GeoAI for Public Health

Andreas Züfle
Emory University
azufle@emory.edu

Taylor Anderson
George Mason University
tander6@gmu.edu

Hamdi Kavak
George Mason University
hkavak@gmu.edu

Dieter Pfoser
George Mason University
dpfoser@gmu.edu

Joon-Seok Kim
Oak Ridge National Laboratory
kimj1@ornl.gov

Amira Roess
George Mason University
aroess@gmu.edu

April 15, 2025

Abstract

Infectious disease spread within the human population can be conceptualized as a complex system composed of individuals who interact and transmit viruses through spatio-temporal processes that manifest across and between scales. The complexity of this system ultimately means that the spread of infectious diseases is difficult to understand, predict, and respond to effectively. Research interest in GeoAI for public health has been fueled by the increased availability of rich data sources such as human mobility data, OpenStreetMap data, contact tracing data, symptomatic online surveys, retail and commerce data, genomics data, and more. This data availability has resulted in a wide variety of data-driven solutions for infectious disease spread prediction which show potential in enhancing our forecasting capabilities. This book chapter (1) motivates the need for AI-based solutions in public health by showing the heterogeneity of human behavior related to health, (2) provides a brief survey of current state-of-the-art solutions using AI for infectious disease spread prediction, (3) describes a use-case of using large-scale human mobility data to inform AI models for the prediction of infectious disease spread in a city, and (4) provides future research directions and ideas.

1 Introduction

Artificial intelligence (AI) models of infectious disease spread are critical decision-support systems to help explain the mechanisms of infectious disease spread, predict the number of cases and deaths, and prescribe effective policy guidelines. Ideally, by observing the *spatiotemporal behavior* of individuals, we can comprehend the uptake of preventative behaviors that help reduce transmission by wearing masks, reducing mobility, limiting social interactions, or getting vaccinated. However, spatiotemporal human behavior is collectively shaped by the biological, environmental, cultural, sociological, and economic makeup of individuals and their spatial environments [18, 59]. As a result, the degree to which individuals engage in preventative behaviors varies over space and time. Efforts to improve this understanding have been limited by the lack of reliable longitudinal spatiotemporal data containing observations that capture the change in human behavior in response to disease spread [23].

This book chapter summarizes some of the important findings of this community and identifies future research directions. We note that the term AI is used more generally in this chapter: Rather than narrowly focusing on neural networks we also include other models that extract patterns from large infectious disease spread related datasets, such as mechanistic models, agent-based models, regression models, and ensemble models. The remainder of this book chapter is organized as follows:

- Section 2 provides a brief overview of data-driven approaches for infectious disease spread prediction before and after the COVID-19 pandemic.

- Section 3 motivates the need for an AI solution in public health by exemplifying the heterogeneity of human behavior related to health in space and time. This heterogeneity means that health-related human behavior is difficult to capture by classic models that assume homogeneous human behavior.

- Section 4 discusses research efforts towards leveraging AI to improve representations of human behavior in models of disease spread by leveraging AI for intelligent and realistic agent decision-making and using AI solutions for estimating simulation model parameters and patterns.

- Section 5 describes a use case of using large-scale human mobility data and AI to realistically parameterize agent human mobility behaviors in an agent-based simulation of the 1.1 million population of a county in the United States.

- Section 6 provides future research directions and ideas.

2   Existing AI Models for Data-Driven Infectious Disease Spread Prediction

Data-driven epidemic forecasting has been a very large research field in the last decade. A recent survey summarizes more than 300 publications in this field [109]. Even before the COVID-19 pandemic, this field was already a focus of the computing community [89] in the context of influenza-like illnesses [5]. For example, ACM KDD has been organizing the International Workshop on Epidemiology meets Data Mining and Knowledge Discovery (EpiDaMiK) since 2018 [4, 6]. The COVID-19 pandemic has brought forth very large sets of human mobility data [46, 57, 103], which enabled new data-driven models. Existing data-driven models to predict the spread of infectious diseases include mechanistic models [7], agent-based models [101, 124], regression models [58], off-the-shelf sequential models [125], graph neural network models [42, 130], density estimation models [29], ensemble models [39, 106], as well as many other types of models [109].

Propelled by the COVID-19 pandemic, researchers have sought interdisciplinary collaborations with researchers in epidemiology and the social sciences to fill this data availability gap and use data-driven approaches that seek to understand and model the relationship between spatiotemporal behavior and disease transmission. Here, we highlight three such collaborative efforts.

After the onset of the COVID-19 pandemic, the ACM SIGSPATIAL community focused its research efforts on understanding the spread of COVID-19 in two special issues on the topic [135, 137]. In this newsletter, Qazi et al. published a dataset of billions of tweets related to COVID-19 [103]. This dataset was later updated in [66]. Gao et al. published a study on the change in human mobility during the COVID-19 pandemic [57] with a dataset of human mobility during the pandemic based on SafeGraph foot traffic data published in [69]. These two datasets were crucial to provide large human mobility datasets to understand the spread of COVID-19 and inform disease spread models. A mobility-informed epidemic simulation platform was published by Fan et al. [48] and extended in [49]. Kim et al. [76] proposed a system to find a consensus among multiple simulations using an ensemble model. Hohl et al. described an algorithm to rapidly detect COVID-19 clusters in space and time using a prospective space-time scan statistic and Bobashev et al. described a machine learning approach using mechanistic models for COVID-19 forecasting [26]. Solutions for contact tracing were discussed in the context of privacy preserving contact tracing [131] and an approach that moves the contact tracing functionality from individual users to facilities [93].

The efforts of the newsletter led to dedicated workshops at the ACM

SIGSPATIAL on Modeling and Understanding the Spread of COVID-19 [15] and on Spatial Computing for Epidemiology [13, 14]. Notable workshop publications related to AI and simulation include simulation-based infection risk estimation models [8, 72, 99], game-theory based COVID-19 simulation [100], a joint pandemic modeling and analysis platform [121], infectious disease case time series prediction [3, 117], analysis of human behavior during the COVID-19 pandemic [36, 47, 127, 132], a tool for pandemic decision-making support [83], mapping and visualization of infectious disease data [32, 112], and real-time detection of COVID-19 clusters [10].

The newsletter and workshops described above led to a special issue of ACM Transactions on Spatial Algorithms and Systems (TSAS) on Understanding the Spread of COVID-19 [136, 137]. The main goal of this special issue was to facilitate interdisciplinary work that included social scientists and epidemiologists. Therefore, the call for papers read: *"This special issue intends to bring together transdisciplinary researchers and practitioners working in topics from multiple areas, including Spatial Data Scientists ... Mathematicians, Epidemiologists, Computational Social Scientists, Medical Practitioners, Psychologists, Emergency Response and Public safety, among others."* This special issue included extended versions of newsletter and workshop papers but also included many new research directions. Notable research directions toward understanding and simulating infectious diseases included infectious disease simulation [17, 118], infectious disease modeling [31, 33, 38, 85], and infectious disease data analysis [21, 50, 90, 133].

3 Spatial and Temporal Heterogeneity of Health-Related Human Behavior

Agent-based models (ABMs) are used to forecast disease spread trajectories and to support policymakers as they prepare for and respond to emerging and re-emerging infectious diseases [45, 64, 116]. ABMs use a bottom-up approach to simulate disease dynamics among a population by representing the mobility, interactions, and subsequent transmission of infectious disease between individuals or "agents." ABMs expand upon traditional assumptions of the compartmental SIR model and its variations to include heterogeneity in the population, the spatial environment and the transmission likelihoods [24, 53]. Therefore, ABMs have been developed to simulate the spread of seasonal influenza [12, 74, 75], H1N1 [35, 61, 80], Ebola [92, 115], smallpox [30], anthrax [37], the pneumonic plague [129], dengue [70], and more recently COVID-19 [27, 43, 64, 116].

However, despite significant advances in ABMs of disease spread, recent commentaries have pointed out that many still lack realistic rep-

resentations of human behavior, a key driver of mobility, physical inter-
action, and subsequent disease transmission [9, 55, 65, 87, 105]. With-
out this, ABMs are limited in making accurate forecasts, especially
over longer prediction horizons. This limitation is best illustrated by
examining some of the models included in the COVID-19 Forecast Hub
[105], a recent collaborative effort between the Center for Disease Con-
trol and Prevention (CDC) and more than 25 scientific teams, each of
which produces a model to forecast the spread of COVID-19. Ray et
al. [105] report that, for each of the models, uncertainty increases and
accuracy decreases in prediction horizons longer than just four weeks
and concludes that to achieve accurate long-term forecasts, models
must incorporate realistic human behavior.

Despite the number of social and behavioral studies that exam-
ine how individuals respond to pandemics and outbreaks of diseases
[57, 78, 107, 123], ABMs of infectious disease spread do little to incor-
porate these findings [55]. Traditionally, ABMs ignore or oversimplify
representations of human health behaviors. For example, it is common
to exogenously impose behaviors upon agents. The modeler may com-
pare disease outcomes in scenarios where 50% or 75% of agents are ran-
domly selected as vaccinated. These approaches assume that human
behavior is temporally stationary, meaning that the change in agent
behavior over time is ignored, even as the risk of infection increases
or new policies are implemented. Furthermore, many ABMs assume
spatial stationarity, ignoring key spatial heterogeneities in human be-
havior produced by local influences related to social norms or culture
[20, 56] Although work has been done to improve representations of
human health behavior in ABMs [9, 19, 41], the encoded behavioral
response is based on broad conceptual theories such as game theory
and economic objective functions rather than empirical observations.
Furthermore, there has been limited effort to acknowledge the spatial
variation of human behavior due to the individual's sociodemographic
profile and their political, social, environmental, historical, and cul-
tural contexts. This is particularly important in locations where many
individuals may be less likely to comply with mitigation strategies, in-
cluding stay-at-home orders, which has a significant impact on disease
spread trajectories.

In the case of the COVID-19 pandemic, the *spatial variation* of
the behavioral response has been empirically observed [57, 128] using
large sets of human mobility that have been made available during
the COVID-19 pandemic that capture billions of tweets [103], foot
traffic [57], and other datasets [94].

Figure 1 shows the variation in stay-at-home behavior in the US
at the census block group level. Here, we can directly observe regional
differences in the fraction of users who stay-at-home on any given day
(Figure 1a). Figure 1b shows a common distinct spatial pattern of

(a) Fraction of stay-at-home users for each census block group across the US

(b) Urban and rural pattern of stay-at-home behavior near Houston, TX

(c) Region having a low fraction of stay-at-home users in Newark, CA

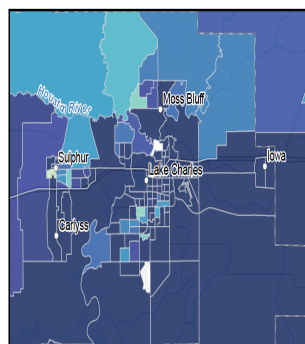(d) Region having a high fraction of stay-at-home users in Lake Charles, LA
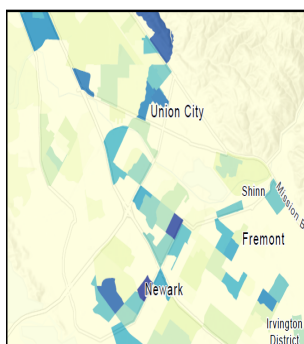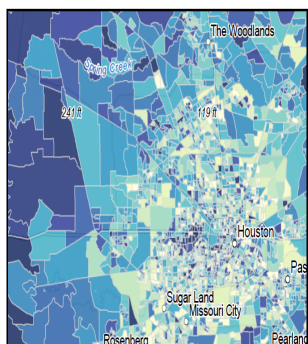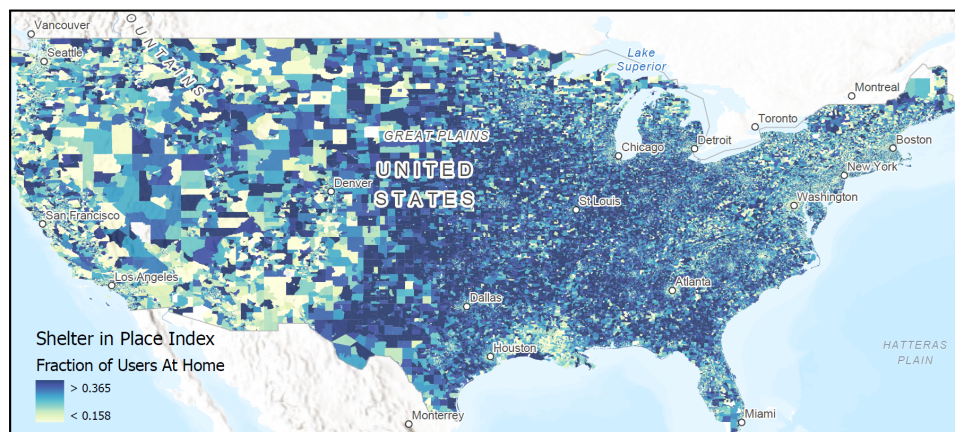
Figure 1: Geographic heterogeneity of stay-at-home behavior

stay-at-home behavior where the fraction of users who stay home is lower in urban areas than in rural areas. Figure 1d and Figure 1c compare two cities with opposite patterns of stay-at-home behavior. Newark, California has a very low fraction of users who stay home, while Lake Charles, Louisiana has a very high fraction.

Although we can observe such spatial differences in behavior, it remains a challenge to understand and explain them sufficiently. Why do people in some places follow social distancing, mask use, and stay-at-home guidelines more strictly than in other places? Why do some support vaccination and not others? What are the driving factors of such behavior? Moreover, what can we do to increase participation in these regions in preventative behavior?

Recent studies have observed a covariate relationship between behavior and the socio-economic and political profile of a region [60, 97,

(a) Percent change in stay-at-home behavior: February–April 2020

(b) Percent change in stay-at-home behavior: February–July 2020

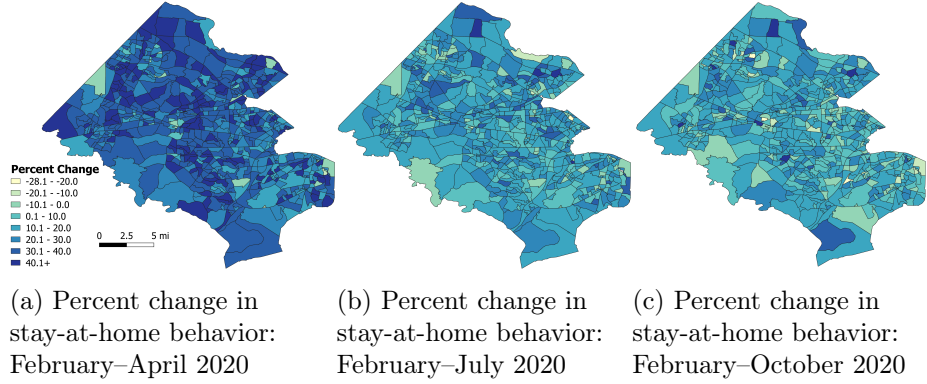(c) Percent change in stay-at-home behavior: February–October 2020

Figure 2: Temporal heterogeneity of stay-at-home behavior

128]. Covariation is a measure of the correlated variation between two variables such that as one variable in a region increases (or decreases), the other also increases (or decreases). Covariate relationships often exhibit spatial trends since individuals who share sociodemographic and thus behavioral commonalities "cluster" together in terms of location. Examining how human behavior varies in geographic space is an emerging area of work in psychology [62, 63].

Covariate relationships between the regional behavioral response to COVID-19 and the corresponding characteristics of those same regions have been explored at various spatial granularities. For example, strong correlations have been observed at the state level between social distancing behavior and income [46, 128], where wealthier states engage in stay-at-home behavior significantly more than lower-income states. Covariation between behavioral response and political tendencies has been observed at the county level, where individuals in Republican counties are less likely to stay home than residents in Democratic counties [46, 60, 97]. Covariation between the number of cases of COVID and behavior has been reported by an early Pew Research Center survey, where states with a higher number of cases have a more significant effect on behavior than states with a lower number of cases [102]. Other variables, including age, race, gender, education, family structure, social connection, religion, and environment, are also likely to be important, but the spatial relationship between these variables and the behavioral response has not yet been quantified. For example, studies have found that young adults are more likely than adults over 30 to attend a party, restaurant, or small gathering [102]. Environmental factors such as temperature and weather are also key in shaping human behavior. Religious gatherings often result in close contact between large numbers of individuals.

In addition to spatial heterogeneity, the behavioral response to a disease outbreak can change over time. Early during the COVID-

19 pandemic, individuals reported being uncomfortable with the idea of attending parties, eating at restaurants, voting, and even going to the grocery store, resulting in increased stay-at-home behavior and decreased mobility [102]. As the disease persists, some individuals have maintained stay-at-home behavior and social distancing, while others' physical, financial or social needs may outweigh the risk of the virus. This behavior change can be directly observed by mapping mobility over time using *SafeGraph's* foot traffic data (Figure 2). In the early stages of the pandemic, that is April 2020 (Figure 2a), the percentage of people who stay home increases compared to February, where February represents "normal mobility". As the pandemic progresses, mobility increases (Figures 2b and 2c) and, in some counties, even returns to normal, where the percent change is close to 0. It is unclear whether this is a function of the individuals' social and financial needs that begin to outweigh the risk of infection, the warmer temperatures that Fairfax County experienced in July and October, the changes in local and state-wide policies over time, or a combination of all of the above.

4   AI for Simulation of Health-Related Human Behavior in Agent-Based Models

More recently, especially in light of the COVID-19 pandemic, the modeling community has worked to improve representations of human behavior in models of disease spread by leveraging AI. As summarized by Brearcliffe and Crooks [28], AI in combination with ABM is used primarily as follows: 1) *to derive behavioral parameter values and patterns* for an ABM to inform dynamics such as human mobility [71, 79, 101] and 2) *as an internal agent decision framework* ranging from simple reactions to complex learning-based decision-making [104]. AI for agent behavior has been integrated in broader applications of ABM and also in applications to ABMs that are used to simulate the spread of diseases.

ABMs of disease spread typically incorporate agents that decide where they want to go (agent mobility) and how they might adjust their behavior, including their mobility, to prevent disease transmission. Health behaviors commonly modeled include vaccination, social distancing, staying at home, and general preventative behavior (usually framed as a combination of several protective behaviors).

As decribed by Russell & Norvig in the book "Artificial Intelligence" [110], there are five types of intelligent agents, increasing in complexity, as follows: reflex agents, model-based reflex agent, goal-based agent, utility-based agent, and learning agent. With each type of agent, behavior and decision-making is self-initatiated, meaning that behavior is endogeneously based on the agent's perception of current,

past, and/or future state of the world which is either complete or limited. This differs from the vast majority of infectious disease models that impose behaviors on agents. Reflex agents, the least complex, have limited intelligence. Agent's simply receive an input, and, using some condition-action rule, will select an action to follow. For example, if someone in the agent's household is sick, then there is some probability that the agent will stay home [86, 96].

Model-based reflex agents perceive the current state of the world and understand how the world evolves, including how the agent's own actions affect the world. Based on this perception, agents use some condition-action to select an action. For example, in Mao [88], agents perceive information about how to protect themselves from disease, the adoption of protective behaviors among their social networks, and the prevalence of local and global disease prevalence. Based on both the fraction of agents in their social network that adopt protective behavior and the fraction that is infected with the disease, agents will decide whether or not to also adopt.

The agent's perception of the world is not always enough to make decisions. For example, a goal-based agent with the goal of moving through a room while maintaining social distancing may have several options in routes to their target [111]. The evaluation of possible routes is binary, meaning that the route either does or does not satisfy the agent's goal. For example, some routes may have points that are too crowded to maintain social distancing, and thus taking these routes to get to the destination will not satisfy the agent's goal. Where goals provide a binary distinction between satisfied and unsatisfied states, economists and computer scientists use utility or "the quality of being useful" as a way to measure the degree of satisfaction that each new potential location would give. The agent will then choose the action that maximizes it's utility or payoff. For example, [95] use an SPIR (susceptible, prophylactic, infectious, recovered) and a rational choice model to represent the choice to adopt prophylactive behavior (handwashing or wearing a face mask) or not. Payoff, rather than utility as a performance measure, is traditionally used to play the "vaccination game" [54], where agents make decisions based on game theory. In this game, vaccines provide immunity against disease. Each individual who gets infected pays the cost of infection, and each individual who decides to get vaccinated pays the cost of vaccination. Thus, the cost paid by a "free-rider" who does not get vaccinated or get infected, is zero. The agent weighs costs with risk of infection based on the number of social connections they have (who might infect them) and how many of those connections have decided to opt for vaccination.

Learning agents are the most complex and adjust their actions over time. For example, Tanaka and Tanimoto [11, 119] use a strategy updating function to incorporate learning into the vaccination game,

where the agent learns from their local or global neighbors about which strategy used in previous years might have the best reward in the current year. Where most of these types of agents implement rule-based intelligence, learning agents sometimes leverage machine learning algorithms such as neural networks, Bayesian networks, reinforcement learning, and genetic algorithms to drive agent behaviors [2]. Outside of disease ABMs, machine learning has been used to simulate agent residential migration decisions [114], spatial optimization of land use allocation [122], land market decisions [1] and conflict management [28]. However, the use of these algorithms for health-related behavior in ABMs of disease spread is somewhat limited. In one example, Abdulkareem [2] uses survey data and Bayesian networks to train and guide agent behavior to simulate risk perception and response to the cholera outbreak. Fuzzy cognitive maps (FCM) model decision making as a system by means of concepts (perceptions of individual health state, local and global disease prevalence, memory, etc.) connected by cause and effect relationships. The FCM can be developed manually or can be learned from data. One such learning method is based on a linear Hebbian learning (NHL) method. Fuzzy cognitive maps have been developed to simulate the protective behaviors of individuals in disease scenarios [84, 91].

5   A Use-Case of Using Agent-Based Modeling to Predict the Spread of an Infectious Disease in Fairfax, VA, USA

This section describes an example agent-based simulation for the spread of an infectious disease using a synthetic population of 1.1 million agents in the county of Fairfax, VA, USA. The following sections describe the datasets used to inform the simulation, how data was abstracted to be digested by the simulation, and the simulation results. Details of this simulation can be found in [101].

5.1   Data Sets

The goal of this use case is to develop an ABM of disease spread based on real-world human mobility patterns rather than relying on simplified and often unrealistic assumptions.

Data from *SafeGraph Inc.*[1] provides unique and valuable insight

---

[1]Attribution: SafeGraph Inc., a data company that aggregates anonymized location data from numerous applications to provide insights into physical presence in places. To enhance privacy, *SafeGraph* aggregates home locations to the census block group level and excludes locations if fewer than five devices visited a POI in a month from a given census

into foot-traffic patterns of large-scale businesses and consumer POIs. This work uses *SafeGraph's* "Weekly Patterns" data, which register GPS-identified visits to POIs (primarily businesses) with an exact location in the United States. For each visit by an individual to a POI, the home census block group (derived from nighttime GPS location) is recorded[2]. Additionally, *SafeGraph* provides a taxonomy of POIs types in a "Core Places" schema, allowing our simulation to test the closure of specific business categories (e.g. restaurants). *SafeGraph* also includes information on the proportion of residents who stay home or leave the house on any given day for each CBG in a separate "Social Distancing Metrics" dataset, allowing us to establish the probabilities of agents leaving their home at the CBG level.

Due to the sheer size of the datasets, all of the data we used was filtered to include only POI and CBG data from Fairfax County, Virginia. We chose to use data from the week that spanned October 28 to November 3, 2019, as a representative sample of typical movement patterns before the onset of COVID-19. We filter POIs only to include those with a large enough sample of aggregate visitors (30 or more) throughout the week-long period. The filtered dataset resulted in 4,130 unique POIs in Fairfax County and 689,731 recorded visits to these POIs.

We also used United States Census data[3] to map the CBGs to their correct geographic locations. This data also facilitates the initialization of agents and agent households.

5.2 Population Initialization

To initialize our simulation, we first generate households according to CBG-level data provided by the US Census, filling each household with its corresponding number of agents. Between one and seven agents are assigned to each household, with "7-or-more person households" being treated as size seven for simplicity. By default, we simulate approximately 10% of the total population of Fairfax County by only generating 10% of households of each size in each CBG, resulting in a simulation of 106,978 agents. In addition to infection status, agents and households are not assigned any other attributes such as age, income, or race. A small percentage (25%) of agents from a single randomly selected CBG are initially infected, resulting in a default of 26 initially infected agents. For consistency, this CBG's *SafeGraph* ID is *510594804023* for all of our trials. We used the integer *1* as the seed for the pseudorandom number generator in all of our trials for

---

block group.

[2]For detailed information, see `https://docs.safegraph.com/docs/weekly-patterns`.
[3]`https://www.census.gov`

reproducibility.

5.3 Representation of Disease Dynamics

Once the agent population initialization is complete, the simulation begins at midnight and runs until the agents are no longer exposed or infected. Each tick in the model represents fifteen minutes, according to the CDC definition of close contact between individuals [51]. The probability that an agent will leave their home location to visit a POI on any given day is based on the *SafeGraph's* "Social Distancing" dataset. We divide the total daily number of people who did not stay home by the total daily number of people in the CBG between October 28 and November 3, 2019. We calculate that the average daily POI visit probability across each CBG in Fairfax County is 74.8%. To roughly approximate the likelihood that an agent would leave the house at each tick, we divide this probability by the number of ticks in a day (96 by default), resulting in a 0.780% average probability. This finding is also consistent with other travel surveys and mobile mobility studies based on cell phones [113]. We consider this probability, calculated based on foot traffic data acquired before the onset of COVID-19 and thus not influenced by the pandemic, as a default parameter of a 100% propensity to leave.

At each 15 minute tick, infectious agents may come into contact with a maximum of five other agents, by default, who are located at the same POI that is not their household. If a susceptible agent comes into contact with an infectious agent, they have a 5% chance of being exposed and subsequently infectious by default [77].

Infectious agents also have the opportunity to spread the virus to susceptible agents in their home. Research indicates that approximately 20.4% of people living in small households (size six or less) will contract the virus if they share a residence with someone infected [68]. This percentage decreases to 9.1% in large households (size seven or larger). Using these numbers and the median infectious period of the virus according to the gamma distributions that we use, as described below, we approximated that susceptible agents have a chance of 4.44% and 1.98% of contracting the virus from an infected household member each day in small and large households, respectively. For simplicity, household infection occurs at midnight each day, even if a household member is visiting a POI.

We represent the dynamics of COVID-19 using a generalized SEIR model [16] that is modified to include subclinical, preclinical and clinical subclasses of the infectious stage. Agents undergo the following stages:

1. Susceptible: An agent who has never been infected or exposed to

the virus but has the potential to become exposed.

2. Exposed: An agent who has caught the virus and will become contagious (infectious) after an incubation period.

3. Infectious: An agent that can infect others and is contagious. We define three subclasses of infectious agents:

   - Subclinical: An asymptomatic infected agent. It is estimated that 40% of infections are subclinical. As these agents will never show symptoms, they are estimated have a 75% relative infectiousness compared to clinical agents.

   - Preclinical: An infected agent who is presymptomatic (not currently symptomatic), but will enter the clinical stage and become symptomatic in the future. All agents entering the clinical stage first pass through the preclinical stage. As preclinical agents do not show symptoms, they are also estimated to have a 75% relative infectiousness compared to clinical agents.

   - Clinical: An infected agent that shows symptoms of the virus and is fully infectious. It is estimated that the remaining 60% infections progress to the clinical stage.

4. Recovered: A previously infected agent that is non-contagious and immune to the virus. An agent is classified as recovered as long as they cannot actively spread the virus, even if they have lasting complications or symptoms.

The duration of each stage in the SEIR model in days is determined by drawing from the following gamma distributions [40]:

- Exposed stage duration: $gamma(\mu = 3.0, k = 4)$

- Subclinical stage duration: $gamma(\mu = 5, k = 4)$

- Preclinical stage duration: $gamma(\mu = 2.1, k = 4)$

- Clinical stage duration: $gamma(\mu = 2.9, k = 4)$

5.4   Foot Traffic Topic Extraction

Latent Topic Modeling

Given the CBG foot traffic data for each POI as obtained from the SafeGraph data, we apply topic modeling using LDA [25] – a generative probabilistic model. Although traditionally used to find K latent topics among a corpus of M text documents containing N words per document, we use LDA in our simulation to find K latent topics among a subset of M CBGs each containing N distinct POI visits. This modeling of CBGs as documents and POIs as words allows us to efficiently generate realistic new POI visits of individuals at a CGB. In using

the LDA approach, each CBG's POI visits are a mixture of underlying latent topics, and each topic has a latent distribution of more and less likely POIs. In that respect, LDA provides two distributions: 1) a topic probability distribution for each visitor home CBG and 2) a POI probability distribution for each topic. These two distributions allow an ABM to be constructed such that 1) agents are generated and assigned a specific LDA topic according to the topic probability distribution of their home CBG and that 2) agents visit POIs based on the POI probability distribution of their assigned topic.

We are now able to generate agents with specific attributes based on their home CBG, providing the foundation for our ABM. First, for each home CBG in Fairfax County that *SafeGraph* provides data for, agents are generated according to the real population of the CBG, each being assigned a topic according to the first distribution. Agents' topics are static and cannot change throughout the simulation.

After agents are assigned topics, we may randomly sample from the second distribution to determine which POI agents will visit if they decide to leave their house. Due to LDA's nature, it is unlikely that two unrelated POIs, such as a nightclub and a library, will have relatively equal weights in this probability distribution, resulting in a vast improvement from the uniform probability distribution used to select POIs in traditional ABMs. However, this probability distribution does not provide information on whether or not an agent will decide to visit a POI in the first place. Details of the LDA model can be found in [101].

Modeling Hourly Visit Patterns

So far, LDA has generated distributions from data that contain the number of visitors from each CBG to each POI over an entire day. However, this approach is slightly flawed because, in reality, the number of visits is dependent on the time of day. For example, a restaurant would likely have higher concentrations of visits at noon and in the evening and lower concentrations during the mid-afternoon. Similarly, visits to a school POI during the evening would be less likely. To remedy this issue, we use additional *SafeGraph* data that provides individual POI visits for each hour over the entire week-long timeframe and create 24 distinct POI probability distributions for every topic, one for each hour of the day. We do this by reweighting the topic's base POI probability distribution 24 times according to each POI's proportion of visits during the given hour. For each topic, a weighted distribution of visits that take place each hour is constructed according to the topic's POI distribution and the number of visits to those POIs that take place during the given hour. This distribution provides the weighted percentage of visits to the topic that takes place in the

given hour compared to the entire day. For example, the midnight hour may have a probability of 0.01, while the noon hour may have a probability of 0.07. By modifying the likelihood, we allow simulating agent POI visits more accurately. Given our simulation's default parameters for the hour of noon example, an agent's chance to visit a POI between 12:15 and 12:30 would be approximately $0.748*0.07/4$, or 1.31%, markedly higher than the generic average probability of leaving each tick of 0.780%.

5.5 Dwell Time Distributions for POIs

The *SafeGraph* data allows us to use a data-driven approach to modeling dwell time by fitting a suitable probability distribution to a POI's bucketed dwell time data (provided by *SafeGraph*) and random sampling the said distribution for every agent that "visits" the POI.

SafeGraph Bucketed Dwell Time Data

SafeGraph provides a "bucketed" version of each POI's dwell times, where only the number of visits within a range of dwell time is quantified, i.e., *"<5 minutes": 266, "5-20 minutes": 4184, "21-60 minutes": 3597, "61-240 minutes": 2492, ">240 minutes": 892.* While providing an initial perspective on the potential probability of dwell times for an individual agent's visit to a POI, the raw *SafeGraph* data is not adequate for direct usage due to their nonspecificity.

Fitting Probability Distributions

To compensate for the bucketed format of the data ("5-20 minutes":4184), we first impute the bucket ranges of each POI's dwell data by random uniform sampling: for a range of 5-20 minutes with 266 visits, we fill the bucket with 266 random uniform samples ranging from 5-20. With a full range of dwell data for each bucket, we move on to methods of sampling. We determine that employing probability distributions allows for the most optimal method of a random sample due to their "smoothing" of minor irregularities that may occur from the random uniform imputation of each POI's dwell time buckets. From these we approximate using the parametric function with the best fit for each POI dwell time distribution. For example, restaurants might have a more normal distribution around a mean stay time of 1 hour. Compared to malls, where a large proportion might drop off or drive by with a large proportion of visits under 5 minutes, so it might be better represented using an exponential curve. For each POI, we test the fit of 10 of SciPy's most common probability distributions for a

15

(a) Intervention: Generic Quarantine   (b) Intervention: Household Quarantine

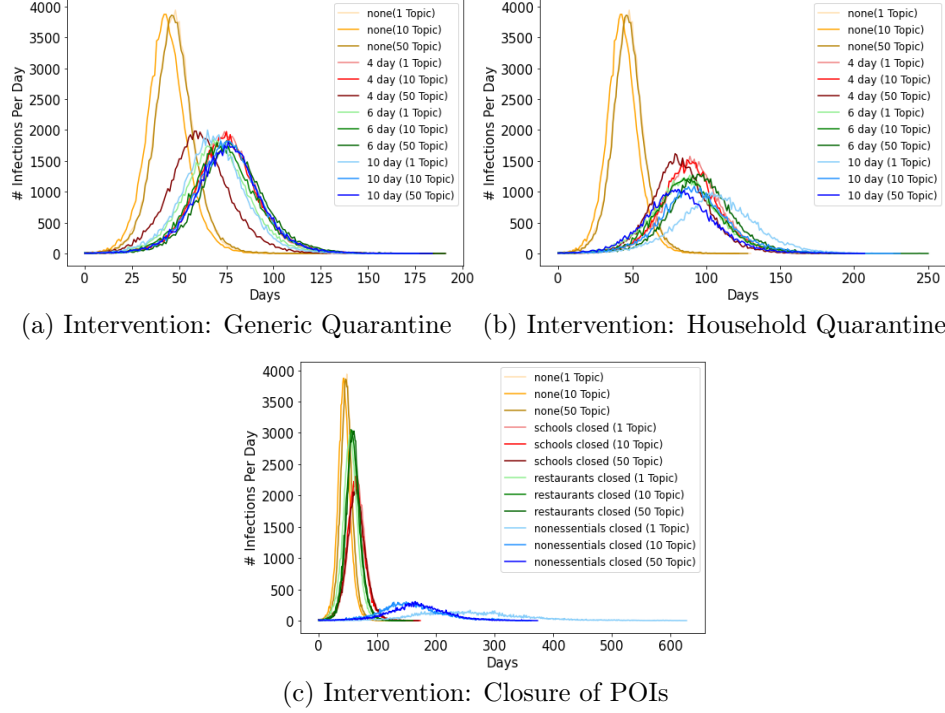(c) Intervention: Closure of POIs

Figure 3: Disease Spread Simulation Results after Prescribing Interventions

continuous random variable—normal, generalized extreme value, exponential, gamma, Pareto, lognormal, double Weibull, beta, Student's t, uniform—and select the most optimal based on the goodness-of-fit test. See [120] for more information on the probability distributions used and the package used. We initialize and cache the fitted distribution for the POI in question; the distribution is randomly sampled five times for each agent visit to said POI. The median is returned, representing an estimated *SafeGraph* data-based dwell time. If the median dwell time is less than one tick, then one tick is returned. Alternatively, if the median dwell time is greater than 16 hours, then 16 hours is returned. The median dwell time is rounded to the nearest tick in all other cases.

5.6   Simulation Results

Various experiments are implemented to test the impact of various public health interventions on the spread of COVID-19 and the subsequent effect on the epidemic curve.

Generic Quarantine

This intervention requires any agent that is infectious and aware of symptoms (in the subclinical stage or beyond) to stay home for a specific number of days. Figure 3a shows the infection curves for zero, four, six and ten days of quarantine. We observe that a quarantine of four days is sufficient to drastically flatten the number of infections per day from 2000. However, since 40% of the agents are subclinical and unaware of their infection, these agents are not quarantined and continue to spread the disease.

Household Quarantine

This intervention additionally requires that all agents that share the same home as the quarantined agent remain at home. This intervention is significantly more effective, further flattening the curve to about 1500 infections per day. This result is intuitive, as agents living in the same household are at highest risk of becoming infected. We also observe that a longer quarantine duration is significantly more effective. Specifically, agents are quarantined before they show symptoms, thus dropping the infections per day to 1,000 using a ten-day quarantine.

Closure of POIs

We tested three interventions, each aimed at closing a specific type of POI including schools, restaurants, and non-essential places. Figure 3c presents the effect of POI closure on the epidemic curve. Instead of visiting the closed POI, the agent will decide to stay at home. We observe that closing restaurants yields a significant reduction of disease spread from a disease peak of 4000 agents per day down to 3000 agents per day. Interestingly, the closure of schools is far more important, reducing the peak to about 2000. This is likely due to the difference in dwell time between schools and restaurants. As agents typically dwell at school POIs for up to eight hours a day, it becomes extremely likely that during any 15-minute tick they are successfully exposed to the virus by an infectious agent. In contrast, dwell times at restaurants are usually less than an hour (this also includes fast food restaurants), drastically reducing the probability of becoming exposed by a collocated infectious agent. We also see that if we choose to close all POIs that are classified as non-essentials (using the POI classification provided by SafeGraph), we observe that the disease is nearly eradicated.

6 Future Direction for GeoAI for Public Health: Prescriptive Analytics

Existing research in GeoAI excels at predictive analytics, such as predicting road traffic [98, 134], rental bike flow [34, 82], and foot traffic [67, 126]. Such solutions excel at finding spatiotemporal patterns to optimize predictions, but lack an understanding of causality that allows one to investigate "what-if" scenarios. For GeoAI to become useful in public health scenarios, we need to go beyond prediction analytics towards prescriptive analytics [52, 75]. Rather than predicting a variable of interest (such as the number of cases of an infectious disease), the goal of prescriptive analytics is to prescribe optimal actions and policies to optimize the variable of interest (such as minimizing the number of cases). Examples of such actions and policies are social distancing measures and the closure of businesses. Prescriptive analytics use machine learning approaches to find the optimal combination of actions and policies that achieve a desired outcome and have been successfully applied in management science and business analytics [22, 81]. A good example of prescriptive analytics for public health was presented at the Prescriptive Analytics for the Physical World (PAPW 2020) workshop held in conjunction with ACM KDD 2020[4]. The workshop featured a programming challenge in which teams would prescribe policies to mitigate the spread of an infectious disease in a simulated scenario. Teams would be able to observe infectious (but not exposed) agents in the simulation and prescribe interventions such as isolation, quarantine, and hospitalization of agents. Teams were then evaluated using a score weighted by the resulting number of infections, but also weighted by the severity of interventions. The top-ranked solutions used diverse machine learning solutions to find optimal policies to minimize infections while also minimizing the severity of interventions [44, 73, 108]. Unfortunately, these machine learning solutions are not applicable to real-world disease spread, as the simulation used in this challenge was overly simple and assumed that all simulated agents behaved the same way and visited the same locations with the same probability.

To use prescriptive analytics for infectious disease mitigation in the real-world, we require a realistic digital twin of a city, county, or even the entire world. Then we could use this simulation to investigate optimal prescriptions (actions or interventions) to mitigate disease outbreaks. Work towards such a realistic simulation has to be interdisciplinary. Social scientists ensure realistic human behavior in a simulation, epidemiologists ensure realistic infectious disease spread for emerging pathogens, and computer scientists leverage efficient algorithms to allow the simulation to scale to large populations despite

---

[4]https://prescriptive-analytics.github.io/

high social and epidemiologic complexity.

Such prescriptions may include specific strategies aimed at isolation, quarantine, vaccination of individuals, closure of sites, working from home, and usage of masks. The union of all prescriptions injected into a single simulation is called a policy and each policy injected into a simulation creates a "what-if" simulation outcome called a "possible world". Such policies may be constrained. For example, at a specific time, only 1000 vaccines may be produced per day, such that the goal is to find the optimal set of people to vaccinate, or the goal might minimize disease spread by closing sites but constrained to an economic loss of no more than $100M$ for a study region.

Once the space of injectable actions and the constraints to define viable policies are defined, GeoAI solutions may be leveraged to search for optimal policies. This can be done by running a massive number of simulations and mining patterns, such as disease hot-spots, from the simulations. Such data mining approaches can identify the different locations that may benefit from the same optimal policies. This approach would allow us to explore the causality between different variables and behavior further. For example, we may find that similar policies are effective in places with similar political or religious beliefs. We can classification models to build supervised models that allow us to predict, for a given place and given mobility and behavior data, what the most effective sets of policies will be. This will give us a broader understanding of how different policies exhibit different degrees of effectiveness among different populations and places. Having experts in epidemiology and policy makers (such as local health departments) in the loop, this optimization may help find optimal policies for expert-defined applications such as the closure of businesses, lock-down, or vaccination rollout.

References

[1] ABDULKAREEM, S. A. *Enhancing Agent-Based Models with Artificial Intelligence for Complex Decision Making.* PhD thesis, University of Twente, 2019.

[2] ABDULKAREEM, S. A., MUSTAFA, Y. T., AUGUSTIJN, E.-W., AND FILATOVA, T. Bayesian networks for spatial learning: a workflow on using limited survey data for intelligent learning in spatial agent-based models. *Geoinformatica 23*, 2 (2019), 243–268.

[3] ABOUBAKR, H. A., AND MAGDY, A. On improving toll accuracy for covid-like epidemics in underserved communities using

user-generated data. In *Proceedings of the 1st ACM SIGSPA-TIAL International Workshop on Modeling and Understanding the Spread of COVID-19* (2020), pp. 32–35.

[4] ADHIKARI, B., SRIVASTAVA, A., PEI, S., KEFAYATI, S., YU, R., YADAV, A., RODRÍGUEZ, A., RAMANATHAN, A., VULLIKANTI, A., AND PRAKASH, B. A. The 4th international workshop on epidemiology meets data mining and knowledge discovery (epidamik 4.0@ kdd2021). In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (2021), pp. 4104–4105.

[5] ADHIKARI, B., XU, X., RAMAKRISHNAN, N., AND PRAKASH, B. A. Epideep: Exploiting embeddings for epidemic forecasting. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (2019), pp. 577–586.

[6] ADHIKARI, B., YADAV, A., PEI, S., SRIVASTAVA, A., KEFAYATI, S., RODRÍGUEZ, A., CHARPIGNON, M.-L., VULLIKANTI, A., AND PRAKASH, B. A. epidamik 5.0: The 5th international workshop on epidemiology meets data mining and knowledge discovery. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (2022), pp. 4850–4851.

[7] ADIGA, A., DUBHASHI, D., LEWIS, B., MARATHE, M., VENKATRAMANAN, S., AND VULLIKANTI, A. Mathematical models for covid-19 pandemic: a comparative analysis. *Journal of the Indian Institute of Science 100*, 4 (2020), 793–807.

[8] AGARWAL, R., AND BANERJEE, A. Infection risk score: Identifying the risk of infection propagation based on human contact. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Modeling and Understanding the Spread of COVID-19* (2020), pp. 1–10.

[9] AGUSTO, F. B., EROVENKO, I. V., FULK, A., ABU-SAYMEH, Q., ROMERO-ALVAREZ, D. D., PONCE, J., SINDI, S., ORTEGA, O., SAINT ONGE, J. M., AND PETERSON, A. T. To isolate or not to isolate: The impact of changing behavior on covid-19 transmission. *medRxiv* (2020).

[10] AJAYAKUMAR, J., CURTIS, A., AND CURTIS, J. A clustering environment for real-time tracking and analysis of covid-19 case clusters. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Spatial Computing for Epidemiology (SpatialEpi 2021)* (2021), pp. 1–9.

[11] ALAM, M., TANAKA, M., AND TANIMOTO, J. A game theoretic approach to discuss the positive secondary effect of vaccination scheme in an infinite and well-mixed population. *Chaos, Solitons & Fractals 125* (2019), 201–213.

[12] ANDERSON, T., AND DRAGIĆEVIĆ, S. Neat approach for testing and validation of geospatial network agent-based model processes: case study of influenza spread. *International Journal of Geographical Information Science* (2020), 1–30.

[13] ANDERSON, T., ROESS, A., KIM, J.-S., AND ZÜFLE, A. *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Spatial Computing for Epidemiology (SpatialEpi 2022)*. ACM, 2022.

[14] ANDERSON, T., YU, J., ROESS, A., KAVAK, H., KIM, J.-S., AND ZÜFLE, A. *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Spatial Computing for Epidemiology (SpatialEpi 2021)*. ACM, 2021.

[15] ANDERSON, T., YU, J., AND ZÜFLE, A. The 1st acm sigspatial international workshop on modeling and understanding the spread of covid-19. *SIGSPATIAL Special 12*, 3 (2021), 35–40.

[16] ARON, J. L., AND SCHWARTZ, I. B. Seasonality and period-doubling bifurcations in an epidemic model. *Journal of theoretical biology 110*, 4 (1984), 665–679.

[17] AZAD, F. T., DODGE, R. W., VARGHESE, A. M., LEE, J., PEDRIELLI, G., CANDAN, K. S., AND CHOWELL-PUENTE, G. Sirtem: Spatially informed rapid testing for epidemic modeling and response to covid-19. *ACM Transactions on Spatial Algorithms and Systems 8*, 4 (2022), 1–43.

[18] BALDASSARE, M. Human spatial behavior. *Annual Review of Sociology 4*, 1 (1978), 29–56.

[19] BARRETT, C., BISSET, K., LEIDIG, J., MARATHE, A., AND MARATHE, M. Economic and social impact of influenza mitigation strategies by demographic class. *Epidemics 3*, 1 (2011), 19–31.

[20] BAUCH, C., D'ONOFRIO, A., AND MANFREDI, P. Behavioral epidemiology of infectious diseases: an overview. In *Modeling the interplay between human behavior and the spread of infectious diseases*. Springer, 2013, pp. 1–19.

[21] BEHERA, S., DOGRA, D. P., AND SATPATHY, M. Effect of migrant labourer inflow on the early spread of covid-19 in odisha: A case study. *ACM Transactions on Spatial Algorithms and Systems 8*, 4 (2022), 1–18.

[22] BERTSIMAS, D., AND KALLUS, N. From predictive to prescriptive analytics. *Management Science 66*, 3 (2020), 1025–1044.

[23] BHARTI, N. Linking human behaviors and infectious diseases. *Proceedings of the National Academy of Sciences 118*, 11 (2021).

[24] BIAN, L. A conceptual framework for an individual-based spatially explicit epidemiological model. *Environment and Planning B: Planning and Design 31*, 3 (2004), 381–395.

[25] BLEI, D. M., NG, A. Y., AND JORDAN, M. I. Latent dirichlet allocation. *Journal of machine Learning research 3*, Jan (2003), 993–1022.

[26] BOBASHEV, G., SEGOVIA-DOMINGUEZ, I., GEL, Y. R., RINEER, J., RHEA, S., AND SUI, H. Geospatial forecasting of covid-19 spread and risk of reaching hospital capacity. *SIGSPATIAL Special 12*, 2 (2020), 25–32.

[27] BOSSERT, A., KERSTING, M., TIMME, M., SCHRÖDER, M., FEKI, A., COETZEE, J., AND SCHLÜTER, J. Limited containment options of covid-19 outbreak revealed by regional agent-based simulations for south africa. *arXiv preprint arXiv:2004.05513* (2020).

[28] BREARCLIFFE, D. K., AND CROOKS, A. Creating intelligent agents: Combining agent-based modeling with machine learning. In *Proceedings of the 2020 Conference of The Computational Social Science Society of the Americas* (2021), Springer, pp. 31–58.

[29] BROOKS, L. C., FARROW, D. C., HYUN, S., TIBSHIRANI, R. J., AND ROSENFELD, R. Nonmechanistic forecasts of seasonal influenza with iterative one-week-ahead distributions. *PLoS computational biology 14*, 6 (2018), e1006134.

[30] BURKE, D. S., EPSTEIN, J. M., CUMMINGS, D. A., PARKER, J. I., CLINE, K. C., SINGA, R. M., AND CHAKRAVARTY, S. Individual-based computational modeling of smallpox epidemic control strategies. *Academic Emergency Medicine 13*, 11 (2006), 1142–1149.

[31] BURTNER, S., AND MURRAY, A. T. Covid-19 and minimizing micro-spatial interactions. *ACM Transactions on Spatial Algorithms and Systems (TSAS) 8*, 3 (2022), 1–17.

[32] CABANA, E., LUTU, A., FRIAS-MARTINEZ, E., AND LAOUTARIS, N. Using mobile network data to color epidemic risk maps. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Spatial Computing for Epidemiology* (2022), pp. 35–44.

[33] CARDOSO, M., CAVALHEIRO, A., BORGES, A., DUARTE, A. F., SOARES, A., PEREIRA, M. J., NUNES, N. J., AZEVEDO, L., AND OLIVEIRA, A. L. Modeling the geospatial evolution of covid-19 using spatio-temporal convolutional sequence-to-sequence neural networks. *ACM Transactions on Spatial Systems and Algorithms* (2021).

[34] CHAI, D., WANG, L., AND YANG, Q. Bike flow prediction with multi-graph convolutional networks. In *Proceedings of the 26th ACM SIGSPATIAL international conference on advances in geographic information systems* (2018), pp. 397–400.

[35] CHAO, D. L., HALLORAN, M. E., OBENCHAIN, V. J., AND LONGINI JR, I. M. Flute, a publicly available stochastic influenza epidemic simulation model. *PLoS Comput Biol 6*, 1 (2010), e1000656.

[36] CHEN, E., AND MCKENZIE, G. Mobility response to covid-19-related restrictions in new york city. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Spatial Computing for Epidemiology (SpatialEpi 2021)* (2021), pp. 10–13.

[37] CHEN, L.-C., CARLEY, K. M., FRIDSMA, D., KAMINSKY, B., AND YAHJA, A. Model alignment of anthrax attack simulations. *Decision Support Systems 41*, 3 (2006), 654–668.

[38] CORO, G., AND BOVE, P. A high-resolution global-scale model for covid-19 infection rate. *ACM Transactions on Spatial Algorithms and Systems (TSAS) 8*, 3 (2022), 1–24.

[39] CRAMER, E. Y., RAY, E. L., LOPEZ, V. K., BRACHER, J., BRENNEN, A., CASTRO RIVADENEIRA, A. J., GERDING, A., GNEITING, T., HOUSE, K. H., HUANG, Y., ET AL. Evaluation of individual and ensemble probabilistic forecasts of covid-19 mortality in the united states. *Proceedings of the National Academy of Sciences 119*, 15 (2022), e2113561119.

[40] DAVIES, N. G., KLEPAC, P., LIU, Y., PREM, K., JIT, M., EGGO, R. M., WORKING GROUP, C. C.-., ET AL. Age-dependent effects in the transmission and control of covid-19 epidemics. *MedRxiv* (2020).

[41] DEL VALLE, S. Y., MNISZEWSKI, S. M., AND HYMAN, J. M. Modeling the impact of behavior changes on the spread of pandemic influenza. In *Modeling the interplay between human behavior and the spread of infectious diseases*. Springer, 2013, pp. 59–77.

[42] DENG, S., WANG, S., RANGWALA, H., WANG, L., AND NING, Y. Cola-gnn: Cross-location attention based graph neural networks for long-term ili prediction. In *Proceedings of the 29th ACM international conference on information & knowledge management* (2020), pp. 245–254.

[43] DIGNUM, F., DIGNUM, V., DAVIDSSON, P., GHORBANI, A., VAN DER HURK, M., JENSEN, M., KAMMLER, C., LORIG, F., LUDESCHER, L. G., MELCHIOR, A., ET AL. Analysing the combined health, social and economic impacts of the corovanvirus pandemic using agent-based social simulation. *arXiv preprint arXiv:2004.12809* (2020).

[44] DONG, Y., YU, C., AND XIA, L. Hierarchical reinforcement learning for epidemics intervention. In *First International KDD Workshop for Prescriptive Analytics for the Physical World (PAPW 2020) (https://prescriptive-analytics.github.io/file/5-YinzhaoDong.pdf)* (2020).

[45] D'ORAZIO, M., BERNARDINI, G., AND QUAGLIARINI, E. How to restart? an agent-based simulation model towards the definition of strategies for covid-19" second phase" in public buildings. *arXiv preprint arXiv:2004.12927* (2020).

[46] ELARDE, J., KIM, J.-S., KAVAK, H., ZÜFLE, A., AND ANDERSON, T. Change of human mobility during covid-19: A united states case study. *PloS one 16*, 11 (2021), e0259031.

[47] ELSAKA, T., AFYOUNI, I., HASHEM, I., AND AL AGHBARI, Z. Correlation analysis of spatio-temporal arabic covid-19 tweets. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Spatial Computing for Epidemiology (SpatialEpi 2021)* (2021), pp. 14–17.

[48] FAN, Z., SONG, X., LIU, Y., ZHANG, Z., YANG, C., CHEN, Q., JIANG, R., AND SHIBASAKI, R. Human mobility based individual-level epidemic simulation platform. *SIGSPATIAL Special 12*, 1 (2020), 34–40.

[49] FAN, Z., YANG, C., ZHANG, Z., SONG, X., LIU, Y., JIANG, R., CHEN, Q., AND SHIBASAKI, R. Human mobility-based

individual-level epidemic simulation platform. *ACM Transactions on Spatial Algorithms and Systems (TSAS) 8*, 3 (2022), 1–16.

[50] FANTICELLI, H. C., RABENJAMINA, S., VIANA, A. C., STANICA, R., DE OLIVEIRA, L. S., AND ZIVIANI, A. Data-driven mobility analysis and modeling: Typical and confined life of a metropolitan population. *ACM Transactions on Spatial Systems and Algorithms* (2022).

[51] FOR DISEASE CONTROL, C., AND PREVENTION. Contact tracing for covid-19. *Centers for Disease Control and Prevention* (Aug 2020).

[52] FRAZZETTO, D., NIELSEN, T. D., PEDERSEN, T. B., AND ŠIKŠNYS, L. Prescriptive analytics: a survey of emerging trends and technologies. *The VLDB Journal 28*, 4 (2019), 575–595.

[53] FRIAS-MARTINEZ, E., WILLIAMSON, G., AND FRIAS-MARTINEZ, V. An agent-based model of epidemic spread using human mobility and social network information. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing* (2011), IEEE, pp. 57–64.

[54] FU, F., ROSENBLOOM, D. I., WANG, L., AND NOWAK, M. A. Imitation dynamics of vaccination behaviour on social networks. *Proceedings of the Royal Society B: Biological Sciences 278*, 1702 (2011), 42–49.

[55] FUNK, S., BANSAL, S., BAUCH, C. T., EAMES, K. T., EDMUNDS, W. J., GALVANI, A. P., AND KLEPAC, P. Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. *Epidemics 10* (2015), 21–25.

[56] FUNK, S., SALATHÉ, M., AND JANSEN, V. A. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of the Royal Society Interface 7*, 50 (2010), 1247–1256.

[57] GAO, S., RAO, J., KANG, Y., LIANG, Y., AND KRUSE, J. Mapping county-level mobility pattern changes in the united states in response to covid-19. *SIGSPATIAL Special 12*, 1 (2020), 16–26.

[58] GINSBERG, J., MOHEBBI, M. H., PATEL, R. S., BRAMMER, L., SMOLINSKI, M. S., AND BRILLIANT, L. Detecting influenza epidemics using search engine query data. *Nature 457*, 7232 (2009), 1012–1014.

[59] GOLLEDGE, R. G. *Spatial behavior: A geographic perspective.* Guilford Press, 1997.

[60] GOLLWITZER, A., MARTEL, C., BRADY, W. J., PÄRNAMETS, P., FREEDMAN, I. G., KNOWLES, E. D., AND VAN BAVEL, J. J. Partisan differences in physical distancing are linked to health outcomes during the covid-19 pandemic. *Nature Human Behaviour* (2020), 1–12.

[61] HALDER, N., KELSO, J. K., AND MILNE, G. J. Developing guidelines for school closure interventions to be used during a future influenza pandemic. *BMC infectious diseases 10*, 1 (2010), 221.

[62] HEHMAN, E., CALANCHINI, J., FLAKE, J. K., AND LEITNER, J. B. Establishing construct validity evidence for regional measures of explicit and implicit racial bias. *Journal of Experimental Psychology: General 148*, 6 (2019), 1022.

[63] HEHMAN, E., OFOSU, E. K., AND CALANCHINI, J. Using environmental features to maximize prediction of regional intergroup bias. *Social Psychological and Personality Science* (2020), 1948550620909775.

[64] HOERTEL, N., BLACHIER, M., BLANCO, C., OLFSON, M., MASSETTI, M., RICO, M. S., LIMOSIN, F., AND LELEU, H. A stochastic agent-based model of the sars-cov-2 epidemic in france. *Nature Medicine* (2020), 1–5.

[65] HOLMDAHL, I., AND BUCKEE, C. Wrong but useful—what covid-19 epidemiologic models can and cannot tell us. *New England Journal of Medicine* (2020).

[66] IMRAN, M., QAZI, U., AND OFLI, F. Tbcov: two billion multilingual covid-19 tweets with sentiment, entity, geo, and gender labels. *Data 7*, 1 (2022), 8.

[67] ISLAM, S., GANDHI, D., ELARDE, J., ANDERSON, T., ROESS, A., LESLIE, T. F., KAVAK, H., AND ZÜFLE, A. Spatiotemporal prediction of foot traffic. In *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Location-based Recommendations, Geosocial Networks and Geoadvertising* (2021), pp. 1–8.

[68] JING, Q.-L., LIU, M.-J., ZHANG, Z.-B., FANG, L.-Q., YUAN, J., ZHANG, A.-R., DEAN, N. E., LUO, L., MA, M.-M., LONGINI, I., ET AL. Household secondary attack rate of covid-19 and associated determinants in guangzhou, china: a retrospective cohort study. *The Lancet Infectious Diseases* (2020).

[69] KANG, Y., GAO, S., LIANG, Y., LI, M., RAO, J., AND KRUSE, J. Multiscale dynamic human mobility flow dataset in the us during the covid-19 epidemic. *Scientific data 7*, 1 (2020), 1–13.

[70] KARL, S., HALDER, N., KELSO, J. K., RITCHIE, S. A., AND MILNE, G. J. A spatial simulation model for dengue virus infection in urban areas. *BMC infectious diseases 14*, 1 (2014), 1–17.

[71] KAVAK, H., PADILLA, J. J., LYNCH, C. J., AND DIALLO, S. Y. Big data, agents, and machine learning: towards a data-driven agent-based modeling approach. In *Proceedings of the Annual Simulation Symposium* (2018), pp. 1–12.

[72] KIAMARI, M., RAMACHANDRAN, G., NGUYEN, Q., PEREIRA, E., HOLM, J., AND KRISHNAMACHARI, B. Covid-19 risk estimation using a time-varying sir-model. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Modeling and Understanding the Spread of COVID-19* (2020), pp. 36–42.

[73] KIM, J.-S., JIN, H., AND ZÜFLE, A. Expert-in-the-loop prescriptive analytics using mobility intervention for epidemics. In *First International KDD Workshop for Prescriptive Analytics for the Physical World (PAPW 2020) (https:// prescriptive-analytics. github. io/ file/ 4-hjin. pdf)* (2020).

[74] KIM, J.-S., KAVAK, H., MANZOOR, U., AND ZÜFLE, A. Advancing simulation experimentation capabilities with runtime interventions. In *SpringSim 2019* (2019), IEEE, pp. 1–11.

[75] KIM, J.-S., KAVAK, H., ROULY, C. O., JIN, H., CROOKS, A., PFOSER, D., WENK, C., AND ZÜFLE, A. Location-based social simulation for prescriptive analytics of disease spread. *SIGSPATIAL Special 12*, 1 (2020), 53–61.

[76] KIM, J.-S., KAVAK, H., ZÜFLE, A., AND ANDERSON, T. Covid-19 ensemble models using representative clustering. *SIGSPATIAL Special 12*, 2 (2020), 33–41.

[77] KLOMPAS, M., BAKER, M. A., AND RHEE, C. Airborne transmission of sars-cov-2: theoretical considerations and available evidence. *JAMA* (2020).

[78] LAATO, S., ISLAM, A. N., FAROOQ, A., AND DHIR, A. Unusual purchasing behavior during the early stages of the covid-19 pandemic: The stimulus-organism-response approach. *Journal of Retailing and Consumer Services 57* (2020), 102224.

[79] LAMPERTI, F., ROVENTINI, A., AND SANI, A. Agent-based model calibration using machine learning surrogates. *Journal of Economic Dynamics and Control 90* (2018), 366–389.

[80] LEE, B. Y., BROWN, S. T., KORCH, G. W., COOLEY, P. C., ZIMMERMAN, R. K., WHEATON, W. D., ZIMMER, S. M., GREFENSTETTE, J. J., BAILEY, R. R., ASSI, T.-M., ET AL. A computer simulation of vaccine prioritization, allocation, and rationing during the 2009 h1n1 influenza pandemic. *Vaccine 28*, 31 (2010), 4875–4879.

[81] LEPENIOTI, K., BOUSDEKIS, A., APOSTOLOU, D., AND MENTZAS, G. Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management 50* (2020), 57–70.

[82] LI, Y., ZHENG, Y., ZHANG, H., AND CHEN, L. Traffic prediction in a bike-sharing system. In *Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems* (2015), pp. 1–10.

[83] LOPES, G. R., DELBEM, A. C., DA SILVA, R. F., JÚNIOR, C. B., DE MATTOS, S. H. V. L., SCATOLINI, D., GHIGLIENO, F., AND SARAIVA, A. M. Multimaps: a tool for decision-making support in the analyzes of multiple epidemics. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Spatial Computing for Epidemiology* (2022), pp. 22–25.

[84] LÓPEZ, L., FERNÁNDEZ, M., AND GIOVANINI, L. Influenza epidemic model using dynamic social networks of individuals with cognition maps. *MethodsX 7* (2020), 101030.

[85] LORCH, L., KREMER, H., TROULEAU, W., TSIRTSIS, S., SZANTO, A., SCHÖLKOPF, B., AND GOMEZ-RODRIGUEZ, M. Quantifying the effects of contact tracing, testing, and containment measures in the presence of infection hotspots. *ACM Transactions on Spatial Algorithms and Systems 8*, 4 (2022), 1–28.

[86] LUKENS, S., DEPASSE, J., ROSENFELD, R., GHEDIN, E., MOCHAN, E., BROWN, S. T., GREFENSTETTE, J., BURKE, D. S., SWIGON, D., AND CLERMONT, G. A large-scale immuno-epidemiological simulation of influenza a epidemics. *BMC public health 14*, 1 (2014), 1–15.

[87] MANFREDI, P., AND D'ONOFRIO, A. *Modeling the interplay between human behavior and the spread of infectious diseases.* Springer Science & Business Media, 2013.

[88] MAO, L. Modeling triple-diffusions of infectious diseases, information, and preventive behaviors through a metropolitan social network—an agent-based simulation. *Applied Geography 50* (2014), 31–39.

[89] MARATHE, M., AND VULLIKANTI, A. K. S. Computational epidemiology. *Communications of the ACM 56*, 7 (2013), 88–96.

[90] MEHRAB, Z., ADIGA, A., MARATHE, M. V., VENKATRAMANAN, S., AND SWARUP, S. Evaluating the utility of high-resolution proximity metrics in predicting the spread of covid-19. *ACM Transactions on Spatial Systems and Algorithms* (2021).

[91] MEI, S., ZHU, Y., QIU, X., ZHOU, X., ZU, Z., BOUKHANOVSKY, A., AND SLOOT, P. M. Individual decision making can drive epidemics: a fuzzy cognitive map study. *IEEE Transactions on Fuzzy Systems 22*, 2 (2013), 264–273.

[92] MERLER, S., AJELLI, M., FUMANELLI, L., GOMES, M. F., Y PIONTTI, A. P., ROSSI, L., CHAO, D. L., LONGINI JR, I. M., HALLORAN, M. E., AND VESPIGNANI, A. Spatiotemporal spread of the 2014 outbreak of ebola virus disease in liberia and the effectiveness of non-pharmaceutical interventions: a computational modelling analysis. *The Lancet Infectious Diseases 15*, 2 (2015), 204–211.

[93] MOKBEL, M., ABBAR, S., AND STANOJEVIC, R. Contact tracing: Beyond the apps. *SIGSPATIAL Special 12*, 2 (2020), 15–24.

[94] MOKBEL, M., SAKR, M., XIONG, L., ZÜFLE, A., ALMEIDA, J., ANDERSON, T., AREF, W., ANDRIENKO, G., ANDRIENKO, N., CAO, Y., ET AL. Mobility data science (dagstuhl seminar 22021). In *Dagstuhl reports* (2022), Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

[95] NARDIN, L. G., MILLER, C. R., RIDENHOUR, B. J., KRONE, S. M., JOYCE, P., AND BAUMGAERTNER, B. O. Planning horizon affects prophylactic decision-making and epidemic dynamics. *PeerJ 4* (2016), e2678.

[96] ORUC, B. E., BAXTER, A., KESKINOCAK, P., ASPLUND, J., AND SERBAN, N. Homebound by covid19: the benefits and consequences of non-pharmaceutical intervention strategies. *BMC public health 21*, 1 (2021), 1–8.

[97] PAINTER, M., AND QIU, T. Political beliefs affect compliance with covid-19 social distancing orders. *Available at SSRN 3569098* (2020).

[98] PAN, Z., LIANG, Y., WANG, W., YU, Y., ZHENG, Y., AND ZHANG, J. Urban traffic prediction from spatio-temporal data using deep meta learning. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (2019), pp. 1720–1730.

[99] PECHLIVANOGLOU, T., ALIX, G., YANIN, N., LI, J., HEIDARI, F., AND PAPAGELIS, M. Microscopic modeling of spatiotemporal epidemic dynamics. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Spatial Computing for Epidemiology* (2022), pp. 11–21.

[100] PEJÓ, B., AND BICZÓK, G. Corona games: Masks, social distancing and mechanism design. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Modeling and Understanding the Spread of COVID-19* (2020), pp. 24–31.

[101] PESAVENTO, J., CHEN, A., YU, R., KIM, J.-S., KAVAK, H., ANDERSON, T., AND ZÜFLE, A. Data-driven mobility models for covid-19 simulation. In *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Advances in Resilient and Intelligent Cities, Seattle, WA, USA* (2020).

[102] PEW RESEARCH CENTER. Most Americans Say Coronavirus Outbreak Has Impacted Their Lives, March 2020 (https://www.pewsocialtrends.org/2020/03/30/most-americans-say-coronavirus-outbreak-has-impacted-their-lives/).

[103] QAZI, U., IMRAN, M., AND OFLI, F. Geocov19: a dataset of hundreds of millions of multilingual covid-19 tweets with location information. *SIGSPATIAL Special 12*, 1 (2020), 6–15.

[104] RAMCHANDANI, P., PAICH, M., AND RAO, A. Incorporating learning into decision making in agent based models. In *EPIA Conference on Artificial Intelligence* (2017), Springer, pp. 789–800.

[105] RAY, E. L., WATTANACHIT, N., NIEMI, J., KANJI, A. H., HOUSE, K., CRAMER, E. Y., BRACHER, J., ZHENG, A., YAMANA, T. K., XIONG, X., ET AL. Ensemble forecasts of coronavirus disease 2019 (covid-19) in the us. *medRxiv* (2020).

[106] REICH, N. G., BROOKS, L. C., FOX, S. J., KANDULA, S., McGOWAN, C. J., MOORE, E., OSTHUS, D., RAY, E. L., TUSHAR, A., YAMANA, T. K., ET AL. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the united states. *Proceedings of the National Academy of Sciences 116*, 8 (2019), 3146–3154.

[107] RIZZO, C., FABIANI, M., AMLÔT, R., HALL, I., FINNIE, T., RUBIN, G. J., CUCUIU, R., PISTOL, A., POPOVICI, F., POPESCU, R., ET AL. Survey on the likely behavioural changes of the general public in four european countries during the 2009/2010 pandemic. In *Modeling the Interplay Between Human Behavior and the Spread of Infectious Diseases*. Springer, 2013, pp. 23–41.

[108] RIZZO, S. G. Balancing precision and recall for cost-effective epidemic containment. In *First International KDD Workshop for Prescriptive Analytics for the Physical World (PAPW 2020) (https://prescriptive-analytics.github.io/file/3-strizzo.pdf)* (2020).

[109] RODRÍGUEZ, A., KAMARTHI, H., AGARWAL, P., HO, J., PATEL, M., SAPRE, S., AND PRAKASH, B. A. Data-centric epidemic forecasting: A survey. *arXiv preprint arXiv:2207.09370* (2022).

[110] RUSSELL, S., AND NORVIG, P. Ai a modern approach. *Learning 2*, 3 (2005), 4.

[111] SAJJADI, S., HASHEMI, A., AND GHANBARNEJAD, F. Social distancing in pedestrian dynamics and its effect on disease spreading. *Physical Review E 104*, 1 (2021), 014313.

[112] SAMET, H., HAN, Y., KASTNER, J., AND WEI, H. Using animation to visualize spatio-temporal varying covid-19 data. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Modeling and Understanding the Spread of COVID-19* (2020), pp. 53–62.

[113] SCHNEIDER, C. M., BELIK, V., COURONNÉ, T., SMOREDA, Z., AND GONZÁLEZ, M. C. Unravelling daily human mobility motifs. *Journal of The Royal Society Interface 10*, 84 (2013), 20130246.

[114] SERT, E., BAR-YAM, Y., AND MORALES, A. J. Segregation dynamics with reinforcement learning and agent based modeling. *Scientific reports 10*, 1 (2020), 1–12.

[115] SIETTOS, C., ANASTASSOPOULOU, C., RUSSO, L., GRIGORAS, C., AND MYLONAKIS, E. Modeling the 2014 ebola virus epidemic–agent-based simulations, temporal analysis and future predictions for liberia and sierra leone. *PLoS currents 7* (2015).

[116] Silva, P. C., Batista, P. V., Lima, H. S., Alves, M. A., Guimarães, F. G., and Silva, R. C. Covid-abs: An agent-based model of covid-19 epidemic to simulate health and economic effects of social distancing interventions. *Chaos, Solitons & Fractals 139* (2020), 110088.

[117] Susarla, A., Liu, A., Thai, D. H., Le, M. T., and Züfle, A. Spatiotemporal disease case prediction using contrastive predictive coding. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Spatial Computing for Epidemiology* (2022), pp. 26–34.

[118] Sydora, C., Nawaz, F., Bindra, L., and Stroulia, E. Building occupancy simulation and analysis under virus scenarios. *ACM Transactions on Spatial Algorithms and Systems (TSAS) 8*, 3 (2022), 1–20.

[119] Tanaka, M., and Tanimoto, J. Is subsidizing vaccination with hub agent priority policy really meaningful to suppress disease spreading? *Journal of Theoretical Biology 486* (2020), 110059.

[120] Taskesen, E. distfit. urlhttps://github.com/erdogant/distfit, 2019.

[121] Thakur, G., Sparks, K., Berres, A., Tansakul, V., Chinthavali, S., Whitehead, M., Schmidt, E., Xu, H., Fan, J., Spears, D., et al. Covid-19 joint pandemic modeling and analysis platform. In *Proceedings of the 1st ACM SIGSPA-TIAL International Workshop on Modeling and Understanding the Spread of COVID-19* (2020), pp. 43–52.

[122] Vallejo, M., Corne, D. W., and Rieser, V. Evolving urbanisation policies-using a statistical model to accelerate optimisation over agent-based simulations. In *ICAART (2)* (2013), Citeseer, pp. 171–181.

[123] Van Bavel, J. J., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M. J., Crum, A. J., Douglas, K. M., Druckman, J. N., et al. Using social and behavioural science to support covid-19 pandemic response. *Nature Human Behaviour* (2020), 1–12.

[124] Venkatramanan, S., Lewis, B., Chen, J., Higdon, D., Vullikanti, A., and Marathe, M. Using data-driven agent-based models for forecasting emerging infectious diseases. *Epidemics 22* (2018), 43–49.

[125] VOLKOVA, S., AYTON, E., PORTERFIELD, K., AND CORLEY, C. D. Forecasting influenza-like illness dynamics for military populations using neural networks and social media. *PloS one 12*, 12 (2017), e0188941.

[126] WANG, X., LIONO, J., MCINTOSH, W., AND SALIM, F. D. Predicting the city foot traffic with pedestrian sensor data. In *Proceedings of the 14th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services* (2017), pp. 1–10.

[127] WANG, Z., AND CRUZ, I. F. Analysis of the impact of covid-19 on education based on geotagged twitter. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Modeling and Understanding the Spread of COVID-19* (2020), pp. 15–23.

[128] WEILL, J. A., STIGLER, M., DESCHENES, O., AND SPRINGBORN, M. R. Social distancing responses to covid-19 emergency declarations strongly differentiated by income. *Proceedings of the National Academy of Sciences 117*, 33 (2020), 19658–19660.

[129] WILLIAMS, A. D., HALL, I. M., RUBIN, G. J., AMLÔT, R., AND LEACH, S. An individual-based simulation of pneumonic plague transmission following an outbreak and the significance of intervention compliance. *Epidemics 3*, 2 (2011), 95–102.

[130] WU, Y., YANG, Y., NISHIURA, H., AND SAITOH, M. Deep learning for epidemiological predictions. In *The 41st international ACM SIGIR conference on research & development in information retrieval* (2018), pp. 1085–1088.

[131] XIONG, L., SHAHABI, C., DA, Y., AHUJA, R., HERTZBERG, V., WALLER, L., JIANG, X., AND FRANKLIN, A. React: Real-time contact tracing and risk monitoring using privacy-enhanced mobile tracking. *SIGSPATIAL Special 12*, 2 (2020), 3–14.

[132] YE, W., AND GAO, S. Understanding the spatiotemporal heterogeneities in the associations between covid-19 infections and both human mobility and close contacts in the united states. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Spatial Computing for Epidemiology* (2022), pp. 1–9.

[133] ZAKARIA, C., TRIVEDI, A., CECCHET, E., CHEE, M., SHENOY, P., AND BALAN, R. Analyzing the impact of covid-19 control policies on campus occupancy and mobility via wifi sensing. *ACM Transactions on Spatial Algorithms and Systems (TSAS) 8*, 3 (2022), 1–26.

[134] ZHAO, L., SONG, Y., ZHANG, C., LIU, Y., WANG, P., LIN, T.,
DENG, M., AND LI, H. T-gcn: A temporal graph convolutional
network for traffic prediction. *IEEE Transactions on Intelligent
Transportation Systems 21*, 9 (2019), 3848–3858.

[135] ZÜFLE, A., AND ANDERSON, T. Introduction to this special
issue: Modeling and understanding the spread of covid-19: (part
i). *SIGSPATIAL Special 12*, 1 (2020), 1–2.

[136] ZÜFLE, A., ANDERSON, T., AND GAO, S. Introduction to the
special issue on understanding the spread of covid-19, part 1,
2022.

[137] ZÜFLE, A., GAO, S., AND ANDERSON, T. Introduction to the
special issue on understanding the spread of covid-19, part 2,
2022.