Simulation of Epidemic Spread using Cell-Phone Call Data: H1N1 Case Study

Enrique Frías-Martínez [‡], Graham Williamson [#], Vanessa Frías-Martínez [‡],

[‡] *Telefónica Research, Madrid – Spain*

[#] *School of Computer Science , University College Dublin – Ireland*
{efm,graham,vanessa}@tid.es

***Abstract*—Ubiquitous computing technologies enable capturing large amounts of human behavioral data. The digital footprints computed from these datasets provide information for the study of social 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. 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.**

I. Introduction

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 has allowed to capture individual human behavior and its inherent fuzziness by representing every person as a software agent.

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

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. An extended description of our system and its evaluation using the 2009 H1N1 outbreak can be found in [1].

We have used call detail records(CDR) to compute: (1) a *mobility user model* and (2) a *social user model* that identifies each agent's 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; and (2) we capture behavioral adaptations to the spread of the disease.

II. 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.

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.

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. We propose a mobility model that divides each day into a set S of i non-overlapping equal-length time slots. The mobility model of agent n, M_n, is defined as:

$$\begin{array}{l} M_n = \{M_n^{wday}, M_n^{wend}\} = \\ \{\{M_n^{wday,0}, .., M_n^{wday,i}\}, \{M_n^{wend,0}, .., M_n^{wend,i}\}\} \quad \forall i \in S \\ M_n^{wday,i} = \{p_n^{wday,i,0}, \ldots, p_n^{wday,i,j}\} \quad \forall j \in B \\ M_n^{wend,i} = \{p_n^{wend,i,0}, \ldots, p_n^{wend,i,j}\} \quad \forall j \in B \end{array} \tag{1}$$

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.

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:

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

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.

The disease model captures the progression of the disease in each agent. We follow a similar approach to that of Barret *et al.* [2] and define a disease model that is composed of two parts: the *between hosts* transmission model and the *within host* progression model. In Figure 1 we observe that the *between hosts* transmission model happens at a probability p_i and represents the probability that an agent goes from Susceptible to Exposed. 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.

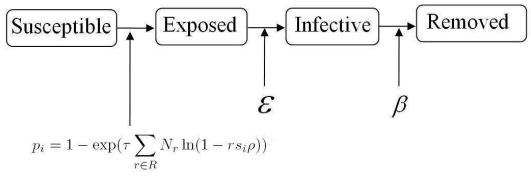

Fig. 1. Disease Model composed of Between hosts and Within hosts models.

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. 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.

III. Experiments: The case of H1N1 in Mexico

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

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.

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. 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.

Generation of Agents

We collected CDRs from January 1^{st} to May 31^{st} of 2009 of one of the most affected Mexican cities. The dataset contains 1 billion CDRs and 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 with a granularity of one hour. Following Song *et al.* [3], we only consider the agents that (1) are assigned to at least two BTSs; (2) have a minimum average calling rate of 0.25 calls/hour; and (3) have at least 20% of the hourly time slots filled. 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.* [4], where the authors examined the effect of the government intervention measures on the epidemic spread using SIR. Details can be found in [1]. Our search determined that the best values were $p_1 = 0.9$ and $p_2 = 0.1$.

To build each agent's disease model, we use the parameters reported in the literature related to the H1N1 outbreak: $R_0 = 1.75$ (Estimated Reproduction number), $\epsilon = 26.4^{-1}$ hours (Expected duration latent period), $\beta = 60^{-1}$ hours, (Expected duration infectious period) and $\rho = 34^{-1}$ hours (Expected time before infecting another agent). We initialize our simulations with one infected agent on April 17th (the first day a case was detected) [4] and run the simulation for 30 days.

Analysis of the Results

In this Section, we compare the results of the *intervention* scenario with the *baseline* scenario from a mobility perspective and from a disease model perspective.

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 2 shows the results.**

Both the *baseline* and the *intervention* plots show similar cyclical changes. However, there are a number of important differences. 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}. To sum up, 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).

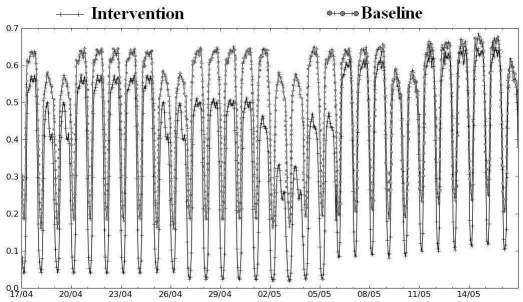

Fig. 2. Percentage of agents that move between BTSs for the *intervention* and *baseline* scenarios. The temporal granularity is 1 hour.

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 3 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. We observe that at the beginning of the simulation all agents are susceptible of being infected. 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 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.

Figure 4 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. The reduction in mobility and the closure of public buildings delayed the peak of the epidemic by 40 hours. Also, 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 [4].

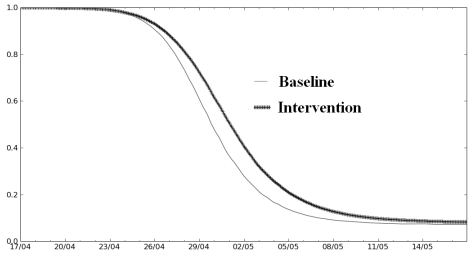

Fig. 3. Fraction of susceptible agents in the population over time. These curves are an average of all simulation runs.

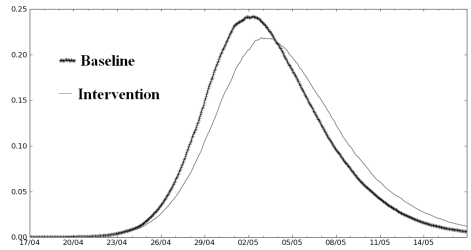

Fig. 4. Fraction of infected agents over time. These curves are an average of all simulation runs.

REFERENCES

[1] E. Frias-Martinez, G. Williamson, and V. Frias-Martinez. An agent-based model of epidemic spread using human mobility and social network information. In *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*, pages 57–64. IEEE, 2011.

[2] Christopher L. Barrett, Keith R. Bisset, Stephen G. Eubank, Xizhou Feng, and Madhav 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.

[3] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–21, 2010.

[4] 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*, 14(26), 2009.