2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing

An Agent-Based Model of Epidemic Spread using Human Mobility and Social Network Information

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 do not consider individual information and hence have limited ability to capture the inherent complexity of the disease spreading process. To overcome this limitation, agent-based models have recently been proposed as an effective approach to model virus spreading. However, most agent-based approaches to date have not included real-life data to characterize the agents' behavior. In this paper we propose an agent-based system that uses social interactions and individual mobility patterns extracted from call detail records to accurately model virus spreading. The proposed approach is applied to study the 2009 H1N1 outbreak in Mexico and to evaluate the impact that government mandates had on the spreading of the virus. Our simulations indicate that the restricted mobility due the government mandates reduced by 10% the peak number of individuals infected by the virus and postponed the peak of the pandemic by two days.**

I. Introduction

Planning for a pandemic (*e.g.*, H1N1, influenza, etc.) is a public health priority 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 recent adoption of agent-based modeling (ABM) approaches to simulate pandemics has allowed to capture individual human behavior and its inherent fuzziness by representing every person as a software agent. The ABM model characterizes each agent with 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. Hence, ABM approaches need realistic data to create agents that effectively capture human behavior. Typically this data is obtained from the census or by means of 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 capturing large scale human behavioral data. These datasets contain information that is critical to accurately model the spread of a virus, such as human mobility patterns or the social network characteristics of each individual [2][1].

In this paper, we propose an ABM system designed to simulate virus spreading using agents that are characterized by their individual mobility patterns and social networks as extracted from cell phone records. We carry out simulations with data collected during the 2009 Mexican H1N1 outbreak and measure the impact that government calls had on the mobility of individuals and the subsequent effect on the spread of the H1N1 virus. To the best of our knowledge, this is the first time that this kind of real-life information is used in an ABM system.

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 describes the infrastructure of a cell phone network and how cell phone records are captured; our proposed ABM architecture is presented in Section IV. Section V presents a case study that evaluates the impact that government mandates regarding mobility restrictions had on the spreading of the 2009 H1N1 virus outbreak in Mexico. Finally, we describe our conclusions and outline our future work in Section VI.

II. Related Work

A. Traditional Epidemic Disease Models

Traditional epidemic disease models are based on the SIR model and 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 or infectious (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 rates are used

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

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 $R_0 \leq 1$ the disease 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 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 bank notes[5].

B. Agent-based Epidemic Models

Compartmental models cannot capture the complexity of human behavior, particularly regarding mobility patterns and social networks. Although metapopulation models attempt to overcome such limitations they still suffer from behavioral generalizations within the 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 already 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 modeled 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-host* disease progression. The within-hosts progression is modeled 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:

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

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 individual i and ρ the basic transmissibility of the disease. This equation represents an intuitive process: the probability of inter-agent transmission increases with the amount of time spent in the presence of an infective individual and the number of 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 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 agent-based epidemic models improve 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 these data might approximate real behavior, it does not account for changes in behavior due to the epidemic itself. The model proposed in this paper aims to achieve a more realistic representation of human behavior which includes 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, called Base Transceiver Stations (BTS), that connect the cell phones to the network. Each BTS has a latitude and a longitude – its geolocation – and gives cellular coverage to an area called a *sector*. We assume that the *sector* 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 process, and for invoice purposes, the information regarding the time and the BTS tower where the user was located when

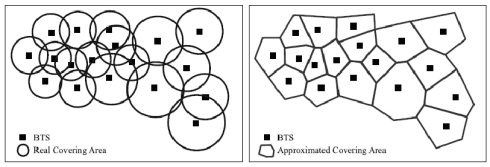


Fig. 1. (Left) Example of a set of BTSs and their coverage and (Right) Approximated coverage obtained applying Voronoi tesselation.

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, and the BTS towers used by the originating and destination cell phone numbers.

We use CDR data to compute the individual mobility and social models that are part of the proposed ABM architecture to model virus spreading. Specifically, we build: (1) a *mobility user model* that estimates the position of each agent at each moment in time and (2) a *social user model* that identifies each agent's social network (in the sense of close relations). Due to the nature of the CDR data available, each agent's mobility model is computed at the 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. Each individual's social network is modeled as the set of close relations obtained from the CDRs. Specifics about its computation are explained in Section IV. Note that this model is critical to determine when the transmission of the virus takes place. We assume that two agents that are part of the same social network are more likely to be physically close than two agents that do not know each other. Hence, 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 previously explained; and (2) we capture behavioral adaptations 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 snapshot 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 used with logs from any other location-based service, such as *e.g.* geolocalized Twitter.

IV. ABM of Virus Spreading using CDRs

We propose an ABM architecture with two main components: (1) a set of agents that are modeled using the information contained in call detail records; and (2) a discrete event simulator (DES) 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 CDR data; (2) a social network model computed from 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 provides the position (at the BTS level) 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 step. The temporal granularity of the mobility model determines 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, 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 B is the number of BTS towers that give coverage to a geographic area; and $p_n^{wday,i,j}$ and $p_n^{wend,i,j}$ denote the probability that agent n may be found at BTS j in timeslot i during a weekday or weekend, respectively. Given a CDR dataset, the mobility model is built by associating with each time slot i the set of BTSs where each person has been *observed* during weekdays or weekends during the period of time under study. Note that each individual might be assigned to more than one BTS in a specific time slot i. In this case, the event simulator assigns the position of the tower with the highest probability, *i.e.,* the BTS that the individual has used the most over the training period. Since people tend to show monotonic behaviors, an average person typically has very few BTS towers in his/her mobility model. In the cases where a time slot contains no data, which typically happens for time slots at night, we assume that the person did not move from the latest predicted location in time.

As shown by Song *et al.* in [10], mobility models computed from CDRs can accurately predict the real locations of users with 93% accuracy. However, two pre-requisites need to be fulfilled in order to achieve this level of accuracy: (1) individuals need to visit more than two locations (BTSs) during the training set; and (2) they need an average call frequency of ≥ 0.5 calls per hour. Additionally, research by Candia *et al.* [11] indicates that there exist relevant behavioral differences

between weekend and weekday behaviors and advocate for mobility models that can capture such differences. We will explain details about the computation of our mobility models that satisfy 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 a person 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 individuals 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. Note that 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:

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

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 its social network. To model physical proximity within a BTS coverage area we define two probabilities: (1) p_1 is the probability that two agents that are in the same BTS at the same time of the simulation and are part of the same social network are physically close enough for the virus to be possibly 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 possibly transmitted. It is expected for p_1 to be larger than p_2 given the social connection. These two probabilities are a novel contribution of our work since previous ABM 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 (explained below).

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 that of 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, given by Eq. 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 ϵ, and from Infected to Removed in period of time β.

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 of ϵ and β in Eq. 1 depend on the disease being modeled and are determined experimentally from epidemiological studies. Details about their computation are given in Section V.

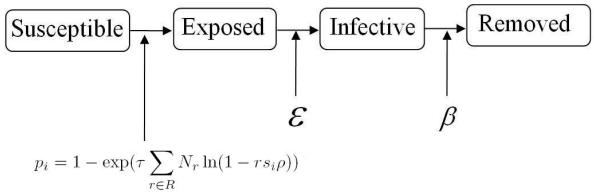


Fig. 2. Disease Model composed of Between hosts and Within hosts models.

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 size of the simulation step is determined by the temporal granularity of the mobility model (see next section for computation details). Specifically, the DES does the following consecutive tasks: (1) It identifies the geographical area (BTS) where each agent is located using the mobility model; (2) it identifies the geographical areas where there is, at least, one Infective agent; (3) for each Infective agent, it 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 its geographical area (not part of its social network), it applies the probability p_2 that they will be physically close for the virus to be transmitted; (5) for the set of agents physically close obtained from steps (3) and (4), it applies 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, it applies the corresponding progression; and at last (7) it removes from the simulation all agents that have reached the Removed state.

These steps are repeated for each simulation step during the overall simulation time.

V. Experiments: The case of H1N1 in Mexico

In case of a pandemic, the World Health Organization (WHO) recommends authoritative bodies to consider the sus-

pension of activities in educational, government and business units as a measure to reduce the transmission of the disease. The actions implemented by the Mexican government to control the H1N1 flu outbreak 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]. From a public health perspective, there are studies that focus on clinical features, incubation times and transmission channels [16]; or on measuring 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 H1N1 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. We overcome these limitations by computing social and mobility models using Call Detail Records collected from a Mexican urban area during the H1N1 flu outbreak. We use these models in the ABM system previously presented and measure the impact that the actions taken by the Mexican government had on human mobility and subsequently on the spread of the virus. Note that 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 induced 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 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 mobility 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 mobility patterns.
closed	27/4 – 31/4	Schools and Universities closed. Normal behavior disrupted as people change their usual mobility patterns.
shutdown	1/5 – 5/5	Closure of all non-essential activities.
reopened	6/5 – 31/5	Restrictions lifted.

TABLE I

TIME PERIODS OF STUDY.

A. Experimental Setting

In order to examine the impact of government restrictions we evaluate changes in the mobility and disease models in five chronological periods. Table I presents the timeline under study. It covers from January 1st, 2009 to May 31st, 2009. Each period is related to specific events that took place during the outbreak *i.e.,* preflu, alert, closed, shutdown and reopening. We generate agents (with corresponding mobility and social models) for each of these time periods. In order to measure behavioral changes, we define two scenarios: a *baseline* scenario and an *intervention* scenario.

The *baseline* scenario is built using the mobility and social models obtained during the pre-flu period, when individuals show normal – not affected by medical alerts – mobility behavior. The *intervention* scenario considers the models that are built with data from the alert, closed, shutdown and reopened periods. In this case, depending on the moment of the simulation, the DES will jump from one set of models to the next. The evaluation is done by comparing the results obtained by both scenarios. Due to the inherent randomness of the spreading process we run each scenario 10 times and average the results.

B. Generation of Agents

To generate realistic agent mobility and social network models, we collected CDRs from January 1^{st} to May 31^{st} of 2009 of one of the most affected Mexican cities. The entire dataset contains around 1 billion CDRs and around 2.4 million unique cell phone numbers. Each cell phone number is associated with one agent and we compute the mobility, social and disease models for both the *baseline* and the *intervention* scenarios.

The mobility models are computed using Eq. 2 with a granularity of one hour. As described in Section IV, we need to fulfill a set of requirements to guarantee that the mobility models computed from CDRs are realistic representations of a human's motion. Following the research carried out by Song *et al.* [10], we filter the individuals such that only those that (1) are assigned to 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.

These requirements narrow down the final number of agents to $25,000$.

We also build 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. In order to compute their values, we make use of the work by Cruz-Pacheco *et al.* [12], where the authors examined the effect of the government intervention measures on the epidemic spread using SIR. We use their simulation to determine the optimal values of p_1 and p_2 as follows: we implement 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 squared error between our curve and the one presented in Cruz-Pacheco *et al.* Our search 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 BTS and in the same social network 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 use the parameters reported in the literature related to the H1N1 outbreak (see Table II). These 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 gives 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 by air.

C. Analysis of the Results

In this Section, we compare the results of the *intervention* scenario with the *baseline* scenario from three different viewpoints: (1) a mobility perspective, by comparing changes in mobility; (2) a disease model perspective, by comparing the number of susceptible and infected agents; and (3) a spatio-temporal perspective, by comparing the geographical evolution of the disease 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 step of the simulation (1 step = 1 hour). Figure 3 shows the results.

The *baseline* plot 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 reduced when compared to weekdays. The *intervention* scenario shows similar cyclical changes. However, there are a number of important differences when compared to the baseline. There is a significant decrease in mobility on April 27^{th}, precisely when the *alert* period finishes and the *close* period starts. This decrease in mobility continues until the beginning of the *shutdown* period. On May 1^{st} and throughout the *shutdown* period, there is an even larger decrease in mobility ($< 30\%$) that lasts until all restrictions are lifted on May 6^{th}. Although the behavioral change during the *shutdown* period is mainly caused by the total closure implemented 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 during the *intervention* scenario there is a reduction in the mobility of the agents of 10% during the alert period and of up to 30% during the closing and shutdown periods, when compared to the baseline. These differences in the agents' mobility disappear once the *reopen* period starts (from May 6^{th} onwards).

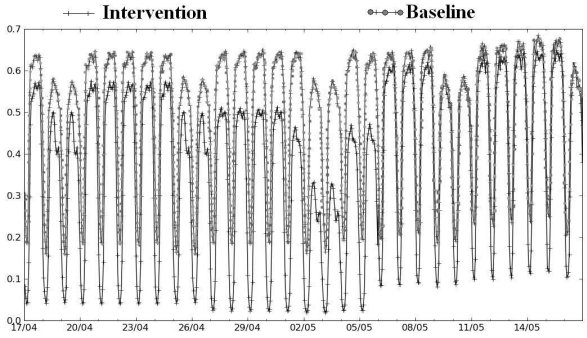


Fig. 3. Percentage of agents that move between BTSs for the *intervention* and *baseline* scenarios. The temporal granularity is 1 hour.

2) Disease Transmission: In this section we study the evolution of the disease focusing on the number of susceptible and infected agents in the *intervention* and *baseline* simulations. Figure 4 displays 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* and the *baseline* scenarios.

In both cases, we observe that at the beginning of the simulation (April 17^{th}) all agents are susceptible of being infected (except for the initial infected agent that starts the simulation). As time passes, the evolution of susceptible agents is described by a sigmoid function. The number of susceptible agents decreases faster in the *baseline* scenario, *i.e.* the number of infected agents grows faster than in the *intervention* scenario. This result supports the hypothesis that the government measures taken during the *intervention* scenario had an impact

on the agents' mobility patterns and hence managed to reduce the number of infected agents (which implies a larger number of susceptible agents) when compared to the *baseline* scenario. The largest difference between both sigmoid functions takes place during the peak of the epidemic, with approximately a 10% less of susceptible agents in the *intervention* scenario. By the end of the outbreak, 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 in the *baseline*, and has a *smaller* absolute value. Delaying the peak of epidemics is a priority in intervention strategies, as the time gained can be used to implement actions such as vaccination campaigns, which have to be delivered before the peak in order to be effective. The reduction in mobility and the closure of public buildings delayed the peak of the epidemic by 40 hours.

Another important objective in intervention strategies focuses on limiting the incidence of a disease (measured in $\%$ of infected agents) at its peak. In our simulations, the total number of infected agents was reduced by 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 this case, the authors, using traditional disease model techniques (SIR), reported a reduction in prevalence as a result of the government restrictive actions of $6\% - 10\%$.

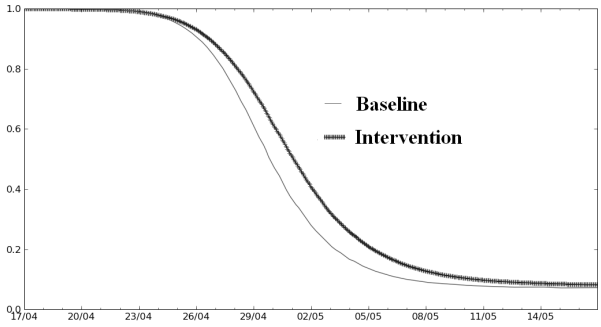


Fig. 4. Fraction of susceptible agents in the population over time. These curves are an average of all simulation runs.

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 the city.

Figure 6(a) displays the main parts of the city under study and some of its landmarks, namely 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. The

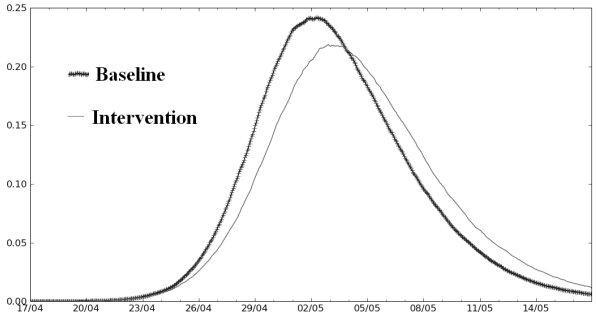


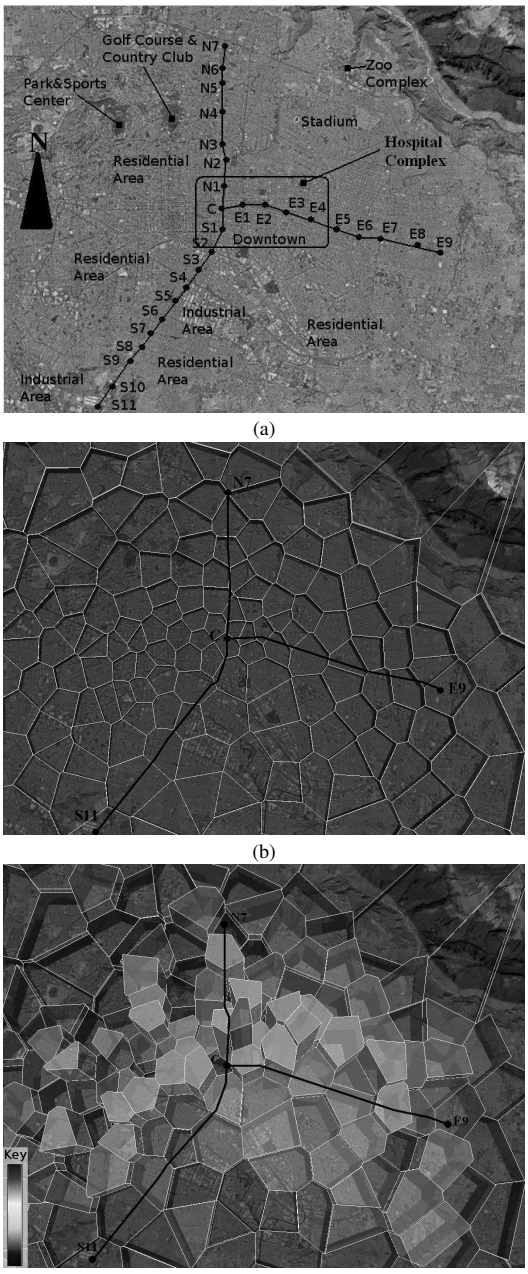
Fig. 5. Fraction of infected agents over time. These curves are an average of all simulation runs.

downtown area is geographically located around C, E1, E2, E3 and E4, where there are university buildings, government offices and commercial areas. Figure 6(b) shows the BTS coverage areas of the cell towers in the city, computed using a Voronoi tessellation.

The spatio-temporal analysis allows to study the spread of the virus 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. Note 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 heavily infected BTS areas, although the number of infected agents is smaller than in the *baseline* scenario. Analogously, the temporal evolution of the transmission follows a similar trend both in the *intervention* and *baseline* scenarios: the spread starts in the airport area (where the first infected agent was) and rapidly evolves towards the city's downtown area, where it peaks, until it dies out as agents turn into the Removed stage. This preliminary spatio-temporal analysis seems to indicate that although behavioral changes due to government restrictions manage to reduce and contain the epidemic, they do not seem to affect its spatio-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 previously proposed ABM systems are able to capture the inherent individuality and randomness of the process, they have not modeled the spatio-temporal dynamics of human behavior and its potential changes due to the alarm situation. This limitation is mainly due to the fact that the agents' behavior is typically built from census or survey data. In this paper, we have introduced an ABM system whose agents' mobility and social network models are built from human behavioral data available in call detail records. As a result, the agents' behavior not only mimics a population's mobility and social patterns, but also the changes of these patterns over time. These changes are critical to



(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 a Voronoi tessellation; 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}.

achieve realistic spread simulations that allow us to measure the real impact of the spread.

We have applied the proposed ABM system to CDR data captured during the H1N1 outbreak of Mexico in 2009. In our experiments, we have found that the spread of the virus was both reduced (by about 10%) and postponed (by about 40 hours) thanks to the government mandates. Our analysis, which focuses on the agents' mobility and social networks, provides a novel approach to ABM simulations based on real behavior.

Future work will focus on enriching the agents' characterization by adding variables such as socio-economic factors and health status that will create even more realistic simulation environments. We also plan to work on formal methods to measure changes in the spread from a spatio-temporal perspective so as to enhance the preliminary results presented in this paper. Finally, we plan to analyze the impact that the location, mobility and social connectedness of the first infected agent has on the spread of the disease.

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.