Microscopic Modeling of Spatiotemporal Epidemic Dynamics

Tilemachos Pechlivanoglou, Gian Alix, Nina Yanin, Jing Li, Farzaneh Heidari and Manos Papagelis
{tipech, gcalix, jliellen, farzanah, papaggel}@eecs.yorku.ca, nina27@my.yorku.ca
York University – Toronto, Ontario, Canada

ABSTRACT

Conventional techniques of epidemic modeling are based on compartmental models, where *population groups* are transitioning from one compartment to another – for example, $\mathcal{S}$, $\mathcal{I}$, or $\mathcal{R}$, (Susceptible, Infectious, or Recovered). Then, they focus on learning *macroscopic properties* of disease spreading, such as the transition rates between compartments. Although these models are useful in studying epidemic dynamics, they lack the granularity needed for analyzing individual behaviors during an epidemic and understanding the relationship between *individual decisions* and *the spread of the disease*. In this paper, we develop *microscopic models of spatiotemporal epidemic dynamics* informed by mobility patterns of individuals and their interactions. In contrast to *macroscopic models*, microscopic epidemic models focus on individuals and their properties, such as their activity level, mobility behaviors, and impact of mobility behavior changes. Our microscopic spatiotemporal epidemic model allows to: (**i**) assess the risk of infection of an individual based on mobility patterns; (**ii**) assess the risk of infection associated with specific geographic areas and points-of-interest (POIs); (**iii**) assess the risk of infection of a trip in an urban environment; (**iv**) provide trip recommendation for mitigating the risk of infection; and (**v**) assess targeted intervention strategies that aim to control the epidemic spreading. Our work provides *an evidence-based data-driven model* to inform individuals about the infection risks associated with their mobility behavior during a pandemic, providing at the same time *safer alternatives*. It can also inform public policy about the effectiveness of *targeted intervention strategies* that aim to contain or mitigate the epidemic spread compared to horizontal measures.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous computing**; • **Computing methodologies** → **Agent / discrete models**.

KEYWORDS

epidemic modeling, mobility, individual variability, COVID-19

ACM Reference Format:
Tilemachos Pechlivanoglou, Gian Alix, Nina Yanin, Jing Li, Farzaneh Heidari and Manos Papagelis. 2022. Microscopic Modeling of Spatiotemporal Epidemic Dynamics. In *The 3rd ACM SIGSPATIAL International Workshop on Spatial Computing for Epidemiology (SpatialEpi '22) (SpatialEpi '22), November 1, 2022, Seattle, WA, USA*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3557995.3566116

SpatialEpi '22, November 1, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9543-4/22/11…$15.00
https://doi.org/10.1145/3557995.3566116

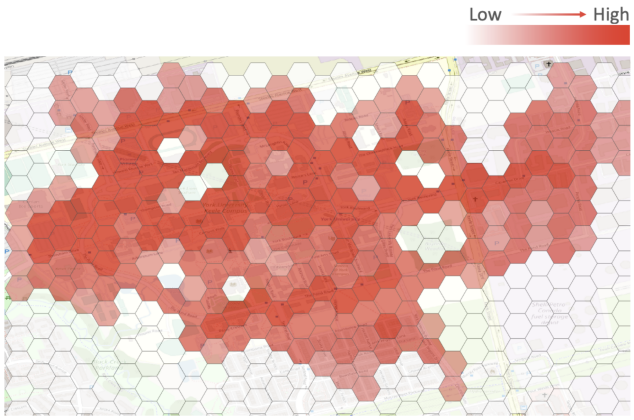


Figure 1: An example overlay risk map of an urban area informed by our microscopic spatiotemporal epidemic model.

1 INTRODUCTION

Motivation. From the Plague of Athens (430 to 426 BC) [45] to the Spanish Flu (1918) [42], pandemics have shaped the human history [30]. In the last 20 years alone, the world had seen many infectious disease outbreaks, such as pandemics caused by the SARS-CoV [48], swine flu [11], MERS-CoV [63], Ebola [7], Zika [12], and recently, SARS-CoV-2 [43]. These pandemics have tested the world's most advanced health systems and have caused an enormous societal and economic damage. We currently witness **two emerging trends** related to the study of epidemics and epidemic response that motivate our research:

(A) advances of digital tracing technologies;
(B) microscopic epidemic modeling that focuses on individuals.

These trends are linked and cannot be considered separately from each other. Digital tracing — despite the privacy concerns — could (potentially) be tolerated by communities as far as it can inform substantially more sophisticated epidemic models and responses, and effective microscopic epidemic modeling is only feasible if continuous digital tracing is in place. We discuss below the premise and risks of these trends as opposed to conventional approaches.

A. Offline versus digital tracing. Conventional non-pharmaceutical intervention (NPI) methods to address the rapid spread of an infectious disease include *physical distancing*, *confinement measures* and *human-based contact tracing* of infected individuals. While one can argue that these measures are effective in controlling the spread of the disease and saving lives [22, 44], they have **well-known drawbacks**: (*i*) they are imposing extreme restrictions and limitations on an individual's activities or freedom, leading to a slowdown of a community's socio-economic activities and side-effects for the individuals themselves; and (*ii*) they depend on offline

human-based contact tracing of infected individuals, which can be cumbersome, expensive, slow and inaccurate. On the other hand, advanced technological responses to the problem have claimed some success in controlling the epidemic based on **digital contact tracing** [26], enabled by GPS technologies and mobile apps [60] and beyond [41]. This approach enables an easy and rapid implementation of infectious disease tracing as it requires to gather and process simple information. Gathering sensitive information might infringe the privacy of individuals [17]. Controlling an infectious disease should not weaken an individual's privacy. We therefore advocate for *privacy-preserving digital tracing* protocols and technologies to protect the privacy of individuals [3, 15, 35, 51, 56].

B. Macroscopic vs microscopic epidemic models. Most existing epidemic models assign *population groups* to compartments – for example, $\mathcal{S}$, $\mathcal{I}$, or $\mathcal{R}$, (Susceptible, Infectious, or Recovered) – and people may progress between compartments [5]. Then, they focus on learning **macroscopic** properties of disease spreading, such as the transition rates between compartments. While these models are useful in studying the spread of epidemics, they lack the granularity needed for analyzing individual behaviors during an epidemic and understanding the relationship between individual decisions and the spread of the disease. Moreover, they assume homogeneous population and full mixing, which is unrealistic (see details in Appendix A). In contrast to macroscopic models, **microscopic epidemic models** focus on individuals and their properties, such as their activity level, mobility behaviors, and impact of mobility behavior changes. Thus, such models can study more realistic heterogeneous populations. See also the discussion in Appendix B about homogeneous and heterogeneous population in the $\mathcal{SEIR}$ epidemic model.

Our approach. The **focus of the current research** is on utilization of digital traces of individuals (mobility data) to inform a more comprehensive analysis of disease spreading through methods of trajectory data mining [62]. We present mathematical models for the spread of the disease in a community that take into account the **mobility patterns of individuals**. We study the effect that **individual variability** (heterogeneity) of mobility behavior has to *individual* and *spatial* risk of infection. As people move in cities following different mobility patterns, they engage in various types of interactions with other people and visit different places. As such, the relative risk of individuals getting infected or infecting others can substantially vary depending on the relative risk of the places they frequent. This observation can have significant consequences on our understanding of how the disease can propagate in a community — from both a social and a spatial perspective — as well as to the intervention strategies designed to control an epidemic.

Contributions. Motivated by these emerging trends and the inherent limitations of conventional epidemiological models (see details in Appendix A and B), we make the following major contributions:

- we present **a data-driven microscopic spatiotemporal epidemic model** that incorporates individual variability due to mobility patterns and allows to: **(i)** assess the relative risk of infection of an individual, **(ii)** assess the relative risk of infection associated with specific geographic areas and points-of-interest (POIs), **(iii)** assess the risk of infection of a (pedestrian) trip in an urban environment;
- we present **a statistical microscopic model** that allows to evaluate the risk of infection associated with a POI as a function of the dispersion observed in indoor mobility;
- we present **a POI recommendation model** that suggests safer trips to POIs; compared to a null model that considers sensible alternatives for selecting a POI to visit, our model's recommendations substantially decrease the risk of infection;
- we present **large-scale simulations** using model parameter values that resemble the recent COVID-19 outbreak and realistic synthetic mobility data in a real urban environment (large University campus and surroundings) that allow for many human-human interactions; the model and algorithms presented generalize to other similar infectious diseases

Organization. The remainder of the paper is organized as follows: Sec. 2 provides preliminaries and formally defines the technical problems of interest. Our microscopic spatiotemporal epidemic model and scientific approach are presented in Sec. 3. Sec. 4 presents an experimental evaluation of the different methods. We review the related work in Sec. 5 and conclude in Sec. 6.

2 PRELIMINARIES AND THE PROBLEM

In this section, we first introduce notation and preliminaries related to our model, then we formally define the problems of interest.

Definition 2.1. (**Map**) Consider a map $\mathcal{M}$ of a finite geographic area of Earth representing the administrative boundaries of a city or a city neighborhood. Since $\mathcal{M}$ is a relatively small region, the Earth surface it represents has a low curvature and is close to flat. We can therefore, for simplicity, assume it represents a finite 2-*dimensional* Euclidean space $\mathbb{R}^2$. This assumption allows to approximate geodesic distances on Earth with Euclidean distances in $\mathbb{R}^2$.

Definition 2.2. (**Block**) Let the map $\mathcal{M}$ consist of a set of non-empty finite disjoint blocks $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_n\}$ covering all its area, where $\mathbf{b}_k \in \mathcal{B}$ is a unit of a geographic area in the map. Assume the block $\mathbf{b}_k$ is a polygon with an arbitrary area a_k. Our work assumes that blocks are *hexagons* that can fill a plane with no gaps, forming a regular tessellation (see example in Fig. 1). We explain in Appendix C the rationale behind this assumption.

Definition 2.3. (**POI**) Let a set of points of interests $\mathcal{POI} = \{POI_1, \ldots, POI_p\}$ located on the map $\mathcal{M}$, representing buildings that individuals visit frequently, such as groceries, restaurants, and so on. We assume that each POI_i consists of a multi-block $H_i \subset \mathcal{B}$ that contains all blocks covered by the area occupied by POI_i.

Definition 2.4. (**Contact**) Let a set of individuals $\mathcal{N} = \{u_1, \ldots, u_N\}$ moving in $\mathcal{M}$ for a finite *observation time interval* $[0, T]$ forming a set of trajectories $\mathcal{P}$. For an individual $u \in \mathcal{N}$, a trajectory is a sequence $\mathcal{P}_u = \{(x_1, y_1, t_1), (x_2, y_2, t_2), \ldots, (x_T, y_T, t_T)\}$, where $t_i \in [0, T]$ and $(x_i, y_i) \in \mathbb{R}^2$. As individuals are moving around, they can at times encounter each other, forming *contacts*. A *contact* between two individuals $u, v \in \mathcal{N}$ occurs when they are both found in the same block $\mathbf{b}_k$ at the same time. Multiple individuals found in the same block at the same time all form contacts with each other.

Problem Definition. Our data-driven microscopic model of spatiotemporal epidemic dynamics addresses the following problems:

(P1) Given a map $\mathcal{M}$, a set of individuals $\mathcal{N}$ and their trajectories $\mathcal{P}$, determine the risk of infection $risk_u$ of each $u \in \mathcal{N}$.

(P2) Given a map $\mathcal{M}$ consisting of a set of blocks $\mathcal{B}$, a set of individuals $\mathcal{N}$ and their trajectories $\mathcal{P}$, determine the $risk_{\mathbf{b}_i}$ of each $\mathbf{b}_i \in \mathcal{B}$. The output can be used to construct a risk map of the infectious disease, such as in Fig. 1. The risk map can be visualized as a heat map, where the deeper the color of a block is, the higher the infection risk associated with it.

(P3) Given a map $\mathcal{M}$ consisting of a set of blocks $\mathcal{B}$, a set of points of interest $\mathcal{POI}$, a set of individuals $\mathcal{N}$ and their trajectories $\mathcal{P}$, determine the risk $risk_{POI_i}$ of each $POI_i \in \mathcal{POI}$. For this problem, we are interested in examining cases where a group of individuals are gathered in a POI_i of maximum capacity m, for varying levels of dispersion (i.e., how individuals are distributed in blocks), measured by the standard deviation σ.

(P4) Given a map $\mathcal{M}$ consisting of a set of blocks $\mathcal{B}$, a set of individuals $\mathcal{N}$, their trajectories $\mathcal{P}$, and a spatio-temporal epidemic spreading model, define block-based intervention strategies to mitigate the risk of infection of individuals.

(P5) Given a map $\mathcal{M}$ consisting of a set of blocks $\mathcal{B}$, a set of points of interest $\mathcal{POI}$ and a query q by user u located at (x, y), recommend a ranked list $\mathcal{L} \subseteq \mathcal{POI}$ to visit that minimizes infection risk.

3 METHODOLOGY

In this section we provide details of our methodology for modeling individual variability in epidemics that is due to mobility patterns. We first present a method for computing the infection risk within each block, as a result of contacts between individuals inside it. Moreover, we use this block risk to formally define the infection risk of each individual, as well as the risk associated with the point of interest the blocks belong to. Finally, we present different block risk-based intervention strategies, as well as a POI risk-based recommendations, with the goal of reducing the infection risk of individuals and the spread of the epidemic. Fig. 2 illustrates a diagram of the process. Note that we do not intend on associating a specific POI name to a particular risk value for privacy reasons.

3.1 Block infection risk

We define the risk of infection of a block $\mathbf{b}$ at time t as the number of all contacts occurring in the block at that time. Let $n_{\mathbf{b}}^t$ represent the number of individuals in $\mathbf{b}$ at time t. Then, the contacts are equal to the number of all pairs of individuals. Formally:

$$brisk_{\mathbf{b}}^t = \frac{n_{\mathbf{b}}^t (n_{\mathbf{b}}^t - 1)}{2} \tag{1}$$

Note that when only a single person is present within a block, the risk of infection is 0, as expected. Next, we define the average risk of a block over the entire observation time $[0, T]$:

$$brisk_{\mathbf{b}} = \frac{\sum_{t=0}^{T} brisk_{\mathbf{b}}^t}{T} \tag{2}$$

Finally, we define the *relative risk* of each block by normalizing the risk over all blocks $\mathcal{B}$ covering the entire observation area:

$$rbrisk_{\mathbf{b}} = \frac{brisk_{\mathbf{b}}}{\sum_{\mathbf{j} \in \mathcal{B}} brisk_{\mathbf{j}}} \tag{3}$$

Algorithm 1: Relative Block Risk Computation

Input: $\mathcal{P} = \{p_1, p_2, ..., p_i, ..., p_n\}$: pedestrian trajectories,
$\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_i, ..., \mathbf{b}_k\}$: blocks

Output: $\mathcal{RBRISK} = \left\{ \left[rbrisk_{\mathbf{b}}^t \right]_{\mathbf{b} \in \mathcal{B}} \right\}_{t=0}^{T}$: relative risk of blocks

$part\mathcal{P} \leftarrow [\,]$; $\mathcal{RBRISK} \leftarrow \{\}$

for $p_i \in \mathcal{P}$ **do**
  $part\mathcal{P}_i \leftarrow p_i$.partition($T$) /* partition traj by time units */
  $part\mathcal{P}$.append($part\mathcal{P}_i$) /* $part\mathcal{P}_i$ = $[p_i^0, p_i^1, \ldots, p_i^T]$ */
$part\mathcal{P}$.transpose() /* arrange traj segments by time */
  /* $part\mathcal{P}^t$ = $[p_0^t, p_1^t, \ldots, p_{n-1}^t, p_n^t]$ */
for $t \in [0, T]$ **do**
  $brisk^t \leftarrow \mathbf{0}_{k \times 1}$
  $activeB \leftarrow \mathcal{B}.intersects(part\mathcal{P}^t)$; /* get blocks containing each trajectory */
  for $\mathbf{b}_i \in activeB$ **do**
    $brisk_{\mathbf{b}_i}^t$ ++;
  $rbrisk^t \leftarrow brisk^t \,/\, sum(brisk^t)$
  $\mathcal{RBRISK} \leftarrow \mathcal{RBRISK} \cup \{rbrisk^t\}$
return $\mathcal{RBRISK}$

Algorithm 1 provides the pseudocode of computing the relative risk $rbrisk_{\mathbf{b}}$ of each block $\mathbf{b}_i \in \mathcal{B}$. Using the $rbrisk_{\mathbf{b}}$ values, we are able to construct an *infection risk map* of $\mathcal{M}$ similar to the example of Figure 1. A risk map provides useful insights into the potential behavior of the epidemic at different locations within the observation area. In the following paragraphs, we make use of the relative risk values of blocks in the risk map $\mathcal{M}$ to provide metrics related to the risk of individuals moving around $\mathcal{M}$, as well as the points of interest (POIs) within it.

3.2 Individual infection risk

In the next step, for a set of trajectories $\mathcal{P}$ and the set of blocks $\mathcal{B}$, we identify the block that each individual is in at every timestamp:

$$\mathbf{b}_{u,t},\ u \in \mathcal{N},\ t \in T_u, \quad \text{such that:} \quad (x_u, y_u) \in \mathbf{b} \text{ for } (x_u, y_u, t)$$

At this point, using the previously calculated *relative block risks* $\mathcal{RBRISK} = \{rbrisk_{\mathbf{b}_1}, rbrisk_{\mathbf{b}_2}, ...\}$ and the block each person has visited, we can define the risk of infection for individuals. To do this, we simply add together the block risk values of all blocks an individual u has visited throughout the entire time they've been present in the system T_u. Formally:

$$risk_{\mathbf{u}} = \sum_{t \in T_u} rbrisk_{\mathbf{b}_{u,t}} \tag{4}$$

Finally, similar to the relative block risk, we calculate the relative individual risks by normalizing over the entire population:

$$rrisk_{\mathbf{u}} = \frac{risk_{\mathbf{u}}}{\sum_{\mathbf{v} \in \mathcal{N}} risk_{\mathbf{v}}} \tag{5}$$

Note that we do not examine the average risk of every individual over the time they were active in the system, but instead we look at their total accumulated risk. These metrics may differ in many cases. For example, a person with high-contact activity for a short time and another with a few contacts but over a longer time could

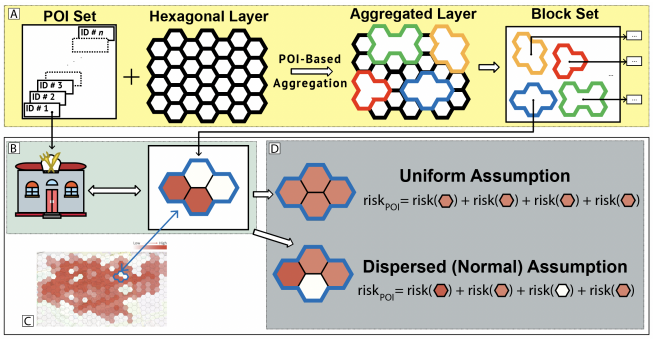


Figure 2: Architecture diagram for calculating the POI Risk. (A) POI-based aggregation of hexagon blocks to form multi-blocks, (B) Associating POI to a multi-block, (C) Risk map, (D) POI risk based on the block risk under varying assumptions.

have the same total risk, but different average risk over time. In this work, we are interested in the total infection risk that an individual experiences throughout the observation time, and as such the examination of average risk over time is outside our scope.

3.3 Multi-block and POI infection risk

Our microscopic epidemic model allows to examine the unique infection interactions happening at the scale of POIs; recall that a POI corresponds to a group of contiguous, adjacent blocks. Therefore, in order to better represent a point of interest, we introduce the concept of a multi-block $H_i \subset \mathcal{B}$, which contains all blocks corresponding to POI_i. An illustration containing two multi-blocks derived from the location of two POIs can be seen in Fig. 3. As observed in the figure, some hexagons do not intersect any POI, and thus cannot be meaningfully consolidated to form a multi-block. Such hexagons are irrelevant to the infection transmission dynamics within a POI. Although we focus on POI-based block aggregation in this work, other aggregation semantics are possible. For example, multiple POIs belonging to the same mall or complex can be further aggregated together to form a larger multi-block. One can also consider aggregation at the level of a city. We are now in position to formally define the infection risk associated with a POI (i.e. the corresponding multi-block) in terms of the aggregated risk of the blocks contained within it. Formally:

$$POIrisk_i = \sum_{\mathbf{b} \in H_i} rbrisk_{\mathbf{b}} \tag{6}$$

3.4 Distribution of individuals in POIs

It is important to note that the POI risk can greatly depend on the distribution of individuals in space. Since the value of $brisk_{\mathbf{b}}$ has an $O(n^2)$ relationship with the number of individuals within the block $\mathbf{b}$, we can infer that the same number of individuals concentrated in a single block of a POI would produce a higher POI risk than the equivalent number of people evenly distributed across every POI block. Therefore, a metric that describes the "spread" or *statistical dispersion* of individuals to POI blocks is particularly insightful.

Note that at this point, the distribution of individuals in such a microscopic scale often is not known and/or considered in relevant research and recommended guidelines [16]. The main factor that is considered is simply the occupancy of a location, presumably along

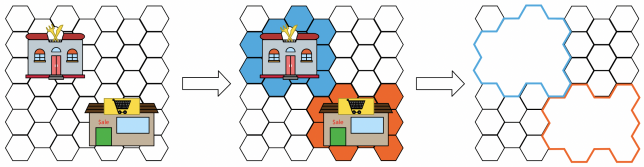


Figure 3: Hierarchical multi-block aggregation of POIs.

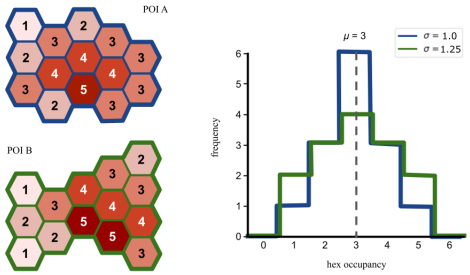


Figure 4: An example illustrating how two different POIs with the same area (14 hexagon blocks) and total occupancy (42 people) may have varying dispersions.

with an arbitrary infection probability multiplier. This inevitably leads to the implicit assumption that all individuals in a given POI are uniformly distributed throughout the POI's area, which is a strong assumption that may not necessarily reflect real conditions. To better model the distribution and dispersion of individuals within POIs, we consider the previously mentioned uniform distribution, along with several instances of the normal distribution with varying levels of statistical dispersion. More specifically, as in this case of a discrete population, this represents a binomial distribution, but we'll refer to it as normal distribution for simplicity.

The normal distribution is popular in probability and statistics and is a natural choice for real-world assumptions. With normal distributions, dispersion is more commonly measured using the *standard deviation*, as opposed to other measures such as the *range* or *interquartile range* [39]. Although standard deviation can be computed experimentally, we assume in our work that domain expertise can provide this number. In particular, we take into account varying levels of dispersion, with standard deviation $\sigma = \{0.5, 1.0, 2.0, 3.0\}$ respectively. The baseline for comparison would involve the uniform case where the standard deviation is not defined (i.e. $\sigma = \infty$). Fig. 4 shows two example distributions of the same population that yield the same mean, but distinctive standard deviation values.

3.5 Spatiotemporal epidemic modeling

Our model of epidemic spreading is based on an agent-based heterogeneous (due to mobility patterns) $\mathcal{SEIR}$ model (see Appendix B for details on the taxonomy of $\mathcal{SEIR}$ epidemic models). At any time step t, an individual can be in any one of the $\mathcal{S}$, $\mathcal{E}$, $\mathcal{I}$, or $\mathcal{R}$ states. First, we briefly describe the characteristics of each state in the spatiotemporal context using semantics of a block:

($\mathcal{S}$) Susceptible. All individuals are initially in this state; each one can get exposed to the infection by any one of the infected individuals that belong to the same block $\mathbf{b}$ as that person, with probability β per timestep.

($\mathcal{E}$) Exposed. Individuals in this state have been infected but have yet to be infectious. Individuals remain in this state for the duration

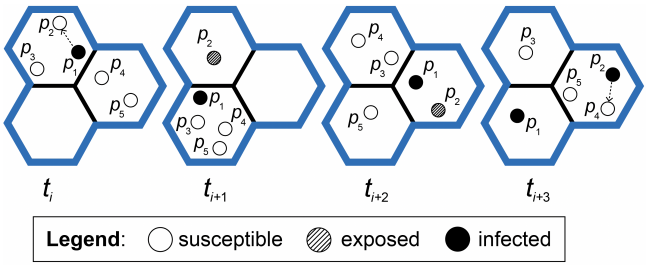


Figure 5: An example illustrating the epidemic spreading dynamics among five individuals in a three-hexagon POI during four observed timesteps.

of the incubation period λ of the disease – which we model as a constant that lasts for $\mathcal{I}_f$ timesteps for simplicity. After this period, the exposed individual is guaranteed to switch to the infected state.

($\mathcal{I}$) Infected. Infected individuals can transmit the infection to anyone that shares the same block $\mathbf{b}$ at time t, with probability β.

($\mathcal{R}$) Recovered. Removed individuals are either infected individuals who have recovered or who have passed away due to the disease. To simplify the model, individuals are removed from the $\mathcal{I}$ state after $\mathcal{I}_r$ timesteps with recovery probability γ. Recovered individuals are assumed to neither catch the disease nor be susceptible to it.

Algorithmic Details of the Stochastic Model. Given a set of (pedestrian) trajectories $\mathcal{P}$, a finite set of regular disjoint blocks $\mathcal{B}$ in the map $\mathcal{M}$, as well as epidemic parameters β, γ, $\mathcal{I}_f$ and $\mathcal{I}_r$, the model allows to keep track of every individual's states over time. Recall that individuals can progress to different compartments (*susceptible* to *exposed* to *infectious* to *removed*). With a population size $\mathcal{N}$, the cumulative number of individuals within each of the states at time t is given by $\mathcal{S}(t)$, $\mathcal{E}(t)$, $\mathcal{I}(t)$, and $\mathcal{R}(t)$, respectively. To represent the size of the epidemic spread, we consider two special sets of infected individuals: (i) the initial seed set $\mathcal{I}_0 = \mathcal{I}(0)$, and (ii) the set of all those individuals infected at the end of the time period $\mathcal{I}_T = \mathcal{I}(T)$. Formally, let individual $u \in \mathbf{b}_k$ be in the *susceptible* compartment and $\mathcal{N}_u$ be the set of all other individuals residing in block $\mathbf{b}_k$ at time t. Each individual in $\mathcal{N}_u$ that is infected has some probability β of transmitting the disease to u (modelled with a biased coin flip). u's state switches to the *exposed* state upon infection (otherwise, u's state remains the same). An exposed individual automatically switches to the *infected* state after $\mathcal{I}_f$ timesteps and an infected individual transitions to the *removed* state after $\mathcal{I}_r$ steps, with probability γ. The pseudocode of the stochastic model of spatiotemporal epidemic spreading is provided in Algorithm 2. It extends the work of [46] by assuming that all individuals residing in the same tessellated block form a fully-connected contact network. This allows to reduce spatiotemporal semantics (block-level) to well-defined network-based epidemic dynamics. Fig. 5 depicts an illustrative example of this epidemic spreading model. Algorithm 2 is a *stochastic* model and can be costly in terms of runtime. Appendix D discusses how this can be alleviated by converting the model to a *deterministic* method based on percolation theory.

3.6 Containment intervention strategies

In this section, we explore intervention strategies focused around social distancing. We present two alternatives, a *targeted intervention* and a randomized one, representing a *null model*.

Algorithm 2: Spatiotemporal Epidemic Spreading Model

Input: $\mathcal{P}$, $\mathcal{B}$, $\mathcal{S}$, $\mathcal{I}$, β, γ, $\mathcal{I}_f$, $\mathcal{I}_r$

Output: $\mathcal{B}^{\mathcal{SEIR}[0,T]} = \left\{\mathcal{B}^{\mathcal{SEIR}(t)}\right\}_{t=0}^{T}$, where $\mathcal{B}^{\mathcal{SEIR}(t)} = \left\{\left[\mathbf{b}_k^{\mathcal{S}(t)}, \mathbf{b}_k^{\mathcal{E}(t)}, \mathbf{b}_k^{\mathcal{I}(t)}, \mathbf{b}_k^{\mathcal{R}(t)}\right]\right\}_{\forall \mathbf{b}_k \in \mathcal{B}}$

$\mathcal{S}(0) \leftarrow \mathcal{S}$; $\mathcal{E}(0) \leftarrow 0$; $\mathcal{I}(0) \leftarrow \mathcal{I}$; $\mathcal{R}(0) \leftarrow 0$; $\mathcal{B}^{\mathcal{SEIR}[0,T]} \leftarrow \{\}$

for $t \in [0, T]$ **do**
  for $\mathbf{b}_k \in \mathcal{B}$ **do**
    $\mathcal{B}^{\mathcal{SEIR}(t)} \leftarrow \{\}$
    for $u \in \mathbf{b}_k$ **do**
      if u.state = $\mathcal{S}$ **then** /* susceptible */
        $\mathcal{N}_u \leftarrow \{p \mid p \in \mathcal{P}$ in block $\mathbf{b}_k$ at time $t\} \setminus \{u\}$
        for $v \in \mathcal{N}_u$ **do**
          if $v \in \mathbf{b}_k^{\mathcal{I}(t)}$ **then**
            v infects u with probability β
            if u *is infected* **then**
              $u.state \leftarrow \mathcal{E}$
              $\mathcal{I}_f^u \leftarrow 0$ /* incubation of u begins */
              $\mathbf{b}_k^{\mathcal{S}(t+1)} \leftarrow \mathbf{b}_k^{\mathcal{S}(t)} \setminus \{u\}$
              $\mathbf{b}_k^{\mathcal{E}(t+1)} \leftarrow \mathbf{b}_k^{\mathcal{S}(t)} \cup \{u\}$
      else if u.state = $\mathcal{E}$ **then** /* exposed */
        $\mathcal{I}_f^u$ ++;
        if $\mathcal{I}_f^u = \mathcal{I}_f$ **then**
          $u.state \leftarrow \mathcal{I}$
          $\mathcal{I}_r^u \leftarrow 0$ /* recovery of u begins */
          $\mathbf{b}_k^{\mathcal{E}(t+1)} \leftarrow \mathbf{b}_k^{\mathcal{E}(t)} \setminus \{u\}$
          $\mathbf{b}_k^{\mathcal{I}(t+1)} \leftarrow \mathbf{b}_k^{\mathcal{I}(t)} \cup \{u\}$
      else if u.state = $\mathcal{I}$ **then** /* infected */
        $\mathcal{I}_r^u$ ++;
        if $\mathcal{I}_r^u = \mathcal{I}_r$ **then**
          $u.state \leftarrow \mathcal{R}$
          $\mathbf{b}_k^{\mathcal{I}(t+1)} \leftarrow \mathbf{b}_k^{\mathcal{I}(t)} \setminus \{u\}$
          u recovers with probability γ
          if u *recovers* **then**
            $\mathbf{b}_k^{\mathcal{R}(t+1)} \leftarrow \mathbf{b}_k^{\mathcal{R}(t)} \cup \{u\}$
      else
        /* removed compartment; do nothing */
    $\mathcal{B}^{\mathcal{SEIR}(t)} \leftarrow \mathcal{B}^{\mathcal{SEIR}(t)} \cup \left\{\left[\mathbf{b}_k^{\mathcal{S}(t)}, \mathbf{b}_k^{\mathcal{E}(t)}, \mathbf{b}_k^{\mathcal{I}(t)}, \mathbf{b}_k^{\mathcal{R}(t)}\right]\right\}$
  $\mathcal{B}^{\mathcal{SEIR}[0,T]} \leftarrow \mathcal{B}^{\mathcal{SEIR}[0,T]} \cup \mathcal{B}^{\mathcal{SEIR}(t)}$
return $\mathcal{B}^{\mathcal{SEIR}[0,T]}$

Targeted intervention. Each individual, using information provided by a risk map $\mathcal{M}$, makes specific local decisions to limit contacts and help contain the infection. This strategy can change the structure of the social interaction network and therefore affect the R_0 number of the spread of infection [9]. Using the risk map, the intervention assumes that an individual is provided by the individual risk of every trip they plan to make within the observation time, and decide to cancel the α most risky ones (i.e. fewer trips). This results in eliminating a fraction α_e of their overall contacts

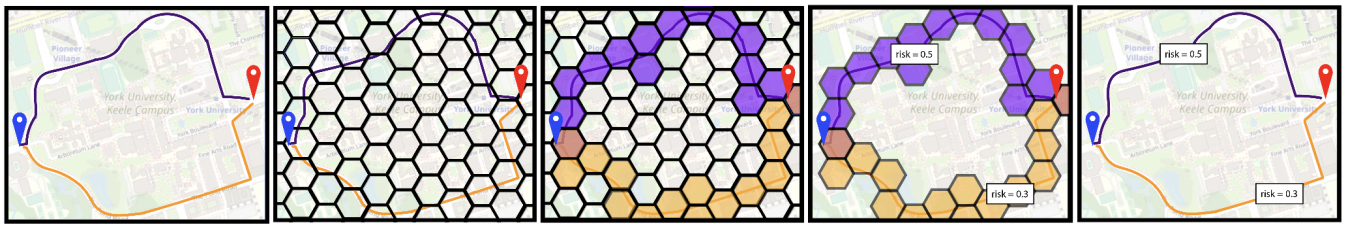


Figure 6: The image depicts the infection risk associated to paths. In the example, the orange path is less risky than the purple one when travelling from the source (blue marker) to the destination (red marker).

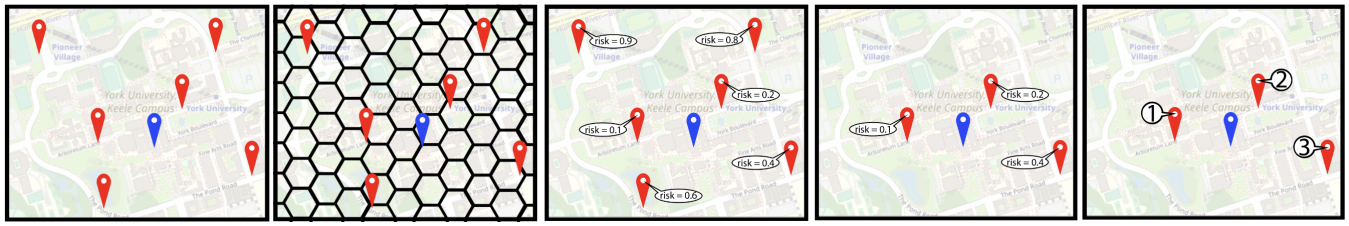


Figure 7: The image depicts the most recommended POIs (red markers) in close proximity to the user (blue location) with the least amount of infection risk. As a disclaimer, we do not intend on displaying the risk associated to a particularly-named POI.

that would have occurred, but eventually did not. This informed decision defines a targeted intervention that results in mitigating the spread of infection (see an example of this process in Fig. 6).

Null model. In order to evaluate the comparative performance of the targeted intervention strategy, we compare it with a null model. The null model assumes that an equal fraction α_e of contacts is *uniformly at random* removed from the overall set of contacts.

The objective is to evaluate how a targeted intervention can be beneficial to mitigating the spread of infection, compared to randomized social distancing strategies (informed by horizontal health policy measures). The evaluation is based on simulating the spread of infection using a stochastic agent-based $\mathcal{SEIR}$ model.

3.7 Safest POI recommendation model

In certain scenarios, evaluating the risk of a trip based on the relative risk of blocks traversed might be insignificant or irrelevant (e.g., driving to a grocery store). In these scenarios, it makes more sense to evaluate the infection risk associated with specific POIs that serve as the intended destination(s) of the trip and provide recommendations for potential better alternatives (see Fig. 7). Formally, given an origin o, a set of $\mathcal{POIs} = \{POI_1, POI_2, POI_3, ...\}$, a minimum social distancing area a, a radius R and a list of maximum occupancy ratios $\Gamma = \{\gamma_1, \gamma_2, \gamma_3, \dots\}$, our objective is to recommend the top-k safest POI(s) within a radius R for different occupancy levels. We consider three recommendation models:

Safest POI: This is our POI recommendation model, where destinations are ranked by *POIrisk* and the safest one is chosen. The pseudocode of the model can be seen in Algorithm 3.

Closest POI: This is a sensible baseline model that effectively represents the expected behavior in the absence of a recommendation service, where a POI is picked based on physical proximity.

Random POI: This model serves as the null model, where a POI is uniformly at random selected as the destination.

For each of the *Safest POI* and *Random POI* models, we consider the possibility that only a percentage ξ of the total population $\mathcal{N}$ actively follows the recommendation, while the rest revert to the default behavior *Closest POI*. We take the relative risks of each of these POIs from each search query and identify the impact of following the recommendation. In particular, we evaluate the *relative risk difference* had the user opted to elect the POI closest to them or some other randomly-selected POI.

Algorithm 3: Safest POI Recommendation

Input: $\mathcal{POIs}$: set of POIs, o: origin, $\Gamma = [\gamma_1, \gamma_2, ..., \gamma_k]$: list of occ. ratios, R: search radius, a: social distance area, σ: std.
Output: *POIrisk*: list of minimum risk POIs within R km

$POIrisk \leftarrow []$
for $POI_p \in \{x \mid x \in \mathcal{POIs}$ **and** $\Delta(o, x) < R\}$ **do**
  $max\_occu \leftarrow \frac{POI_p.AreaSize}{a}$
  for $\gamma \in \Gamma$ **do**
    $ns \leftarrow Normal\{\sigma, POI_p.numHex\} \times max\_occu \times \gamma$
    $risk \leftarrow 0$
    for $n \in ns$ **do**
      $risk \leftarrow risk + \frac{n \cdot (n-1)}{2}$ /* compute POI risk */
    $POI_p.nrisk.append(\frac{risk}{POI_p.numHex})$
  $POIrisk.append(POI_p)$
$POIrisk \leftarrow (POIrisk.min(axis = col))$
return $POIrisk$

4 RESULTS AND DISCUSSION

In this section, we present a comprehensive experimental evaluation of the methods and models introduced and discuss the consequences of the results obtained. We begin by listing the research questions/scenarios we aim to explore. Then, we present details of the data employed. Finally, we elaborate on process we followed to experimentally evaluate each research question, along with the results and discussion of broader insights. Table 1 provides a summary of the parameters used in the experiments. Our experiments aim at answering the following questions:

(Q1) Effect of POI visitor distribution on risk. How visitors' dispersion within a POI affect POI's infection risk?

Table 1: Summary of Parameters

	Parameter	Symbol	Values
Traj.	#individuals		3,000
	#data points		259,200,000
Constant	Total timesteps	T	96 timesteps*
	Simulation runs	$\mathbb{I}$	100 iterations
	POI visitor count	$\mathcal{N}$	500 people
	Ave. block count $^{w}/_{n}$ POI	$\lvert\mathbb{H}\rvert$	57 hexagons
	Hexagonal area	A	4 m^2
	Max duration of POI visit	τ	15 timesteps
Varying	Initial infection seeds	$\lvert\mathcal{I}_0\rvert$	$\{1, 3, 5, 7, 10\}\%$ of $\mathcal{N}$
	Max occ. in the POI	m	$\{10, 20, ..., 100\}\%$
	Occ. std. for POI	σ	$\{0.5, 1.0, 2.0, 3.0, \infty\}$
Epidemic-related	Transmission rate	β	0.01
	Incubation rate	λ	1.0
	Recovery rate	γ	1.0
	Exposure period	T_{exp}	2 days
	Infection period	T_{inf}	4 days
	Recovery period	T_{rec}	7 days

* One timestep represents five minutes.

(Q2) Effect of POI visitor distribution, occupancy and initial infected seed size on direct infections. How many people come in direct contact with infected individuals for distributions of visitors with varying levels of dispersion? What happens when the number of visitors approaches the max capacity of the POI? How do varying proportions of initially infected individuals impact long term infection counts?

(Q3) Impact of targeted & non-targeted interventions. How does each strategy affect the progress of the epidemic?

(Q4) Impact of recommendation policy. How POI recommendations can decrease the risk of infection?

Data generation. In our experiments, we developed a parametric, stochastic agent-based synthetic data generator. We discuss our data generation techniques in each of the following experiments, and further provide more details on this in Appendix E.

4.1 Q1 Effect of POI visitor distribution on risk

For this experiment, we simulate the movement of individuals within POIs using a stochastic process. For each timestamp $t \in [0, T]$, individuals freely move around within the POI – moving from one block to another or perhaps remain in the same block. The destination block is selected based on a predefined spatial distribution (conceptually, this corresponds to less or more "interesting" places within the POI). This spatial distribution corresponds to either the uniform distribution or the normal distribution with prespecified σ values. Individuals remain within the POI up until a maximum duration τ, while new individuals may enter the POI as long as the maximum allowed capacity m has not been reached. Entry times of new individuals are also randomized and follow a normal distribution, which can be parametrically configured.

In order to examine the block risks inside POIs for different visitor dispersion values, we generate a population of POI visitors with a fixed number of individuals $\mathcal{N}$, corresponding to a percentage m of the POI's max occupancy, as well as various spatial distributions with corresponding standard deviation values σ. In Fig. 8a, we can observe that the uniformly-distributed occupancies will tend to a flat horizontal line in our plot as occupancy counts in the POI's hexagons are approximately the same at each timestep – which implies that the number of contacts, and consequently the relative risks of each of these hexagons, would roughly be the same. We can observe the risk of infection increases as the dispersion decreases (i.e. lower σ, people are more concentrated at the same spots). This is in agreement with intuition: as social distancing implies that people are more distributed in space, the risk of infection is lower.

4.2 Q2 Effect of POI visitor distribution, OCC. and infection seed on direct infections

For this experiment, we use the same model as Q1 – however now we select a portion $|\mathcal{I}_0|$ of the population to be initially infected and act as seed for the epidemic. These individuals, when coming to direct contact with others (i.e., occupy the same block), transmit the infection to others who became exposed. Note that newly exposed individuals are unable to become infected and cannot infect others as this is a short-term observation time over the course of one day.

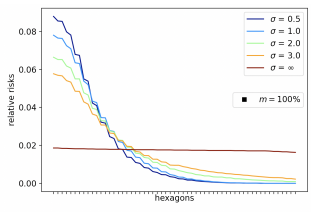
We utilize this method to execute three variants of this experiment: (1) varying distributions of individuals, including the uniform ($\sigma \to \infty$) and normal distributions ($\sigma \in [0.5, 1.0, 2.0, 3.0]$); (2) varying maximum occupancy ($m \in [0\%, 10\%, 20\%, \ldots, 100\%]$) and (3) varying sizes of initial seed ($|\mathcal{I}_0| \in [1\%, 3\%, 5\%, 7\%, 10\%]$). In all cases, the non-varying values are fixed to $\sigma = 1.0, m = 100\%$ and $|\mathcal{I}_0| = 10\%$, with constant population size $\mathcal{N} = 500$, observation time $T = 8$ hrs $= 96$ timestamps, and visit duration $\tau = 15$ timestamps. The results for these can be seen in Figs. 8b, 8c and 8d, respectively.

As anticipated, situations that lead to increasingly dense concentrations of people (high occupancy, low dispersion) or larger numbers of infected individuals (high occupancy, large seed size), lead to higher direct infection counts. Although increasing the occupancy contributes to both of these scenarios, the highest infection counts are achieved by decreasing the dispersion in the population. This may likely be due to the quadratic relationship between the block occupancy and number of contacts, as discussed in Sec. 3.4.

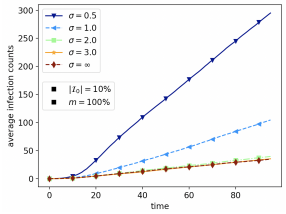
4.3 Q3 Impact of targeted and non-targeted intervention strategies

Now we address the proposed research question without making the strong assumption of homogeneous mixing. Instead, we use synthetically generated dataset of simulated pedestrian trajectories in the York University, Keele campus in Toronto, with $\mathcal{N} = 3000$ individuals, 1.5 million contacts and 260 million individual position measurements (Table 1). We then model the spread of an epidemic using an agent-based $\mathcal{SEIR}$ model, with description as presented in Appendix B. It is worth noting that this experiment demonstrates interventions in a microscopic simulation process; more sophisticated intervention strategies can be developed similarly.

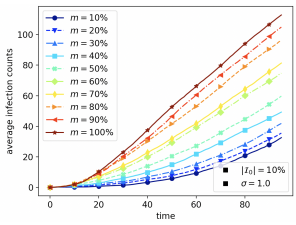
After using the $\mathcal{SEIR}$ model to simulate the course of an infection on the given population over the period of a month ($T = 30$), we obtained an average of the final infected individual count at any time in $[0, T]$ (i.e. everyone in the $\mathcal{E}$, $\mathcal{I}$, $\mathcal{R}$ groups) over 20 executions. The final value was approximately 2449 individuals



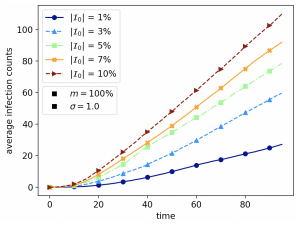
(a) block risks for different σ, $m = 100\%$



(b) #infected for different σ, $m = 100\%$, $|\mathcal{I}_0| = 10\%$



(c) #infected for different m, $\sigma = 1.0$, $|\mathcal{I}_0| = 10\%$



(d) #infected for different $|\mathcal{I}_0|$, $m = 100\%$, $\sigma = 1.0$

Figure 8: (a) The POI visitor distribution versus the risk of blocks (Q1); the effect of (b) POI visitor distribution, (c) the maximum occupancy, and (d) the initial infections on the long-term direct infection counts (Q2)

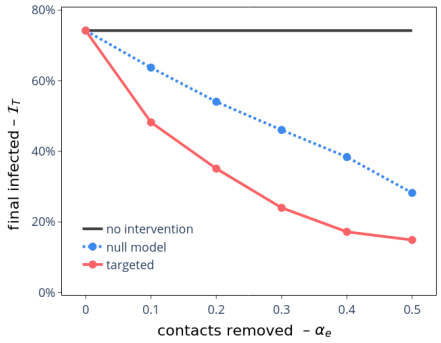


Figure 9: Comparison of interventions. This plot illustrates how targeted interventions are more effective in practice, compared to no or non-targeted intervention strategies (Q3).

– around 75% of the total population. Next, we repeated the simulation, after first applying *targeted* intervention strategy with $\alpha \in [0.1, 0.2, ..., 0.5]$ values as well as applying the *null-model* intervention strategy at the corresponding α_e values. The final results can be seen in Fig. 9. It is evident that the null model achieves an approximately linear reduction in final number of infections, with the targeted intervention performing better. The proposed method achieved twice as high as the amount of prevented infections as the null model, although there appears to be diminishing returns after removing > 40% of contacts in the dataset.

4.4 Q4 Impact of recommendation policy

In this experiment, we made use of a dataset provided by SafeGraph[1], which contains information about the location and floor area of POIs, and generated queries originating at one of 380 locations O in Toronto, Canada. For each query, all relevant POIs within $R = 5$ km belonging to a specific category (such as "restaurants") were retrieved, along with their $POIrisk_i$ from a generated risk map $\mathcal{M}$. These are then ranked based on the criteria from Sec. 3.7.

In Fig. 10a, we examine the relative risk difference between the safest option and the null model for both uniform and normal ($\sigma = 1.0$) distributions. Risks were calculated based on a max occupancy definition of 4 m^2 per person as per post-COVID-19 social distancing regulations in Ontario[2]. In the uniform case, people are sufficiently dispersed; hence, close contacts do not occur. Thus, all destinations pose the same individual risk value of 0. However, in the more realistic case of individuals distributed normally, the relative risk difference is negative for both recommended and null-model options. As expected, the POIs recommended by our model are safer than both the closest and the randomly sampled POIs.

We also examine scenarios where different percentages ξ of people follow the recommendation model, while the remaining people ($100\% - \xi$) select the closest POIs, with a max occupancy of 4 m^2 per person. In Fig. 10b, it is evident that as more people follow our recommendations, the relative risk value goes down. The effect is more evident for higher values of maximum occupancy.

Finally, we use the recommendation framework to evaluate how the new post-COVID19 maximum occupancy regulations affect the relative risk value, compared with the pre-COVID-19 value of 2 m^2 per person. In Figure 10c, we report the difference between the two for different occupancy values. It is evident that in realistic, normal distribution scenarios, the post-COVID-19 maximum occupancy guidelines have prevented a significant amount of infection risk.

5 RELATED WORK

Our research is mostly related to spatiotemporal epidemic modeling and mobility-based epidemics spreading. Several key ideas on these topics have already been cited throughout the paper. Here we briefly elaborate on other aspects relevant to our research.

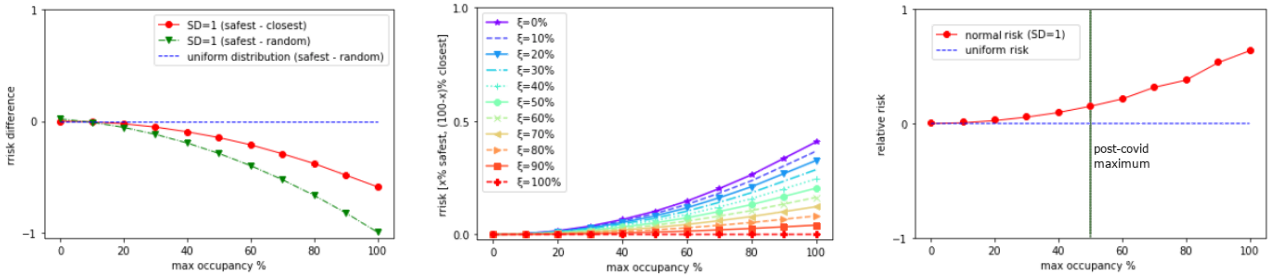
Trajectory data mining. Computational methods for mining spatiotemporal, trajectory, and mobility data, have been extensively studied by data mining and database communities [6, 62]. Of particular interest are problems related to trajectory similarity [55], trajectory clustering [36], anomaly detection in moving objects [21], computing node centrality in trajectory networks [47], and mining interactions among moving objects overtime (eg. pedestrian group trajectory detection [52–54]). Recently, deep learning approaches for spatiotemporal learning have gained increasing attention [58].

Digital contact tracing. In our research, we assume that human mobile traces are available. This can be enabled by existing digital contact tracing technologies [14, 18, 25]. For instance, Aleta et al. [1] synthesized contact networks and modeled SARS-CoV-2 transmission in the Boston metropolitan area using census and contact tracing data. They showed how vital contact tracing is in preventing 2nd wave spreading when complete isolation is relaxed.

Microscopic vs macroscopic epidemic models. Several works have shown the limitations of macroscopic models over microscopic models. For instance, Cui et al. [19] showed how high-risk exposed individual counts in 50 U.S. states can be estimated at the micro-level, but not in the macro-level. Liu [37] used *multi-agent*

[1]https://www.safegraph.com

[2]www.ontario.ca/laws/regulation/r21520



(a) Risk difference between recommended and closest restaurant destinations

(b) Effect of a percentage of population ξ following recommendations on risk

(c) Post-Covid - pre-Covid maximum occupancy relative risk evaluation

Figure 10: Experimental results for Q4

reinforcement learning to conclude the same. Cantin et al. [13] presented a hybrid model that integrates micro- with macro- attributes of epidemic models, and showed that microscopic behaviors can strongly influence the dynamics of macroscopic epidemic models.

Epidemic spreading in complex networks. Mathematical modeling of epidemic spreading in networks is a popular topic that can help to study and control the emergence of infectious diseases in a population. For instance, Weitz et al. [59] designed epidemiological interventions that exploit the idea of *'shield immunity'*. However, methods cannot easily translate to health policy or individual-level recommendations. In our study, we used an *agent-based $\mathcal{SEIR}$ model*, which has previously been employed to study epidemic spread in dynamic networks. For instance, Perez and Dragicevic [49] proposed a spatially explicit epidemiological model of infectious disease for understanding the diffusion of a disease in a network of human contacts. In their model, interactions are not fully dynamic, but are determined by the location of individuals at certain times.

Non-pharmaceutical interventions. Instead of lockdown restrictions, Block et al. [9] proposed more moderate distancing strategies including limiting contacts to *similar*, *community-based* or *repetitive* contacts. More accurate microscopic epidemic modeling is useful in designing intervention strategies. Miralles-Pechuán et al. [40] used reinforcement learning to suggest high-level intervention strategies rewarding fewer infections and less severe lockdowns. Fan et al. [24] showed that mobility inspired interventions, such as limited long distance trips, can notably reduce epidemic spread.

Population heterogeneity in epidemics due to mobility. Hébert-Dufresne et al. [31] argued the importance of heterogeneity in mobility for predicting an infectious disease's outbreak size through digital contact tracing technologies. Changruenngam et al. [16] studied the effects of individual human mobility on disease transmission dynamics. Lloyd-Smith et al. [38] showed how the basic R_0 in traditional epidemic analyses is only a population-level estimate; thus, more targeted control interventions would be more effective.

Trip recommendations and epidemics. Part of our work focuses on providing trip recommendations that satisfy domain-specific constraints, such as in the works of [20, 50]. With recent pandemic, factoring in infection risk in recommendation models has attracted interest. For example, Fotsing et al. [27] designed an epidemic-aware socio-spatial POI recommender model. Anastasiou et al.'s [4] graph-based path search algorithm reduces an individual's COVID-19 exposure risk taking into account accessibility constraints, outdoor exposure thresholds, and congestion tolerance. Similarly, Alix et al. [2] presented a system for recommending safe trips to POIs on a geographic map based on risk evaluation of alternative trips.

6 CONCLUSIONS

We presented a data-driven microscopic spatiotemporal epidemic model that takes into account the variability of individual mobility patterns. Through this, we made analytic and insightful observations to the individuals' infection risk, the relative risk of infection of associated POIs and geographic areas, and the relative risk of pedestrian trips. Our statistical microscopic model demonstrated that the risk of infection increases with lower values of dispersion. We further presented a model that suggests the safest POIs based on low infection risks. We showed in our experiments the impact of our model by exhibiting how much infection risk is reduced by selecting our recommendation over sensible baselines. We also established this by demonstrating that a high percentage of individuals observing this policy can significantly reduce infection risks. Finally, our experiment on reducing pedestrian trips demonstrated that the overall infection count drops remarkably due to the decline in number of contacts, and consequently the amount of interactions. As a result, we established that our targeted interventions can outperform generic intervention strategies. Our research on microscopic modeling of spatiotemporal epidemic dynamics offers new insights and has potential for significant socioeconomic impact.

REFERENCES

[1] A. Aleta, D. Martin-Corral, A. P. y Piontti, M. Ajelli, M. Litvinova, M. Chinazzi, N. E. Dean, M. E. Halloran, I. M. Longini Jr, S. Merler, et al. Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the COVID-19 epidemic. *medRxiv* (2020).
[2] G. Alix, N. Yanin, T. Pechlivanoglou, J. Li, F. Heidari, and M. Papagelis. 2022. A Mobility-based Recommendation System for Mitigating the Risk of Infection during Epidemics. In *23rd IEEE Intl. Conf. on Mobile Data Management (MDM).*
[3] H. Alsdurf, Y. Bengio, T. Deleu, P. Gupta, D. Ippolito, R. Janda, M. Jarvie, T. Kolody, S. Krastev, T. Maharaj, et al. COVI White Paper. *arXiv:2005.08502* (2020).
[4] C. Anastasiou, C. Costa, P. K. Chrysanthis, C. Shahabi, and D. Zeinalipour-Yazti. ASTRO: Reducing COVID-19 Exposure through Contact Prediction and Avoidance. 8, 2, Article 11 (Dec 2022).
[5] R. M. Anderson, B. Anderson, and R. M. May. 1992. *Infectious diseases of humans: dynamics and control.* Oxford University Press.
[6] G. Atluri, A. Karpatne, and V. Kumar. Spatio-temporal data mining: A survey of problems and methods. *ACM Computing Surveys (CSUR)* 51, 4 (2018), 1–41.
[7] S. Baize, D. Pannetier, L. Oestereich, T. Rieger, L. Koivogui, N. Magassouba, B. Soropogui, M. S. Sow, S. Keïta, H. De Clerck, et al. Emergence of Zaire Ebola virus disease in Guinea. *New England Journal of Medicine* 371, 15 (2014), 1418–1425.
[8] C. P. Birch, S. P. Oom, and J. A. Beecham. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological modelling* 206, 3-4 (2007), 347–359.

[9] P. Block, M. Hoffman, I. J. Raabe, J. B. Dowd, C. Rahal, R. Kashyap, and M. C. Mills. Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. *Nature Human Behaviour* (2020), 1–9.

[10] F. Brauer, C. Castillo-Chavez, and C. Castillo-Chavez. 2012. *Mathematical models in population biology and epidemiology*. Vol. 2. Springer.

[11] D. Butler. Swine flu goes global: New influenza virus tests pandemic emergency preparedness. *Nature* 458, 7242 (2009), 1082–1084.

[12] G. S. Campos, A. C. Bandeira, and S. I. Sardi. Zika virus outbreak, bahia, brazil. *Emerging infectious diseases* 21, 10 (2015), 1885.

[13] G. Cantin, C. Silva, and A. Banos. Mathematical analysis of a hybrid model: Impacts of individual behaviors on the spreading of an epidemic. *Networks & Heterogeneous Media* (2022).

[14] G. Cencetti, G. Santin, A. Longa, E. Pigani, A. Barrat, C. Cattuto, S. Lehmann, M. Salathe, and B. Lepri. Digital Proximity Tracing in the COVID-19 Pandemic on Empirical Contact Networks. *medRxiv* (2020).

[15] J. Chan, S. Gollakota, E. Horvitz, J. Jaeger, S. Kakade, T. Kohno, J. Langford, J. Larson, S. Singanamalla, J. Sunshine, et al. Pact: Privacy sensitive protocols and mechanisms for mobile contact tracing. *arXiv:2004.03544* (2020).

[16] S. Changruenngam, D. J. Bicout, and C. Modchang. How the individual human mobility spatio-temporally shapes the disease transmission dynamics. *Scientific Reports* 10, 1 (2020), 1–13.

[17] H. Cho, D. Ippolito, and Y. W. Yu. Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs. *arXiv:2003.11511* (2020).

[18] F. W. Crawford, S. A. Jones, M. Cartter, S. G. Dean, J. L. Warren, Z. R. Li, J. Barbieri, J. Campbell, P. Kenney, T. Valleau, and O. Morozova. Impact of close interpersonal contact on COVID-19 incidence: Evidence from 1 year of mobile device data. *Science Advances* 8, 1 (2022).

[19] Z. Cui, M. Cai, Y. Xiao, Z. Zhu, M. Yang, and G. Chen. Forecasting the transmission trends of respiratory infectious diseases with an exposure-risk-based model at the microscopic level. *Environmental Research* (2022).

[20] M. Debnath, P. K. Tripathi, A. K. Biswas, and R. Elmasri. 2018. Preference aware travel route recommendation with temporal influence. In *Proc. of the 2nd ACM SIGSPATIAL/LBSN Workshop*. 1–9.

[21] S. Dodge, R. Weibel, and E. Forootan. Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems* 33, 6 (2009), 419–34.

[22] K. T. Eames and M. J. Keeling. Contact tracing and disease control. *Proc. of the Royal Society of London. Series B: Biological Sciences* 270, 1533 (2003), 2565–71.

[23] D. J. Earn, P. Rohani, B. M. Bolker, and B. T. Grenfell. A simple model for complex dynamical transitions in epidemics. *Science* 287, 5453 (2000), 667–670.

[24] C. Fan, R. Lee, Y. Yang, and A. Mostafavi. Fine-grained data reveal segregated mobility networks and opportunities for local containment of COVID-19. *Scientific Reports* 11, 1 (2021), 1–6.

[25] K. Farrahi, R. Emonet, and M. Cebrian. Epidemic contact tracing via communication traces. *PloS one* 9, 5 (2014), e95133.

[26] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* 368, 6491 (2020).

[27] C. P. K. Fotsing, Y.-W. Teng, G.-S. Lee, C.-Y. Shen, Y.-S. Chen, and D.-N. Yang. 2022. On Epidemic-aware Socio Spatial POI Recommendation. In *23rd IEEE Intl. Conf. on Mobile Data Management (MDM)*.

[28] H. Frisch and J. Hammersley. Percolation processes and related topics. *Journal of the society for industrial and applied mathematics* 11, 4 (1963), 894–918.

[29] 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* 198, 1-2 (2006), 115–126.

[30] J. N. Hays. 2005. *Epidemics and pandemics: their impacts on human history*. Abc-clio.

[31] L. Hébert-Dufresne, B. M. Althouse, S. V. Scarpino, and A. Allard. Beyond R0: Heterogeneity in secondary infections and probabilistic epidemic forecasting. *medRxiv* (2020).

[32] H. W. Hethcote. The mathematics of infectious diseases. *SIAM review* 42, 4 (2000), 599–653.

[33] N. Hoertel, M. Blachier, C. Blanco, M. Olfson, M. Massetti, F. Limosin, and H. Leleu. Facing the COVID-19 epidemic in NYC: a stochastic agent-based model of various intervention strategies. *medRxiv* (2020).

[34] M. J. Keeling and B. T. Grenfell. Individual-based perspectives on R0. *Journal of theoretical biology* 203, 1 (2000), 51–61.

[35] W. Kim, H. Lee, and Y. D. Chung. Safe contact tracing for COVID-19: A method without privacy breach using functional encryption techniques based-on spatio-temporal trajectory data. *PLOS ONE* 15 (Dec 2020), 1–12.

[36] J.-G. Lee, J. Han, and K.-Y. Whang. 2007. Trajectory clustering. In *Proc. of the 2007 ACM SIGMOD Intl. Conf. on Management of Data – SIGMOD07*.

[37] C. Liu. 2020. A Microscopic Epidemic Model and Pandemic Prediction Using Multi-Agent Reinforcement Learning.

[38] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz. Superspreading and the effect of individual variation on disease emergence. *Nature* 438, 7066 (2005), 355–359.

[39] S. Manikandan. Measures of dispersion. *Journal of Pharmacology & Pharmacotherapeautics* 2, 4 (2011), 315.

[40] L. Miralles-Pechuán, F. Jiménez, H. Ponce, and L. Martínez-Villaseñor. 2020. A methodology based on deep q-learning/genetic algorithms for optimizing covid-19 pandemic government actions. In *Proc. of the 29th ACM Intl. Conf. on Information & Knowledge Management*. 1135–1144.

[41] M. F. Mokbel, S. Abbar, and R. Stanojevic. Contact Tracing: Beyond the Apps. *arXiv:2006.04585* (2020).

[42] D. M. Morens and A. S. Fauci. The 1918 influenza pandemic: insights for the 21st century. *The Journal of infectious diseases* 195, 7 (2007), 1018–1028.

[43] W. H. Organization et al. Coronavirus disease 2019 (COVID-19): situation report, 72. *WHO situation reports* (2020).

[44] W. H. Organization, C. for Disease Control, Prevention, et al. 2015. *Implementation and management of contact tracing for Ebola virus disease: emergency guideline*. Technical Report. World Health Organization.

[45] M. J. Papagrigorakis, C. Yapijakis, P. N. Synodinos, and E. Baziotopoulou-Valavani. DNA examination of ancient dental pulp incriminates typhoid fever as a probable cause of the Plague of Athens. *Intl. Journal of Infectious Diseases* 10, 3 (2006), 206–214.

[46] T. Pechlivanoglou, J. Li, J. Sun, F. Heidari, and M. Papagelis. Epidemic Spreading in Trajectory Networks. *Big Data Research* 27 (2022).

[47] T. Pechlivanoglou and M. Papagelis. 2018. Fast and Accurate Mining of Node Importance in Trajectory Networks. In *2018 IEEE Intl. Conf. on Big Data (Big Data)*. 781–790.

[48] J. S. Peiris, K. Y. Yuen, A. D. Osterhaus, and K. Stöhr. The severe acute respiratory syndrome. *New England Journal of Medicine* 349, 25 (2003), 2431–41.

[49] L. Perez and S. Dragicevic. An agent-based approach for modeling dynamics of contagious disease spread. *Intl. journal of health geographics* 8, 1 (2009), 50.

[50] T. Qian, B. Liu, Q. V. H. Nguyen, and H. Yin. Spatiotemporal Representation Learning for Translation-Based POI Recommendation. 37, 2 (Jan 2019).

[51] L. Reichert, S. Brack, and B. Scheuermann. Privacy-Preserving Contact Tracing of COVID-19 Patients. *IACR Cryptol. ePrint Arch.* 2020 (2020), 375.

[52] A. Sawas, A. Abuolaim, M. Afifi, and M. Papagelis. 2018. Tensor methods for group pattern discovery of pedestrian trajectories. In *19th IEEE Intl. Conf. on Mobile Data Management (MDM)*. 76–85.

[53] A. Sawas, A. Abuolaim, M. Afifi, and M. Papagelis. 2018. Trajectolizer: Interactive analysis and exploration of trajectory group dynamics. In *2018 19th IEEE Intl. Conf. on Mobile Data Management (MDM)*. IEEE, 286–287.

[54] A. Sawas, A. Abuolaim, M. Afifi, and M. Papagelis. A versatile computational framework for group pattern mining of pedestrian trajectories. *GeoInformatica* 23, 4 (2019), 501–531.

[55] K. Toohey and M. Duckham. Trajectory similarity measures. *SIGSPATIAL Special* 7, 1 (2015), 43–50.

[56] C. Troncoso, M. Payer, J.-P. Hubaux, M. Salathé, J. Larus, E. Bugnion, W. Lueks, T. Stadler, A. Pyrgelis, D. Antonioli, et al. Decentralized privacy-preserving proximity tracing. *arXiv:2005.12273* (2020).

[57] A. R. Tuite, D. N. Fisman, and A. L. Greer. Mathematical modelling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada. *CMAJ* 192, 19 (2020), E497–E505.

[58] S. Wang, J. Cao, and P. Yu. Deep learning for spatio-temporal data mining: A survey. *IEEE Transactions on Knowledge and Data Engineering* (2020).

[59] J. S. Weitz, S. J. Beckett, A. R. Coenen, D. Demory, M. Dominguez-Mirazo, J. Dushoff, C.-Y. Leung, G. Li, A. Măgălie, S. W. Park, et al. Modeling shield immunity to reduce COVID-19 epidemic spread. *Nature medicine* (2020), 1–6.

[60] S. Woodhams. 2020. COVID-19 Digital Rights Tracker.

[61] D. Yao and R. Durrett. Epidemics on Evolving Graphs. *arXiv:2003.08534* (2020).

[62] Y. Zheng. Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)* 6, 3 (2015), 1–41.

[63] A. Zumla, D. S. Hui, and S. Perlman. Middle East respiratory syndrome. *The Lancet* 386, 9997 (2015), 995–1007.

APPENDIX

A BASIC REPRODUCTIVE NUMBER R_0 LIMITATIONS AND BEYOND

The basic reproductive number R_0 (sometimes called basic reproduction ratio), is the most widely used parameter in epidemiology. It can be thought of as the expected number of new infections caused by a single infected individual. Commonly used epidemiological models suggest that $R_0 = 1$ is a critical value. On one hand when $R_0 < 1$, each infected person produces less than one new case in expectation, therefore the size of the outbreak is constantly trending downwards, until eventually the disease dies off. On the

other hand, when $R_0 > 1$, each infected person produces more than one new case in expectation, therefore the size of the outbreak is constantly trending upwards. In principle, the larger the value of R_0, the more challenging it is to control the epidemic. Despite its usefulness as an approximate indication of the spreading power of the disease, many studies have stressed **the limitations of** R_0. An underlying assumption of R_0 is that the disease is spreading in a network that resembles a *regular tree network* – a special type of a network topology that has no cycles and that each internal node has a constant number of children defined by a branching factor d. However, real-world communities do not resemble regular trees, as it is common for people to have common friends (forming triangles or cycles) and some people may have more contacts than others. It is also easy to see how the basic computation of R_0 breaks down when we consider transmission of infection to be a stochastic process involving discrete individuals [34]. In addition, the computation of R_0 assumes the homogeneous mixing of individuals that do not take into account any mobility-based individual variation.

B TAXONOMY OF $\mathcal{SEIR}$ EPIDEMIC MODELS

Epidemic compartmental models partition the population into several compartments where people transition between them according to their infection status. Classical models employ the $\mathcal{SIR}$ model [5, 10] (*susceptible*, *infected*, *removed* compartments). However our work instead utilizes the $\mathcal{SEIR}$ model, allocating an additional *exposed* compartment specifically designed for individuals that have been infected, yet not been infectious themselves. This is particularly useful for infectious diseases that comes with a significant latent (incubation) time period. There are two main classes of $\mathcal{SEIR}$ models, depending on the assumptions made of population structure and transmission progression:

Homogeneous Population. This class assumes homogeneous mixing of population, in which individuals progress between compartments at certain transition rates as described by ordinary differential equations [23, 32]:

$$\frac{\mathrm{d}\mathcal{S}}{\mathrm{d}t} = -\frac{\beta \mathcal{I} \mathcal{S}}{N}; \qquad \frac{\mathrm{d}\mathcal{E}}{\mathrm{d}t} = \frac{\beta \mathcal{I} \mathcal{S}}{N} - \lambda \mathcal{E}; \qquad \frac{\mathrm{d}\mathcal{I}}{\mathrm{d}t} = \lambda \mathcal{E} - \gamma \mathcal{I}; \qquad \frac{\mathrm{d}\mathcal{R}}{\mathrm{d}t} = \gamma \mathcal{I}$$

where β, λ and γ are the transmission, incubation, and recovery rates, respectively. As this model is *deterministic*, a fixed parameter set and model initialization ($t = 0$), each simulation generates the same result. Although this model can provide insights on the state of epidemic spread and of health policy [57], it is also limited by the assumption that individuals share the same characteristic.

Heterogeneous Population. This class is based on an agent-based $\mathcal{SEIR}$ model, where each agent represents an individual [33, 49]. Through this approach, individual characteristics and behaviors (such as mobility and contact patterns overtime) can be modelled towards the epidemic. Various mobility patterns result in complex spatiotemporal social interactions among people in the community [47, 61]. These models pose a challenge of analysis and interpretability that is due to their dependence on *probabilistic* epidemic spread. Such models are more complex [29], yet are more realistic. In addition, they can aid in better understanding a disease due to individual behaviors, as well as provide opportunities for designing targeted intervention strategies that better translates with health policies.

C HEXAGONAL BLOCK REPRESENTATION

We elected hexagons to represent our blocks, as these are preferred over other shapes (such as squares or triangular tessellations) for various reasons. For one, the "circular nature" of regular hexes allows such polygon to tessellate a map, forming evenly-spaced grid. Moreover, the nature of its "circularity" enables a natural representation of curvatures in (mobility) trajectory data [8].

D STOCHASTIC TO DETERMINISTIC MODEL USING PERCOLATION THEORY

Algorithm 2 exhibits a probabilistic process that is random in nature. With the same initial conditions (eg. the same sets of infected individuals, etc.) and parametric values, multiple independent simulations of epidemic spreading process can generate several various outputs, in terms of the total number of infected individuals at the end of the observation period. This is due to the *biased coin-flipping mechanism* that occurs at each timestep to decide whether the infection would be transmitted to another individual. Fortunately, a model based on *percolation theory*[3] presents itself to be an equivalent deterministic model that allows faster simulations than that of a stochastic model. To bring in this idea of percolation in the model of epidemic spread, we can make a decision for each edge "interaction" between any two given individuals in $\mathcal{P}$ at the beginning of the observation process, as opposed to deferring this decision at runtime. In practice, we simply have to flip the biased coin with probability β for each edge as many times as the duration of contact (expressed in time units); and then decide whether to keep or remove it from the network. At the end of the whole process, we have constructed a smaller interaction network.

E DETAILS ON DATA GENERATION

To model individuals' mobility at a microscopic-level, having a small-scale high-fidelity trajectory data is vital. Such data can easily be obtained through camera motion tracking or motion sensors. However, acquiring real trajectory datasets is challenging and not a trivial task. This is mostly due to their sensitive nature, as using real data comes with considerable privacy concerns – which we have briefly covered in Sec. 1. While simulated datasets are not always accurate of a real scenario, we have made an effort to create "realistic" synthetic data with the goal of achieving results as close as possible to experiments performed on real data. We believe that the lack of sensitive real datasets should not refrain us from working on interesting problems where we can demonstrate the validity and effectiveness of models, provide support to health policy and decision making, and have a potentially high social and economic impact. As such, we instead developed a parametric, stochastic agent-based synthetic data generator through a mobility simulator SUMO[4]. This allows us to produce data reflecting both realistic and extreme scenarios, while enables us to fully explore the problem space. Furthermore, this tool could enable a domain expert to produce highly customized datasets tailored to specific conditions that would better reflect reality despite being synthetic.

[3] *Percolation theory* in physics can help explain how fluids can flow through certain types of porous material [28]. In network science, percolation can help describe the behavior of a network when nodes or links are removed.
[4] https://www.eclipse.org/sumo/