A Design-Methodology for Epidemic Dynamics via Time-Varying Hypergraphs

Alessia Antelmi
Università degli Studi di Salerno
Fisciano, Italy
aantelmi@unisa.it

Gennaro Cordasco
Università della Campania "Luigi Vanvitelli"
Caserta, Italy
gennaro.cordasco@unicampania.it

Carmine Spagnuolo*
Università degli Studi di Salerno
Fisciano, Italy
cspagnuolo@unisa.it

Vittorio Scarano
Università degli Studi di Salerno
Fisciano, Italy
vitsca@unisa.it

ABSTRACT

In epidemiology science, the importance to explore innovative modeling tools for acutely analyzing epidemic diffusion is turning into a big challenge considering the myriad of real-world aspects to capture. Typically, equation-based models, such as SIS and SIR, are used to study the propagation of diseases over a population. Improved approaches also include human-mobility patterns as network information to describe contacts among individuals. However, there still is the need to incorporate in these models information about different types of contagion, geographical information, humans habits, and environmental properties. In this paper, we propose a novel approach that takes into account: 1. direct and indirect epidemic contagion pathways to explore the dynamics of the epidemic, 2. the times of possible contagions, and 3. human-mobility patterns. We combine these three features exploiting *time-varying hypergraphs*, and we embed this model into a design-methodology for agent-based models (ABMs), able to improve the correctness in the epidemic estimations of classical contact-network approaches. We further describe a diffusion algorithm suitable for our design-methodology and adaptable to the peculiarities of any disease spreading policies and/or models. Finally, we tested our methodology by developing an ABM, realizing the SIS epidemic compartmental model, for simulating an epidemic propagation over a population of individuals. We experimented the model using real user-mobility data from the location-based social network *Foursquare*, and we demonstrated the high-impact of temporal direct and indirect contagion pathways.

KEYWORDS

Epidemiology, Agent-based Model, Time-Varying Hypergraph, Direct and indirect infection, Location-based Social Network

*Corresponding author.

1 INTRODUCTION

With the increasing amount of global human-mobility, designing, implementing and evaluating public health policies has become a crucial challenge, notably for epidemiology science. In this research area, most of the well-known adopted models are based on math equations and have proved their ability to mimic the epidemic spreading in a population of individuals [15, 17]. However, these models assume that the population behavior and individual contact types are homogeneous [3]. The aforementioned is a serious limitation for real-world scenarios, as it reduces the modeling effectiveness in describing different sub-populations behaviors, and/or urban mobility patterns, geographic information systems (GIS), and so on [18]. Furthermore, equation-based models do not provide an easy way to model different types of contacts [12], even though many epidemic contagion operate in two ways: *direct* contagion (person-to-person infection) and *indirect* contagion (infection via an intermediary, such as another person, furnishings, clothing).

Agent-based Models (ABMs) are a modeling tool able to easily incorporate features related to population and society and are widely adopted in epidemiology. ABMs allow researchers to naturally include human-mobility data to model humans interactions between the environment and/or other individuals. Typically, many epidemic ABMs also exploit networks to define possible agents interactions. Nowadays, online social networks (OSNs) users can share their real-time location (Foursquare), geo-tagging media post (Facebook and Instagram), reviewing businesses (Yelp). The growing popularity of these online platforms and the ubiquitous online access provide gold data for studying users habits, lifestyle, and mobility patterns to be included in ABMs.

In this paper, we focus on modeling direct and indirect contacts by integrating them into a common design-methodology for the analysis of epidemic spreads over a population of individuals, based on an ABM that includes OSNs information to describe human-to-human and human-to-environment interactions. Here, we exploit OSNs as input to model interactions, but our methodology is suitable for any kind of *human-mobility data*. We consider classical epidemic models (i.e. equation-based, such as SIR and SIS), integrated into a simulation that mimic agents interaction by exploiting OSNs data [21]. In our work, we propose an innovative modeling approach to study the propagation of an epidemic over a set of agents considering many-to-many relationships by exploiting *hypergraphs*.

A hypergraph is a generalization of the graph mathematical model, where links (edges) represent not only pairwise relationships but involve an arbitrary group of nodes (hyperedges). Thanks to their ability to capture more complex scenarios than graphs, hypergraphs are being increasingly used in many disciplines, like in social sciences [35]. We based our research initiative on an extended version of the hypergraph structure that includes time information about contacts among individuals. As for the case of time-varying graphs (*TVG*), we named it *time-varying hypergraph (TVH)* or temporal hypergraphs. Our approach is thus focused on allowing scientists to easily model in their ABM models direct and indirect contacts, by exploiting the design power of TVHs. The major contributions of this paper are summarized as follows.

- We formally define TVHs. For our literature review, we present a novel definition of hypergraphs including time information on hyperedges, and a suitable to model contact-networks.

- We introduce direct and indirect interactions between agents to study the epidemic spreading through a TVH contact-network. We assemble these aspects into an ABM design-methodology.

- We propose an epidemic diffusion algorithm on TVHs that considers direct and indirect contagion pathways and enables to reduce the over-estimation of an epidemic propagation analysis.

- We developed the SIS compartmental equation-model into an ABM that exploits our methodology to simulate interactions between agents and locations. We tested and analyzed this model with real-world data from *Foursquare*. We further compared the effect of direct and indirect contacts, and in particular, we showed the importance of time in diseases diffusion.

The paper is organized as follows. In Section 2, we define temporal hypergraphs both for modeling population mobility patterns and to explore epidemic dynamics via direct and indirect contacts. Section 3 reviews some relevant literature about epidemic analysis. Our ABM design-methodology and the epidemic diffusion algorithm are introduced in Section 4. Section 5 presents our experiments of an ABM that uses the SIS model to propagate infections in an agents population defined by real OSNs data. Finally, Section 6 details the conclusion and future works.

2 PRELIMINARIES

Propagation of contagious diseases is a complex dynamic process that holds abounding human behavior aspects. Thus, to correctly resemble real diffusion dynamics, modeling tools with a high level of expressiveness are required. By employing a *TVH* to describe a contact network, we are able to minimize the effect of *time* and the presence of only *direct contacts*. Not considering these two features would mean over-estimating the whole disease propagation as a link between two agents will exist even though they have been in the same place but far in time. Many studies have analyzed how the timing feature of contacts is crucial in epidemics [22]. *TVGs* [7, 11] are a variant of the graph model, where a link between two nodes is valid only for a given time interval. A contact-network can be easily extended to include the time dimension by using TVG. However, a TVG is not a definitive solution as the epidemic spreading process may still be over-estimated for particular diseases. Individuals that are in the same place at the same time come in contact with each

other and, as a result, there will be a link per each connection. In this case, the information that ties a group of persons at the same time in a particular geo-location is lost. To further enrich the TVH model using important social dynamics, we include this approach in an ABM simulation.

Agent-based models. An ABM is a class of computational models that provides a bottom-up design approach to define a complex system. As presented by Tracy et al in [27], ABMs are widely adopted in epidemiology science. Frequently, they are used to integrate GIS to simulate the spread of an epidemic in a particular environment, as a result of individuals interactions generated by their mobility over a geographical space [18]. ABMs are also suitable for simulating the interactions of autonomous agents, and they can describe a complex system at a micro-scale level [8]. By adopting an ABM in this context, we are able to capture several important aspects of the epidemic dynamics, such as real human behaviors and complex interactions. For instance, we need to simulate both the indirect-contagion process among agents and environments, and a direct-contagion process to accurately reproduce the spreading phenomenon in the transmission of an airborne disease. The duration of a contact is another crucial aspect in these processes where persons frequently change their habits according to socials events and places popularity, as well as for pathogens proprieties that exhibit different infection times for direct/indirect contacts. Generally speaking, many social and information contagions may be cast in the epidemic methapor [23]. Also in these cases, ABMs are a precious tool to examine influence diffusion and opinion formation in OSNs [10, 13, 20, 29, 31].

Epidemic compartmental models on networks. Modeling a disease transmission introduces a trade-off between a high level of details and computability. As described by Brauer in [5], there are many mathematical models for studying and analyzing diseases spreading. *Compartmental* is a class of models in which the study of the transmission and population is divided into compartments and assumptions on the nature and time rate of the spreading between compartments themselves are made. Popular compartmental models - appropriate for most diseases transmitted by contact - are *SIR* and *SIS*. In these models, the population is studied by partitioning it into three classes labeled S(usceptible), I(nfected), and R(ecoverd). In the SIR model, an individual belongs to one of these possible classes, corresponding to an individual that is susceptible, infected or recovered to the disease at a particular time. If the disease confers no immunity against infection, individuals are not able to move their status to recovered, and they will come back in the susceptible class; in such cases, SIS can be adopted. The transmission patterns of epidemics are defined by the pathogen proprieties, such as its contagiousness, the length of its infectious period, and its severity; on the other hand, these patterns are also defined by the network structure within the population it is affecting, as described by Easley et al. in [11]. The opportunities for a disease to spread from one individual to another is given by a *contact network*, defining the relationships between agents in an ABM model. A contact network can be modeled by a graph. Formally, $\mathcal{G} = (V, E)$, where V is the set of individuals and E is the set of binary relationships between two agents, which indicates whether two agents have come in contact in a particular interval of time.

Location-based social networks (LBSNs). LSBNs are OSNs in which a geo-location dimension is added on top of the social structure. As described by Zheng in [34], a LBSN allows users to share their locations, add geo-tags to their media, define a geographical point of interest, and share their trajectory in travel/sports experience. In this work, we are particularly interested in LBSNs where users can share their position at a particular time. This kind of social interaction is named *check-in in a place* and corresponds to the event that a user v is in a location ℓ at time t. For our purposes, a LBSN is formally defined as a graph structure $\mathcal{G} = (V, E)$, where V is the set of agents and E is the set of edges that connects two agents if they have been somewhere together.

2.1 Time-Varying Hypergraphs

Hypergraphs. Modeling a LBSN using the graph structure is not enough expressive to easily simulate direct and indirect contacts among individuals. Furthermore, a graph does not consider the time a contact happens, feature proved to be a crucial aspect in epidemic dynamics. To correctly model an epidemic propagation in a many-to-many fashion and capturing that individuals, moving through different locations, form a community in a given time and space, we adopted *hypergraphs*. A hypergraph is a mathematical model that generalizes the well-known graph model. Formally, it is an ordered pair $\mathcal{H} = (V, E)$ where V is a set of vertices and E is a set of edges. Each edge is a non-empty subset of vertices; that is, $E \subseteq 2^V \setminus \{\emptyset\}$, where 2^V is the power set of V. We will refer to $n = |V|$ and $m = |E|$ for referring to the size of the vertex set and the edge set, respectively.

Time-varying hypergraphs. To better mimic an epidemic spreading, we included the time dimension within a contact network. For this reason, we extended the definition of *TVGs*, presented by Casteigts et al. in [7], to hypergraphs. Casteigts et al. define a TVG considering a set of entities V (or vertices), and a set of relation E between vertices. They further consider an alphabet L that defines a particular propriety of the relationship between the entities, such as the intensity of a relation in a social network or a location where two individuals are placed. Accordingly, E is defined as $E \subseteq V \times V \times L$. A TVG model evolves over time; therefore, relationships take place over a time span $\mathcal{T} \subseteq \mathbb{T}$. For our purposes, we considered a simpler version of the model, and we adapted it to hypergraphs.

Definition 2.1. A **TVH** is a hypergraph $\mathcal{H} = (V, E, \mathcal{T}, \rho)$, where $\rho: E \times \mathcal{T} \to \{0, 1\}$ is an existing function, indicating if a given hyperedge is accessible in a given time span.

For each $t \in \mathcal{T}$, we refer to the hypergraph $\mathcal{H}_t = (V, E_t)$ as the hypergraph corresponding to a particular time t (i.e., $E_t = \{e \in E : \rho(e, t) = 1\}$). The set of accessible hyperedges in a time span t is $\bigcup(e \in E : \rho(e, t) = 1)$. Figure 1 presents an instance of a TVH (on the left) compared to its corresponding clique-graph or two-section graph (on the right). It illustrates a trivial TVH made up by 5 individuals (nodes), $V = \{a, b, c, d, e\}$, and 3 geographical locations (hyperedges), $E = \{P_0, P_1, P_2\}$. Each hyperedge is labeled with its corresponding availability time span $t = [t_s, t_e)$, $\mathcal{T} = \{[1, 2), [1, 3), [3, 5), [4, 5)\}$. As described by Bretto in [6], the two-section view of $\mathcal{H}$, denoted with $[\mathcal{H}]_2$, is a graph whose vertices are the vertices of $\mathcal{H}$, and where two vertices form an edge if they are in the same hyperedge. It is worth noting that the $[\mathcal{H}]_2$ representation
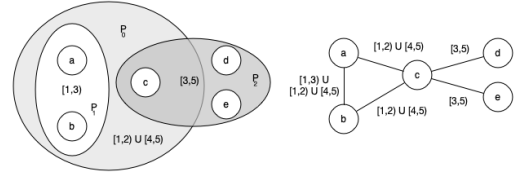

Figure 1: A simple TVH $\mathcal{H}$ (left) and its clique representation $\mathcal{G} = [\mathcal{H}]_2$ (right). Each hyperedge/edge is labeled with the corresponding availability time span.

introduces a loss of information in the contact network. For instance, it is not possible to recognize which is the time span when the individuals a and b were both in the venue P_0 or in the venue P_1.

Definition 2.2. **A TVH for an epidemic diffusion** is a hypergraph $\mathcal{H} = (V, E, \mathcal{T}, \omega)$, where

- V is the set of n vertices a (users/agents);
- E is the set of m hyperedges ℓ (geo-places or locations);
- $\mathcal{T}$ is the lifetime of the system;
- $\omega: V \times E \times \mathcal{T} \to \{0, ct_{v,\ell}\}$ is a function mapping whether a given vertex v has visited the location e in a given time span t. The value $ct_{v,\ell}$ is the last check-in time of v in ℓ.

The ω function only keeps track of the last check-in time $ct_{v,\ell}$ for a vertex v in a given location ℓ during a time span. If v has not checked-in in the same location ℓ or in another location ℓ' in a next time span t', the value of $ct_{v,\ell}$ is still considered a valid check-in time in the current timestamp t. We store the last check-in time $ct_{v,\ell}$ of an vertex v in a location ℓ as the weight of v in the hyperedge representing ℓ. This modeling strategy allows us to easily simulate direct and indirect contagion processes over time: for each time interval, we can effortlessly know in which place the user is. The simulation time is split into fixed-width intervals of length Δ. Given a total sampling time Φ of a LBSN, the total number of time intervals considered is $|\mathcal{T}| = \lceil \frac{\Phi}{\Delta} \rceil$.

3 RELATED WORK

A review of epidemic processes in complex networks is given by Satorass et al. [23]. After the analysis of the fundamental results in classical epidemic modeling and the characterization of complex networks, they discuss different methodologies to understand the dynamics of contagion processes in the case of heterogeneous connectivity patterns. In particular, they highlight the importance of unveiling the interplay between epidemic processes and real-world networks, whose structure is inherently dynamic, heavy-tailed and community-based. In this context, taking into account the time dimension is fundamental when studying contacts among individuals [16, 26, 28] and immunization strategies [19, 24].

Accurately modeling interactions between persons and environments and keeping into account not only their contact network but also the locations they have visited is another key-point towards the understanding of epidemic dynamics. Jindal et al. [18] study the evolution of mosquito-borne diseases through a generalized agent-based model that integrates both geographic information and census data to account for the spatial movement of infections, and

climate data to capture the temporal nature of an epidemic. Singh et al. [25] address the problem of modeling realist agent behaviors for informing public health policies by developing a methodology to create and calibrate and agent decision making problem for a multi-agent simulation using survey data. Pathogen transmission can involve direct and/or indirect pathways, as illustrated by Cortez et al. in [9]. They describe how the transmission pathway can result in quantitatively different epidemiological dynamics and to what extent those differences can be used to identify the transmission pathway from population-level time series.

Many works model the contact network as a graph and run the SIS epidemic model on it. For instance, Peng et al [30] study the SIS epidemic model with vaccination in a dynamical contact network of moving individuals. Others exploit hypergraphs as a more accurate structure to model contact networks. Bodò et al. [4] highlight how a real model of an epidemic outbreak has to take into account two factors: community structure and infection pressure. First, a person may visit different places during the day - home, workplace, etc. Even though it is known that the community structure has a strong impact on the spread of the epidemic, it is important to consider that the community itself is built up from smaller units. This idea is translated into practice using different contagion probabilities according to the place. Second, the probability that a susceptible individual becomes infected in a unit is not proportional to the number of infected individuals. To both model the community structure and the non-linear dependence of the infection pressure, they developed the theory of epidemic propagation on hypergraphs, where each node is an individual and each hyperedge is a unit. Furthermore, they discuss how a graph is not a well-suited structure to capture this kind of relationship. They show that using a non-linear function to model the infection pressure is crucial to do not over-estimate the epidemic propagation.

We developed our epidemic model starting from the model proposed by Bodò et al. [4], but we enhanced and generalized it by including direct and indirect contacts exploiting TVHs (that extend the notion of temporal networks) and combining them into a design-methodology for ABM.

4 EPIDEMIC DYNAMICS

4.1 An Agent-based Modeling Approach

An ABM is able to catch many aspects of human behaviors by modeling individuals nature, and by including environments properties according to assumptions given by several disciplines, such as psychology, social and natural science. Our work is based on the intuition of exploiting TVH to simulate different kinds of interactions between agents (direct contact) and environments (indirect contact). By using our approach, the user can thus design an ABM capable of handle both the time contacts happen and the infections transmitted through the environment. As the spread of an epidemic in a population is governed by social and structural dynamics, combining a TVH vision of interactions into an ABM allows facing both aspects in a single design-methodology. At each simulation step, the researcher is naturally aware of who has been in contact with whom, and consequently the probability the epidemic will spread towards different locations. This approach also permits to model health interventions at different scales. An example is answering

whether is better vaccinating people or carrying out health policies in specific locations.

4.2 The Design-Methodology

We based the idea of our ABM design-methodology on the assumption that an epidemic process is governed by humans interactions, and that it is regularized by two spreading policies: *direct* and *indirect* contagions (or contacts) between individuals and environments. A direct contact implies a pairwise interaction between two individuals in the same community or location. On the other hand, an indirect contagion embodies the interchanges that may happen between agents and locations. These two types of contacts are a natural consequence of the daily activities and commuting routes of each person. For instance, when an agent moves from its home to its workplace, it may be either infected by touching some furniture or simply breathing contaminated air (indirect interaction), or by a face to face talk with another agent (direct interaction).

In this work, we propose a diffusion algorithm whose spreading process is designed in a discrete-events fashion. We exploit the TVH structure to discover whether direct and/or indirect interactions may happen. The time Φ is divided into discrete time intervals: during each interval - corresponding to a simulation step - agents are simulated according to their scheduling policy. Then, our diffusion algorithm is performed. Specifically, time is discretized using a given value Δ, corresponding to the sampling time-span length of the TVH. In the case of a LBSN, this process is translated in a given number of independent agents moving during the time over a set of geographical location L. At each time interval t, each user may freely move in another location. As a consequence, the epidemic has the chance of spreading from one location to another. Simultaneously, the outbreak may still spread across agents located in the same place at a particular interval of time. When an agent moves to a location ℓ_i, it may thus be infected by direct or indirect contacts. While direct contaminations require the co-presence of agents, indirect contacts happen between agents and the environment and the co-presence is not required.

Formal definitions. In the following, we provide a formal definition of the concepts adopted throughout the rest of the paper.

- Φ is the time-span of the LBSN data sampling.

- $\mathcal{T}$ is the set of time intervals describing the evolution of the relationships between agents and geo-locations.

- Δ is a real value (minutes, hours or days) corresponding to the time discretization parameter. It further refers to the time-span when *indirect* contagions may happen.

- δ is real - small - value (milliseconds, seconds or minutes) defining when direct infections may take place. A direct contagion is established if two agents arrive in the same location within a time difference less than δ.

- t is the current simulation time interval $t \in \mathcal{T}$.

- $\mathcal{H} = (V, E, \mathcal{T}, \omega)$ is the TVH representing the LBSN (see def. 2.2).

- Γ_t and N_t define the neighborhood functions of an agent $a \in V$ in a given simulation time t. Specifically,

$$\Gamma_t(a) = \{\ell \in E : \omega(a, \ell, t) = 1\},$$

is the set of locations visited by a during the interval t.

$N_t(a)$ is the set of neighbors of a during the simulation time t, which corresponds to the agents that visited at least one of the locations visited by a. Formally,

$$N_t(a) = \bigcup_{\ell \in \Gamma_t(a)} V_t(\ell),$$

where $V_t(\ell)$ denotes the set of agents that visited the location ℓ during the interval t.

- $\Upsilon(a, \ell)$ is a time function which provides the last check-in time of the agent a in the venue ℓ. In other words, it returns the weight of a in ℓ in the hypergraph $\mathcal{H}$.

- $T_t(a)$ and $T_t(\ell)$ denote the infection state of an agent or a location at a given time t, respectively.

- $X_t(a, b)$ is a direct contact function. Given two agents a and b, it returns 1 if they have a direct contact in the time span t, 0 otherwise. Formally,

$$X_t(a, b) = \begin{cases} 1, & \text{if } \exists \ell \in \Gamma_t(a) \cap \Gamma_t(b) \text{ AND } |\Upsilon(a, \ell) - \Upsilon(b, \ell)| < \delta \\ 0, & \text{otherwise.} \end{cases}$$

In the following, we will remove the subscript t when the simulation time is clear from the context.

The diffusion algorithm. We considered the simulation proceeding in $|\mathcal{T}|$ discrete steps. At each simulation step i, every agent independently runs its *step function* and updates its internal state, which will be effective in the next simulation phase $i+1$. It is worth noting that computing whether a direct contact takes place in a fixed time interval $t \in \mathcal{T}$, information from the previous $(t-1)$ and next $(t+1)$ intervals is required as well. Figure 2 details a snapshot of an example hypergraph for a a time interval i (about 3 hours), containing 11 agents, and 4 locations (hyperedges). In this example, dotted lines represent direct relationships. Hyperedges (drawn with an oval shape) define indirect contacts. To easily compute direct contagions, we stored the last check-in time of an agent as the weight of that agent in the hyperedge corresponding to the location it checked-in (see Section 2). During each time step, our diffusion process proceeds in three contagious phases.

1. *Agent-to-Environment.* The first phase simulates the *environment infectiveness*. For all non contaminated locations, (i.e., $\ell \in E : T(\ell) = 0$), we compute the number of infected agents that have visited that location:

$$I^e(\ell) = \sum_{a \in V(\ell)} T(a).$$

This value is then used to update the contagiousness level of ℓ as expressed by the following:

$$T(\ell) = \begin{cases} 1, & \text{infected according to the value} f^e(I^e(\ell)) \\ 0, & \text{not infected,} \end{cases}$$

where $f^e()$ is a non-linear function, typically adopted to govern the behavior of the epidemic outbreak over the hyperedges [4]. In our experiments (see Section 5.2), we considered the following regularization function:

$$f^e(x) = \begin{cases} x, & \text{if } 0 \le x \le c \\ c, & \text{if } x > c, \end{cases}$$
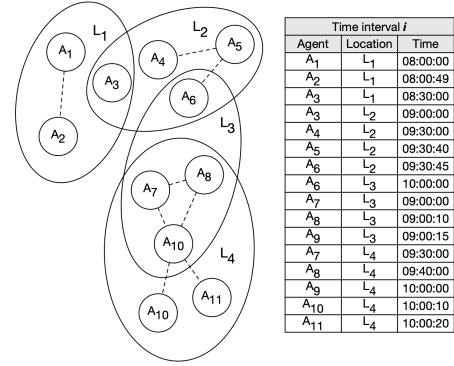
where c is a constant value given as parameter.

Figure 2: An example of temporal direct/indirect contagion (contacts) modeled with a hypergraph.

2. *Agent-to-Agent.* The second phase simulates the *direct* propagation process. For all non infected agents (i.e., $a \in V : T(a) = 0$), the total number of infected neighbors is computed. Formally,

$$I^d(a) = \sum_{b \in N(a)} T(b)X(a, b).$$

This value is then used to update the infection state of a, as

$$T(a) = \begin{cases} 1, & \text{infected according to the value } I^d(a) \\ 0, & \text{not infected.} \end{cases}$$

3. *Environment-to-Agent.* The third and last phase simulate the *indirect* propagation process. For all non infected agents, (i.e., $a \in V : T(a) = 0$), we compute the number of infected locations visited. Formally,

$$I^i(a) = \sum_{\ell \in \Gamma(a)} T(\ell).$$

This value is then used to update the infection state of a, as

$$T(a) = \begin{cases} 1, & \text{infected according to the value } I^i(a) \\ 0, & \text{not infected.} \end{cases}$$

It worth highlighting that the discussed diffusion algorithm can be easily extended by implementing additional phases (before or after the algorithm execution) describing supplementary agent behaviors. Therefore, the algorithm is suitable for other typologies of ABMs.

5 EXPERIMENTS AND ANALYSIS

To evaluate the proposed approach and to what extent the described design-methodology is capable of resembling the epidemic spreading, we developed an ABM simulation running the SIS model over a population of individuals. In particular, we tested the model expressiveness in distinguishing the epidemic diffusion via direct and indirect contagion pathways (see Section 5.2). We further analyzed the effect of time when modeling contacts by varying the values of Δ and δ. This experiment allowed us to explore the dynamic of the epidemic by increasing/decreasing the amount of direct/indirect contacts (see Section 5.3). We experimented our model using real-data from the social network Foursquare.
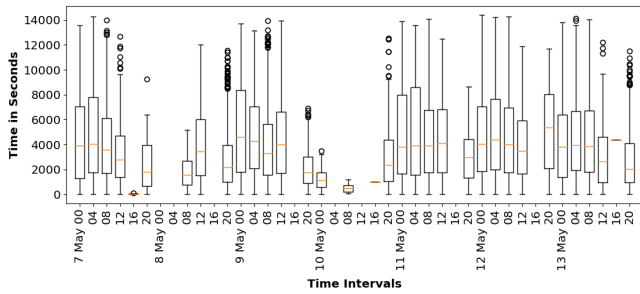
Figure 3: Time difference distribution of check-ins within the same place in 7 **days and** $\Delta = 4$ **hours. Foursquare dataset.**

5.1 The Foursquare Data set

In the ABM model we developed, each agent moves between geo-locations over time and comes in contact, via direct or indirect pathways, with other agents and with different environments (geo-locations). To model individuals (agents) mobility patterns, we used the Foursquare social network data [33] originated from the city of Tokyo and crawled from 12 April 2012 to 16 February 2013. The dataset contains $573, 703$ check-ins, of $2, 293$ users and $61, 858$ locations (such as restaurants, plaza, cinema, sports and so on).

To estimate the value of the parameter δ, i.e., the time window within which a direct contagion may occur, we examined the time difference of check-ins happening in the same place over a period 7 days. We fixed $\Delta = 4$ hours and analyzed the most crowded week in the 10 months available ($7 - 14$, May, 2012). Figure 3 presents the distribution of the time difference evaluated pairwise over check-ins within the same time interval. As shown, the data contain some time windows where no check-ins are available (meaning that the epidemic has no-probability to propagate). Generally, the majority of the intervals exhibit the same trend with a median value of about 1 hour. Having estimated $\delta = 1$ hour, we inspected the number of direct contacts within each place using this value. Figure 4 shows the distribution of direct contacts per location within each time interval. Direct contacts are evenly distributed over the whole week, with a median value of about $1, 750$. Finally, we considered the distribution of indirect contacts. Figure 5 reveals how many different places users have visited within each time interval. The plot shows that users tend to visit (or, at least, checking-in) just one venue or a very small number of them. Several outliers visit nearly 20 venues, but no more than 30. As described in several work [2, 14, 32], this may represent a typical kind of power-law behavior, where few users post a lot while the majority posts few actions.

5.2 Experiment Formulation

In a typical SIS model, the infection and recovery states are ruled by a Poisson process. Thus, either a susceptible individual or location becomes (directly or indirectly) infected with a probability $1 - e^{-\beta_* f(n)}$. Here, β_* denotes the infection rate per-contact (considering either direct or indirect contacts), n is the number of infected entities (either individuals or locations), and $f()$ is a non-linear function used to bound the infection pressure for large value of n, as described in Section 3. Similarly, a contaminated individual (location) recovers with probability $1 - e^{-\gamma_*}$, where γ_* denotes
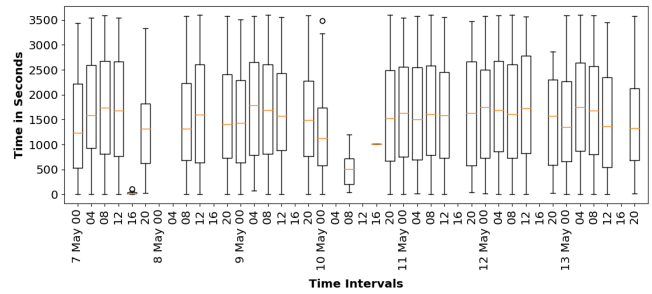

Figure 4: Number of *direct* contacts for each person in 7 **days,** $\Delta = 4$ **hours, and** $\delta = 1$ **hour. Foursquare dataset.**
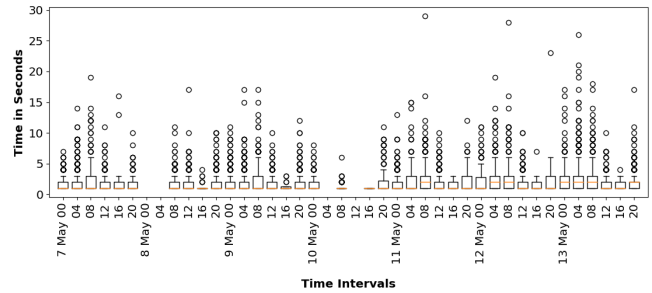

Figure 5: Number of *indirect* contacts for each person in 7 **days, and** $\Delta = 4$ **hours. Foursquare dataset.**

the recovery rate for agents (environments). We embedded the described SIS variant inside our diffusion algorithm. The parameters in input to our model are listed in Table 1. To govern the epidemic propagation via direct and indirect pathways, we used three values for β and two values for γ. The constant c bounds the number of considered contacts when evaluating $f()$.

Listing 1 shows the pseudo-code of the ABM simulation process, which exploits TVHs to consider both direct and indirect contagions between agents and locations (see Section 4.2) when running the SIS model. At line 1, a random number generator is created. Then, the algorithm proceeds for a number of steps equal to the number of intervals in $\mathcal{T}$. In each step, the algorithm computes the hypergraph corresponding to the current time t using the function ξ (line 3). Right after, it runs the three epidemic diffusion phases. In the first phase, *Agent-to-Environment* (lines $4 - 10$), the infection status of every location is updated. In the second phase, *Agent-to-Agent* (lines $11 - 14$), the epidemic is propagated through the network using direct contacts and updating the infection status of the agents. Finally, in the last phase, *Environment-to-Agent* (lines $15 - 21$), the epidemic is propagated using indirect contacts, and the infection state of the agents is updated again. It is worth noticing that the recovery process is computed only a single time for both agents and locations.

5.3 Direct *vs* Indirect Contagions

According to our design-methodology and the previously described SIS model, we developed an epidemic diffusion ABM model exploiting the *Julia* language and the *SimpleHypergraph.jl* [1] library. We

Parameter	Description
β_d	Probability that an agent a_i is infected by another agent a_j via a direct-contact in *Agent-to-Agent*
β_i	Probability that an agent a is infected via an indirect-contact due to a location ℓ in *Environment-to-Agent*
β_e	Probability that a location ℓ is infected by an agent in *Agent-to-Environment*
γ_a	Probability that an agent a spontaneously recovers
γ_e	Probability that a location ℓ is sanitized
c	Number of contact in *Agent-to-Environment*

Table 1: Diffusion algorithm parameters.

tested to what extent our proposed model is able to distinguish the epidemic spreading according to the direct and indirect contagion pathways. In this experiment, we run our ABM simulation according to two different parameters configuration of the SIS model. In the first configuration (*Low*), we used $\beta_d = 0.2$, $\beta_i = 0.1$, $\beta_e = 0.06$, $\gamma_e = 0.06$, $\gamma_a = 0.1$, and $c = 5$. In the second configuration, (*High*), we used $\beta_d = 0.8$, $\beta_i = 0.4$, $\beta_e = 0.26$, $\gamma_e = 0.06$, $\gamma_a = 0.1$, and $c = 5$. Basically, we selected the value β_d and computed $\beta_i = \frac{\beta_d}{2}$, and $\beta_e = \frac{\beta_d}{4}$. We chose the value of γ according to typical values of the SIS model. We performed three kinds of experiments by varying the parameter setting of the two configurations to exploit either direct contacts, indirect contacts, or both. To have reliable results, we run each experiment 10 times and we present their average values. According to previous work [4], we set $c = 5$. We fixed the value of $\Delta = 4$ hours and $\delta = 1$ minutes, selecting these values according to a realistic spreading policy for airborne disease transmission. This setting suggests that indirect contacts may still happen if the pathogen is deposited in the environment 4 hours before the indirect contact, and direct contact can happen if two persons stay in the same time of one minute. In each test, we initiated the state of 80% of the agents to *susceptible* and the remaining 20% to *infected*. We used the most crowded month of the Foursquare data set.

Figure 6 illustrates how the number of infected agents evolves according to the SIS spreading policies defined above. The first interesting result regards the percentage of infected nodes using both direct and indirect contacts. In both parameter configurations (*low* and *high*), this value is not statistically different from the series obtained considering only indirect contagion pathways. This finding suggests that indirect contacts have much more importance in spreading the epidemic and, consequently, they should be investigated when studying epidemic diffusion processes that do not necessarily involve a direct contact between two agents. To further support this argument, we have to consider what happens when examining only direct contacts. In both parameter configurations, the epidemic diffusion will drop out, and eventually, all nodes will be susceptible again. This result highlights how the epidemic dynamic - for this study - is extremely governed by the temporal network evolution, and, even more, that the spreading opportunity is strongly related to how human-mobility patterns are modeled.

5.4 Modeling the Effect of Time

The epidemic diffusion process and its impact on the population are strictly related to the pathogens life cycle, and their survival time in

---

Algorithm 1 Epidemic Diffusion Simulation

1: r ▷ a random number generator $\in [0, 1]$.
2: **for** $t \in \mathcal{T}$ **do**
3:     $\mathcal{H} \leftarrow \xi(t)$
4:     **for** $\ell \in E$ **do**   ▷ Agent-to-Environment.
5:         **if** $T_t(\ell) == 0$ **then**
6:             **if** $r_{next} < 1 - e^{-\beta_e f^e(I^e(\ell))}$ **then**
7:                 $T_{t+1}(\ell) = 1$
8:             **else if** $r_{next} < 1 - e^{-\gamma_e}$ **then**
9:                 $T_{t+1}(\ell) = 0$
10:     **for** $a \in V$ **do**   ▷ Agent-to-Agent.
11:         **if** $T_t(a) == 0$ **then**
12:             **if** $r_{next} < 1 - e^{-\beta_d I^d(\ell)}$ **then**
13:                 $T_{t+1}(a) = 1$
14:     **for** $a \in V$ **do**   ▷ Environment-to-Agent.
15:         **if** $T_t(a) == 0$ **then**
16:             **if** $r_{next} < 1 - e^{-\beta_i I^i(\ell)}$ **then**
17:                 $T_{t+1}(a) = 1$
18:             **else if** $r_{next} < 1 - e^{-\gamma_a}$ **then**
19:                 $T_{t+1}(a) = 0$

---

the environment. In this experiment, we investigated how different modeling of time intervals - within which a direct (δ) or indirect (Δ) contagion may happen - impact on the epidemic spreading in the network. It is worth noting that a smaller value of δ corresponds to higher accuracy in computing direct contacts. Similarly, a smaller value of Δ correlates to higher accuracy in computing indirect contacts. We ranged the value of δ from 1 to 60 minutes, while the value of Δ from 4 to 24 hours. We configured the model infection parameters as in the previous experiment (described in Section 5.3), applying the same two SIS configurations (*Low* and *High*) and starting the simulation with a 20% of infected agents. As before, we used the most crowded month of the Foursquare data set.

Figure 7 details how the percentage of infected nodes change according to the different time intervals used to model contacts. The top-left corner of the image shows the most accurate configuration ($\delta = 1$ minute and $\Delta = 4$ hours), the bottom right corner the less accurate. As highlighted in the figure, direct contact pathways become more relevant for the epidemic diffusion as the value of δ increments. However, initializing δ with a value too large would
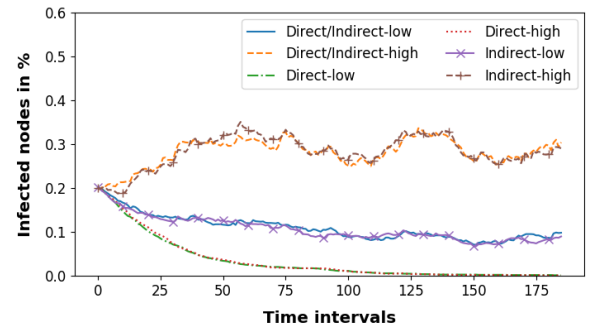

Figure 6: Direct *vs* Indirect contacts within a period of 7 days, setting $\Delta = 4$ hours, and $\delta = 1$ minute.
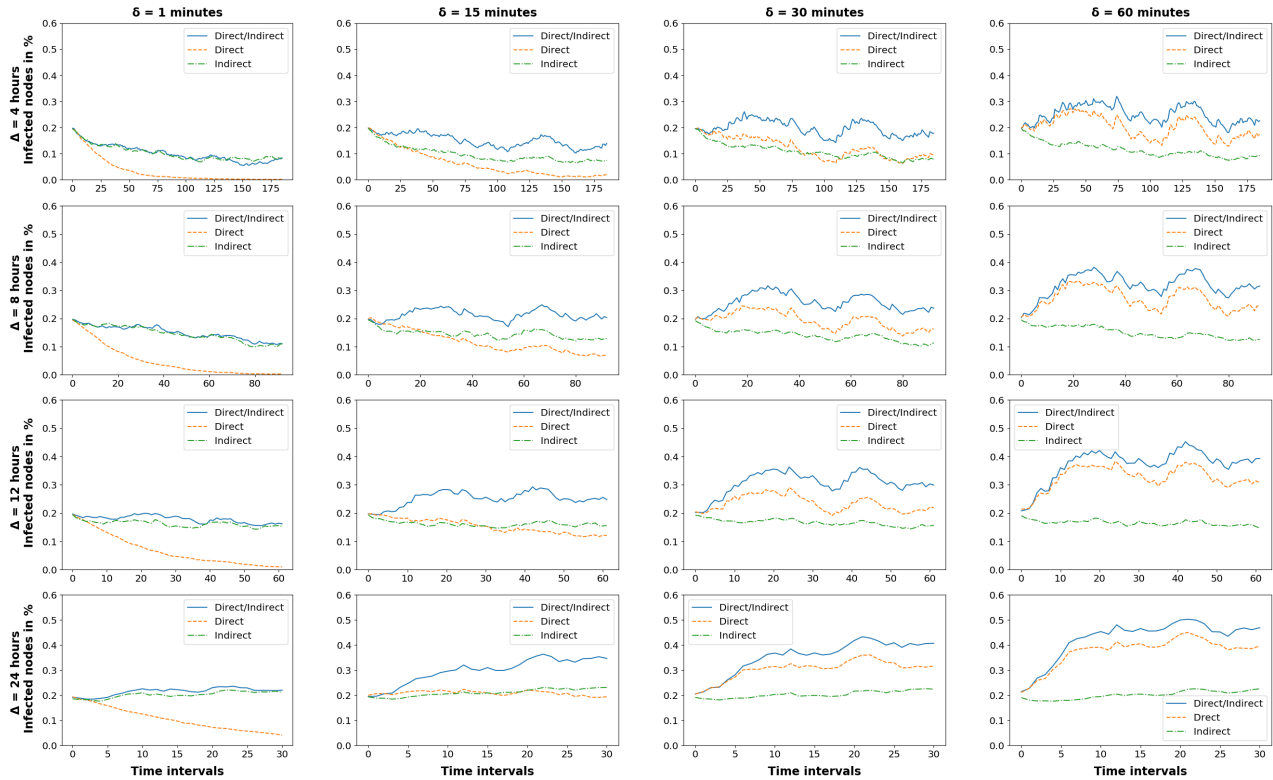
Figure 7: Epidemic evolution over 1 month, varying the length of Δ (vertically) and of δ (horizontally).

mean computing too many direct contacts comparing to a real-life system, and, therefore, over-estimating the epidemic propagation. In practice, the contact network would be modeled as a graph rather than a hypergraph. On the other hand, increasing the value of Δ implies computing too many indirect contacts, and assuming a long pathogens life. Determining the proper values of δ and Δ according to given disease properties has fundamental importance to correctly estimate the epidemic propagation using our design-methodology.

6 CONCLUSION AND FUTURE WORKS

In this work, we proposed a new approach to model an epidemic propagation by adopting *TVHs* to design an ABM simulation able to exploit direct and indirect contacts between individuals and environments. We introduced the new notion of TVH to describe a complex population, where individuals move through different locations over time, and where the TVH structure defines interactions between agents and environments evolving over time in an ABM simulation. We suggested a design-methodology which integrates the idea of infection spreading via direct and indirect contacts within an epidemic ABM. We also presented a diffusion algorithm suitable for our methodology, which allows us to reduce over-estimation in the epidemic dynamics. Furthermore, we developed - in *Julia* language - an epidemic ABM on top of our design-methodology, and we used real-world data from the Foursquare social network to model agents mobility patterns over a set of geographical locations. We performed several tests to validate the

methodology ability to both distinguishing epidemics spreading via direct and indirect contagion, and estimating the impact of time when modeling agent-to-agent and agent-to-environment interactions. We proved that we could improve the accuracy of the epidemic diffusion estimation by exploiting TVHs. Additionally, we demonstrated the crucial importance of correctly model the time interval of contagion to compute direct and indirect epidemic diffusion pathways. The major limitation of our approach lies in the intrinsic complexity of dealing with hypergraphs, as ad-hoc algorithms and more computational resources are needed to execute and validate ABM simulations.

We are currently working towards the development of immunization and vaccination (for agents), and sanitization (for environments) strategies considering direct and indirect contacts. Future work will focus on developing other classical compartmental models using our framework, and experimenting with the epidemic ABM model on other real-world data, such as a more accurate mobility pattern data as the users mobility trajectory in the Microsoft GeoLife data [1].

---
[1]http://bit.ly/geolife

REFERENCES

[1] A. Antelmi, G. Cordasco, B. Kamiński, P. Prałat, V. Scarano, C. Spagnuolo, and P. Szufel. 2019. SimpleHypergraphs.jl—Novel Software Framework for Modelling and Analysis of Hypergraphs. In *Algorithms and Models for the Web Graph*, K. Avrachenkov, P. Prałat, and N. Ye (Eds.). Springer International Publishing, Cham, 115–129. https://doi.org/10.1007/978-3-030-25070-6_9

[2] A. Antelmi, D. Malandrino, and V. Scarano. 2019. Characterizing the Behavioral Evolution of Twitter Users and The Truth Behind the 90-9-1 Rule. In *Companion Proceedings of The 2019 World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 1035–1038. https://doi.org/10.1145/3308560.3316705

[3] S. Bansal, B. T. Grenfell, and L. A. Meyers. 2007. When individual behaviour matters: homogeneous and network models in epidemiology. *J. R. Soc. Interface* 4, 16 (2007), 879–891. https://doi.org/10.1098/rsif.2007.1100

[4] Á. Bodó, G. Y. Katona, and Pé. L. Simon. 2016. SIS Epidemic Propagation on Hypergraphs. *Bulletin of Mathematical Biology* 78 (2016), 713–735. Issue 4. https://doi.org/10.1007/s11538-016-0158-0

[5] F. Brauer. 2017. Mathematical epidemiology: Past, present, and future. *Infectious Disease Modelling* 2, 2 (2017), 113–127. https://doi.org/10.1016/j.idm.2017.02.001

[6] A Bretto. 2013. *Hypergraph Theory: An Introduction.* Springer International Publishing, New York, NY (USA). https://doi.org/10.1007/978-3-319-00080-0

[7] A. Casteigts, P. Flocchini, W. Quattrociocchi, and N. Santoro. 2011. Time-Varying Graphs and Dynamic Networks. In *Ad-hoc, Mobile, and Wireless Networks*, H. Frey, X. Li, and S. Ruehrup (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 346–359. https://doi.org/10.1007/978-3-642-22450-8_27

[8] G. Cordasco, R. De Chiara, F. Raia, V. Scarano, C. Spagnuolo, and L. Vicidomini. 2013. Designing computational steering facilities for distributed agent based simulations. *SIGSIM-PADS 2013 - Proceedings of the 2013 ACM SIGSIM Principles of Advanced Discrete Simulation* (2013), 385–390. https://doi.org/10.1145/2486092.2486147

[9] M. H. Cortez and J. S. Weitz. 2013. Distinguishing between Indirect and Direct Modes of Transmission Using Epidemiological Time Series. *The American Naturalist* 181, 2 (2013), E43–E52. https://doi.org/10.1086/668826 arXiv:https://doi.org/10.1086/668826

[10] A. De, S. Bhattacharya, and N. Ganguly. 2018. Shaping Opinion Dynamics in Social Networks. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '18)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1336–1344.

[11] D. Easley and J. Kleinberg. 2010. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World.* Cambridge University Press, Cambridge (UK). https://doi.org/10.1017/CBO9780511761942

[12] N. Ganguly, T. Krueger, A. Mukherjee, and S. Saha. 2014. Epidemic spreading through direct and indirect interactions. *Phys. Rev. E* 90 (2014), 032808. Issue 3. https://doi.org/10.1103/PhysRevE.90.032808

[13] J. Gao, G. Schoenebeck, and F. Yu. 2019. The Volatility of Weak Ties: Co-Evolution of Selection and Influence in Social Networks. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '19)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 619–627.

[14] W. Gong, E.-P. Lim, and F. Zhu. 2015. Characterizing silent users in social media communities. *Proceedings of the 9th International Conference on Web and Social Media, ICWSM 2015* (2015), 140–149. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10462

[15] R. A. Hammond. 2015. Considerations and Best Practices in Agent-Based Modeling to Inform Policy. In *Committee on the Assessment of Agent-Based Models to Inform Tobacco Product Regulation. Board on Population Health and Public Health Practice; Institute of Medicine*, R. Wallace, A. Geller, and V.A. Ogawa (Eds.). National Academic Press (US), Washington DC, WA, US, Appendix A, 161–193. https://www.ncbi.nlm.nih.gov/books/NBK305917/

[16] P Holme and J. Saramäki. 2012. Temporal networks. *Physics Reports* 519, 3 (2012), 97 – 125. https://doi.org/10.1016/j.physrep.2012.03.001 Temporal Networks.

[17] B. Hunter, E.and Mac Namee and J. D. Kelleher. 2017. A Taxonomy for Agent-Based Models in Human Infectious Disease Epidemiology. *Journal of Artificial Societies and Social Simulation* 20, 3 (2017), 2. https://doi.org/10.18564/jasss.3414

[18] A. Jindal and S. Rao. 2017. Agent-Based Modeling and Simulation of Mosquito-Borne Disease Transmission. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems (AAMAS '17)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 426–435.

[19] S. Lee, L. E. C. Rocha, F. Liljeros, and P. Holme. 2012. Exploiting temporal network structures of human interaction to effectively immunize populations. *PloS one* 7 (2012), e36439 – e36439. Issue 5. https://doi.org/10.1371/journal.pone.0036439 Temporal Networks.

[20] W. Li, Q. Bai, M. Zhang, and T. D. Nguyen. 2018. Modelling Multiple Influences Diffusion in On-Line Social Networks. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '18)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1053–1061.

[21] M. Nadini, K. Sun, E. Ubaldi, M. Starnini, A. Rizzo, and N. Perra. 2018. Epidemic spreading in modular time-varying networks. *Scientific Reports* 8 (2018), 2352–2363. Issue 1. https://doi.org/10.1038/s41598-018-20908-x

[22] P. E. Paré, C. L. Beck, and A. Nedić. 2018. Epidemic Processes Over Time-Varying Networks. *IEEE Transactions on Control of Network Systems* 5, 3 (2018), 1322–1334. https://doi.org/10.1109/TCNS.2017.2706138

[23] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. 2015. Epidemic processes in complex networks. *Rev. Mod. Phys.* 87 (2015), 925–979. Issue 3. https://doi.org/10.1103/RevModPhys.87.925

[24] B. A. Prakash, H. Tong, N. Valler, M. Faloutsos, and C. Faloutsos. 2010. Virus Propagation on Time-Varying Networks: Theory and Immunization Algorithms. In *Machine Learning and Knowledge Discovery in Databases*, J. L. Balcázar, F. Bonchi, A. Gionis, and M. Sebag (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 99–114. https://doi.org/10.1007/978-3-642-15939-8_7

[25] M. Singh, A. Marathe, M. V. Marathe, and S. Swarup. 2018. Behavior Model Calibration for Epidemic Simulations. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '18)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1640–1648.

[26] M. Tizzani, S. Lenti, E. Ubaldi, A. Vezzani, C. Castellano, and R. Burioni. 2018. Epidemic spreading and aging in temporal networks with memory. *Phys. Rev. E* 98 (2018), 062315. Issue 6. https://doi.org/10.1103/PhysRevE.98.062315

[27] M. Tracy, M. Cerdá, and K. M. Keyes. 2018. Agent-Based Modeling in Public Health: Current Applications and Future Directions. *Annual Review of Public Health* 39 (2018), 77–94. https://doi.org/10.1146/annurev-publhealth-040617-014317

[28] C. L. Vestergaard and M. Génois. 2015. Temporal Gillespie Algorithm: Fast Simulation of Contagion Processes on Time-Varying Networks. *PLoS computational biology* 11 (2015), e1004579–e1004579. Issue 10. https://doi.org/10.1371/journal.pcbi.1004579

[29] B. Wilder and Y. Vorobeychik. 2018. Controlling Elections through Social Influence. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '18)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 265–273.

[30] P. Xiao-Long, Z. Ze-Qiong, Y. Junyuan, and J. Zhen. 2019. An SIS epidemic model with vaccination in a dynamical contact network of mobile individuals with heterogeneous spatial constraints. *Comm. in Nonlinear Science and Numerical Simulation* 73 (2019), 52 – 73. https://doi.org/10.1016/j.cnsns.2019.02.004

[31] A. Yadav, H. Chan, A. Xin Jiang, H. Xu, E. Rice, and M. Tambe. 2016. Using Social Networks to Aid Homeless Shelters: Dynamic Influence Maximization under Uncertainty. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems (AAMAS '16)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 740–748.

[32] C. Yang, X. Shi, L. Jie, and J. Han. 2018. I Know You'll Be Back: Interpretable New User Clustering and Churn Prediction on a Mobile Social Application. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)*. Association for Computing Machinery, New York, NY, USA, 914–922. https://doi.org/10.1145/3219819.3219821

[33] D. Yang, D. Zhang, V. W. Zheng, and Z. Yu. 2015. Modeling User Activity Preference by Leveraging User Spatial Temporal Characteristics in LBSNs. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 45, 1 (2015), 129–142. https://doi.org/10.1109/TSMC.2014.2327053

[34] Y. Zheng. 2011. *Location-Based Social Networks: Users.* Springer New York, New York, NY, 243–276. https://doi.org/10.1007/978-1-4614-1629-6_8

[35] J. Zhu, J. Zhu, S. Ghosh, W. Wu, and J. Yuan. 2019. Social Influence Maximization in Hypergraph in Social Networks. *IEEE Transactions on Network Science and Engineering* 6, 4 (2019), 801–811. https://doi.org/10.1109/TNSE.2018.2873759