100, rep = 25$.

5.3.2. Emergence of a viral disease through micromobility system

In this experiment, we 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, we count the number of times an outbreak occurs after 50 replications. We 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, we conduct a multi-variable sensitivity analysis for $R_0 = [0.5, 1, 1.5, 3, 5]$ and $\rho = [2, 5, 10]$ 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, we include R_0 values under one in the series of experiments.

Figure 5.a 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 $\rho = 2$. 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, outbreaks are likely to happen. Out of 50 replications, one outbreak has occurred in a scenario with $\rho = 2$ when the $R_0 = 1$.

5.3.3. Dynamics of viral disease in micromobility systems

In order to study the impact of micromobility on dynamics of the viral disease, we design a worst-case scenario with high basic reproductive rate ($R_0 = 2.5$) and low relative transmission factor ($\rho = 2$). Then, we 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 (Rodrigue et al., 2016). Since our model has considered individual trips it is possible to directly measure the infections that occur in micromobility system. We implement the scenario with and without micromobility transportation. In this experiment, we 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. We use the current bicycles and trips in the data set to impute new bikes and trips. To create new bikes, we 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, we 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. We design four experiments with 50%, 100%, 300%, and 500% increase in the number of bicycles, respectively

Table 4 shows the infected population and their percentage increase compared to the controlled experiment. The maximum increase in the infected population (ΔP) is 28,921 new cases which is less than one percent increase in the extent of the overall infected population. For the Divvy system with current number of bicycles and trips, 5,409 (0.121%) new infections are added by

the end of the epidemic. While these ΔP values 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 4 shows the number of infected individuals in the micromobility system at the end of the epidemic (also see figure 5.b). Comparing these numbers with ΔP (column 4) reveals that the number of infected individuals can be higher by 14% for scenario with six times more bicycles. This finding is key in our study 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 public transportation systems. Compared to the experiment without micromobility, the epidemic reaches its half course less than one day earlier (ΔT_{mid}<one day) which is negligible.

Table 4. Comparison of population infected in the presence and absence of micromobility.

No. Bike	Total infection from micromobility	Total infection in micromobility & baseline population (P)	ΔP	$\Delta P(\%)$
0	0	4,466,785 (±186.4)	-	-
5,874 (current)	5,417 (±47.8)	4,472,194 (±196.8)	5,409	0.121
current × 0.5	8,110 (±47.5)	4,475,245 (±239.9)	8,460	0.189
current × 2	10,803 (±101.0)	4,477,143 (±252.2)	10,358	0.232
current × 4	21,127 (±89.7)	4,485,945 (±222.9)	19,160	0.429
current × 6	32,902 (±186.0)	4,495,706 (±314.6)	28,921	0.647

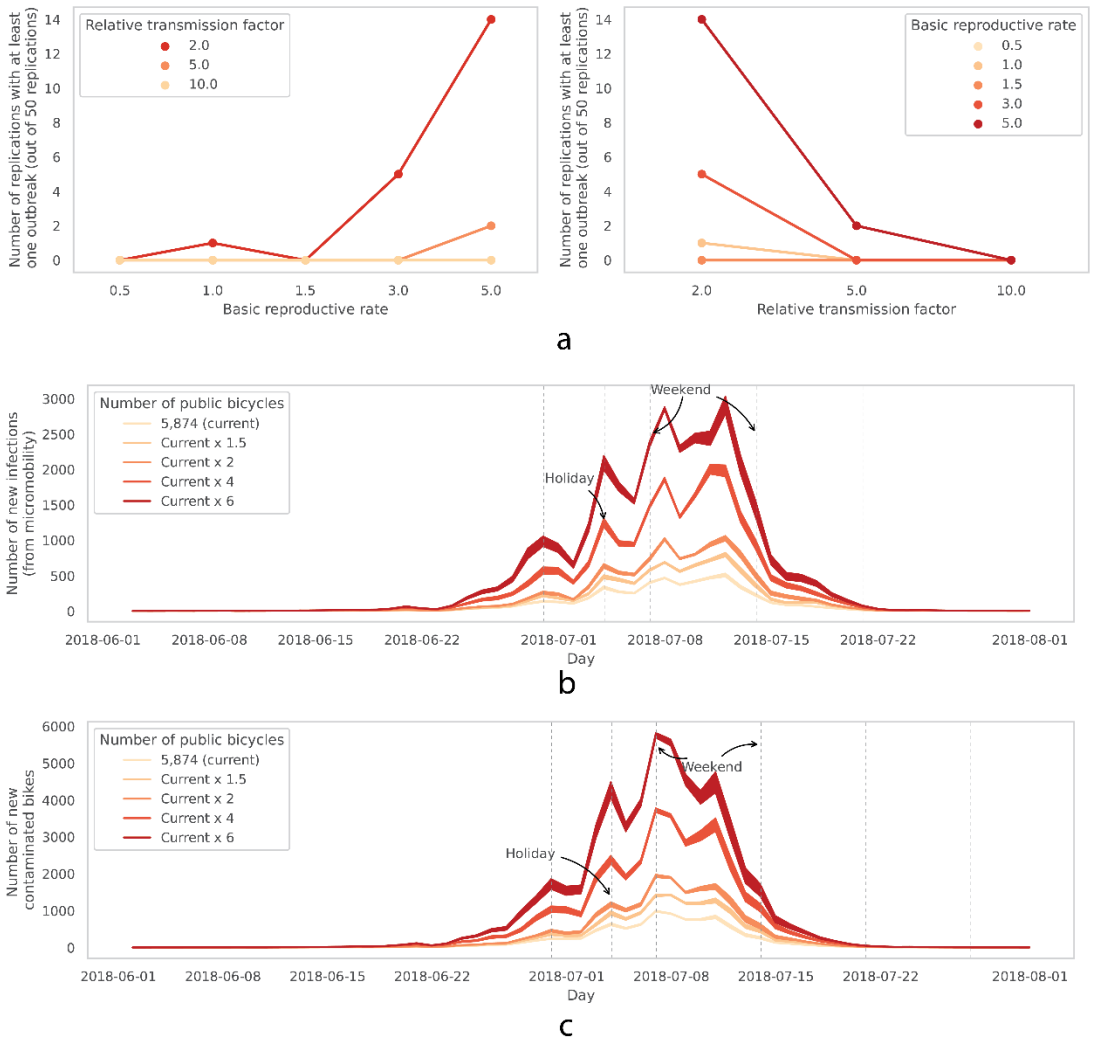


Figure 6. Emergence (a) and dynamics (b and c) of viral disease through micromobility. (a) Emergence of outbreaks through micromobility. Initial conditions: $R_0 = (0.5, 1, 1.5, 3, 5), \rho = (2, 5, 10), \alpha = one\ day, \gamma = 0.2, I_0^H = 0, I_0^M = 1, rep = 50$. (b and c) Daily new infections in micromobility and daily new bike contaminations. Initial conditions: $R_0 = 2.5, \rho = 2, \alpha = one\ day, \gamma = 0.2, I_0^H = 100, I_0^M = 0, rep = 25$.

5.3.4. Risk of exposure to viral disease in micromobility

We studied the spatial pattern of exposure risk to viral contamination in MVs. To study the spatial pattern of exposure, we 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, $R_0 = 2.5$, $\rho = 2$, $\gamma = 0.2$, and survival time of one day. Then, we rank the stations based on the cumulative number of contaminated bikes. Figure 6 shows the spatial distribution of the epidemic in terms of the cumulative number of contaminated bikes visiting each station at the end of epidemic. 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, or have no trips, which means they are out of service for the period of study (green service areas). The spatial pattern 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.

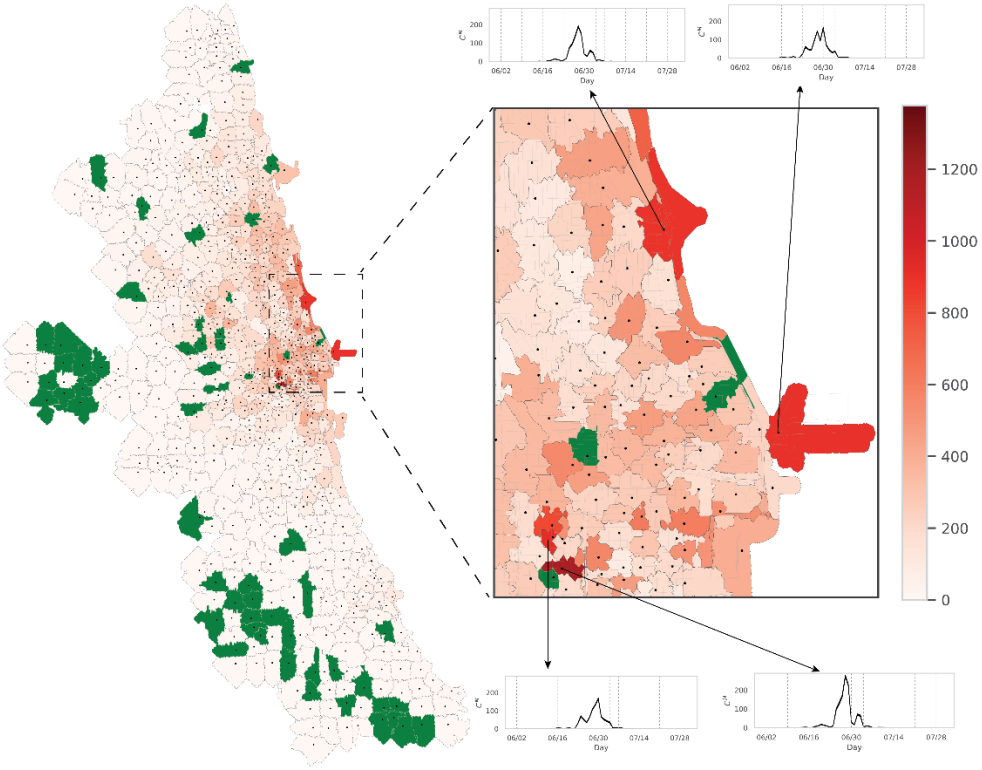


Figure 7. Spatial and temporal patterns of exposure risk to contaminated bicycles. The service areas in green have no exposure to contaminated bicycles by the end of epidemic. These stations have either a small number of trips or are out of service for the period of study. Stations at high risk of exposure are concentrated around downtown and the central business district. Initial conditions: $R_0 = 2.5, \rho = 2, \alpha = one\ day, \gamma = 0.2, I_0^H = 100, I_0^M = 0, rep = 25$.

6. Discussion and concluding remarks

This study aimed to examine the role of micromobility transportation in the spread of viral diseases in dense urban areas, focusing on the Chicago Divvy bikesharing system. We developed a spatially-explicit agent-based model to simulate disease transmission. Our findings showed that even with just a single contaminated bicycle in Cook County, an outbreak can occur. Although the overall impact of the Divvy system on infection incidence is relatively small, neglecting the role of micromobility as a disease vector may underestimate its contribution by 14%. This underscores the need for individual-based models to accurately assess the impact of public

transportation systems. Exposure risk to viruses through micromobility followed the pattern of trips between origins and destinations, particularly in central business districts and downtown areas.

From a public transportation perspective, this research provides a platform to study disease spread within micromobility systems by considering vehicles as potential vectors. The model offers a more accurate measure of the role of public transportation in disease transmission and can be adapted to other transportation modes.

From a modeling perspective, we 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 MVs. We 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 us to incorporate actual spatial and temporal dynamics (e.g., seasonality) of MVs. 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 individual behavior can now be entered as an input to ABMs. Third, we 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.

At the same time, there are some limitations in the model. Since we 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. We 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 section 6 in supplementary materials).

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.

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

With respect to the human contact network, we 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 may be 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 data.

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.

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.

References

Tiago P. Peixoto, “The graph-tool python library”, figshare. (2014) DOI: 10.6084/m9.figshare.1164194 [sci-hub, @tor]

Hagberg, A., Swart, P., & S Chult, D. (2008). *Exploring network structure, dynamics, and function using NetworkX* (No. LA-UR-08-05495; LA-UR-08-5495). Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

Aitken, C., & Jeffries, D. J. (2001). Nosocomial spread of viral disease. *Clinical Microbiology Reviews*, *14*(3), 528–546.

Ajelli, M., Gonçalves, B., Balcan, D., Colizza, V., Hu, H., Ramasco, J. J., Merler, S., & Vespignani, A. (2010). Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infectious Diseases*, *10*(1), 1–13.

Baker, R. E., Mahmud, A. S., Miller, I. F., Rajeev, M., Rasambainarivo, F., Rice, B. L., Takahashi, S., Tatem, A. J., Wagner, C. E., Wang, L. F., Wesolowski, A., & Metcalf, C. J. E. (2022). Infectious disease in an era of global change. In *Nature Reviews Microbiology* (Vol. 20, Issue 4, pp. 193–205). Nature Research. https://doi.org/10.1038/s41579-021-00639-z

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., & Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. In *PNAS December* (Vol. 22, Issue 51).

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

Borjigin, S. G., He, Q., & Niemeier, D. A. (2023). COVID-19 transmission in U.S. transit buses: A scenario-based approach with agent-based simulation modeling (ABSM). *Communications in Transportation Research*, *3*, 100090. https://doi.org/10.1016/j.commtr.2023.100090

Bóta, A., Gardner, L. M., & Khani, A. (2017). Identifying Critical Components of a Public Transit System for Outbreak Control. *Networks and Spatial Economics*, *17*(4), 1137–1159. https://doi.org/10.1007/s11067-017-9361-2

Burger, A., Oz, T., Crooks, A., & Kennedy, W. G. (2017). Generation of realistic mega-city populations and social networks for agent-based modeling. *Proceedings of the 2017 International Conference of The Computational Social Science Society of the Americas*, 1–7.

Chen, C., Feng, T., Gu, X., & Yao, B. (2022). Investigating the effectiveness of COVID-19 pandemic countermeasures on the use of public transport: A case study of The Netherlands. *Transport Policy*, *117*, 98–107. https://doi.org/10.1016/j.tranpol.2022.01.005

Cromley, E. K., & McLafferty, S. L. (2011). *GIS and public health*. Guilford Press.

Crooks, A. T., & Hailegiorgis, A. B. (2014). An agent-based modeling approach applied to the spread of cholera. *Environmental Modelling & Software*, *62*, 164–177.

Dommar, C. J., Lowe, R., Robinson, M., & Rodó, X. (2014). An agent-based model driven by tropical rainfall to understand the spatio-temporal heterogeneity of a chikungunya outbreak. *Acta Tropica*, *129*(1), 61–73. https://doi.org/10.1016/j.actatropica.2013.08.004

Du, Z., Fox, S. J., Holme, P., Liu, J., Galvani, A. P., & Meyers, L. A. (2018). Periodicity in movement patterns shapes epidemic risk in urban environments. *ArXiv Preprint ArXiv:1809.05203*.

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

Germann, T. C., Kadau, K., Longini, I. M., & Macken, C. A. (2006). Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences*, *103*(15), 5935–5940.

Harland, K., Heppenstall, A., Smith, D., & Birkin, M. H. (2012). Creating realistic synthetic populations at varying spatial scales: A comparative critique of population synthesis techniques. *Journal of Artificial Societies and Social Simulation*, *15*(1).

Hobeika, A. (2005). TRANSIMS fundamentals: Chapter 3: Population synthesizer. *US Department of Transportation*.

Keeling, M. J., & Rohani, P. (2011). *Modeling infectious diseases in humans and animals*. Princeton university press.

Kerr, C. C., Stuart, R. M., Mistry, D., Abeysuriya, R. G., Rosenfeld, K., Hart, G. R., Núñez, R. C., Cohen, J. A., Selvaraj, P., Hagedorn, B., George, L., Jastrzębski, M., Izzo, A. S., Fowler, G., Palmer, A., Delport, D., Scott, N., Kelly, S. L., Bennette, C. S., … Klein, D. J. (2021). Covasim: An agent-based model of COVID-19 dynamics and interventions. *PLoS Computational Biology*, *17*(7). https://doi.org/10.1371/journal.pcbi.1009149

Kraay, A. N. M., Hayashi, M. A. L., Hernandez-Ceron, N., Spicknall, I. H., Eisenberg, M. C., Meza, R., & Eisenberg, J. N. S. (2018). Fomite-mediated transmission as a sufficient pathway: a comparative analysis across three viral pathogens. *BMC Infectious Diseases*, *18*(1), 1–13.

Lazarus, J., Pourquier, J. C., Feng, F., Hammel, H., & Shaheen, S. (2020). Micromobility evolution and expansion: Understanding how docked and dockless bikesharing models complement and compete – A case study of San Francisco. *Journal of Transport Geography*, *84*. https://doi.org/10.1016/j.jtrangeo.2019.102620

Lei, H., Li, Y., Xiao, S., Yang, X., Lin, C., Norris, S. L., Wei, D., Hu, Z., & Ji, S. (2017). Logistic growth of a surface contamination network and its role in disease spread. *Scientific Reports*, *7*(1), 1–10.

Lenormand, M., & Deffuant, G. (2012). Generating a synthetic population of individuals in households: Sample-free vs sample-based methods. *ArXiv Preprint ArXiv:1208.6403*.

L'Huillier, A. G., Tapparel, C., Turin, L., Boquete-Suter, P., Thomas, Y., & Kaiser, L. (2015). Survival of rhinoviruses on human fingers. *Clinical Microbiology and Infection*, *21*(4), 381–385.

Li, S., Eisenberg, J. N. S., Spicknall, I. H., & Koopman, J. S. (2009). Dynamics and control of infections transmitted from person to person through the environment. *American Journal of Epidemiology*, *170*(2), 257–265.

Linard, C., Ponçon, N., Fontenille, D., & Lambin, E. F. (2009). A multi-agent simulation to assess the risk of malaria re-emergence in southern France. *Ecological Modelling*, *220*(2), 160–174.

Meyers, L. (2007). Contact network epidemiology: Bond percolation applied to infectious disease prediction and control. *Bulletin of the American Mathematical Society*, *44*(1), 63–86.

Mo, B., Feng, K., Shen, Y., Tam, C., Li, D., Yin, Y., & Zhao, J. (2021). Modeling epidemic spreading through public transit using time-varying encounter network. *Transportation Research Part C: Emerging Technologies*, *122*. https://doi.org/10.1016/j.trc.2020.102893

Newman, M. E. J., & Watts, D. J. (1999). Renormalization group analysis of the small-world network model. *Physics Letters A*, *263*(4–6), 341–346.

Nicolaides, C., Avraam, D., Cueto-Felgueroso, L., González, M. C., & Juanes, R. (2020). Hand-Hygiene Mitigation Strategies Against Global Disease Spreading through the Air Transportation Network. *Risk Analysis*, *40*(4), 723–740. https://doi.org/10.1111/risa.13438

Nita Bharti. (2021). Linking human behaviors and infectious diseases. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(11). https://doi.org/10.1073/pnas.2005241118

Pastor-Satorras, R., Castellano, C., Van Mieghem, P., & Vespignani, A. (2014). *Epidemic processes in complex networks*. https://doi.org/10.1103/RevModPhys.87.925

Przybylowski, A., Stelmak, S., & Suchanek, M. (2021). Mobility behaviour in view of the impact of the COVID-19 pandemic-public transport users in gdansk case study. *Sustainability (Switzerland)*, *13*(1), 1–12. https://doi.org/10.3390/su13010364

Rahmandad, H., & Sterman, J. (2008). Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management Science*, *54*(5), 998–1014.

Riley, S., Eames, K., Isham, V., Mollison, D., & Trapman, P. (2015). Five challenges for spatial epidemic models. *Epidemics*, *10*, 68–71. https://doi.org/10.1016/j.epidem.2014.07.001

Rodrigue, J.-P., Comtois, C., & Slack, B. (2016). *The geography of transport systems*. Routledge.

Salathé, M., Kazandjieva, M., Lee, J. W., Levis, P., Feldman, M. W., & Jones, J. H. (2010). A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences*, *107*(51), 22020–22025.

Seto, W. H., Tsang, D., Yung, R. W. H., Ching, T. Y., Ng, T. K., Ho, M., Ho, L. M., Peiris, J. S. M., of Expert SARS group of Hospital Authority, A., & others. (2003). Effectiveness of precautions against droplets and contact in prevention of nosocomial transmission of severe acute respiratory syndrome (SARS). *The Lancet*, *361*(9368), 1519–1520.

Shaheen, S., & Chan, N. (2016). *MOBILITY AND THE SHARING ECONOMY 573 BUILT ENVIRONMENT VOL 42 NO 4 Mobility and the Sharing Economy: Potential to Facilitate the First-and Last-Mile Public Transit Connections*.

Shaheen, S., Cohen, A., & Broader, J. (2021). *What's the "big" deal with shared micromobility? Evolution, curb policy, and potential developments in North America*.

Sun, L., Axhausen, K. W., Lee, D. H., & Huang, X. (2013). Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(34), 13774–13779. https://doi.org/10.1073/pnas.1306440110

Truszkowska, A., Behring, B., Hasanyan, J., Zino, L., Butail, S., Caroppo, E., Jiang, Z. P., Rizzo, A., & Porfiri, M. (2021). High-Resolution Agent-Based Modeling of COVID-19 Spreading in a Small Town. *Advanced Theory and Simulations*, *4*(3). https://doi.org/10.1002/adts.202000277

Voas, D., & Williamson, P. (2000). An evaluation of the combinatorial optimisation approach to the creation of synthetic microdata. *International Journal of Population Geography*, *6*(5), 349–366.

Xiao, S., Li, Y., Sung, M., Wei, J., & Yang, Z. (2018). A study of the probable transmission routes of MERS-CoV during the first hospital outbreak in the Republic of Korea. *Indoor Air*, *28*(1), 51–63.

Zhang, N., Tang, J. W., & Li, Y. (2019). Human behavior during close contact in a graduate student office. *Indoor Air*, *29*(4), 577–590. https://doi.org/10.1111/ina.12554

Supplemental materials

S1. Pseudo code:

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 other three variables are transmission rates.*

Output: (C^M, I^{HM}, I^{HH}) *where* C^M *records the newly contaminated MVs each day,* I^{HM} *records newly infected hosts due to riding contaminated MVs each day, and* I^{HH} *records newly infected hosts in the baseline population each day.*

Temporary variables: $(health_{state}, contamination_{state})$

```
# initialize the environment (Λ)
for t = 0 to T do:
  for each person v ∈ G do
    # update locaiton of v according to day of week and time of day.
  for each trip f ∈ 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.
    for each rider i ∈ F(day) do
      # get intervention A and apply.
      # calculate probability of infection for rider i using β^MH and number of contaminated MVs
        used by rider i.
    for each micromobility vehicle b ∈ F(day) do
      # get intervention A and apply
      # calculate probability of contamination for MV b using β^HM and number of infectious
        individuals that used b.
    for each infectious node u ∈ G do
      for each contac w of node u do
        if w is susceptible then
          # calculate probability of infection for individual w using using β^HH and number
          # of infectious contacts of w.
    for each infectious node u ∈ G do
      if individual u passed infectious period then
        # update health_state u to recovered.
    for each contaminated micromobility b do
      if duration of contamination b > α then
        # update contamination_state b to clean.
Output(C^M, I^HM, I^HH)
```

Figure S1. Pseudo code.

S2. Study area:

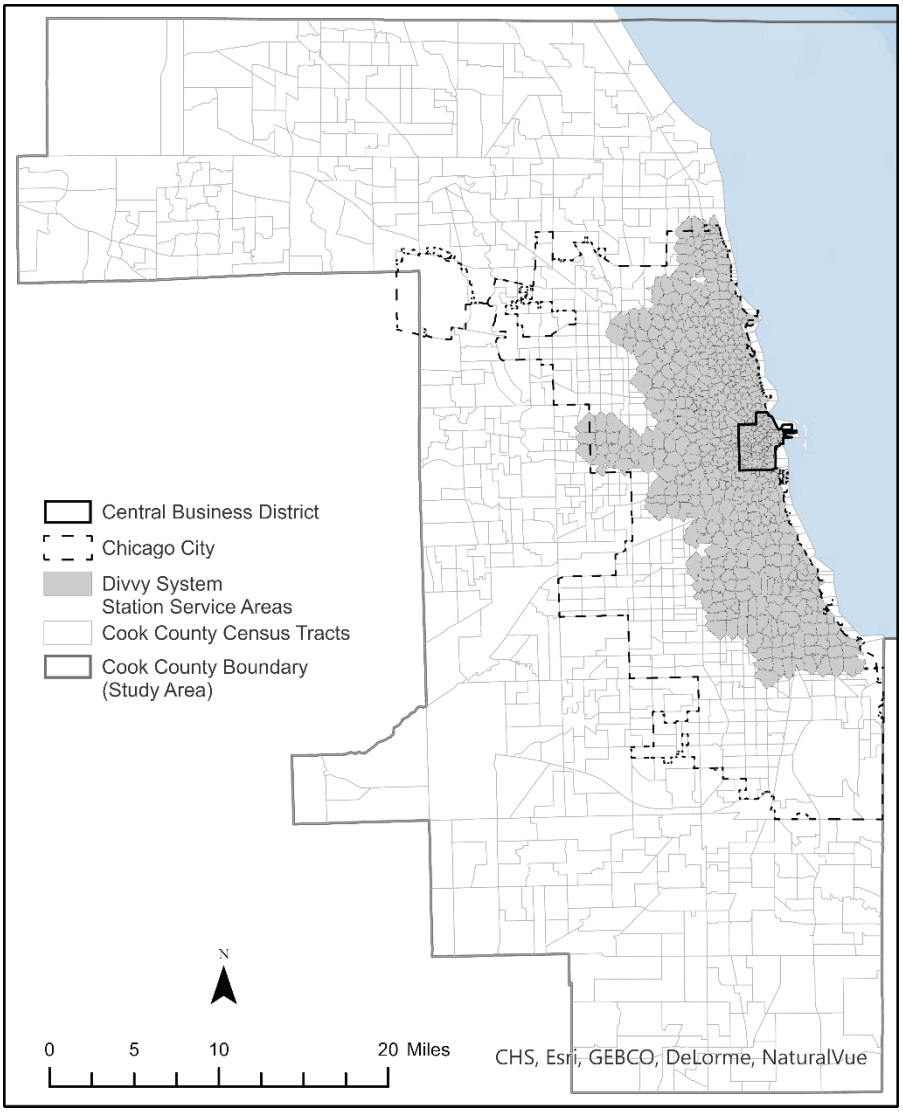


Figure S2. Study area.

S3. Micromobility trips:

We 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. We 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 zero (26,973 trips). We 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. We also calculate the speed of trips. It is less likely that speed is faster than 25 km/h, and we remove those trips from the database as well (14 trips). 308,359 records do not include the age of the cyclist. We use the age distribution by gender from the rest of the database to impute age values for these records. We 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 S1 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, we demonstrate the hourly distribution of starting Divvy trips only for July 2018 in figure S3(a) and figure S3(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 S3(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 individual's trip from the data set.

S4. 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 study period. If a station has no record during the time period of our study, we check if trips are available outside of that period within 2018. If the trips are available before and after, we keep the station in the micromobility stations table. Otherwise, we 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 S2.

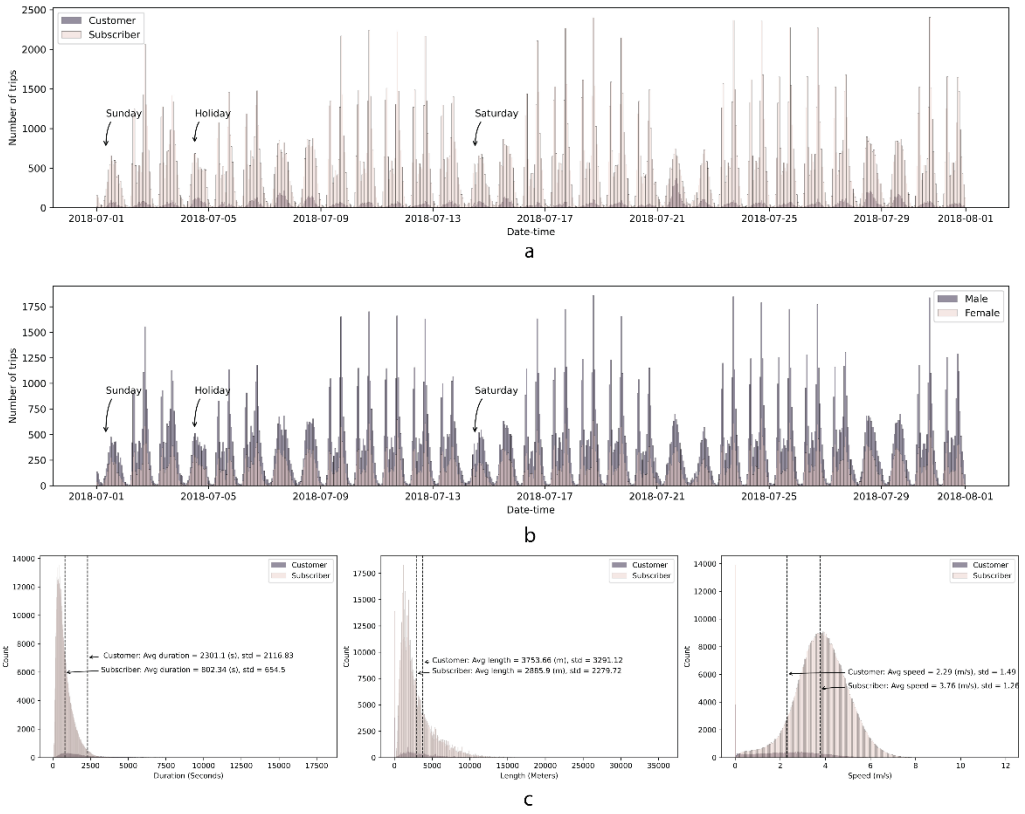


Figure S3. Divvy micromobility system: (a) hourly distribution of starting Divvy trips by type of users in July 2018. (b) hourly distribution of starting Divvy trips by gender of users for July 2018. (c) duration (left), length (center), and speed (right) of the trips for the entire period of study.

S5. UML class diagram of the agent based model

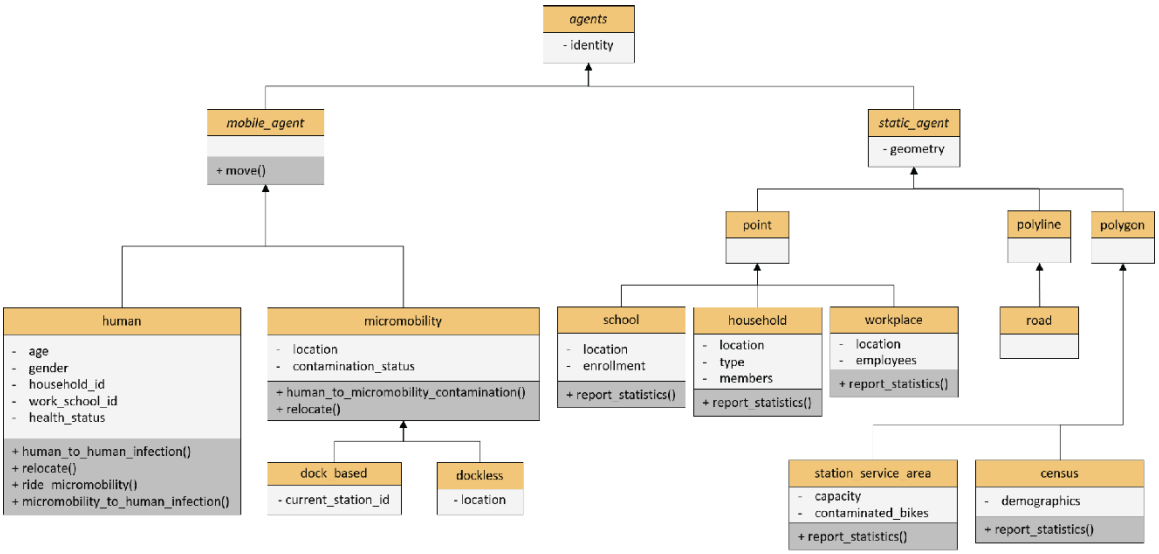


Figure S4. UML class diagram of the agent based model.

S6. Generating synthetic population with their social ties

We used a variation of the methodology proposed by (Burger et al., 2017) to generate a realistic large scale synthetic baseline population with social ties using the census data, work and school locations and commuting patterns. Here, we describe in detail the process of generating the synthetic contact network with social ties for Cook County, IL.

The synthetic contact network is created in four steps:

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

S6.1. Generating and assigning individual agents to households

In order to create synthetic population, one needs to simulate every individual and then place them in the households. We 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, the 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 (7-11). These 11 household types together with group quarters (both group quarter types are considered together as category 12) are generated for the Cook. For simplicity, we 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[6], 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 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).

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

For the households with children under 18, the number of children is stochastically 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. We 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 belonging to this group out of 11,309 residents. However, we do not expect individuals in a jail to have connections with the outside world. There is no way in the model to account for that unless we know the types of group quarter.

S6.2. Creating the Geographic Space

S6.2.1 House unit locations

House locations are created along the road network in Cook County[7]. The number of occupied houses is 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)[8]. When the population density is high they are placed on top of each other.

S6.2.2 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 Employment Statistics (LODES). The CBP database is available through the US Census Bureau County Business Patterns survey (CBP)[9]. 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 we need to downscale the data into census tract level. Origin-Destination Employment Statistics (LODES) data from Longitudinal Employer-Household Dynamics (LEHD) program[10] is used for this purpose. LEHD provides the residence and workplace flow among blocks for the years between 2002-2019 (for this case study we 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[11]. Jobs totals are associated with home and work Census Blocks. First the jobs totals are aggregated into the census tract level. Then, we 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

[7] https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html
[8] See technical documentation of the TIGER data set, Appendix E, MAF/TIGERFeature Class Code (MTFCC) Definitions at https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html
[9] https://www.census.gov/programs-surveys/cbp.html
[10] https://lehd.ces.census.gov/
[11] https://ideas.repec.org/p/cen/wpaper/14-38.html

($nworkplaces_{tract}$), one can calculate the number of workplaces within each census tract using the following equation:

$$nworkplaces_{tract} = nworkplaces_{county} * \frac{njobs_{tract}}{njobs_{county}} \text{ (Eq. S.1)}$$

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

S6.2.3 School locations

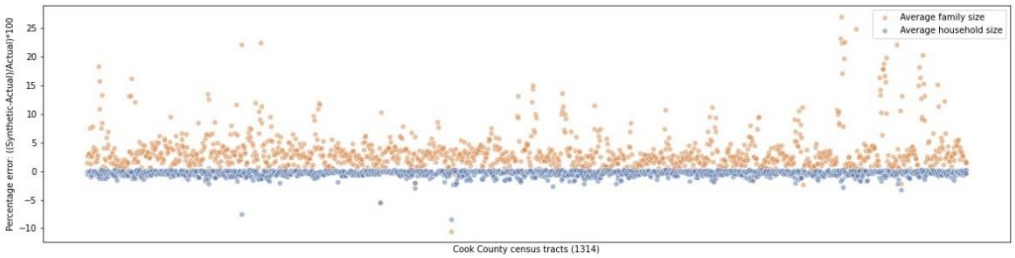
School locations are extracted from the United States Environmental Protection Agency (EPA), Office of Environmental Information[12]. The data set is called ORNL_Education, which is collected by Oak Ridge National Laboratory and includes colleges and universities, supplemental colleges, private schools, 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 S4 shows some example census tracts with simulated house, workplace, and school, as well as actual public bicycle stations in Cook County, IL.

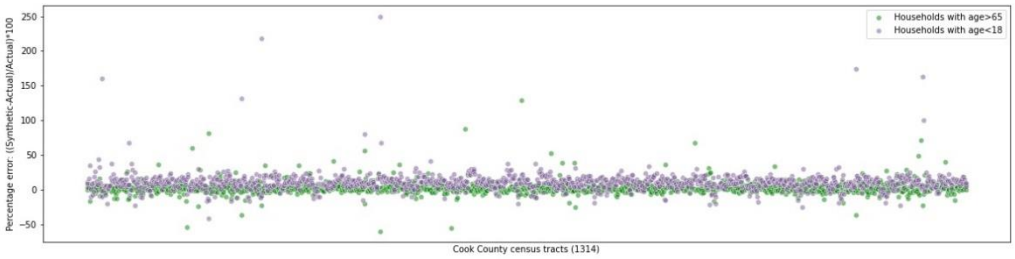
[12] https://geodata.epa.gov/arcgis/rest/services/OEI/ORNL_Education/MapServer

Figure S5. Example census tracts with house, work, school, and micromobility station locations in Cook County.

S 6.3. Evaluating the synthetic population



a



b

Figure S6. Accuracy assessment of the synthetic population. (a) Average family size and average household size. (b) Households with age>65 and households with age<18.

S6.4. 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[13]. 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.

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

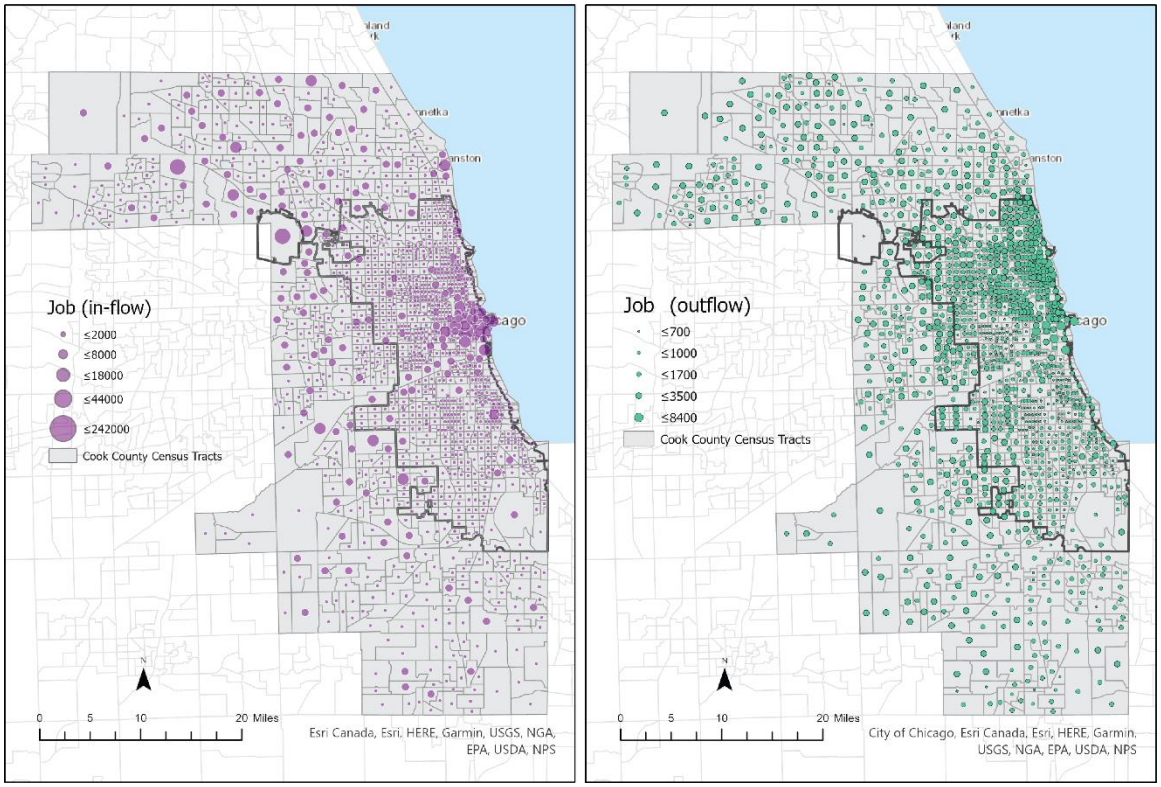


Figure 7. Origin-destination employment statistics.

S6.5. Creating a network of social contacts

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 (Salathé et al., 2010). When the number of individuals in households (including group quarters), schools, or workplaces are larger than the average number of contacts k for an individual in these groups, a Newman–Watts–Strogatz small-world graph (Newman & Watts, 1999) is created indicating each individual makes contacts with k closest members in that group. Otherwise, a complete contact network is created for the group. We assume this contact network stays the same for the workdays of the week. For weekends and holidays, however, we assume the work and school connections are removed. In this study the value of k for different venues are extracted from the literature. (Salathé et al., 2010) measured the close proximity interactions of the students at an American high school using wireless sensor networks. Analyzing their data shows each student, on average, has 151, 68, 31, and 20 daily contacts longer than 20s, 1min, 5min, and 15min, respectively. For schools and households (including group quarters), we assume five minutes as the reasonable amount of time for viruses to transmit and we consider k=31. For workplaces, we assume k=9 based on measured close interactions using video cameras in an office space (Zhang et al., 2019).