Agent-Based Modelling of Epidemic Spreading using Social Networks and Human Mobility Patterns

Enrique Frías-Martínez [‡], Graham Williamson [1#], Vanessa Frías-Martínez [‡],
[‡] *Telefónica Research, Madrid – Spain*
[#] *School of Computer Science , University College Dublin – Ireland*
`{efm,graham,vanessa}@tid.es`

Abstract— **The recent adoption of ubiquitous computing technologies has enabled capturing large amounts of human behavioral data. The digital footprints computed from these datasets provide information for the study of social and human dynamics, including social networks and mobility patterns, key elements for the effective modeling of virus spreading. Traditional epidemiologic models are not able to capture the inherent complexity of the spreading process because no individual information is considered. Recently, agent-based modeling have been used as an effective approach to model virus spreding as it captures the inherent individuality and complexity of the process. One of the main limitations of the current approaches is the lack of relevant data to create agents. In this paper we present an agent based system that uses the information extracted from call detail records to model social interactions and mobility patterns of individuals in order to accurately model virus spreading. The approach is applied to study the** 2009 **H1N1 outbreak in Mexico and to evaluate the impact that government calls had on the spreading of the virus. Our simulations indicate that the restrictive mobility implemented by the government reduced the peak of the virus by 10% and postponed it by two days.**

I. INTRODUCTION

Planning for a pandemic (H1N1, influenza, etc.) is one of the main public health priorities of any government. Traditional epidemiological approaches base their solutions on using differential equations that divide the population into subgroups based on socio-economic and demographic characteristics. Although these models fail to capture the complexity and individuality of human behavior, they have been extremely successful in guiding and designing public health policies. The adoption of agent-based modeling (ABM) approaches to simulate pandemics has allowed to capture individual human behavior and its inherent fuzziness by representing every single person (agent) as a software individual. The ABM model assigns to each agent a variety of variables that are considered relevant to model virus spreading such as mobility patterns, social network characteristics, socio-economic status, health status, etc. Nevertheless, ABM approaches suffer from the need of relevant data to create agents than effectively capture human behavior. Typically this data is generated from the census or using surveys [1].

The adoption of ubiquitous computing technologies by very large portions of the population (e.g. GPS devices, ubiquitous cellular networks or geolocated services) has enabled the capture of large scale human behavioral data. These datasets can reveal information that is critical for the accurate modeling of virus spreading such as human mobility patterns or the social network characteristics of each individual [2][1]. This information represents a critical improvement over ABM approaches that borrow information from the census to design agents for virus spreading simulations.

In this paper we present an ABM system to simulate virus spreading with agents that are characterized by their individual mobility patterns and social networks extracted from cell phone records. We also evaluate the ABM system with data collected during the latest Mexican H1N1 outbreak to measure the impact that government calls had on individual user mobility and the effect on the virus spreading.

The remainder of this paper is organized as follows: Section II discusses the related work regarding traditional disease models and ABM simulation environments; Section III presents the infrastructure of a cell phone network and how cell phone records are captured and Section IV-A describes our proposed architecture for ABM modelling. After that, Section V presents a case study that evaluates the impact that government mandates regarding mobility restrictions had in the virus spreading of the 2009 H1N1 outbreak in Mexico. Section VI describes conclusions and future work.

II. RELATED WORK

A. Traditional Epidemic Disease Models

Traditional epidemic disease models are based on the SIR model or some of its variations (SI, SIR, SIS, SEIR, etc.) [3]. These approaches, called *compartmental models*, split the population into compartments that represent the different stages of a disease. The most general approach is the SIR model that typifies the disease progression as follows: (1) S, represents the susceptible (S) portion of the population *i.e.* those yet to be infected; (2) I, represents those that are currently infective (I); and (3) R, represents individuals that have recovered (R) from the disease and no longer take an active part in the disease spread. Other models like the SEIR, add an intermediate stage (E) which represents a latent state in which individuals have been exposed to the disease but are not yet infective, *i.e.* the individuals in this stage have the virus but can not infect others. All these models represent the virus transmission by a set of nonlinear ordinary differential equations (ODEs) that associate a transition rate to the mobility of agents between compartments. These transition

---

[1] Work done while author was an intern at Telefónica Research, Madrid.

rates are used by the models to define a reproductive rate R_0 that represents the number of people in a susceptible population that could be infected by an infective agent. In general, if $R_0 > 1$ the disease spreads epidemically and when $R_0 < 1$ the infection dies out.

One of the main restrictions of the original compartmental models is that they assume that all members within one compartment are identical to each other. Recent literature has evolved the SIR/SEIR models to overcome such homogeneity by creating the metapopulation models. Metapopulation models extend the traditional epidemiological approaches to differentiate types of population within each epidemic state (S,E,I,R). For example, Balcan et al. differentiate subgroups within the population based on vaccinations received, symptomatic versus asymptomatic individuals, citizens that travel versus those who do not, natural immunity to diseases, etc. [4]. Similarly, Brockmann et al. define different metapopulations based on their mobility patterns inferred from the movements of US banks across the world [5].

B. Agent-based Epidemic Models

Compartmental models cannot capture the complexity of human behavior including mobility patterns and social networks. Although metapopulation models attempt to overcome such limitations they still suffer from behavioral generalizations within metapopulations. In this context, agent-based epidemic models (ABMs) are designed to capture the behavior of each unique individual (agent). As a result, agent-based epidemiological simulations are more powerful than metapopulation models to represent the spreading of viruses given their granularity and capability to model behavior and interactions individually [6].

Although this research line is quite novel, the literature reports some relevant results. Apolloni et al. propose *Simdemics*, an integrated modeling environment that aids public health officials in pandemic planning [7]. Simdemics is an agent-based simulator that defines four models to evolve the epidemic spread: (1) a statistical model of the population (based on age, gender or geographical density), (2) a social interaction model, (3) a disease model, that accounts for the impact that demographic or socio-economic factors might have on epidemic spreading, and (4) intervention models *e.g.,* public policy changes, agent behavioral changes, etc. In their conclusions, the authors advocate for the necessity to have accurate human behavioral models that reveal mobility and interaction patterns.

Barrett et al. present an agent-based simulator called *EpiSimdemics* [8]. The authors build a synthetic population from the United States Census characterizing each individual (agent) with 163 different variables. Individuals are mapped to geographically located housing units, and their daily activities are modelled from a wide arrange of datasets like education statistics to model school attendance or transport surveys to model mobility patterns. The disease model consists of two parts: the *between-hosts* disease transmission and the *within-hosts* disease progression. The within-hosts progression is modelled as a finite state machine with probabilistic transitions (PTTS) that determines the evolution through the various disease states. The between-hosts transmission is modeled as follows:

$$p_i = 1 - \exp(\tau \sum_{r \in R} N_r \ln(1 - r s_i \rho)) \tag{1}$$

where p_i is the probability that an infection is triggered in a susceptible agent i, τ is the duration of exposure, R is the set of infective agents and N_r the number of such agents with infectivity r. s_i is the susceptibility of the individual i and ρ the basic transmissibility of the disease. This equation represents an intuitive process: the probability of inter-agent transmission increases based on the amount of time spent in the presence of an infective individual and the number of the infectious agents (and their infectivity) present at a given location. This approach is specially relevant when the transmission is mainly by direct contact, which is the case of H1N1.

ABM simulations, specially if done for large populations, require large amounts of memory and time. Recent literature has also explored how to effectively compute such ABM models. Parker et al. present the *Global-Scale Agent Model*, GSAM, which focuses on achieving high performance while computing realistic agents [9]. The GSAM system can generate over a billion distinct agents with models that include daily interactions. Additionally, the authors show how to use GSAM system to model epidemic evolutions at a planetary scale.

In general, although ABM improves traditional epidemiological approaches, all the solutions implemented so far face the same limitation: the information used to model human mobility and social networks is extracted from census data and surveys. Although in principle this data might approximate the real behavior, it does not account for the changes in behavior due to the epidemic itself. For that reason, we aim to achieve both a more realistic representation of human behavior and one that accounts for the behavioral changes that might take place during the epidemic.

III. PRELIMINARIES

In order to capture realistic human mobility patterns and social dynamics, we use the ubiquitous infrastructure provided by a cell phone network. Cell phone networks are built using a set of cell towers, also called Base Transceiver Stations (BTS), that are in charge of communicating cell phones with the network. Each BTS has a latitude and a longitude, indicating its geolocation, and gives cellular coverage to an area called a *cell*. We assume that the *cell* of each BTS is a 2-dimensional non-overlapping polygon, and we use a Voronoi tessellation to define its coverage area. Figure 1(left) shows a set of BTSs with the original coverage area of each cell, and Figure 1(right) presents its approximated coverage computed using Voronoi.

Call Detail Record (CDR) databases are generated when a mobile phone connected to the network makes or receives a phone call or uses a service (e.g., SMS, MMS, etc.). In the

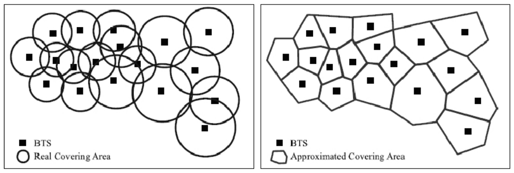


Fig. 1. (Left) Example of a set of BTSs and their coverage and (Right) Approximated coverage obtained applying Voronoi Tesselation.

process, and for invoice purposes, the information regarding the time and the BTS tower where the user was located when the call was initiated is logged, which gives an indication of the geographical position of a user at a given moment in time. Note that no information about the exact position of a user in a cell is known. From all the data contained in a CDR, our study only uses the encrypted originating number, the encrypted destination number, the time and date of the call, the duration of the call, the BTS tower used by the originating cell phone number and the BTS used by the destination cell phone number when the interaction happened.

Our objective is to use CDR data to compute the individual mobility and social models that are part of the ABM architecture to model virus spreading. Specifically, we aim to build: (1) a mobility user model that predicts the position of an agent at each moment in time and (2) a social user model that identifies the social network (in the sense of close relations) of each agent. Due to the nature of the CDR data available, each agent's mobility model is computed at a BTS level *i.e.,* the ABM system will be able to determine, at each moment in time, the BTS coverage area where an agent is located. The position of the agent within the coverage area of the BTS is unknown. As a result of that limitation, the ABM system will provide more accurate mobility models in areas with high densities of towers (urban areas) where coverage areas per BTS are smaller in size. Regarding the social network of the individual, we model it as the set of close relations obtained from the call detail records. Specifics about its computation are explained in Section IV-A. However, it is important to highlight that this model is critical to determine when a virus transmission happens. In fact, if we assume that two agents that are part of the same social network will be more prone to be physically close than two agents that do not know each other, we can then infer that whenever two agents are in the same coverage area (BTS), the probability of infection between the two will be higher if they are part of the same social network.

This approach of capturing and modeling agent behavior from CDRs, sets our work apart from others because: (1) we model agents from real individual data and not from census or surveys as explained in the related work; and (2) we capture behavioral adaptions to the spread of the disease *i.e.,* changes in mobility patterns or in the social network of the agents as the disease spreads over time. In fact, census or survey data give a one-shot view of a society's behavioral patterns. However, cell phone data is collected in real time and provides an accurate daily representation of the agents' behaviors and their changes due to external events. Finally, note that although the ABM system we present is designed for cell phone records, a similar approach could be potentially used with logs from any other location-based service like, for example, geolocalized twitter.

IV. ABM OF VIRUS SPREADING USING CELL PHONE DATA

We propose an ABM architecture with two main components: (1) the generation of agents from the information contained in call detail records and (2) the discrete event simulation process that simulates the virus propagation over time based on the agents' models.

A. Agent Generation

We define the behavior of each agent with three models: (1) a mobility model extracted from the CDR data; (2) a social network model computed from the CDR data and (3) a disease model that characterizes the progression of the disease through its various states in each agent.

1) Mobility Model: The mobility model captures the position (at the level of BTS) where the agent is at each moment in time. This model is used by the event simulation process to predict the location of each agent at each simulation tick. The granularity of the mobility model will also determine the granularity of the simulation steps *e.g.,* if the mobility model computes hourly distributions of locations, the simulation step will be one hour.

We propose a mobility model that divides each day into a set S of i non-overlapping equal-length time slots. Formally, being B the number of BTS towers that give coverage to a geographic area, the mobility model of agent n, M_n, is defined as:

$$\begin{array}{l} M_n = \{M_n^{wday}, M_n^{wend}\} = \\ \{\{M_n^{wday,0}, .., M_n^{wday,i}\}, \{M_n^{wend,0}, .., M_n^{wend,i}\}\} \quad \forall i \in S \\ M_n^{wday,i} = \{p_n^{wday,i,0}, \ldots, p_n^{wday,i,j}\} \quad \forall j \in B \\ M_n^{wend,i} = \{p_n^{wend,i,0}, \ldots, p_n^{wend,i,j}\} \quad \forall j \in B \end{array} \tag{2}$$

where $p_n^{wday,i,j}$ indicates the probability that agent n may be found at BTS j in timeslot i during a weekday, and $p_n^{wend,i,j}$ indicates the probability that agent n may be found at BTS j in timeslot i during a weekend. Given a CDR dataset for a period of time T, the ABM computes these models by associating to each time slot i the set of BTSs where the agent has been *seen* during weekdays or weekends over the period of time under study. The discrete event simulator will use the mobility model to determine the geographical location of each individual at a moment in time. Given that for a specific time slot i the user might have more than one BTS tower associated to its behavior, the event simulator assigns a position using the probability based distributions of the towers. Since individuals tend to show monotonic behaviors, an average user typically has very few BTSs associated to its mobility model. In the cases where a time slot contains no data, which typically happens for time slots at night, we assume that the user did not move from the latest predicted location in time.

In order to show that mobility models computed from CDRs can accurately represent the real mobility of an individual, we refer the reader to the research by Song et al. [10] and Candia et al. [11]. The authors show that mobility models computed from CDR data can predict the real locations of users with a 93% of correctness. However, they determine two pre-requisites that need to be fulfilled in order to achieve that accuracy: (1) users need to visit more than two locations (BTSs) during the training set and (2) users need an average call frequency of ≥ 0.5 calls per hour. Additionally, their research also indicates that there exist relevant behavioral differences between weekend and weekday behaviors and advocate for mobility models that can capture such diferences. We will explain details about the computation of our mobility models with these requirements in Section V.

2) Social Network Model: The social network of an individual plays a key role in virus spreading because it identifies the set of individuals with whom an agent has a close relationship. This is specially relevant for viruses that are transmitted by direct physical contact, like H1N1. We compute the social network of an agent as the set of agents with whom there was at least one reciprocal contact during the time period under study. By contact, we mean any type of communication: call, SMS or MMS, and does not need to be the same type to imply bidirectionality. Thus, an agent can be a member of more than one social network. Additionally, given that humans show clear different behavioral patterns between weekday and weekend, we compute two social networks per agent. Formally speaking, the social network S_n of agent n is computed as:

$$\begin{aligned}
&S_n = \{S_n^{wday}, S_n^{wend}\} = \\
&S_n^{wday} = \{list\ of\ reciprocal\ contacts\ in\ wdays\} \\
&S_n^{wend} = \{list\ of\ reciprocal\ contacts\ in\ wends\}
\end{aligned}$$

where S_n^{wday} is the social network during the weekdays and S_n^{wends} the social network during the weekends. Given the social networks of an agent, we assume that the probability of being physically close to another agent will be higher if that other agent is part of his social network. To model physical proximity we define two probabilities: (1) p_1 is the probability that two agents that are in the same BTS at the same moment of the simulation and are part of the same social network are physically close enough for the virus to be probably transmitted; and (2) p_2 the probability that two agents that are in the same BTS and are not in the same social network at the same moment in time are physically close for the virus to be probably transmitted. It is expected that p_1 will be much higher than p_2 given the social connection. These two parameters are a novel contribution to ABM systems since previous approaches did not have access to real behavioral data. It is important to clarify that p_1 and p_2 define the probability of two agents being physically close when they are in the same BTS at the same moment in time. The probability for the infection to occur between those agents will be defined by the disease model.

3) Disease Model: The disease model captures the progression of the disease in each agent. This model, together with the mobility and social models, is used by the discrete event simulator to reproduce the evolution of the disease at a global scale. We follow a similar approach to Barret et al. [8] and define a disease model that is composed of two parts: the *between hosts* transmission model and the *within host* progression model. In Figure 2 we observe that the *between hosts* transmission model happens at a probability p_i as described in Formula 1, and represents the probability that an agent goes from Susceptible to Exposed. In our model, we assume that all agents have the same initial susceptibility and infectivity *i.e.,* $r_i = 1$ and $s_i = 1 \forall i$. The *within host* model represents the evolution from Exposed to Infective in a given period of time $p = \epsilon$, and Infected to Removed in period of time $p = \beta$. Once an agent reaches the Removed state, it is considered to be protected from the virus and thus is removed from the simulation. The specific values for Formula 1, ϵ and β depend on the disease being modeled and are determined experimentally from epidemiological studies. Details about their computation are given in Section V.

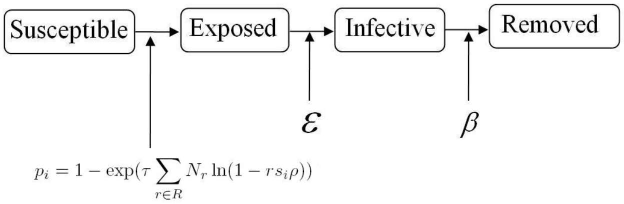


Fig. 2. Disease Model composed of Between hosts and Within hosts models.

B. Discrete Event Simulator

The Discrete Event Simulator (DES) simulates the evolution of the epidemic spreading for a set of agents over a specific period of time. To bootstrap the epidemic spreading, we assume that an initial agent is Infected and starts the transmission. The DES has a global clock and evaluates, at each simulation step, the state of all the agents in terms of mobility, social network and disease model. The value of the simulation step is determined by the granularity of the mobility models' time slots (see next section for computation details). Specifically, the DES does the following tasks: (1) identifies, for each agent, the geographical area where it is located (BTS) using the mobility model; (2) identifies the geographical areas where there exists, at least, one Infective agent; (3) for each Infective agent, takes all the Susceptible agents of his social network that are located in the same geographical area (BTS coverage) and applies probability p_1 that they will be physically close for the virus to be transmitted; (4) for each Infective agent and the rest of Susceptible agents included in his geographical area (not part of its social network), apply the probability p_2 that they will be physically close for the virus to be transmited; (5) for the set of agents phisically closed obtained from steps (3) and (4), apply the *between hosts transmission probability* to go from Susceptible to Exposed; (6) for the agents that are already in the Exposed or Infective state of the disease model apply the corresponding progression; and at last (7) remove from the simulation all agents that have reached the Removed state.

V. EVALUATING THE EFFECTS OF GOVERNMENT MANDATES FOR VIRUS CONTAINMENT: THE CASE OF H1N1 IN MEXICO

In case of a pandemic, the World Health Organization (WHO) recommends authoritative bodies to assess the suspension of activities in educational, government and business units as a measure to reduce the transmission of the disease. The preventive actions implemented by the Mexican government to control the H1N1 flu outbreak (Influenza A) of April 2009 constitute an illustrative example. The actions consisted of alerts and/or mandates aimed at reducing mobility, and where issued in three stages: (a) a medical alert issued on Thursday, April 16th, which was triggered by the diagnosis of the first H1N1 flu cases; followed by (b) the closing of schools and universities, enacted from Monday April 27th through Thursday, April 30th; and (c) the suspension of all non essential activities, implemented from Friday, May 1st to Tuesday, May 5th.

The Mexican H1N1 outbreak has been investigated in a number of recent papers, using analytical SIR models [12], agent-based approaches [13], [14] or metamodels [15]. Also from a public health perspective, there are studies that focus on clinical features, incubation times and transmission channels [16]; or on meassuring the impact of interventions such as anti-viral drugs [4], [17] or vaccination campaigns [15]. However, research into the impact that the Mexican government mandates had on the spread of the virus and on the mobility of the population is limited [12]. This is mainly due to the lack of large scale data about social and mobility behavioral patterns. In this section, we overcome these limitations by computing social and mobility models using Call Detail Records collected from a mexican urban area. These models, together with the ABM system presented in the previous section, will allow us to measure the impact that the actions taken by the Mexican government had on the spread of the virus and on human mobility. Implicitelly we assume that changes in human behavior are exclusively caused by government mandates. Although it is probably the main cause, there might be other reasons – such as fear produced by the media– that could also have influenced behavioral changes and that are not considered in our simulations. Next, we describe the experimental setting, the generation of the agents and our result analysis.

A. Experimental Setting

In order to examine the impact of government restrictions we evaluate changes in the mobility and disease models in five differentiated chronological periods. Table I presents the timeline under study. It covers from the 1^{st} of January 2009 until the 31^{st} of May 200. Each period is related to specific events that took place during the outbreak *i.e.,* preflu, alert, closed, shutdown and reopening. Agents (with mobility and social models) were computed for each one of these time periods. In order to measure behavioral changes we defined two scenarios: a baseline scenario and an intervention scenario. Given that the preflu period represents a time when individuals show normal behavior not affected by medical alerts, we build our baseline using the mobility and social models obtained using the preflu behavior. The intervention scenario wil consider the models defined by the alert, closed, shutdown and reopened time periods. In this case, depending on the moment of the simulation DES will jump from one set of models to the next. The evaluation is done by comparing the results obtained by both scenatios. Due to the inherent randoness of the spreading process we run each scenario 10 times and average the results.

Period	**Date Range**	**Description**
preflu	1/1 – 16/4	Period before any H1N1 case has been discovered. Agents will move largely unaffected and showing their usual patterns.
alert	17/4 – 26/4	April 16th - Diagnosis of H1N1 cases and medical alert triggered the following day. People may be reacting to the news and modify their usual patterns.
closed	27/4 – 31/4	Schools and Universities closed. Normal behaviour disrupted as people change their usual patterns.
shutdown	1/5 – 5/5	Closure of all non-essential activities.
reopened	6/5 – 31/5	Restrictions lifted.

TABLE I

TIME PERIODS TO EVALUATE BEHAVIORAL CHANGES.

B. Generation of Agents

To generate realistic agent models for mobility and social networks, we collected call records from January 1^{st} to May 31^{st} of 2009 for one of the most affected mexican cities. The entire dataset contains around 1 billion CDRs and around 2.4 million unique cell phone numbers. We associate each cell phone number to one agent and compute the mobility, social and disease models for both the *baseline* and the *intervention* scenarios. To compute the mobility models we use Formula 2 and select a granularity of one hour. As described in Section IV-A, we need to fulfill a set of requirements to guarantee that the mobility models computed from CDRs are realistic representations of human's motion. Following the research carried out by Song *et al.* [10], we filter the agents such that only those that (1) have at least two BTSs throughout the time periods; (2) have a minimum average calling rate of 0.25 calls/hour and (3) have at least 20% of the hourly time slots filled, are considered. Finally, since we want to measure behavioral changes during the outbreak, we only take into account agents that are active during the five time periods under study. This condition, together with the mobility filters, leaves us with a final number of $25,000$ agents.

We also built the social network models for the *baseline* and the *intervention* scenarios. As part of these models, we needed to define values for the contact probabilities p_1 and p_2. Given that these probabilities are new to our approach, it was impossible to find any reference in the literature. Thus, in order to compute their values, we made use of the work by Cruz-Pacheco et al. [12]. In this work, the authors examined the effect of the intervention measures on the epidemic spread using SIR. For that purpose, they represented the number of infective agents over the simulation time. We use their simulation to determine the optimal values of p_1 and p_2 as follows: we implemented an exhaustive search in the range $[0-1]$ over all combinations of p_1 and p_2, using .1 increments. For each pair of values tested, we run the simulation and obtain a curve representing the number of infected agents. We select as final p_1 and p_2 values the ones that minimize the mean squaredg error between our curve and the one presented in Cruz-Pacheco et al. Our searched determined that the best values were $p_1 = 0.9$ and $p_2 = 0.1$ *i.e.,* the probability that two agents that are in the same social network and in the same BTS are physically close for the infection to be transmitted is 0.9, and 0.1 if the agents are not in the same social network.

To build each agent's disease model, we used the parameters reported in the literature related to the H1N1 outbreak, see Table II. These set of parameters are common to both scenarios. Balcan *et al.* [4] used maximum likelihood analysis of epidemic simulations to derive values of $R_0 = 1.75$, an infectious period of 60 hours $\beta = 60^{-1}$ and a latent state (Exposed) of 1.1 days ($\epsilon = 26.4$ hours). Finally, we compute the value of ρ using R_0 as explained in [13]: $R_0 = \frac{\rho}{\beta}$; which gave a final value of $\rho = 34^{-1}$.

Parameter	Value	Description
R_0	1.75	Estimated Reproduction number.
ϵ	26.4^{-1} hours	Expected duration latent period.
β	60^{-1} hours	Expected duration infectious period.
ρ	34^{-1} hours	Expected time before infecting another agent.

TABLE II

PARAMETERS OF THE DISEASE MODEL.

Once all the agent models have been computed for both scenarios, we are ready to run both simulations. We initialize our simulations with one infected agent on April 17th (the first day a case was detected) [12] and run the simulation for 30 days. The initial agent infected was chosen to have a median connectivity (size of its social network), and located in one of the coverage areas that gives service to the airport to simulate a spread started by an agent that had just arrived to the city.

C. Result Analysis

In this Section, we compare the results of the *intervention* scenario with the *baseline* scenario, from three different perspectives: (1) from a mobility perspective, comparing changes in mobility; (2) from a disease model perspective, by comparing the number of susceptible and infected agents; and (3) from a spatial-temporal persective of the geographical evolution of the spread.

1) Agent Mobility: In order to measure the changes in mobility due to government mandates, we computed for each scenario the percentage of agents that moved from one BTS coverage area to another one at each time slot of the simulation ($1tick = 1hour$). Figure 3 shows the results.

The *baseline* clearly shows a cyclical day/night behavior throughout the simulation period. In general, it can be observed that at mid-day, more than 60% of the agents change BTS; whereas that number decreases to less than 20% during night hours. We also observe a cyclical behavioral change during the weekends, where the mobility is also reduced when compared to week days. The *intervention* scenario shows similar cyclical changes, however there exist relevant differentes when compared to the baseline. We can observe a decrease in mobility on the 27^{th} of April, precisely when the *alert* period finishes and the *close* period starts. In fact, we observe that this change is extended until the *shutdown* period starts. On May 1^{st} and throughout the *shutdown* period, we can observe an even larger decrease in mobility ($< 30\%$) that lasts until all restrictions are lifted on the 6^{th} of May. Although this behavioral change during the *shutdown* period is mainly caused by the total closure implemente by the government, it is important to note the following facts: (1) the *shutdown* period includes a weekend, which as observed in the baseline, always implies reduced mobility; and (2) May 1^{st} and May 5^{th} were national holidays in Mexico (Labor Day and Cinco de Mayo), which from a mobility perspective should show a behavior similar to the weekends baseline. To sum up, we can conclude that –in general– the *intervention* scenario shows a reduction in the number of agents that move when compared to its baseline during the alert, closing and shutdown periods of up to 30%. These differences dissappear once the *reopen* period starts (from the 6^{th} of May onwards).

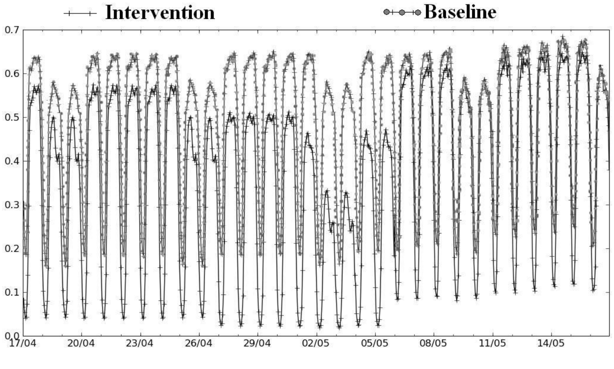


Fig. 3. Percentage of agents that move between BTSs for the *intervention* and *baseline* scenarios. The mobility granularity is $1hour$.

2) Disease Transmission: In this section we study the evolution of the disease focusing on the number of susceptible and infected agents for the *intervention* and *baseline* simulations. Figure 4 presents the percentage of the population that is in the susceptible stage of the disease model for a specific date and time. Results are shown for both the *intervention* scenario and the *baseline*. In both cases, we observe that at the beggining of the simulation (17^{th} of April) all agents are susceptible of being infected (except for the initial agent infected to start the simulation). As time passes, the evolution of susceptible agents is described by a sigmoid function. We observe that the number of susceptible agents decreases faster in the *baseline* scenario. This fact indicates that the government measures taken during the *intervention* scenario had an impact on the agents' behaviors and managed to reduce the number of infected agents (which implies more susceptible agents) when compared to the *baseline* scenario. We observe the largest difference during the peak of the epidemic, with approximately a 10% less of susceptible agents. By the end of the outbreak, we see that the number of susceptible agents is lower in the *baseline* than in the *intervention* scenario (i.e., more agents were infected in the baseline scenario).

Figure 5 shows the percentage of infected agents during the simulation for both scenarios. We observe that the peak of the epidemic in the *intervention* scenario happens later in time than the *baseline*, and has a smaller value. Delaying the peak of epidemics is a priority for intervention strategies, as the time gained can be used for implementing actions such as vaccination campaigns, which have to be delivered before the peak in order to be effective. In this case, we see that the reductions in mobility and the closure of public buildings delayed the peak of the epidemic by 40 hours. On the other hand, limiting the incidence of a disease at its peak prevalence, is also an important objective in intervention strategies. In our simulations, we observe that the number of infected agents was reduced around 10% in the peak of the epidemic in the *intervention* scenario when compared to the *baseline* scenario. These results are in agreement with the ones reported in [12]. In fact, using traditional disease model techniques, the authors reported a reduction in prevalence as a result of the government restrictive actions of $6\% - 10\%$. Our approach, not only confirms such numbers, but also extends their work by incorporating agent models for geographical mobility and social networks.

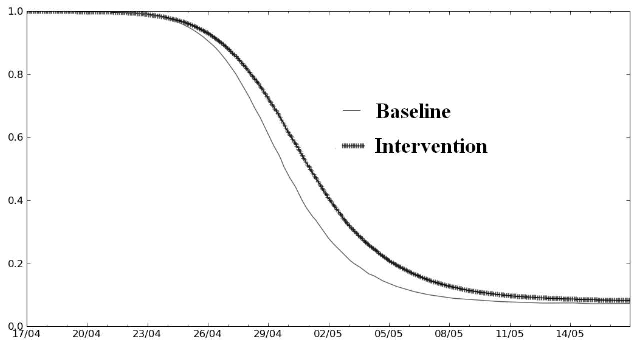


Fig. 4. Fraction of susceptible agents in the population over time. These curves are an average of all simulation runs.

3) Spatio-Temporal Evolution: The combination of the mobility and disease models provides us with a spatio-temporal representation of the spread of the virus. In fact, we can analyze the spread not only at a global scale –as done in the previous section– but at a BTS level. Such analysis gives an understanding of the geographical and chronological transmission of the spread throughout a city.

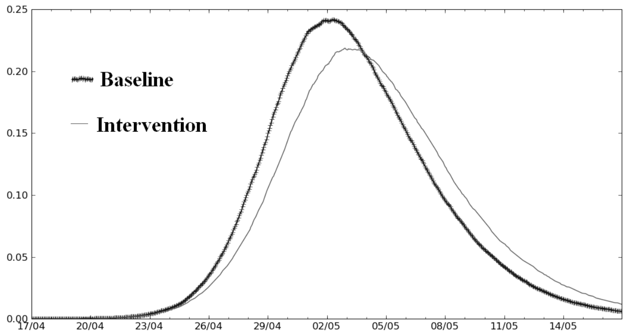
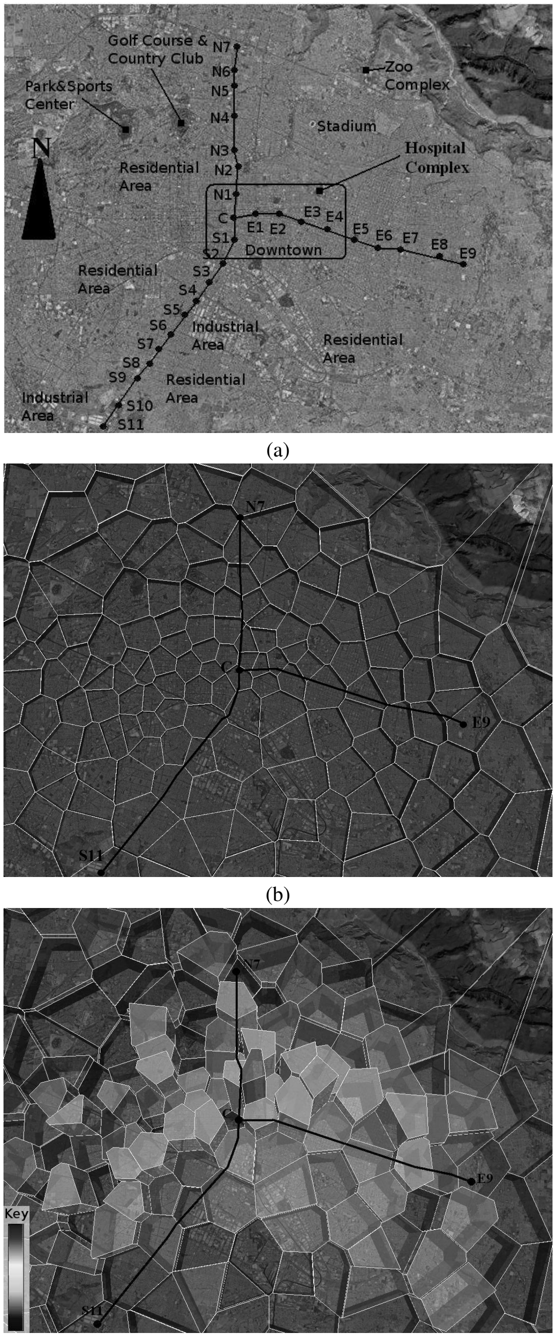


Fig. 5. Fraction of infective agents over time. These curves are an average of all simulation runs.

Figure 6(a) presents the main part of the city under study and some of its landmarks, mainly the subway system which consists of two lines: L1, runs East-West and L2, which runs North-South (L2) with one central station in common, C. Downtown is geographically located around C, E1, E2, E3 and E4, where we can find university buildings, government offices and commercial areas. Figure 6(b) shows the BTS coverage areas of the cell towers in the city, computed using Voronoi. The spatio-temporal dimension allows to study the spread of viruses in this lattice. Figure 6(c) depicts the number of infected agents per BTS at $12am$ on May 2^{nd} (at the peak of the spreading) in the *baseline* scenario. It can be observed that the downtown area contains the largest number of infected agents, although residential areas located to the west of the city are also heavily infected, specially when compared to other residential areas. The *intervention* scenario shows a similar geographical distribution of the heavily infected BTS areas, although the numbers are smaller as explained in the agent mobility analysis. Analogously, the temporal evolution of the transmission follows a similar trend both in the *intervention* and *baseline* scenarios: the spread starts at the airport area and rapidly evolved towards downtown, where it peaks, until it dies out as agents move to the Removed stage. This preliminary spatio-temporal analysis give some indications that although behavioral changes due to government restrictions manage to reduce and contain the epidemic, they do not seem to affect its spatial or temporal evolution.

VI. CONCLUSIONS AND FUTURE WORK

The ability to model and predict the evolution of a virus spreading is a critical issue for governments and health organizations. Although ABM solutions implemented so far capture the inherent individuality and randomness of the process, they do not fully represent the spatio-temporal dynamics and the way humans change their behavior and adapt to situations. This limitation is mostly due to the fact that human behavior is typically modelled from census data or surveys. In this



(a)

(b)

(c)

Fig. 6. (a) Map of the city under study with the subway system and reference landmarks; (b) Division of the city into the BTS coverage areas using Voronoi; and (c) Number of infected agents (represented by the height and color of the bars in each coverage area) in the *baseline* scenario at 12am on May 2^{nd}.

paper, we have introduced an ABM system whose agents are generated with the behavioral information extracted from call detail records. As a result, the agents contain not only mobility and social patterns, but also the changes of these patterns over time. These behavioral changes are critical to achieve realistic spread simulations that allow us to measure the real impact of the spread on society. We have presented a case study of the H1N1 outbreak of Mexico in 2009, identifying that the spread was both reduced and postponed due to the government actions. Our analisys, which focuses on the mobility and social networks of the agents, provides a novel approach to simulations based on real behavior. Future work will focus on enriching the agents' characterization by adding variables such as socio economic factors or health status that will create even more realistic simulation environments. Finally, we also plan to work on formal methods to measure changes in the spread from a spatio-temporal persective so as to enhance the preliminary results presented in this paper.

REFERENCES

[1] J. Epstein, D. Goedecke, F. Yu, R. Morris, D. Wagener, and G. Bobashev, "Controlling pandemic flu: the value of international air travel restrictions," *PLoS One*, vol. 2, no. 5, p. e401, 2007.

[2] S. Riley, "Large-scale spatial-transmission models of infectious disease," *Science*, vol. 316, no. 5829, p. 1298, 2007.

[3] W. Kermack and A. McKendrick, "Contributions to the Mathematical Theory of Epidemics. II. The Problem of Endemicity," *Proceedings of the Royal society of London. Series A*, vol. 138, no. 834, p. 55, 1932.

[4] D. Balcan, H. Hu, B. Goncalves, P. Bajardi, C. Poletto, J. J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van Den Broeck, V. Colizza, and A. Vespignani, "Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility." *BMC medicine*, vol. 7, p. 45, 2009.

[5] D. Brockmann, L. Hufnagel, and T. Geisel, "The scaling laws of human travel." *Nature*, vol. 439, no. 7075, pp. 462–5, 2006.

[6] A. Apolloni, V. Kumar, M. Marathe, and S. Swarup, "Computational epidemiology in a connected world," *Computer*, vol. 42, no. 12, pp. 83–86, 2009.

[7] A. Apolloni, V. A. Kumar, M. V. Marathe, and S. Swarup, "Computational Epidemiology in a Connected World," *Computer*, vol. 42, no. 12, pp. 83–86, 2009.

[8] C. L. Barrett, K. R. Bisset, S. G. Eubank, X. Feng, and M. V. Marathe, "EpiSimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks," in *SC'08: Proceedings of the 2008 ACM/IEEE conference on Supercomputing*, 2008.

[9] J. Parker and J. Epstein, "Distributed Platform for Global-Scale Agent-Based Models of Disease Transmission (in press)," *ACM Trans. Model. Comput. S.*, 2011.

[10] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility." *Science*, vol. 327, no. 5968, pp. 1018–21, 2010.

[11] J. Candia, M. González, P. Wang, T. Schoenharl, G. Madey, and A. Barabási, "Uncovering individual and collective human dynamics from mobile phone records," *Journal of Physics A: Mathematical and Theoretical*, vol. 41, p. 224015, 2008.

[12] G. Cruz-Pacheco, L. Duran, L. Esteva, A. A. Minzoni, M. López-Cervantes, P. Panayotaros, A. Ahued Ortega, and I. Villaseñor Ruíz, "Modelling of the influenza A(H1N1)V outbreak in Mexico City, April-May 2009, with control sanitary measures," *Eurosurveillance*, vol. 14, no. 26, 2009.

[13] D. Brockmann, V. David, and A. M. Gallardo, *Human Mobility and Spatial Disease Dynamics*. Leipzig: Leipziger Universitätsverlag, 2009.

[14] J. M. Epstein, "Modelling to contain pandemics." *Nature*, vol. 460, no. 7256, p. 687, 2009.

[15] P. Bajardi, C. Poletto, D. Balcan, H. Hu, B. Goncalves, J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van Den Broeck, V. Colizza, and A. Vespignani, "Modeling vaccination campaigns and the Fall/Winter 2009 activity of the new A(H1N1) influenza in the Northern Hemisphere," *Emerging Health Threats Journal*, vol. 2, 2009.

[16] V. Shinde, C. B. Bridges, T. M. Uyeki, B. Shu, and et al., "Triple-reassortant swine influenza A (H1) in humans in the United States, 2005-2009." *The New England journal of medicine*, vol. 360, no. 25, pp. 2616–25, 2009.

[17] C. Barrett, K. Bisset, J. Leidig, A. Marathe, and M. Marathe, "Estimating the Impact of Public and Private Strategies for Controlling an Epidemic: A Multi-Agent Approach," in *Innovative Applications of Artificial Intelligence*, 2009, pp. 34–39.