Epidemic Spreading in Trajectory Networks

Tilemachos Pechlivanoglou *, Jing Li, Jialin Sun, Farzaneh Heidari, Manos Papagelis

Lassonde School of Engineering, York University, Toronto, Canada

A B S T R A C T

Epidemics of infectious diseases, such as the one caused by the rapid spread of the coronavirus disease 2019 (COVID-19), have tested the world's more advanced health systems and have caused an enormous societal and economic damage. The mechanism of contagion is well understood. As people move around, over time, they regularly engage in social interactions. The *spatiotemporal network* representing these interactions constitutes the backbone on which an epidemic spreads, causing outbreaks. At the same time, advanced technological responses have claimed some success in controlling the epidemic based on *digital contact tracing technologies*. Motivated by these observations, we design, develop and evaluate *a stochastic agent-based $\mathcal{SEIR}$ model* of epidemic spreading in spatiotemporal networks informed by mobility data of individuals (trajectories). The model focuses on *individual variation* in mobility patterns that affects the degree of exposure to the disease. Understanding the role that individual nodes play in the process of disease spreading through network effects is fundamental as it allows to (*i*) assess *the risk of infection* of individuals, (*ii*) assess *the size of a disease outbreak* due to specific individuals, and (*iii*) assess *targeted intervention strategies* that aim to control the epidemic spreading. We perform a comprehensive analysis of the model employing COVID-19 as a use case. The results indicate that simple individual-based intervention strategies that exhibit significant network effects can effectively control the spread of an epidemic. We have also demonstrated that targeted interventions can outperform generic intervention strategies. Overall, our work provides *an evidence-based data-driven model* to support decision making and inform public policy regarding intervention strategies for containing or mitigating the epidemic spread.

© 2021 Elsevier Inc. All rights reserved.

1. Introduction

From the Plague of Athens (430 to 426 BC) [1,2] to the Spanish Flu (1918) [3,4], pandemics have had a significant impact on human society [5]. In the last 20 years alone, the world has seen many infectious disease outbreaks. Notorious examples include the pandemics caused by the Severe Acute Respiratory Syndrome coronavirus (SARS-CoV) [6], the influenza A virus subtype H1N1 (swine flu) [7], the Middle East Respiratory Syndrome coronavirus (MERS-CoV) [8], the Ebola virus (EVD) [9], the Zika virus (ZIKV) [10], and most recently the Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2) [11]. These pandemics have tested the world's more advanced health systems and have caused an enormous societal and economic damage. Conventional methods to address the rapid spread of an infectious disease include *physical distancing*, *confinement measures* and *human-based contact tracing* of infected individuals. These describe some of the common policies imposed by governing authorities and jurisdictions that aim to contain or slow down the spread of the virus to levels that can be managed by healthcare units and socio-political institutions. While these easily understood policies can be effective in controlling the spread of the disease and saving lives [12,13], they have **well-known drawbacks**: (**i**) they are imposing extreme restrictions or limitations on individuals' activities or freedom, leading to a slowdown of social and economic activities of the community and to socioeconomic side-effects for the individuals themselves; (**ii**) they depend on human-based contact tracing of infected individuals that are cumbersome, expensive, slow and inaccurate; and (**iii**) they do not provide the means of a controlled transition to an immune community through well-defined intervention strategies that can easily translate to health policy and potentially ameliorate the socioeconomic impact.

More recently, advanced technological responses to the problem based on **digital contact tracing** have claimed some success in controlling the epidemic [14]. *Digital contact tracing* or *proximity tracing*, enabled by GPS-enabled devices, mobile apps [15] and beyond [16], represents the ability to track and reconstruct the close contacts that an individual had with other people within a time period. The way that proximity tracing can have an impact in

---

containing or slowing down the disease spread is straightforward. Individuals that are known to be infected, can inform (via cloud services) recent close contacts that have been digitally traced, who can then take precautions and avoid further contacts with other people by isolating themselves or seeking expert advice. The process can involve third parties, such as governing authorities and/or health experts responsible for the containment of the disease.

The **focus of the current research** is on utilization of GPS-enabled digital contact traces of individuals (i.e., mobility data or trajectories) to inform a more comprehensive analysis and modeling of disease spreading through methods of graph mining [17] and trajectory data mining [18]. In particular, we present a data-driven model for the spread of the disease in a community that take into account the **mobility patterns of individuals**. As people move in cities, they engage in different types of interaction with other people, resulting in different mobility patterns. As such, the relative risk of them being infected or infecting others can be substantially different. We systematically study the effect of the **individual variability** of mobility behavior to the risk of infection of an individual. This observation can have significant consequences to a model's accuracy of how the disease propagates in a community, as well as to the intervention strategies that can be designed to control the epidemic.

Contributions. Motivated by the feasibility of digital contact tracing technologies [19] and the inherent limitations of traditional epidemiological models (see Section 2), this paper presents data-driven models of infectious disease spreading that incorporate individual variability due to individuals' mobility patterns. Our study aims to clarify how differences in mobility patterns can inform infectious epidemic dynamics and determine the impact of various intervention strategies. In summary, the major contributions of this work are as follows:

- we present novel data-driven models for assessment of the risk of infection of an individual based on mobility patterns and the amount of time they spend in proximity with others ("**individual risk assessment**");
- we present a stochastic agent-based Susceptible-Exposed-Infected-Removed (SEIR) network model for infectious disease spreading in trajectory networks ("**community risk assessment**");
- we design and evaluate novel individual-based intervention strategies for containing (or mitigating to an acceptance rate) the spread of an infectious disease in trajectory networks ("**containment intervention strategies**");
- we design and evaluate novel individual-based immunization strategies for providing a controlled and safe transition to an immune community ("**targeted immunization**");
- we present a large-scale **case study** 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 allows for many human-human interactions; the model and algorithms presented generalize to other similar infectious diseases;
- we provide **source code** and **data** to encourage reproducibility of results.

The remainder of this paper is organized as follows: Section 2 provides background information, introduces notation and provides definitions of the technical problems of interest in this paper. Section 3 presents our epidemic model, algorithmic details of epidemic spreading in trajectory networks and descriptions of disease containment intervention strategies. Section 4 presents an experimental evaluation of the different models and methods for varying settings. We review the related work in Section 5 and conclude in Section 6.

2. Background and problem definitions

In this section, we introduce notation and preliminaries of the problem of interest, as well as formal problem definitions. The background mostly relates to the definition of a *contact network* or a *trajectory network* as defined in [20]. We also provide background information related to the basic reproductive number R_0 and its limitations, as well as information about the SEIR epidemic model and its variations, as we employ it in our study.

2.1. Observation area: Earth surface versus Euclidean space

We assume monitoring of mobility data of individuals within a finite observation area $\mathcal{A}$. For the needs of our study, this area typically represents the administrative boundaries of a city or a city neighborhood where daily human contacts occur. Since $\mathcal{A}$ is a relatively small region, the Earth surface it represents has a low curvature and is close to flat. We can therefore, for simplicity and without loss of generality, assume that individuals move in a finite 2-*dimensional* Euclidean space $\mathbb{R}^2$ and not on the surface of the Earth. This assumption allows to approximate geodesic distances on Earth with Euclidean distances in $\mathbb{R}^2$, a common practice in many real-world algorithms and services.

2.2. Contacts and events

Consider a set of objects $\mathcal{N} = \{u_1, u_2, \ldots, u_N\}$ moving in an observation area $\mathcal{A}$, defined as a finite 2-*dimensional* Euclidean space $\mathbb{R}^2$ for a finite *observation time interval* $[0, T]$, forming a set of trajectories $\mathcal{P}$. We formally define a *trajectory* as follows.

Definition 2.1 *(Trajectory)*. A trajectory $\mathcal{P}_u$ of a moving object $u \in \mathcal{N}$ 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 $(x, y) \in \mathcal{A} \subseteq \mathbb{R}^2$ represent latitude and longitude coordinates in the 2D Cartesian system. We assume that an object might appear and disappear multiple times during the observation time interval $[0, T]$.

As individuals move in $\mathcal{A}$, they can encounter each other, forming *contacts*. Following Pechlivanoglou and Papagelis [20], we define a *contact* as follows.

Definition 2.2 *(Contact)*. A *contact* $c_{u,v}$ between two moving individuals $u, v \in \mathcal{N}$ occurs when their physical proximity (spatial distance) $d_{u,v}$ is smaller than or equal to a threshold τ (i.e. $d_{u,v} \leq \tau$).

Several approaches exist to estimate the spatial distance of two points in Euclidean plane. We employ its simplest form, the Euclidean distance, given by:

$$d_{u,v} = \sqrt{(x_u - x_v)^2 + (y_u - y_v)^2}$$

where (x_u, y_u) and (x_v, y_v) are the spatial coordinates of individuals u and v at a time t, where $0 \leq t \leq T$ respectively. The two individuals u and v are considered to be in contact for as long as their spatial distance remains consistently smaller than a proximity threshold τ. We extend the concept of a *contact* to include its temporal dimension and formally define an *event* as follows.

Definition 2.3 *(Event)*. An event $e_{u,v}$ between two moving objects $u, v \in \mathcal{N}$, represents a contact $c_{u,v}$ that lasted for a time interval $[t_s, t_e]$, where t_s represents the time point of the beginning of the contact and t_e represents the time point the contact ended. An event is represented by the quadruple $e_{u,v} = (u, v, t_s, t_e)$. We also define the duration of the event as $\delta(e_{u,v}) = t_e - t_s$.

Note that, in our setting, we do not preclude the case that two individuals are in contact multiple times. In this case, the contact information between two moving individuals u and v is represented by a sequence of events $E_{u,v} = \{e_{u,v}^1, e_{u,v}^2, \ldots, e_{u,v}^n\}$ or $E_{u,v} = \{(u, v, t_s^1, t_e^1), (u, v, t_s^2, t_e^2), \ldots, (u, v, t_s^n, t_e^n)\}$. We also define the duration of all events as $\Delta(E_{u,v}) = \sum_i^n \delta_{e_{u,v}^i} = \sum_i^n (t_e^i - t_s^i)$. Furthermore, in this paper we employ a universal proximity threshold τ, so the contacts will always be *reciprocal*, meaning that (u, v, t_s, t_e) is equivalent to (v, u, t_s, t_e). Indeed, this is sufficient for the case of human-to-human interactions we examine in this work.

2.3. Trajectory networks and node importance

A network that is constructed by connecting pairs of individuals that are close to each other based on physical proximity is called a *proximity network*. However, a proximity network is static and does not capture well the idea of individuals moving in space. When individuals are moving, the temporal dimension of interactions must be considered, and the resulting network can be thought of as a *temporal network*, also referred to as a *time-varying network*. Most characterizations of temporal networks discretize time by grouping together temporal information into a sequence of T network "snapshots" $G_t(V_t, E_t), t \in \{1, 2, \ldots, T\}$. Each snapshot contains the vertices V_t and edges E_t, representing the individuals and their contacts, respectively, within a basic time unit t (e.g., second, minute, hour, etc.). The resulting data structure can be thought of as either a single aggregation graph with varying vertices and edges, or a sequence of proximity graphs. In either case, we refer to $G_t(V_t, E_t)$ as a *trajectory network* $G(V, E)$ for the rest of the manuscript.

There are many possible metrics to determine the *importance* (or influence) of an individual (or node) in a temporal network. Note that the term *node centrality* refers to node importance that is common in static network analysis, and isn't applicable for trajectory networks. This is because measures of node centrality in the traditional setting of a static network are commonly based on *shortest paths* (e.g., betweenness centrality [21,22]), but shortest paths in temporal networks take a different character [23]. For example, in [24], the authors define *minimum temporal paths* to capture the different characterizations of time-constraint shortest paths including cases of earliest-arrival paths, latest-departure paths, or fastest paths. It is possible to evaluate a notion of temporal betweenness [25], but in our setting, we focus on notions of importance that are critical in the context of epidemic modeling in the trajectory network. Similarly to Pechlivanoglou and Papagelis [20], we define metrics that relate to *the temporal node degree* and the *duration* of events, and use these metrics to construct *node profiles* that describe the behavior of each individual.

Definition 2.4 (*Trajectory network node degree*). We define the following metrics related to node degree in the trajectory network:

- C_u: a set of all contacts of u during the observation time interval $[0, T]$.
- $D_{deg_u}(k)$: a distribution that represents the fraction of the time steps $t_i \in [0, T]$ that u has node degree k.

2.4. Problem definitions

In this paper, we are interested in the assessment and mitigation of the risk of infectious disease spreading in trajectory networks based on mobility data. In particular, we aim to address the following problems:

Problem 1. Given an observation area $\mathcal{A}$, an observation time interval $[0, T]$, a set of individuals $\mathcal{N}$ and their trajectories $\mathcal{P}$, determine the risk of infection $risk_u$ of each individual in the trajectory network $G(V, E)$.

Problem 2. Given a trajectory network $G(V, E)$, a seed set of initially infected individuals $\mathcal{I}_0 \subseteq \mathcal{N}$ and the risk $risk_u$ of any individual $u \in \mathcal{I}_0$, determine the size of the epidemic spread $\mathcal{I}_T \subseteq \mathcal{N}$ at time $t = T$, where $\mathcal{I}_0 \subseteq \mathcal{I}_T$.

Problem 3. Given a trajectory network $G(V, E)$, and the parameters of an emerging infectious disease, determine the impact of various epidemic containment intervention strategies that can easily translate to health policy. The focus is on comparative analysis of the impact of targeted individual-based interventions against a null model (informed by less sophisticated horizontal measures).

2.5. Limitations of the basic reproductive number R_0

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. 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 cases 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 perfect mixing network (i.e., a complete graph) or a *regular tree network* – a special type of a network topology that has no cycles and each internal node has a constant number of children, defined by a branching factor d. However real-world communities do not resemble a complete graph or regular trees, since some people have more contacts than others and it is common for people to have common friends (forming triangles or cycles). 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 [26].

For the purposes of this work, when we refer to R_0 for individuals, we define it as "the expected number of secondary cases produced, in a completely susceptible population, produced by a typical infected individual" [27].

2.6. The $\mathcal{SEIR}$ epidemic model

Compartmental models of epidemic modeling divide the population into separate divisions (compartments) and people transition between them based on their health status during an epidemic. For instance, in the classic $\mathcal{SIR}$ model [28,29], people progress between three compartments: *susceptible* ($\mathcal{S}$), *infectious* ($\mathcal{I}$) and *removed/recovered* ($\mathcal{R}$). For many infectious diseases, there is a significant latent period (incubation) during which susceptible individuals have been infected, but are not yet infectious themselves. During this period an individual is considered to be in a compartment labeled as *exposed* ($\mathcal{E}$), and the model is known as $\mathcal{SEIR}$. The current research employs the $\mathcal{SEIR}$ model for modeling the spread of a virus in the community. Depending on assumptions of population structure and transmission progression, there are two main classes of the $\mathcal{SEIR}$ model studied in the literature.

Homogeneous population. The first class, assumes a large, homogeneously mixing population where individuals move between compartments at certain transition rates described by ordinary differential equations [30,31]:

$$\frac{d\mathcal{S}}{dt} = -\frac{\beta \mathcal{I}\mathcal{S}}{N} \qquad \frac{d\mathcal{E}}{dt} = \frac{\beta \mathcal{I}\mathcal{S}}{N} - \sigma \mathcal{E}$$

$$\frac{d\mathcal{I}}{dt} = \sigma \mathcal{E} - \gamma \mathcal{I} \qquad \frac{d\mathcal{R}}{dt} = \gamma \mathcal{I}$$

where β is the transmission rate, σ is the incubation rate, and γ is the recovery rate, respectively. This is a *deterministic model*, so for a fixed set of parameter values and $\mathcal{S}\mathcal{E}\mathcal{I}\mathcal{R}$ model initialization ($t = 0$), it produces the same outcome at each simulation. This model can inform about the state of the epidemic spread in the community and provide insights about future trends as well as inform health policy at large [32]. However, there are certain **limitations** of this model. Its results and usefulness are limited by the inherent assumption that all individuals share the same characteristics.

Heterogeneous population. The second class, assumes heterogeneity of population and is based on an agent-based $\mathcal{S}\mathcal{E}\mathcal{I}\mathcal{R}$ model, where each agent is representing an individual [33,34]. This approach allows to model individual characteristics and behavior towards the epidemic. In our research we focus on heterogeneity that is attributed to different mobility and contact patterns of individuals, over time. Different mobility patterns, lead to complex spatio-temporal social interactions between people in the community [20,35]. These models are more challenging to analyze and interpret as they depend on *a stochastic (probabilistic) process* of epidemic spreading that increases the complexity [36]. However, they are more realistic and can help to better understand the emergence of a disease due to different individual behaviors. In addition, since they operate (simulate) on individual-level behavior, they provide an opportunity to design targeted intervention strategies that can more easily translate to health policy. In section 3 we present the details of the agent-based $\mathcal{S}\mathcal{E}\mathcal{I}\mathcal{R}$ model.

3. Methodology

Recent studies on epidemic modeling highlight the importance of individual variability in modeling the spread of an infectious disease in a community and predicting its relevant outcome. For example, Gomes et al. [37] studied the effect of the biological variation in susceptibility of individuals and their physical exposure to infection. And, Britton et al. [38] studied how population heterogeneity affects herd immunity. In our research, we study the effect of individual variability in epidemic modeling that is due to mobility patterns. We first present a method for computing the risk of infection of any individual in the community, as a result of their spatiotemporal interactions. Then, we present a stochastic agent-based epidemic model (and optimizations) that can better capture the dynamic disease spreading in a community.

3.1. Modeling individual risk of infection

We integrate individual variation by modeling the risk of infection of an individual in relation to its mobility patterns and contacts over a time period. Intuitively, we would like to model that the more contacts an individual has and the more time they spent with each other, the higher the risk of infection. Formally, given a trajectory network $G(V, E)$, an individual $u \in \mathcal{N}$ and its contacts C_u during $[0, T]$, we model the risk of infection $risk_u$ of an individual by the following three methods, each offering a different level of analysis.

$$risk_u^{(1)} = \sum_{i=1}^{|C_u|} 1 = |C_u| \tag{1}$$

$$risk_u^{(2)} = \sum_{i=1}^{|C_u|} \Delta(E_{u,i}) \tag{2}$$

$$risk_u^{(3)} = \sum_{i=1}^{|C_u|} (1 - (1 - \beta)^{\Delta(E_{u,i})}) \tag{3}$$

Out of the three definitions, $risk_u^{(1)}$ is the simplest one as it is based on the node degree in the aggregation network (i.e., the network defined by aggregating the edges of a temporal network over $[0, T]$); $risk_u^{(2)}$ takes into consideration both the number of contacts of u and the total duration of these contacts (due to potentially multiple events); $risk_u^{(3)}$ models the risk of infection as a probability of getting infected by any of its contacts factoring the total duration of these contacts (due to potentially multiple events), where β is the transmission probability of the disease. In particular, we use a geometric function to represent the risk attributed to each distinct contact. The outcome is a regularized metric for risk (capped at 1), so that specific contacts with a very long duration do not dominate the overall risk of an individual.

Relative risk of infection: While the actual value of an individual's risk of infection does not hold any natural interpretation, it is important for our analysis to represent the **relative risk** $rrisk_u$ of u to other individuals in the network. We therefore normalize each risk metric by the aggregated risk of all N individuals in the network to get the relative risk of $u \in \mathcal{N}$, as follows:

$$rrisk_u^{(i)} = \frac{risk_u^{(i)}}{\sum_{u=1}^{N} risk_u^{(i)}} \tag{4}$$

Note that it is $rrisk_u^{(1)} \in [0, 1]$, $rrisk_u^{(2)} \in [0, 1]$, $rrisk_u^{(3)} \in [0, 1]$ and that $\sum_{u=1}^{N} rrisk_u^{(i)} = 1$. We utilize the relative risk in our experimental analysis.

3.2. Epidemic spreading in trajectory networks

We present **a stochastic agent-based $\mathcal{S}\mathcal{E}\mathcal{I}\mathcal{R}$ network model** for epidemic spreading in a trajectory network, where nodes represent individuals and edges represent contacts of nodes. According to the epidemic model, each node can be at one of the following infection states, at any discrete time t:

- *Susceptible ($\mathcal{S}$)*. This is the initial state of all nodes; a node can get exposed to the infection by any of its infected neighbors with probability β, per time step.
- *Exposed ($\mathcal{E}$)*. A node is in this state if it has been infected by one of its neighbors, but it is not yet infectious itself. A node stays in this state for as long as the incubation period of the disease lasts, which for simplicity we model as a constant that lasts $\mathcal{I}_f$ time steps. After that period, the node becomes infectious and switches to state $\mathcal{I}$ with certainty. Depending, on the disease we aim to model, the certainty can be relaxed by incorporating a parameter to control the probability of a node switching to $\mathcal{I}$ (or to $\mathcal{S}$).
- *Infected ($\mathcal{I}$)*. A node is in this state if it is infectious, therefore can transmit the disease to any of its neighbors with probability β.
- *Removed ($\mathcal{R}$)*. A node is in this state if it has been removed, meaning either has passed away or has recovered. Nodes that are in $\mathcal{I}$ will be removed after $\mathcal{I}_r$ time steps with a recovery probability γ. The recovered nodes are neither infectious anymore nor susceptible to the infection.

Given a trajectory network $G(V, E)$ and parameter values β, γ, $\mathcal{I}_f$ and $\mathcal{I}_r$ the model allows us to monitor the state of every individual over time. Given a population of $\mathcal{N}$ individuals, the cumulative number of individuals within each of the disease states at time t is given by $\mathcal{S}(t)$, $\mathcal{E}(t)$, $\mathcal{I}(t)$, and $\mathcal{R}(t)$, respectively. We also define two special sets of infected nodes: (i) the initial seed set of infected nodes $\mathcal{I}_0 = \mathcal{I}(0)$, and (ii) the set of the infected nodes at the end of the process $\mathcal{I}_T = \mathcal{I}(T)$, which represents the size of the epidemic spread.

3.2.1. Algorithmic details of the stochastic model

We describe here algorithmic details of the stochastic model. Recall that individuals move between infection states $\mathcal{S}$, $\mathcal{E}$, $\mathcal{I}$ and $\mathcal{R}$ based on a stochastic process. At each discrete time step t, each $\mathcal{S}$(usceptible) node has a chance to switch to $\mathcal{E}$(xposed), $\mathcal{E}$(xposed) nodes might switch to $\mathcal{I}$(nfected), and $\mathcal{I}$(nfected) nodes might be $\mathcal{R}$(emoved). Formally, let $u \in \mathcal{S}$ and let $\mathcal{N}_u$ be the set of neighbors of u at time t. Each neighbor $v \in \mathcal{N}_u$ such that $v \in \mathcal{I}$, flips a biased coin with a bias equal to the transmission probability β to determine whether it will infect u. If u is infected, then it switches its infection state to $\mathcal{E}$(xposed), otherwise its infection state remains $\mathcal{S}$(usceptible). Similarly, an $\mathcal{E}$(xposed) node will switch to $\mathcal{I}$(infected) after $\mathcal{I}_f$ steps and an $\mathcal{I}$(nfected) node will switch to $\mathcal{R}$(ecovered) after $\mathcal{I}_r$ steps, with a probability γ. The pseudocode of the stochastic model of epidemic spreading is given in Algorithm 1. In our analysis, each time step in the discrete time simulation corresponds to a minute (60 secs), so negligible contacts (interactions of less than a minute) are not considered.

Studies on infectious diseases have showed that prolonged exposure of a susceptible node to an infected node increases the likelihood of infection [39]. It is easy to see that in the epidemic model presented in Algorithm 1, an infected node u has multiple chances to propagate the disease. Formally, given a trajectory network $G(N, V)$, the probability $p_{u,v}$ of a susceptible node $u \in V$ being infected by a neighboring infected node $v \in V$ after k independent trials is given by the cumulative distribution function of the geometric distribution:

$$p_{u,v} = 1 - (1 - \beta)^k \tag{5}$$

where β is the transmission probability of the disease. Eq. (5) represents the complementary probability of u not being infected after k independent trials. It is easy to see that k depends on the duration of the contact between an infected node v and u (i.e., one chance per time unit) and that $0 \leq p_{u,v} \leq 1$.

3.3. Conversion of the stochastic model to a deterministic model

The epidemic spreading model we described in Algorithm 1, is a stochastic process that possesses some inherent randomness. Starting with the same initial conditions (i.e., the same sets of Susceptible and Infected nodes) and parameter values, multiple independent simulations of the epidemic spreading process can produce outputs that vary a lot, in terms of the total number of nodes infected at the end of the process. This is because the final outcome depends on flipping a biased coin at every time step to decide whether the disease will diffuse from one node to another in the network.

Interestingly, there is an equivalent deterministic model that offers a static view of the network and is more practical, as it allows for faster simulations than the stochastic model. We describe here a method that given a stochastic model of epidemic spreading in the trajectory network, converts it to a deterministic model based on *percolation theory* [40,41]. In mathematics and physics, percolation theory is used to explain the flow of fluids through certain types of porous material. Similarly, in network science, it is used

---

Algorithm 1: Epidemic spreading in trajectory networks.

Input: $G(V, E)$, $\mathcal{S}$, $\mathcal{I}$, β, γ, $\mathcal{I}_f$, $\mathcal{I}_r$, q
Output: $\mathcal{SEIR}_{[0,T]}$ = {$[\mathcal{S}(0), \mathcal{E}(0), \mathcal{I}(0), \mathcal{R}(0)]$, ... , $[\mathcal{S}(T), \mathcal{E}(T), \mathcal{I}(T), \mathcal{R}(T)]$}

$\mathcal{S}(0) \leftarrow \mathcal{S}$; $\mathcal{E}(0) \leftarrow 0$; $\mathcal{I}(0) \leftarrow \mathcal{I}$; $\mathcal{R}(0) \leftarrow 0$;

for $t \in [0, T]$ **do**
  for $u \in V$ **do**
    switch $u.state$ **do**
      case $u \in \mathcal{S}(t)$ **do**
        $N_u \leftarrow u.neighbors()$;
        for $v \in \mathcal{N}_u$ **do**
          if $v \in \mathcal{I}(t)$ *and* $v.Q \neq 1$ **then**
            v infects u with probability β;
            if u *is infected* **then**
              $u.state \leftarrow \mathcal{E}$;
              $u.Q \leftarrow 1$ with probability q;
              $\mathcal{I}_f^u = 0$;   /* incubation period of u
               begins */
              $\mathcal{S}(t+1) \leftarrow \mathcal{S}(t) \setminus u$;
              $\mathcal{E}(t+1) \leftarrow \mathcal{S}(t) \cup u$;
        break;
      case $u \in \mathcal{E}(t)$ **do**
        $\mathcal{I}_f^u$ ++;
        if $\mathcal{I}_f^u = \mathcal{I}_f$ **then**
          $u.state \leftarrow \mathcal{I}$;
          $\mathcal{I}_r^u = 0$; /* recovery period of u begins */
          $\mathcal{E}(t+1) \leftarrow \mathcal{E}(t) \setminus u$;
          $\mathcal{I}(t+1) \leftarrow \mathcal{I}(t) \cup u$;
        break;
      case $u \in \mathcal{I}(t)$ **do**
        $\mathcal{I}_r^u$ ++;
        if $\mathcal{I}_r^u = \mathcal{I}_r$ **then**
          $u.state \leftarrow \mathcal{R}$;
          $\mathcal{I}(t+1) \leftarrow \mathcal{I}(t) \setminus u$;
          u recovers with probability γ;
          if u *recovers* **then**
            $\mathcal{R}(t+1) \leftarrow \mathcal{R}(t) \cup u$;
        break;
      case $u \in \mathcal{R}(t)$ **do**
        break;   /* do nothing */
  $\mathcal{SEIR}_{[0,T]} \leftarrow \mathcal{SEIR}_{[0,T]}.append([\mathcal{S}(t), \mathcal{E}(t), \mathcal{I}(t), \mathcal{R}(t)])$;

  return $\mathcal{SEIR}_{[0,T]}$;

---

to describe the behavior of a network when nodes or links are removed.

To utilize this idea in the epidemic spreading model, recall that each infected node in the network has a probability β to infect each of its neighboring nodes at every time step t, by flipping a biased coin with a probability β. At the end of the interaction, the infected node has either infected the neighboring node, in which case we consider the edge to be "active", or not, in which case we consider the edge to be "removed or blocked". The idea of percolation is that instead of deferring the decision of whether an edge will be "active" or "removed" at runtime, we can make a decision for each edge of the trajectory network $G(V, E)$ at the very beginning of the whole process. In practice, for each edge in the network, we just need to flip a biased coin with probability β as many times as the duration of the contact (expressed in time units), and decide whether to keep it or remove it from the network. At the end of the process a smaller network $G'(V, E')$ is constructed, such that $E' \subseteq E$.

In terms of the correctness of the epidemic spreading process itself, it does not matter if the decision to keep or remove an edge is made at runtime or early in the process. In terms of runtime cost, percolation allows to work on a smaller network (since many edges are already removed) and allows simulations to fin-
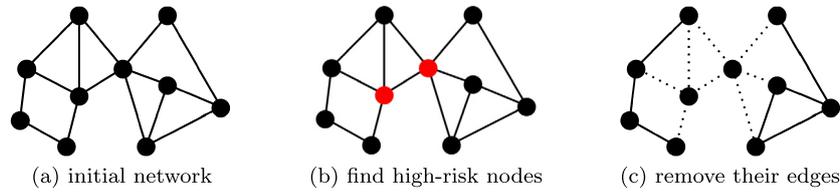
(a) initial network (b) find high-risk nodes (c) remove their edges

Fig. 1. Example of removing high-risk individuals.

(a) examine neighbors (b) remove edges to high-risk (c) repeat for entire network

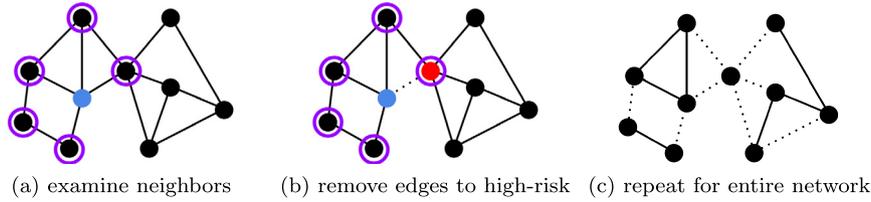**Fig. 2.** Example of removing high-risk contacts of individuals.

ish faster. We therefore employ percolation in the relevant set of experiments.

3.4. Containment intervention strategies

In this section, we explore various network-based intervention strategies that aim to contain an epidemic [42]. These interventions change the structure of the trajectory network — the backbone on which an epidemic spreads over time — and eventually affect the size of the set $\mathcal{I}_T$ of the infected nodes at the end of the process. The strategies relate to *node immunization* (network node removal) or *breaking of social ties* (network edge removal) and are actuated either at a *network-level* (governing authority decision) or a *node-level* (individual decision). Our goal is to design *targeted models* of intervention and evaluate them against sensible *null models*. Details of these strategies and models are presented below, along with discussion on their feasibility and their implications to health policy.

Strategy 1: node immunization. Based on this strategy, we remove a fraction α_n of all nodes in the network. Formally, given a set $S \subseteq V$ of nodes to be removed, where $|S| = \alpha_n |V|$, the infectious disease now spreads in the induced subgraph $G'(V', E')$ of G whose vertex set is $V' = V \setminus S$ and whose edge set E' consists of all of the edges in E that have both endpoints in V'. The real-world interpretation of this strategy is that some individuals are quarantined (i.e., they are in a state of isolation where no contacts occur) or develop immunization because of a vaccine. The network effect is that a contagious disease cannot spread through their contacts anymore.

Null model: A fraction α_n of nodes is removed uniformly at random.

Targeted model (removing high-risk individuals): Nodes are ranked based on their relative risk of infection $rrisk_u$, in a descending order. Then, a fraction α_n of the nodes with the largest risk $rrisk_u$ are removed. Ties are resolved uniformly at random. An illustration of this intervention can be seen on Fig. 1.

It is important to note that this is a network-level intervention strategy, where a national authority determines a set of individuals to immune (or request to quarantine) based on an estimate of their relative risk $rrisk_u$. Such an intervention, is resource-intensive, but also might infringe the privacy of individuals. It also carries a risk of discriminating against individuals with specific mobility patterns (i.e., super-spreaders). As a result, the feasibility of this intervention strategy is rather weak for large communities.

Strategy 2: breaking of social ties. Based on this strategy, we remove a fraction α_e of edges adjacent to each node (contacts).

Formally, given a node $u \in V$ and its set of neighbors $\Gamma(u)$, we remove $|H_u| = \alpha_e |\Gamma(u)|$ edges, where $H_u \subseteq \Gamma(u)$. The total number of edges removed from the network is $|H| = \sum_{i=1}^N |H_i|$, $H \subseteq E$. The infectious disease now spreads in the subgraph $G'(V, E')$ of G, where $E' = E \setminus H$. The real-world interpretation of this strategy is that individuals have some understanding of the mobility patterns of their contacts and they can make decisions about who to avoid. The network effect is that a contagious disease cannot spread through some specific contacts anymore.

Null model: For each node u, a fraction α_e of its contacts to neighboring nodes $\Gamma(u)$ are removed, uniformly at random.

Targeted model A (removing high-risk contacts): Nodes $v \in \Gamma(u)$ in the neighborhood of u are ranked based on their relative risk of infection $rrisk_v$, $v \in \Gamma(u)$, in a descending order. Then, a fraction α_e of contacts to the neighboring nodes with the largest risk $rrisk_v$ are removed from the network. Ties are resolved uniformly at random. An illustration of this intervention can be seen on Fig. 2.

It is important to note that this is an individual-level intervention strategy, where each individual makes a local decision about who to avoid, based on some understanding of the relative risk $rrisk_u$ associated with each of its contacts. The model assumes that individuals are in position to understand that they should avoid contacts that are frequently and regularly interacting with many others (e.g., due to their occupation or mobility habits). Such an intervention is easier and not resource-intensive to implement, due to its distributed nature and does not infringe on the privacy of individuals. As a result, the feasibility of this intervention strategy is rather high for large communities.

Targeted model B (removing non-community contacts): Nodes $v \in \Gamma(u)$ in the neighborhood of u are ranked based on the number of shared friends $s(u, v) = |\Gamma(u) \cap \Gamma(v)|$, in an ascending order. Then, a fraction α_e of edges to the neighbors with the smallest $s(u, v)$ is removed from the network. Ties are resolved uniformly at random. An illustration of this intervention can be seen on Fig. 3.

This model resembles a "social bubble" policy practiced by many, where an individual maintains contact with only family members and a few close friends. This way, potential "network bridges" between different well-knit communities in the network are eliminated and the infectious disease finds it hard to cross between them. This is an individual-level intervention strategy, where each individual makes a local decision about who to keep in its social bubble, based on some understanding of how many friends they have in common. This is a relatively easier assumption to make (than the one made by the *targeted model A*). Such an intervention
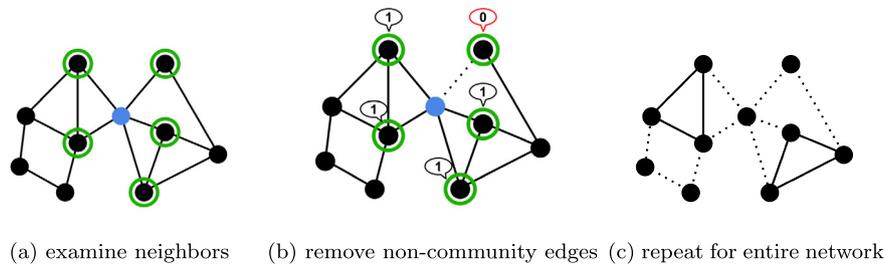
(a) examine neighbors (b) remove non-community edges (c) repeat for entire network

Fig. 3. Example of removing non-community contacts of individuals.

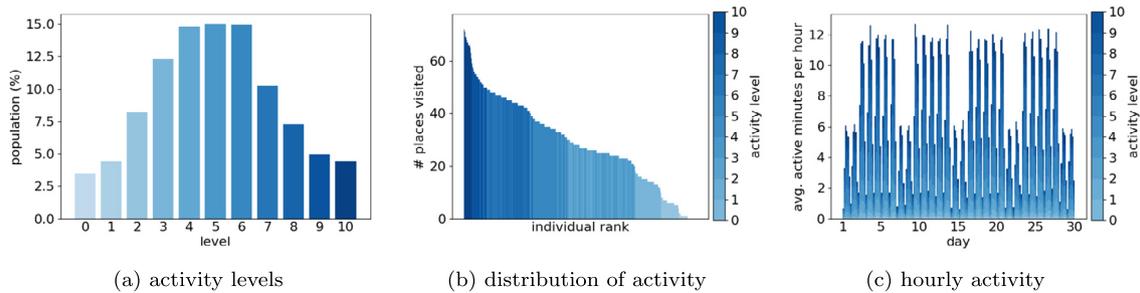(a) activity levels (b) distribution of activity (c) hourly activity

Fig. 4. Indicators and distribution of generated individuals' activity.

Table 1
Summary of trajectory data.

Parameter	Value
#INDIVIDUALS	2-10k
AREA (km^2)	1
OBSERVATION TIME (second)	2,592,000
#POINTS OF INTEREST	1,330
ACTIVITY DISTRIBUTION	$A_{level} \sim \mathcal{N}(\mu = 5.5, \sigma^2 = 2.5), A_{level} \in [0, 10]$

is once again easier and not resource-intensive to implement, due to its distributed nature and does not infringe on the privacy of individuals. As a result, the feasibility of this intervention strategy is rather high for large communities.

4. Experimental evaluation

In this section, we provide details of our experimental evaluation. We first present our synthetic data generator and describe the characteristics of the contact networks produced. Then, we present a COVID-19 use case, by specifying the parameters and refining the research questions we aim to explore. For each research question, we outline the experimental scenario and process followed to effectively address it. Finally, we discuss the results and any implications.

4.1. Data generation

In order to evaluate our stochastic agent-based $\mathcal{SEIR}$ epidemic model, we had to rely on large-size data representing trajectories of individuals or their spatiotemporal contacts. Moreover, for simulations to be reliable, the data needs to be (almost) *complete*; if significant amount of information about people's mobility or contacts is missing, then any underlying analysis related to community structure and individual behavior could be significantly affected. At the same time, mobility data is *highly sensitive*; many contact tracing applications rely on privacy-preserving proximity data, making the collection of real-world data impossible. With these factors in mind, we opted to use *synthetically generated data*. On the other hand, the benefit of generating synthetic data is that all parameters could be tuned and therefore analysis can be more comprehensive.

We generated synthetic data that simulates the activity of people living and working within a specified urban area over the course of *a month*. We defined an observation area $\mathcal{A}$ of approximately 1 km^2 including the York University Keele campus and surrounding neighborhoods in Toronto, Canada. Each individual in the simulation is randomly assigned a *home location* and frequents a number of *favorite places* (out of a predefined set of places), following a normal distribution. Moreover, each person is assigned an *activity level* parameter that determines how "active" they are by controlling the number of hours they may spend outside their home every day and the number of places they are likely to visit. Based on existing research on daily activity [43], each individual was assigned between 0 and 12 active daily hours, determined by their activity level. Table 1 presents the parameters of the data generator and Fig. 4 presents descriptive analytics of the generated individual mobility data, including the distribution of activity levels, the distribution of places visited and the hourly activity over the course of a month by individuals of different activity level.

We combine all previous parameters to generate a set of destinations and daily schedules for a specified number of individuals. Afterwards, the exact movement and trajectory traces of these people are simulated using Eclipse Simulation of Urban MObility (SUMO) [44], an open source, highly portable, microscopic and continuous multi-modal traffic simulation package. SUMO is capable of modeling accurate and highly realistic movement of vehicles but also pedestrians, including movement through pedestrian crossings and crowded sidewalks. The end result is synthetic but reliable, complete datasets representing the daily movement of individuals in the observation area $\mathcal{A}$, over a period of a month.

4.2. Use case: parameters for the COVID-19 epidemic

We used the synthetic data generator to model a population of 2,000, 3,000, 5,000 and 10,000 individuals moving in the same campus area. Of course, the resulting datasets correspond to different population densities. This allows us to examine the progression of an epidemic in urban areas with different population densities while controlling the rest of the parameters in the problem.

Table 2
Summary of datasets.

Name	2K	3K	5K	10K
#INDIVIDUALS	2,000	3,000	5,000	10,000
#DATA POINTS	172,800,000	259,200,000	432,000,000	864,000,000
#CONTACTS	230,170	1,493,344	9,890,294	56,232,058
CLUSTERING COEFF.	0.1896	0.2737	0.5899	0.6914

Focusing on the COVID-19 epidemic use case, we can transform the trajectory traces we obtained into trajectory networks to model the spread of infection. To do this, we need to determine a specific distance threshold where two individuals are considered in contact. Prior research on SARS-CoV-2 transmission through air droplets has shown that individuals at a close physical distance of $\leq$1-2 m have a high probability of transmission, while there still exists a lower probability when within 2-9 m [39,45]. As we require a single cutoff value, we selected the conservative threshold value $\tau = 2m$. We used the tools developed by [20] to construct the corresponding trajectory networks. The properties of the generated network datasets can be seen in Table 2.

A factor that is somewhat uncertain in related research is the duration required for two individuals to be considered in contact and, subsequently, the transmission probability per time unit β. Studies that examine definitions of contact duration typically consider the case of 1-2 m for 15 minutes or more [19]. With the 12.8% transmission probability from [39], this would result in $\beta \approx 0.85\%$ per minute. Furthermore, there are studies of transmission times in different environments such as airplanes [46] or ventilated spaces [47]. These provide values of 1.8% per minute (quadrupled for conservative results) when within 1 m and 1% per minute when in a well-ventilated space without masks, respectively. In our work, we use $\beta = 1\%$ per minute in most experiments, but we also explore the progress of an epidemic with different values of β.

Regarding the infection's progress, we follow the example of well-established prior research on COVID-19 [32] and utilize the SEIR model with exposure period of 3 days, infectious period of 6 days and recovery period of 10 days. The recovery probability γ helps to understand the severity of a disease in long term, since together with transmission rate β, it determines R_0. However, analysis of varying values of parameter γ is out of the scope of our model that focuses on heterogeneity due to mobility patterns and targeted intervention strategies. In the experiments, we therefore fix the recovery probability to $\gamma = 1$ (i.e., 100%).

With these parameters selected, we aim to answer the following questions:

- **Q1 Estimation of infection risk** $risk_u$. What is the distribution of relative infection risks, for each of the proposed estimation metrics?
- **Q2 Properties of infection seed** $\mathcal{I}_0$. How does the estimated risk or the number of initial infected individuals affect the progress of the epidemic?
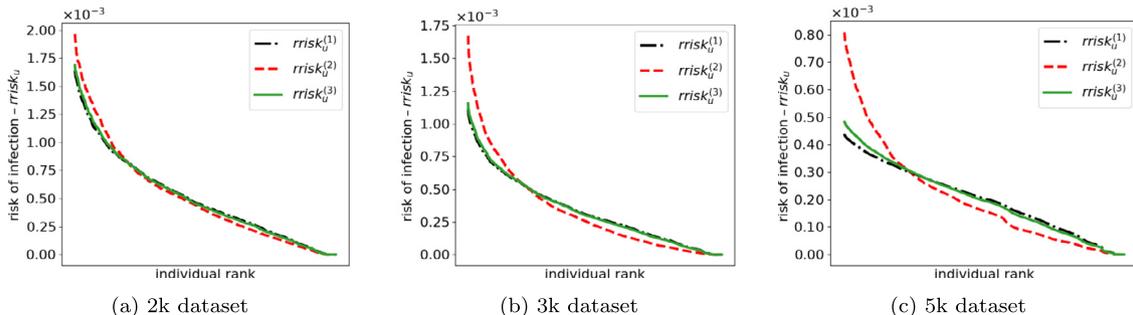
- **Q3 Effect of transmission probability** β. How does the probability of transmission influence the progress of the epidemic?
- **Q4 Effect of quarantine** q. How does the quarantine of infectious individuals influence the progress of the epidemic?
- **Q5 Effect of intervention strategies.** When each intervention strategy is applied, how is the progress of the epidemic affected?
- **Q6 Comparative analysis of intervention strategies.** When the same number of contacts are removed, how does the effect of different intervention strategies compare?

Throughout the experimental evaluation, experiments are repeated for 10 iterations with different random seeds, reporting the average values of results.

4.3. Estimation of infection risk

We utilize each of the three proposed methods to estimate the relative risk of infection for the population sample in all datasets. The resulting distribution of risks can be seen in Fig. 5. As can be seen, the duration-based $risk_u^{(2)}$ gives a higher risk to a smaller number of individuals than the degree-based $risk_u^{(1)}$. The geometric interaction-based $risk_u^{(3)}$ produces an estimate that is balanced between the two other metrics, and we use this in all remaining experiments. The reason why we employ $risk_u^{(3)}$ is that it naturally captures the dynamics of the infection transmission process. In particular, the risk model needs to capture the following characteristics:

- the more contacts an individual has the higher the risk;
- the longer the duration of an interaction, the higher the risk; and
- the risk of infection due to a singular contact should not increase infinitely but it should plateau once it reaches a probability close to 1 (i.e., certain infection).

$risk_u^{(3)}$ uses a geometric function to naturally represent the risk due to these characteristics. Note that the probability of u infecting v after n attempts is increasing for every time unit (i.e., minute), demonstrates diminishing returns and it is eventually plateauing out as it approaches to 1 (i.e., 100%). If we were not considering a geometric function (or similar diminishing returns function), then the risk of certain individuals would grow continuously as a factor of the duration of the contact and would lead to disproportional large risk to certain individuals (due to certain lengthy interactions).

4.4. Infection transmission characteristics

Fig. 6a presents the progress of the SEIR model over the period of 30 days for the case of the 10K dataset; we report the number of individuals that are found in each of the four compartments
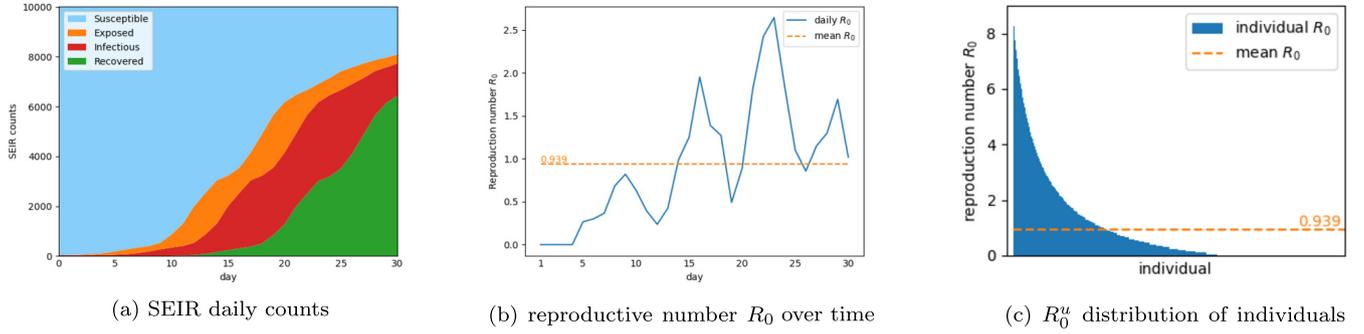

(a) 2k dataset   (b) 3k dataset   (c) 5k dataset

Fig. 5. Infection risk distribution of individuals.

(a) SEIR daily counts

(b) reproductive number R_0 over time

(c) R_0^u distribution of individuals

Fig. 6. Infection progress and characteristics.

(a) daily total infected for seed types
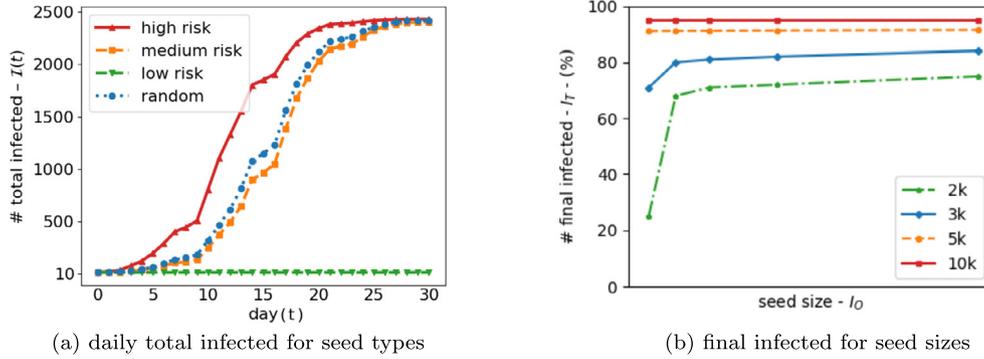
(b) final infected for seed sizes

Fig. 7. Effect of initial seed type and size.

of the model (*Susceptible*, *Exposed*, *Infected* and *Recovered*, respectively).

Furthermore, Fig. 6b shows the basic reproductive number R_0 over time. It can be seen that the R_0 fluctuates over time with values ranging from 0.0 to 2.5, while its 30-day moving average is equal to 0.939 (dashed line). Note that the R_0 is assuming a perfect mixing network (i.e., a complete graph). However, real-world communities do not resemble a complete graph. Our micro-scale analysis of infections allows to monitor the *direct* and *secondary* infections attributed to an individual u, and therefore allows to report the reproductive number R_0^u of $u \in \mathcal{N}$. In contrast to the basic reproduction number R_0 that represents the expected number of infections directly generated by an infectious individual, the R_0^u represents the exact number of people infected by the specific individual u. Now, instead of relying on the R_0 we are in position to provide the distribution of the individual R_0^u values in the population. In Fig. 6c, we show the distribution of R_0^u for the 10k dataset, along with the mean $R_0 = \sum_u^{\mathcal{N}} R_0^u$. It is evident that there is significant variation between individuals, something that the mean R_0 fails to capture.

4.5. Properties of infection seed

In Fig. 7a we can see the infected count $\mathcal{I}$ over the period of a month for an initial "seed" $risk_u^{(3)}$, all of them with high, medium, low, or random risks $risk_u^{(3)}$. In the case of low-risk individuals, the infection never spreads to other people as the initial ones have very limited or no contact with anyone else. In all other cases however, we can see that after a month the vast majority of the population has been infected, with that conclusion arriving faster or slower depending on the initial seed risk. Similarly, in Fig. 7b we can see the final infected counts $\mathcal{I}_T$ for initial seeds of random risk but different size $\mathcal{I}_O$, for all 3 datasets. While there is some reduction in final infections when $\mathcal{I}_O = 1$, all other seed sizes lead to the same result, determined by the population density.

An explanation for this behavior can be found when examining the contribution of each individual in the spread of the infection, and the role of super-spreaders. As mentioned above, we use R_0^u to define the set of individuals that were directly infected by their contact with individual u. We define as R_1^u those infected by any person $v \in R_0^u$, R_2^u those infected by $v \in R_1^u$, and so on. Furthermore, we define $R^u = \{R_0^u, R_1^u, \ldots\}$, i.e. all individuals that were infected directly or indirectly because of u. In Fig. 8a we can see the distribution of R_0^u for the entire population, along with each person's relative risk $rrisk_u^{(2)}$. As expected, high-risk individuals are responsible for the vast majority of direct disease transmissions. However, in Fig. 8b we display the equivalent distribution for R_u. There, we can see that many medium or low-risk individuals are actually responsible for a lot of the secondary, indirect infections. This means that, even when a person with few contacts is infected, a single contact with a high-risk super-spreader is enough for the disease to quickly propagate across the community.

The rest of the experiments use $\mathcal{I}_O = 10$ individuals of random risk $risk_u^{(3)}$.

4.6. Effect of transmission probability

In Fig. 9a we can see the progress of an epidemic in a population of 10,000 when the probability of transmission β has different values. Furthermore, in Fig. 9b we can see the final infected counts for those same values of β for the different datasets. Any value above 1-2% all but guarantees the rapid infection of the entire population, and values below 1% result in significantly reduced counts when the population density isn't too high. It is evident that the transmission probability has a significant impact on the spread of an epidemic; face masks and any other means of reducing it can have critical benefits, as the vast majority of existing research also indicates [45].
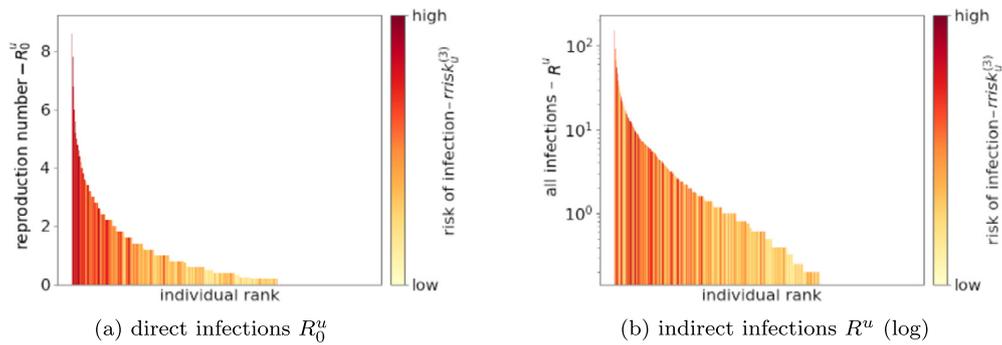
(a) direct infections R_0^u

(b) indirect infections R^u (log)

Fig. 8. Direct and indirect infection counts for different individuals.

(a) daily total infected for different β (10K dataset)

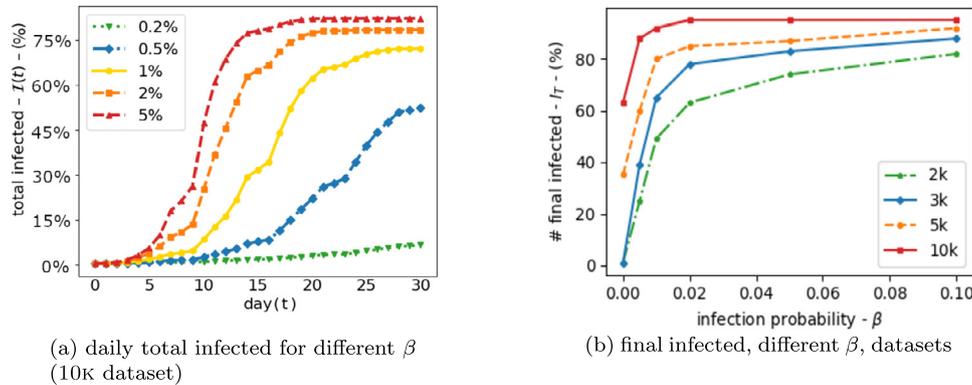(b) final infected, different β, datasets

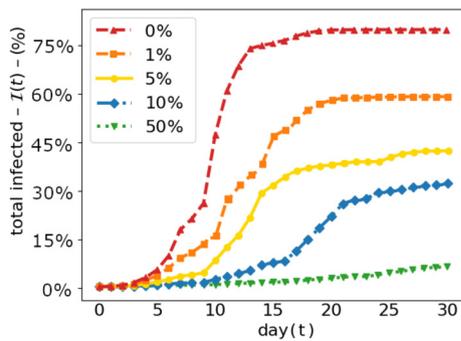**Fig. 9.** Effect of transmission probability β.

Fig. 10. Effect of quarantine on final infected (10K dataset).

4.7. Effect of quarantine

Quarantine policies, which include social distancing, home confinement and centralized quarantine, have been widely used to break the transmission chain of epidemic spread [48]. Even though some human-rights and socioeconomic issues have been raised in the process, this public health measure has proved effective in controlling disease spread [49,50]. Ideally, only those who are infected should be quarantined, while others can travel as they wish. However, this imposes another challenge as recent studies have shown that many infected individuals are asymptomatic or only have mild symptoms [51,52]. These individuals are most likely not aware they are infected, still able to transmit the virus to others, and therefore in need of quarantine.

To evaluate the effect of quarantine, we incorporate a quarantine parameter q in the experiments that represents the proportion of infected individuals that will have no contact with any others until they fully recover; we assume that the rest $1 - q$ proportion of the infected nodes will not quarantine and will continue interacting with others (e.g., due to no symptoms). Given

that the quarantine of every infectious individual corresponds with the removal of many probable-transmission contacts, the expectation for this case is that the effect of the quarantine parameter q on the epidemic will be significant. Indeed, as can be seen in Fig. 10, even small values can result in greatly reduced infection numbers, which greatly highlights the importance of quarantine measures. For the remainder of this work we wish to examine other properties in isolation, and we therefore set the parameter $q = 0$.

4.8. Effect of intervention strategies

To answer this research question, we examine the progress of the epidemic after applying the proposed intervention strategies on the 3k dataset. For each strategy, we report intervention and null-model results of $\mathcal{I}$ for an intervention proportion $\alpha = 0.2$, and the final infected $\mathcal{I}_T$ for different α values.

Fig. 11a shows the results of removing 20% of the network nodes, i.e. individuals based on their risk. As mentioned earlier, this corresponds with targeted immunization or isolation of selected individuals. We can see that this has a substantial effect, greatly reducing the spread of the epidemic. The result is much less pronounced when removing medium or random risk individuals, and removing low-risk individuals has almost no effect. In Fig. 11b we can see the outcome for different proportions α. In this case, simply removing 30% of the highest-risk individuals practically eliminates the spread of infection completely.

Figs. 12a and 12b show the same results for the high-risk contact removal intervention. As mentioned in Section 3, this is equivalent to every person avoiding high-risk individuals among their contacts. We can see that this time the targeted strategy performs only slightly better than the null model, although that difference grows for higher intervention proportions. The result is not surprising. This is because the null model employed is not necessarily
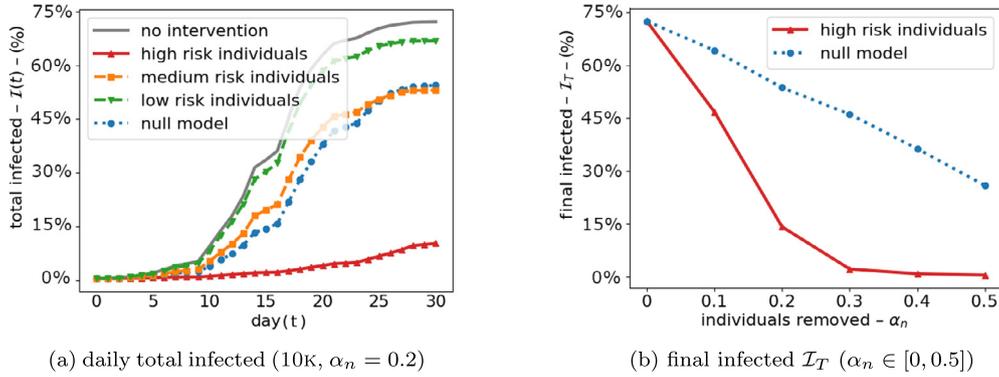
(a) daily total infected (10ᴋ, $\alpha_n = 0.2$)

(b) final infected $\mathcal{I}_T$ ($\alpha_n \in [0, 0.5]$)

Fig. 11. Results after high risk node immunization intervention.

(a) daily total infected (10ᴋ, $\alpha_e = 0.2$)

(b) final infected $\mathcal{I}_T$ ($\alpha_e \in [0, 0.5]$)
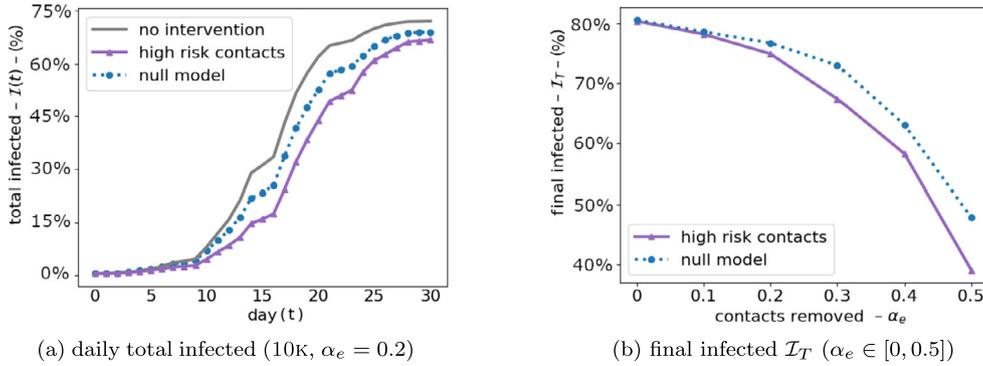
Fig. 12. Results of removing high-risk individuals intervention.

(a) daily total infected (10ᴋ, $\alpha_e = 0.2$)

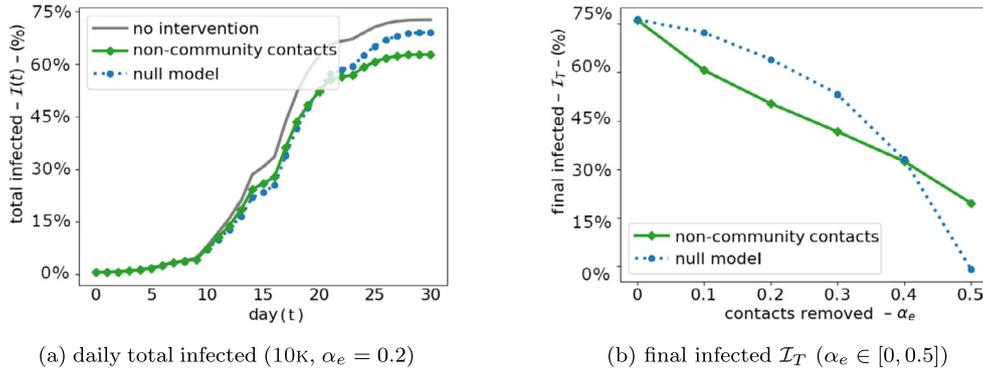(b) final infected $\mathcal{I}_T$ ($\alpha_e \in [0, 0.5]$)

Fig. 13. Results of non-community contacts removal intervention.

representing a bad strategy; by randomly selecting nodes and removing edges, one may remove high-risk, medium-risk or low-risk contacts. Furthermore, the aggregated network of the 3ᴋ dataset represents a small-world graph. It is well-known that when only a small proportion of edges are removed from that network (i.e. *alpha* is small), the impact on the connectivity of the network might not be significant due to the high clustering property of this type of graph.

Finally, Figs. 13a and 13b show the $\alpha = 20\%$ progress and different-proportion final counts $\mathcal{I}_T$ for the non-community contact removal intervention strategy. As mentioned earlier this is equivalent to the "social bubble" concept, where each person only maintains contact with their close friends and family, avoiding people from other groups. We can see that this strategy is notably more effective than the null-model one for $\alpha < 40\%$. Above that value, it is more beneficial to simply reduce each person's contacts overall, rather than targeting the more sporadic and (brief) contacts with people outside their community.

4.9. Comparative analysis of intervention strategies

In order to compare the effectiveness of the proposed intervention strategies we have to take into account their relative impact on the population. Specifically, in high-risk node immunization with $\alpha_n = 0.1$ the percentage of removed edges α_e is much higher. In order to present a fair comparison of the intervention strategies, we report their results based on the number of removed contacts α_e. The results for this experiment can be seen on Fig. 14a.

In order to compare the effectiveness of the proposed intervention strategies, we have to take into account their relative impact on the population. Specifically, in high-risk node immunization with $\alpha_n = 0.1$, the percentage of removed edges α_e is much higher. Even with the same strategies (e.g., breaking of social ties), the number of edges to be removed can be different. As such, in order to present a fair comparison of the intervention strategies, we normalized the number of edges being removed in the experiments and report their results based on the number of
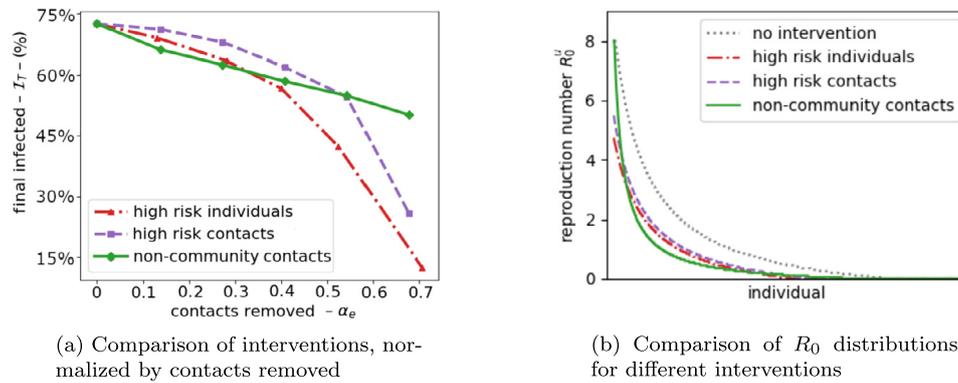
(a) Comparison of interventions, normalized by contacts removed

(b) Comparison of R_0 distributions for different interventions

Fig. 14. Distribution of R_0 value for each individual.

removed contacts α_e. The results for this experiment can be seen on Fig. 14a.

We can see that the "social bubble" intervention yields slightly better results than the high-risk individual removal one, when 30% or fewer contacts are removed. Afterwards, targeted immunization/isolation of individuals is significantly more effective. However, as mentioned in Section 3, we also need to take into consideration the feasibility of each approach. In that case, the uncommon edge intervention isn't only the most easily implemented one, but also performs very well up until a full 50-60% of all contacts have been removed. Any intervention above such magnitudes would significantly impact the movement and interactions of the population, and may be unrealistic for larger communities.

Fig. 14b provides the R_0^u distributions after applying each of the interventions. This allows for a more in-depth examination of each intervention's effects and better interpretation of the performance results. In this plot, the total area under each curve corresponds to the sum of all transmissions of the infection for that case. Therefore, it is expected that all interventions will result in curves covering smaller areas than the non-intervention case. However, each intervention accomplishes this in a different way. The node immunization and high-risk edge removal interventions simply remove most of the highest-risk contacts, corresponding to many of the highest R_0^u values in the distribution, resulting in "shifting" the entire distribution to the left. On the other hand, the non-community edge removal intervention removes values throughout the distribution, resulting in a steeper curve.

5. Related work

Our research is related to (*i*) *trajectory data mining*, (*ii*) *dynamic network analysis*, and (*iii*) *epidemic spreading in complex networks*. These topics have been active research directions for a long time, so there is a broad spectrum of related literature. However, it is only recently that due to the technological advancement in geolocation tracking devices (e.g., GPS-enabled mobile devices), mobility data is becoming more accessible. Mining patterns in large amounts of human digital traces provides an opportunity for designing more accurate epidemic spreading models and more effective network intervention strategies than before. We cover below some of the most significant efforts relevant to that goal. Note, as well that some of the related work has already been cited throughout the manuscript to keep the discussion focused, so we omit it here.

5.1. Trajectory data mining

Computational methods for mining spatiotemporal data, including trajectory/mobility data, have been extensively studied by the data mining and database communities. Two comprehensive surveys are provided by Zheng [18] and Atluri et al. [53]. Closely related to the problem of interest in this paper are problems that focus on mining the *interactions among moving objects, over time*, such as detecting pedestrian groups in trajectories [54–56] or determining the node centrality of moving objects in trajectory networks [20]. More recently, deep learning approaches for learning from spatiotemporal data and spatiotemporal networks have gained increasing attention [57,58].

Digital contact tracing: In our research, we assume that human mobile traces are available. This can be enabled by existing digital contact tracing technologies [59–61]. For instance, Aleta et al. [62] synthesized contact networks and modeled SARS-CoV-2 transmission in the Boston metropolitan area. They found that effective testing and contact tracing plays an important role in preventing second-wave spreading when complete isolation is relaxed. Other technologies of contact tracing have also been proposed [63].

Privacy concerns of digital contact tracing: Digital contact tracing enables an easy and rapid implementation of infectious disease tracing, as it requires to gather and process simple information. However, gathering sensitive information might infringe the privacy of individuals [64]. We believe that controlling an infectious disease should not lead to a weakening of the privacy of individuals. We therefore advocate for protocols and technologies for "privacy-preserving proximity tracing" that protect the privacy of individuals [65,66]. Connectivity technology available in mobile devices and newly developed connectivity protocols can make a decisive contribution to efficiently and widely support proximity tracing enabled by bluetooth connection and/or GPS location. Both approaches protect the privacy of the user through dynamic pseudo-IDs [67]. For instance, Apple (iOS) and Google (Android) are providing privacy-preserving cross-platform contact tracing via an open API and an opt-in Bluetooth-based proximity tracking.[1] In addition, the Pan-European Privacy-Preserving Proximity Tracing (PEPP-PT) protocol[2] and the Decentralized Privacy-Preserving Proximity Tracing (DP-3T) [68] protocol[3] have recently been proposed for providing a secure and decentralized privacy-preserving proximity tracing system.

5.2. Dynamic networks analysis

The problem of objects being dispersed in space and interacting with each other if they are in close vicinity has been intensely explored in graph theory. Graph theory concepts, such as *proximity*

---

graphs [69] and *geometric intersection graphs* [70] are characteristic examples. For instance, relative neighbor graphs [71] and Gabriel graphs [69] connect nearest neighbors if no other vertexes are nearby, while Delaunay triangulations [72] maximize the minimum angles of all triangles formed. These graphs, however, mostly deal with static data, while in our problem we are studying cases of multiple proximity graphs, one for each time unit. There has also been significant research on *dynamic networks*, such as *time-varying networks* or otherwise *temporal networks*. With the addition of temporal information, several concepts of static networks are not valid anymore, so they need to be adapted/studied in the context of dynamic networks. Due to its importance, dynamic network analysis is therefore an emergent discipline of network science focusing on network dynamics. Example problems include the computation of temporal node centrality [25] or computing metrics of network reachability [73], shortest paths [24], motifs [74] and other [75]. The dynamic nature of these systems introduces additional complexity and computational challenges. Dynamic networks have also been studied using machine learning models in the context of gradually evolving networks, where nodes/edges are added/removed over time [76,77]. A comprehensive survey of this line of research can be found in [78]. Related to the current research, Leitch et al. [79] presented a review towards epidemic thresholds on temporal networks; they pointed out that temporal networks engage dynamics of real-world contacts, so their study is of great importance for understanding disease spreading processes. Statistical approaches or computer simulations are often necessary to explore the evolution of these external processes over evolving networks.

5.3. Epidemic spreading in complex networks

Mathematical modeling of epidemic spreading in networks can help to study and control the emergence of infectious diseases in a population. Based on the traditional $\mathcal{SIR}$ model, Weitz et al. [80] designed epidemiological interventions that can exploit the idea of *'shield immunity'*. The main idea of the model is to deploy recovered individuals as focal points for sustaining safer interactions via interaction substitution. This method, however, cannot easily translate to health policy and/or individual level recommendations. In our study, we employ an *agent-based $\mathcal{SEIR}$ model*. The agent-based $\mathcal{SEIR}$ model has previously been employed to study the spread of epidemics in dynamic networks. For instance, Perez and Dragicevic [33] proposed a spatially explicit epidemiological model of infectious disease for understanding of the diffusion of a disease in a network of human contacts. In their model, human interactions are not fully dynamic, but are determined by the geographic area they are found at the same time. Yang et al. [81] proposed a flow-based edge betweenness method that detects important "bottleneck" edges in contact networks. They show that targeting those edges can contain the epidemic spread more than state-of-the-art edge betweenness methods. Their model does not easily translate to individual level policy, but offers high-level guidance for the network containment problem. Agent based epidemic spreading models are also very similar to the study of the epidemic spreading driven by random walks and one can translate to the other. Similar to the $\mathcal{SEIR}$ models, in the random walk based epidemic models, the infection is spread to the neighbors with a probability. This probability is determined by the transition matrix of the random walker. Pu et al. [82] proposed a biased random walk based spreading model. They show that the average node degree and homogeneity of the node degree plays an important role in the number of the infected nodes. More recently, Bestehorn et al. [83] derive an upper bound for the reproduction number based on a discrete-time Markovian random walk model of the infection spreading. New advances in deep learning and network represen-

tation learning have also been used to model the epidemics [84]. Change et al. [85] propose an epidemic model based on the embedding of the mobility network and show the role of the social and economical disparities in the spread of the epidemics and the network structure. The main limitation of these models is that they model the epidemics on a static network and evolving aspect of the network is not analyzed. Heidari and Papagelis [76,77] have proposed evolving network representation learning methods that lie in the intersection of these two topics and can be used to model the epidemic spreading in evolving networks.

Population heterogeneity in epidemics: Britton et al. [38] studied the effect of population heterogeneity on achieving the epidemiological objective of '*herd immunity*' to the COVID-19 disease. In their study, they found that herd immunity can be achieved with less per cent of the population being infected than it was thought to be required (i.e., 60%), if one considers the age and social activity into the model. Similar ideas of evaluating the effect of *individual variability* in the disease spreading process have been considered by others. Our research is individual variability based on the structural information of a network; information that is related to the differences of nodes in the network based on basic graph/network characteristics (e.g., based on node centrality in the dynamic network, etc.). Related to our main thesis, Hébert-Dufresne et al. [86] argued that using R_0 alone to predict epidemic size is not enough in real-world outbreaks. They pointed out the necessity of considering heterogeneity in secondary infections to predict outbreak size. In practice, this can be done automatically, fairly cheaply, and highly accurately, by a wide-spread deployment and adoption of digital contact tracing technologies. Moreover, Lloyd-Smith et al. [87] pointed out that the basic reproduction number R_0 in the traditional epidemic analyses is a population-level estimate. Through a theoretical and statistical analysis, they showed that individual variation greatly affects epidemic growth rates, and therefore targeted control interventions would be more effective than population-wide ones. On a similar basis, Changruenngam et al. [88] studied the effect of human mobility on disease transmission dynamics in two contrasted countries. In the study, they incorporated individual human mobility in the $\mathcal{SEIR}$ model, which helped better describe infection spreading dynamics. In particular, based on population data of two areas, human mobility was modeled by obtaining the probability for an individual exploring new locations, from which the human mobility landscape could also be captured. Rocha and Masuda [89] augmented the $\mathcal{SIR}$ model by incorporating an individual-based approximation that captures the evolution of the probability that an individual is infected by another individual in the network. These studies highlighted the importance of incorporating individual variation in the epidemic model.

Network containment interventions: Instead of a complete or near-complete lockdown, P. Block et al. [42] proposed more moderate distancing strategies inspired by network science, including limiting contacts to *similar*, *community-based* or *repetitive* contacts. This study is based on *a static network*, while in a real-world situation, the network is dynamic informed by how individuals move around and interact with each other. Incorporating individual variation in the model has also the advantage that network interventions can be designed at an individual's level. For instance, Zhou et al. [90] reviewed several studies related to network immunization and concluded that vaccination of targeted nodes (e.g., nodes with large node degree), outperforms random immunization. Similarly, Torres et al. [91] evaluated different immunization strategies and found that node degree is a very strong measurement in determining node importance when considering the targeted nodes. In [92], authors study the spread of epidemics on static and temporal networks. Epidemics on temporal networks is closely related to our

work as it follows the same assumption in the network construction: the network evolves faster than the spread of the pathogen. However, that work assumes same degree for all the nodes in the temporal network. In our case, node degrees are determined by the mobility patterns of individuals.

6. Conclusions

Effective modeling of an emerging infectious disease has the potential to improve or save human lives. It can also minimize the societal and economic damage caused by physical distancing and confinement measures imposed by governing authorities to control an epidemic. Towards that end, we have presented a data-driven model of infectious epidemic spreading in spatiotemporal networks informed by mobility data of individuals. We designed and evaluated simple individual-based intervention strategies that exhibit network effects and can significantly control the spread of an infectious disease. We have also demonstrated that these targeted interventions can outperform generic intervention strategies. These strategies are easy to understand and translate to public health policy. While COVID-19 serves as a use case in this research, the same methodology can be used to model and mitigate any emerging infectious disease. An inherent limitation of our model is that it assumes availability of trajectories of individuals through digital tracing technologies. While these data are typically available to telecom and other third-parties through privacy-preserving techniques, our research relied only on realistic synthetic data sets.

Reproducibility: We make source code and data sets used in the experiments publicly available[4] to encourage reproducibility of results.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C.F. Smith, et al., Thucydides: History of the Peloponnesian War, vol. 108, W. Heinemann, 1962.

[2] M.J. Papagrigorakis, C. Yapijakis, P.N. Synodinos, E. Baziotopoulou-Valavani, Dna examination of ancient dental pulp incriminates typhoid fever as a probable cause of the plague of Athens, Int. J. Infect. Dis. 10 (3) (2006) 206–214.

[3] J.K. Taubenberger, D.M. Morens, 1918 influenza: the mother of all pandemics, Rev. Biomed. 17 (1) (2006) 69–79.

[4] D.M. Morens, A.S. Fauci, The 1918 influenza pandemic: insights for the 21st century, J. Infect. Dis. 195 (7) (2007) 1018–1028.

[5] J.N. Hays, Epidemics and Pandemics: Their Impacts on Human History, ABC-CLIO, 2005.

[6] J.S. Peiris, K.Y. Yuen, A.D. Osterhaus, K. Stöhr, The severe acute respiratory syndrome, N. Engl. J. Med. 349 (25) (2003).

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

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

[9] 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, N. Engl. J. Med. 371 (15) (2014) 1418–1425.

[10] G.S. Campos, A.C. Bandeira, S.I. Sardi, Zika virus outbreak, Bahia, Brazil, Emerg. Infect. Dis. 21 (10) (2015) 1885.

[11] W.H. Organization, et al. Coronavirus disease 2019 (covid-19): situation report, 72 (2020).

[12] K.T. Eames, M.J. Keeling, Contact tracing and disease control, Proc. R. Soc. Lond. B, Biol. Sci. 270 (1533) (2003) 2565–2571.

[13] World Health Organization & Centers for Disease Control and Prevention (U.S.), Implementation and management of contact tracing for Ebola virus disease: emergency guideline, Tech. rep., World Health Organization, 2015.

[14] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, C. Fraser, Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing, Science 368 (2020).

[15] S. Woodhams, Covid-19 digital rights tracker, 2020.

[16] M.F. Mokbel, S. Abbar, R. Stanojevic, Contact tracing: beyond the apps, preprint, arXiv:2006.04585, 2020.

[17] D. Chakrabarti, C. Faloutsos, Graph mining: laws, generators, and algorithms, ACM Comput. Surv. 38 (1) (2006) 2.

[18] Y. Zheng, Trajectory data mining: an overview, ACM Trans. Intell. Syst. Technol. 6 (3) (2015) 1–41.

[19] M.J. Keeling, T.D. Hollingsworth, J.M. Read, The efficacy of contact tracing for the containment of the 2019 novel coronavirus (covid-19), https://doi.org/10.1101/2020.02.14.20023036, 2020.

[20] T. Pechlivanoglou, M. Papagelis, Fast and accurate mining of node importance in trajectory networks, in: 2018 IEEE International Conference on Big Data, Big Data, IEEE, 2018, pp. 781–790.

[21] M.E. Newman, A measure of betweenness centrality based on random walks, Soc. Netw. 27 (1) (2005) 39–54.

[22] U. Brandes, A faster algorithm for betweenness centrality, J. Math. Sociol. 25 (2) (2001) 163–177.

[23] D. Kempe, J. Kleinberg, A. Kumar, Connectivity and inference problems for temporal networks, in: Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, ACM, 2000, pp. 504–513.

[24] H. Wu, J. Cheng, S. Huang, Y. Ke, Y. Lu, Y. Xu, Path problems in temporal graphs, Proc. VLDB Endow. 7 (9) (2014) 721–732.

[25] H. Kim, R. Anderson, Temporal node centrality in complex networks, Phys. Rev. E 85 (2) (2012) 026107.

[26] M.J. Keeling, B.T. Grenfell, Individual-based perspectives on r0, J. Theor. Biol. 203 (1) (2000) 51–61.

[27] O. Diekmann, J. Heesterbeek, J. Metz, On the definition and the computation of the basic reproduction ratio r 0 in models for infectious diseases in heterogeneous populations, J. Math. Biol. 28 (4) (1990), https://doi.org/10.1007/bf00178324.

[28] R.M. Anderson, B. Anderson, R.M. May, Infectious Diseases of Humans: Dynamics and Control, Oxford University Press, 1992.

[29] F. Brauer, C. Castillo-Chavez, C. Castillo-Chavez, Mathematical Models in Population Biology and Epidemiology, vol. 2, Springer, 2012.

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

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

[32] A.R. Tuite, D.N. Fisman, A.L. Greer, Mathematical modelling of covid-19 transmission and mitigation strategies in the population of Ontario, Canada, CMAJ, Can. Med. Assoc. J. 192 (19) (2020) E497–E505.

[33] L. Perez, S. Dragicevic, An agent-based approach for modeling dynamics of contagious disease spread, Int. J. Health Geogr. 8 (1) (2009) 50.

[34] N. Hoertel, M. Blachier, C. Blanco, M. Olfson, M. Massetti, F. Limosin, H. Leleu, Facing the covid-19 epidemic in nyc: a stochastic agent-based model of various intervention strategies, medRxiv.

[35] D. Yao, R. Durrett, Epidemics on evolving graphs, preprint, arXiv:2003.08534, 2020.

[36] 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, Ecol. Model. 198 (1–2) (2006) 115–126.

[37] M.G.M. Gomes, R. Aguas, R.M. Corder, J.G. King, K.E. Langwig, C. Souto-Maior, J. Carneiro, M.U. Ferreira, C. Penha-Goncalves, Individual variation in susceptibility or exposure to sars-cov-2 lowers the herd immunity threshold, medRxiv, 2020.

[38] T. Britton, F. Ball, P. Trapman, A mathematical model reveals the influence of population heterogeneity on herd immunity to sars-cov-2, Science (2020).

[39] N.R. Jones, Z.U. Qureshi, R.J. Temple, J.P.J. Larwood, T. Greenhalgh, L. Bourouiba, Two metres or one: what is the evidence for physical distancing in covid-19?, BMJ, Br. Med. J. 370 (2020), https://doi.org/10.1136/bmj.m3223.

[40] A. Barbour, D. Mollison, Epidemics and random graphs, in: Stochastic Processes in Epidemic Theory, Springer, 1990, pp. 86–89.

[41] H. Frisch, J. Hammersley, Percolation processes and related topics, J. Soc. Ind. Appl. Math. 11 (4) (1963) 894–918.

[42] P. Block, M. Hoffman, I.J. Raabe, J.B. Dowd, C. Rahal, R. Kashyap, M.C. Mills, Social network-based distancing strategies to flatten the covid-19 curve in a post-lockdown world, Nat. Hum. Behav. (2020) 1–9.

[43] C.R. Bhat, S.K. Singh, A comprehensive daily activity-travel generation model system for workers, Transp. Res., Part A, Policy Pract. 34 (1) (2000) 1–22, https://doi.org/10.1016/s0965-8564(98)00037-8.

[44] P.A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, E. Wießner, Microscopic traffic simulation using sumo, in: The 21st IEEE International Conference on Intelligent Transportation Systems, IEEE, 2018, https://elib.dlr.de/124092/.

[45] D.K. Chu, E.A. Akl, S. Duda, K. Solo, S. Yaacoub, H.J. Schünemann, A. El-harakeh, A. Bognanni, T. Lotfi, M. Loeb, et al., Physical distancing, face masks, and

---

[4] To be announced.

eye protection to prevent person-to-person transmission of sars-cov-2 and covid-19: a systematic review and meta-analysis, Lancet 395 (10242) (2020) 1973–1987.

[46] V.S. Hertzberg, H. Weiss, L. Elon, W. Si, S.L. Norris, Behaviors, movements, and transmission of droplet-mediated respiratory diseases during transcontinental airline flights, Proc. Natl. Acad. Sci. 115 (14) (2018) 3623–3627, https://doi.org/10.1073/pnas.1711611115.

[47] B.L. Augenbraun, Z.D. Lasner, D. Mitra, S. Prabhu, S. Raval, H. Sawaoka, J.M. Doyle, Assessment and mitigation of aerosol airborne sars-cov-2 transmission in laboratory and office environments, J. Occup. Environ. Hyg. 17 (10) (2020) 447–456.

[48] G.F. Gensini, M.H. Yacoub, A.A. Conti, The concept of quarantine in history: from plague to sars, J. Infect. 49 (4) (2004) 257–261.

[49] E. Tognotti, Lessons from the history of quarantine, from plague to influenza a, Emerg. Infect. Dis. 19 (2) (2013) 254.

[50] Y. Wang, L. Shi, J. Que, Q. Lu, L. Liu, Z. Lu, Y. Xu, J. Liu, Y. Sun, S. Meng, et al., The impact of quarantine on mental health status among general population in China during the covid-19 pandemic, Mol. Psychiatry (2021) 1–10.

[51] Z. Gao, Y. Xu, C. Sun, X. Wang, Y. Guo, S. Qiu, K. Ma, A systematic review of asymptomatic infections with covid-19, J. Microbiol. Immunol. Infect. 54 (1) (2021) 12–16.

[52] G.-u. Kim, M.-J. Kim, S.H. Ra, J. Lee, S. Bae, J. Jung, S.-H. Kim, Clinical characteristics of asymptomatic and symptomatic patients with mild covid-19, Clin. Microbiol. Infect. 26 (7) (2020) 948–e1.

[53] G. Atluri, A. Karpatne, V. Kumar, Spatio-temporal data mining: a survey of problems and methods, ACM Comput. Surv. 51 (4) (2018).

[54] A. Sawas, A. Abuolaim, M. Afifi, M. Papagelis, Tensor methods for group pattern discovery of pedestrian trajectories, in: 2018 19th IEEE International Conference on Mobile Data Management, MDM, IEEE, 2018.

[55] A. Sawas, A. Abuolaim, M. Afifi, M. Papagelis, Trajectolizer: interactive analysis and exploration of trajectory group dynamics, in: 2018 19th IEEE International Conference on Mobile Data Management, MDM, IEEE, 2018, pp. 286–287.

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

[57] S. Wang, J. Cao, P. Yu, Deep learning for spatio-temporal data mining: a survey, in: IEEE Transactions on Knowledge and Data Engineering, 2020.

[58] S. Mehmood, M. Papagelis, Learning semantic relationships of geographical areas based on trajectories, in: 2020 21st IEEE International Conference on Mobile Data Management, MDM, IEEE, 2020, pp. 109–118.

[59] K. Farrahi, R. Emonet, M. Cebrian, Epidemic contact tracing via communication traces, PLoS ONE 9 (5) (2014) e95133.

[60] G. Cencetti, G. Santin, A. Longa, E. Pigani, A. Barrat, C. Cattuto, S. Lehmann, M. Salathe, B. Lepri, Digital proximity tracing in the covid-19 pandemic on empirical contact networks, 2020, medRxiv.

[61] F.D. Sahneh, C. Scoglio, Epidemic spread in human networks, in: 2011 50th IEEE Conference on Decision and Control and European Control Conference, IEEE, 2011, pp. 3008–3013.

[62] A. Aleta, D. Martin-Corral, A.P. y Piontti, M. Ajelli, M. Litvinova, M. Chinazzi, N. Dean, M.E. Halloran, I. 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, 2020, medRxiv.

[63] Y. Luo, C. Zhang, Y. Zhang, C. Zuo, D. Xuan, Z. Lin, A.C. Champion, N. Shroff, Acoustic-turf: acoustic-based privacy-preserving covid-19 contact tracing, preprint, arXiv:2006.13362, 2020.

[64] H. Cho, D. Ippolito, Y.W. Yu, Contact tracing mobile apps for covid-19: privacy considerations and related trade-offs, preprint, arXiv:2003.11511, 2020.

[65] B. Hoh, M. Gruteser, H. Xiong, A. Alrabady, Preserving privacy in gps traces via uncertainty-aware path cloaking, in: Proceedings of the 14th ACM Conference on Computer and Communications Security, 2007.

[66] L. Reichert, S. Brack, B. Scheuermann, Privacy-preserving contact tracing of covid-19 patients, IACR Cryptol. ePrint Arch. 2020 (2020) 375.

[67] 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, 2020.

[68] 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, 2020.

[69] K.R. Gabriel, R.R. Sokal, A new statistical approach to geographic variation analysis, Syst. Zool. 18 (3) (1969) 259.

[70] T. Erlebach, K. Jansen, E. Seidel, Polynomial-time approximation schemes for geometric intersection graphs, SIAM J. Comput. 6 (2005).

[71] G.T. Toussaint, The relative neighbourhood graph of a finite planar set, Pattern Recognit. 12 (4) (1980) 261–268.

[72] B. Delaunay, Sur la sphère vide. a la mémoire de Georges Voronoï, Bull. Acad. Sci. URSS 6 (1934) 793–800.

[73] P. Holme, Network reachability of real-world contact sequences, Phys. Rev. E 71 (4) (2005) 046119.

[74] L. Kovanen, M. Karsai, K. Kaski, J. Kertész, J. Saramäki, Temporal motifs in time-dependent networks, J. Stat. Mech. Theory Exp. 2011 (11) (2011) P11005.

[75] V. Nicosia, J. Tang, C. Mascolo, M. Musolesi, G. Russo, V. Latora, Graph metrics for temporal networks, in: Temporal Networks, Springer, 2013.

[76] F. Heidari, M. Papagelis, Evonrl: evolving network representation learning based on random walks, in: International Conference on Complex Networks and Their Applications, Springer, 2018, pp. 457–469.

[77] F. Heidari, M. Papagelis, Evolving network representation learning based on random walks, Appl. Netw. Sci. 5 (1) (2020) 1–38.

[78] Z. Zhang, P. Cui, W. Zhu, Deep learning on graphs: a survey, in: IEEE Transactions on Knowledge and Data Engineering, 2020.

[79] J. Leitch, K.A. Alexander, S. Sengupta, Toward epidemic thresholds on temporal networks: a review and open questions, Appl. Netw. Sci. 4 (1) (2019) 105.

[80] 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 (2020).

[81] S. Yang, P. Senapati, D. Wang, C.T. Bauch, K. Fountoulakis, Targeted pandemic containment through identifying local contact network bottlenecks, preprint, arXiv:2006.06939, 2020.

[82] C. Pu, S. Li, J. Yang, Epidemic spreading driven by biased random walks, Physica A 432 (2015) 230–239.

[83] M. Bestehorn, A.P. Riascos, T.M. Michelitsch, B.A. Collet, A markovian random walk model of epidemic spreading, in: Continuum Mechanics and Thermodynamics, 2021, pp. 1–15.

[84] W.L. Hamilton, R. Ying, J. Leskovec, Representation learning on graphs: methods and applications, preprint, arXiv:1709.05584, 2017.

[85] S. Chang, E. Pierson, P.W. Koh, J. Gerardin, B. Redbird, D. Grusky, J. Leskovec, Mobility network models of covid-19 explain inequities and inform reopening, Nature 589 (7840) (2021) 82–87.

[86] L. Hébert-Dufresne, B.M. Althouse, S.V. Scarpino, A. Allard, Beyond r0: heterogeneity in secondary infections and probabilistic epidemic forecasting, medRxiv, 2020.

[87] J. Lloyd-Smith, S. Schreiber, P.E. Kopp, W. Getz, Superspreading and the effect of individual variation on disease emergence, Nature 438 (2005).

[88] S. Changruenngam, D.J. Bicout, C. Modchang, How the individual human mobility spatio-temporally shapes the disease transmission dynamics, Sci. Rep. 10 (1) (2020) 1–13.

[89] L.E. Rocha, N. Masuda, Individual-based approach to epidemic processes on arbitrary dynamic contact networks, Sci. Rep. 6 (2016) 31456.

[90] Z. Tao, F. Zhongqian, W. Binghong, Epidemic dynamics on complex networks, Prog. Nat. Sci. 16 (5) (2006) 452–457.

[91] L. Torres, K.S. Chan, H. Tong, T. Eliassi-Rad, Node immunization with non-backtracking eigenvalues, preprint, arXiv:2002.12309, 2020.

[92] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks, Rev. Mod. Phys. 87 (2015).