# EpiGraph: A Scalable Simulation Tool for Epidemiological Studies

**Gonzalo Martín, Maria-Cristina Marinescu, David E. Singh and Jesús Carretero**
Computer Science Department
Carlos III University of Madrid
28911 Leganés, Spain

**Abstract**— *This paper presents a novel approach to modeling the propagation of the flu virus throughout a realistic interconnection network based on actual individual interactions which we extract from social networks. We allow the individual interconnections to change during the propagation by making them time-dependent. We have implemented a scalable, fully distributed simulator and we validated the epidemic model by comparing the simulation results against those of another epidemic simulator, with similar prediction values and better performance. We then performed an extensive analysis of the effects of the new features of our approach on the results of the simulations.*

**Keywords:** simulation, epidemiology, social networks, distributed algorithms

## 1. Introduction

Modeling the evolution of an epidemics involves both modeling the specific infectious agent as well as the actual social structure of the population under study. The purpose of the work we present in this paper is to accurately model the evolution of an epidemics in specific populations over a short to medium time span. Using an actual social model as input for the epidemic model promises more accurate results then either using probability distributions or synthetically generating the interaction graphs. Our approach approximates an actual social model by a realistic model based on real demographic information and actual individual interactions extracted from social networks. To the extent of our knowledge ours is the first attempt to model the connections within a population at the level of an individual based on information extracted from virtual social networks such as Enron or Facebook. Additionally, we allow modeling the characteristics of each individual as well as customizing his daily interaction patterns based on the time and the day.

We implemented EpiGraph, a simulator which takes as inputs the social model and an epidemic model specific to the influenza virus. The implementation is distributed and fully parallel; this allows simulating large populations of the order of millions of individuals in execution times of the order of minutes. We compared the results for our simulations in terms of the effects of the epidemics with the results obtained by InfluSim in [1]. We show that the simulators predict similar results. We further perform an extensive study of the effects of the features specific to our approach on the disease propagation. For instance, we study how different social models affect the disease propagation and we investigate the effects of introducing different vaccine or quarantine programs at different stages of the epidemic.

*Our contributions:* The specific contributions of this work are the following: (1) We use real demographic data to model group types with different characteristics. We leverage data extracted from social networks to model the interaction patterns between individuals pertaining to the same social group; (2) We allow modeling individual characteristics such as profession, age, gender, etc. We also allow customizing individual behavior based on the time of day for every type of interaction between individuals; (3) We implement a scalable, fully distributed simulator and we evaluate its performance on two platforms; (4) We validate the results of the simulation against another epidemic simulator. We additionally perform an extensive analysis of the effects of the features specific to our approach on the results of the simulations.

The rest of the paper is organized as follows. Section 2 discusses the related work. Section 3 describes the modeling task and the simulation algorithm. Section 4 presents an study of the performance and simulation results of EpiGraph. Section 5 summarizes the paper with the conclusions and some directions for future work.

## 2. Related work

*Interconnection networks:* The majority of human-transmitted infectious diseases use physical contact as the main transmission mean. For this reason the dynamics of the propagation is tightly related to the structure and the characteristics of the network of connections between the individuals within a population [2], [3], [4], [5], [6]. Typically epidemiological models are compartmental in the sense that they model the dynamics of the epidemics by nonlinear differential equations and do not model the topology of the contact network. The assumption is that individuals in a population are homogeneously connected, which means that all individuals have the same probability of infecting other individuals [5]. In reality each person has specific, possibly very different, interaction patterns. This makes the interconnection network be heterogeneous [7], [5]. Additionally, there tend to be few people who have many connections, some strong but most of them weak—these are

the "core groups"—while most of the individuals have few connections [8], [9].

The typical way to approximate a heterogeneous contact network is to build a contact graph in which the individuals are nodes and edges represent connections [10], [11], [12]. A straightforward model implements the graph as an adjacency matrix. We use a more sophisticated model in which each matrix cell holds a value that represents the type of social interconnection: study, work, leisure, or family. The patterns of interactions depend on whether they occur between individuals within the same group or from different groups. We additionally allow the type of interconnection to change depending on a time parameter to reflect the fact that we may interact with individuals from different group types at different times during the day. This approach allows to more accurately model the heterogeneity of the actual contact network.

Work such as HPCgen and Epigrass [13], [14] take the approach of modeling actual populations; FastGen and CL-model [15], [16] choose instead to generate a random adjacency matrix. HPCgen uses actual demographic data from census data and interviews, and introduces the idea of generating the contact network based on social structures with arbitrary degree distributions following a Poisson distribution. To work well HPCgen requires a very high accuracy when modeling the social contacts for a specific population. The contact network is fully static in the sense that the interconnections between individuals cannot change during simulation. Experiments have shown that such a model is accurate in the case that the propagation rate of the infection is high relative to the rate with which the interconnections may change in the network [17], but would break down otherwise.

[18] presents a large-scale simulator based on a stochastic model for influenza. It uses a molecular dynamic algorithm for modeling the interactions between individuals. Their approach is computationally expensive, requiring extended simulation times and a large number of processors to complete. In contrast, EpiGraph has lower computational requirements and can simulate single individuals with specific characteristics and dynamically evolving interactions.

A different approach is followed by BioWar [19]. Biowar is a multiagent network model for simulating the effects of epidemic outbreaks due to bioterrorism attacks. It takes into account several input models such as disease, geography, weather, attack and communication technology, also it models the population behavior distributed in social group types with real census data. InfluSim [1] extends the SEIR epidemic model. It uses demographic information from real census data and it models the social structure based on different age groups. InfluSim uses differential equations to model the transmission of the disease and does not take into account time-dependent individual interactions, such as EpiGraph does.

*Epidemic models:* The typical mathematical model for simulating epidemics is the SIR model [20]. The SIR model is usually appropriate for infectious diseases which confer immunity to recovered individuals and it works best if demographic effects may be neglected. Our work focuses on the propagation of the influenza virus over short to medium time spans. Work in [21] extends the mathematical model with latent, asymptomatic, and dead states, as well as the possibility of introducing a vaccine program. The latent state corresponds to the incubation state in which an individual is infected but has not yet developed symptoms. A relatively small percent of the population will never develop them, passing into an asymptomatic state. All asymptomatic individuals, together with a high percentage of infected individuals recover and become immune. The rest of them pass to the dead state. EpiGraph builds on this model and extends it to introduce a new hospitalized state.

[22] proposes a more detailed model for the dissemination of the influenza virus. In their approach the susceptible cases first go to a latent stage that is non-infective. This can transition either to an asymptomatic stage which leads to removal, or to a second latent stage with some contagion degree, followed by two contagious stages with different contagion degrees. Treatment is applied only during the first infective stage.

# 3. The modeling task

EpiGraph consists of two main components: (1) a model for the population under study with the patterns of contact between individuals within this population, and (2) a model of how the participating agents spread the disease. This work focuses on the dissemination of the flu virus over a short to medium length time span. Our goal is to facilitate the understanding and prediction of how the virus spreads within specific populations with possibly dramatically different interaction patterns over short and medium time spans. We do *not* focus on extended time periods during which qualitatively different parameters may make a difference. For instance, in our model there is no entry into or departure from the population, except possibly through death from the disease. This is a reasonable hypothesis in case of short to medium time spans. On the other hand we are modeling interaction features that may have a large impact in the case of a single epidemic outbreak but whose effects level out over time. Generally diseases transmitted by viral agents confer immunity so the assumption is that if an infected individual recovers he will acquire immunity for a time period at least as extended as the simulation time for the infection.

In the social model each graph node models a single individual and may have specific characteristics such as gender, age, role, as so on. Each graph edge represents an interaction between two individuals and depends on the time of the day. That is, EpiGraph can capture heterogeneity

features at the level of both the individual and each of his interactions.

The social model is based on two data sources: actual demographic information, as well as a realistic model of social interactions. These are used to build graphs for both intra- and inter-group interactions. A group is a collection of individuals of the same group type as extracted from the demographic information. The complete graph is then used as an input for the epidemic model. This model captures the characteristics that are important in the process of spreading a contagious agent, is specific to the agent under study, and needs to make assumptions such as what is the subset of susceptible individuals that an infected individual may pass the agent to. Rather than assuming a distribution or generating synthetic interaction graphs, we use real information from social networks to model the social interaction patterns. The interaction network is built statically to reflect the existence of communication between individuals but abstracts away the timing for these interactions. To recover the dynamic nature of these interactions we introduce a time component depending on which an individual may interact with any number of other individuals following his own patterns.

## 3.1 Modeling the population

To most faithfully simulate the effects of an infectious agent spreading through a specific population we decided to use real instead of synthetic data. We use real demographic information obtained from the Primary Metropolitan Statistical Area of Philadelphia [23] to determine the distribution of the population in group types; these typically show different patterns in terms of social interactions. The group types which we extracted from the census and which we are modeling are the following: (1) school-age children and students, (2) workers, (3) stay-home parents, and (4) retired individuals. The population is split into many groups of each of these types—a structure which reflects the way individuals tend to associate with each other in terms of social contacts. Each individual has contacts within his own group as well as with individuals from other groups. Let's take the example of a worker. She's going to interact frequently with people from the same work group during work hours, with friends during leisure hours, and with family during evening/night hours. We therefore model three kinds of interactions: (1) between individuals of the same group, (2) between individuals of different groups, and (3) between members of the same family. Each of these kinds of interactions is assigned to a specific daily time frame depending on the schedule for the main activity—work, study, etc—, for leisure activities, and for family time. This makes the simulation more realistic, particularly over short time periods.

*Intra-group connections:* Which specific group an individual belongs to determines the actual number and patterns of interactions with other individuals from his own group. One of the contributions of our work is that we model intra-group communications by scaling down real interaction graphs extracted from the Social Networks (SN) of Enron and Facebook. The idea is to exploit the connectivity that exists in real business and leisure SNs. The graph extracted from the Enron email database consists of 70,578 nodes and 312,620 connections, while Facebook has 250,000 nodes and 3,239,137 connections. We use Enron's SN to model the worker and retired groups and Facebook's to create the school and stay-home groups. Note that the SNs are bigger than the generated groups. We scale each down by selecting as many random entries of the SN as group members, then connecting the nodes following the same patterns as those in the SN. The selection of random entries of the SN allows us to create different structures for each group. This approach is more realistic than either synthetically generating the interaction graphs or using discrete probability distributions to approximate the number of individual interactions.

*Inter-group and family connections:* We create a number of intergroup contacts per individual based on the group characteristics which the individual belongs to. Mostly the inter-group contacts occur in the hours between finishing one's main daily activity—such as work or study—and going home in the evening, or during weekends. These reflect daily activities which occur in public places such as parks, gyms, public transport, coffee shops, where one generally interacts with unknown people or friends pertaining to a different group. The connections of inter-group contacts are generated at the level of the group based on a set of percentages which reflect the degree to which groups of specific types are connected. There are two types of connections between pairs of groups: strong and weak. Probabilistic parameters decide whether two groups are strongly connected, weakly connected, or are not connected at all. In addition to intra- and inter-group contacts we also model a different type of social interaction: the contacts one has with members of his family. These may be pertaining to the same or to a different group and one has contacts with them from late night to morning, and during the weekends.

*Strong vs. weak ties:* Interactions between groups may be either strong or weak. This reflects the degree to which the connection may serve as a channel for spreading the infectious agent. Strongly coupled groups tend to be the ones who spend many hours in contact, either for affinity, family, or work related reasons. On the other hand, weak connections are between groups that only share a few contacts. It reflects occasional or casual contacts between individuals.

*Data structures:* EpiGraph models interactions between individuals via a graph. To represent it we are using sparse matrices in Compressed Sparse Column format which enables both optimized matrix operations and an efficient way to distribute and access the matrices in parallel.

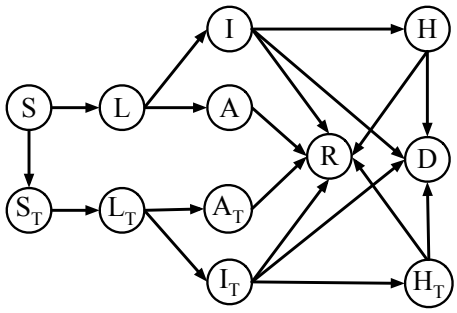


Fig. 1: State diagram for the epidemic model.

## 3.2 Modeling the infectious agent

The basic epidemic model is based on the principles of the SIR model as it is described in [20] and extended for the case of the flu virus by [21]. The extended model consists of a set of additional states—latent, asymptomatic, and dead—which reflect real possible stages during the development of the infection within a host. We further enhance the model with a hospitalized state in which an individual's contacts are severed. Having such a state is important when simulating realistic cases where hospitalization may be needed in order to curb the effects of the epidemics.

Figure 1 consists of two symmetrical subgraphs; the upper part has states with non-subscripted names, the lower part consists of subscripted states. Let's focus on the non-subscripted subset of the states for the time being. A susceptible individual in state `S` may be infected by another individual and pass to the latent—or incubating—state `L`. From here he normally goes to the infective state `I`, but may also become asymptomatic and go to state `A`. Individuals which are asymptomatic will always recover and go to state `R`; infective individuals may recover, get hospitalized, or die. A hospitalized individual in state `H` either recovers or dies. In the case of the flu virus we assume that recovery implies immunity over short and medium time spans such that a recovered individual will not get infected again during the time of the simulation.

The epidemic model for influenza has many parameters, some of the most important being the basic reproduction number $R_0$ (average number of secondary cases of infections which produces an infected individual), the time an individual spends in each of the states, the probability that an individual will take a transition from a source state into each of the target states, and so on. The time each individual spends in a given state is generated following a Gaussian distribution to faithfully simulate the time ranges which are specific to the stages of a flu infection. The probability of infecting another individual while incubating depends on whether the specific connection is low or high risk. A high risk interaction reflects a contact between individuals which has high probability to transmit the infection. For instance, these may be interactions between members of the same work team in a company or between friends in a classroom. On the other hand, low risk connections are related to contacts that have a low probability for disease transmission. For instance, these may be contacts between members of different work teams in the same company.

We adopted most of the concrete values for the model parameters from the existing literature on flu epidemics [21], [24], [25]. The epidemic model also receives as an input the social model constructed in the previous step.

*Vaccination:* Our simulator provides for the possibility of vaccinating a subset of individuals either before the outbreak of the epidemics or at any other point during the outbreak. The lower half of Figure 1 consists of subscripted states which reflect the susceptible, latent, asymptomatic, infectious, and hospitalized states for the case of vaccinated individuals. The figure contains a transition from state `S` to state $S_t$ which reflects the adoption of a vaccination policy for susceptible individuals. Since in case of the flu virus no symptoms are evident during the latent period it is in reality possible to vaccinate individuals either in the latent or in the asymptomatic state. We assume that getting vaccinated when are states `L` or `A` does not make any difference with respect to the individual's response to infection. Vaccination has specific implications such as: reducing the susceptibility of getting infected at the time of contact with an infected individual, reducing the probability of infecting another individual, reducing the recovery time, and reducing the possibility of becoming symptomatic. Vaccination is implemented such that it is possible to control the number of vaccines available and the probability of it succeeding when applied to a specific individual. Due to the fact that only part of the population is susceptible as result of a vaccination program we now use for the subscripted cases a control reproduction number $R_v$ instead of the basic reproduction number $R_0$.

In case of an epidemics the period of time between its onset and the time when a vaccine becomes available is usually problematic because of the lack of understanding of the effects of the timing when the vaccine is administrated. Our simulator allows analyzing the effects of implementing a vaccination program at different times throughout the dissemination of the infectious agent. One of the advantages of our epidemic model is that it is possible to monitorize the effect of interventions such as vaccination or hospitalization for each individual. It is therefore possible to simulate various scenarios like vaccinating or insolating a specific collective, for instance, the members of a specific company or school, or a given city area.

## 3.3 The simulation algorithm

Our simulation algorithm uses as inputs both the social model as well as the epidemic model. The social model provides the intra-group connections for each individual; these are the paths through which the infectious agent may propagate and they may be either low-risk or high-risk. The

epidemic model captures the states that each individual goes through during an epidemics and the probabilities for taking transitions from a given source to a specific destination state. The simulation algorithm processes each connection of every individual to generate a probability with which the connection will serve for transmitting the infection. This probability depends on: (1) The connection type and current time: the connection types are intra-group, inter-group, and family, and each of them corresponds to a specific daily time slice; and (2) The current state of the individual: this is the current state in the epidemic model plus other factors like the group which he belongs to, age, etc.

### 3.4 Performance issues

EpiGraph has been designed as a fully parallel application. It employs MPI [26] to perform the communication and synchronization for both components of the simulator: the contact network model and the epidemic model. This approach has two main advantages. First, it can be executed efficiently both on shared memory architectures—for instance multicore processors—and on distributed memory architectures—such as clusters. On both platforms EpiGraph successfully exploits the hardware resources and achieves a significant reduction in execution time relative to a sequential implementation. The second advantage is that the simulator scales with the available memory. Given that all the data structures are evenly distributed, the size of the problems that can be simulated grows with the number of computational resources.

## 4. Results

Our main simulation scenario is the population of the Primary Metropolitan Statistical Area of Philadelphia, U.S. We used [23] to extract statistical demographic data for the city and we created a *basic scenario* with the following characteristics. The city has 3,849,647 inhabitants with the following distribution: 27.95% school-age children, 43.62% workers, 14.52% stay-home parents, and 13.92% retired individuals. The interconnection graph has 160 millions of contacts, on average 41 per inhabitant. Working hours are from 9am to 5pm, leisure time is from 5pm to 7pm, and time spent at home—family and sleep time—is from 7pm to 9am. We consider 13,181 groups of workers; 8,513 groups of school-age children corresponding to classrooms; 4,192 groups of stay-home parents corresponding to friends that share activities such as shopping or walking; and 4,314 groups of retired individuals. We use Gaussian distributions to assign a size to each group; the mean size for each of the four group types is 261, 39, 12 and 8.

Figure 2 displays in logarithmic scale the number of individuals in each epidemic state during a simulation of 200 days for our basic scenario. This scenario includes the following parameters extracted from [24]: the basic reproduction number $R_0 = 1.373$, the factor by which the

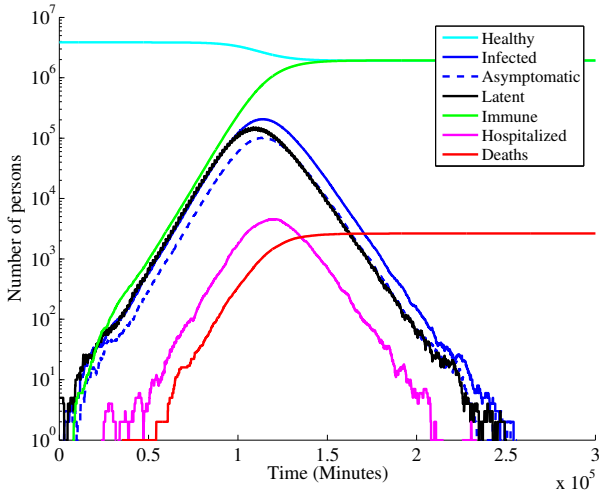


Fig. 2: Epidemic propagation for the basic scenario and a 200-day simulation.

infectivity of asymptomatic individuals is reduced $\delta = 0.5$, the probability that susceptibles become asymptomatic $p = 0.33$, the latent period for influenza 1.9 days, the infective period for influenza 4.1 days, and the hospitalization period 3 days. We can observe that the infection lasts approximately 175 days and its peak is around day 82.

We have performed a number of experiments in order to evaluate the strengths of EpiGraph. These experiments address three different properties of the simulator: (1) the prediction accuracy of the mathematical epidemic model, (2) the ability to accurately model highly heterogeneous scenarios where each individual and her connections may be customized, and (3) the performance and scalability of the simulator.

### 4.1 Validation of the EpiGraph model

In order to evaluate the accuracy of our mathematical model we compare the simulation results of EpiGraph with those of InfluSim [1]. In order to perform a comparison we used in both simulators the population and epidemic parameters of the basic scenario. Table 1 shows the number of susceptible, immune and dead individuals for each simulator. Results show deviations of 3.30%, 2.97% and 8.04% in the number of susceptible, immune, and dead individuals. Another aspect that we have considered is the numerical stability of EpiGraph under different conditions. More specifically, we have analyzed the variability of the results for two cases: when EpiGraph is executed several times with the same input parameters and when it is executed using different time step durations.

Table 1: InfluSim and EpiGraph results.

| State | InfluSim | EpiGraph | Deviation |
|---|---|---|---|
| Susceptible | 2,023,187 | 1,930,773 | 3.30% |
| Immune | 1,837,305 | 1,916,226 | 2.97% |
| Deaths | 2,362 | 2,647 | 8.04% |

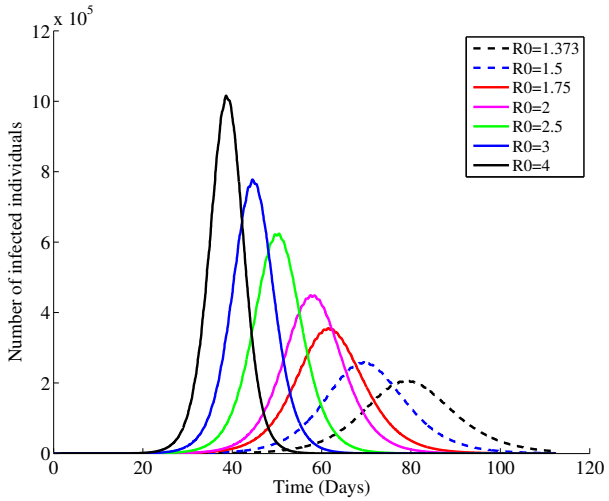


Fig. 3: Impact of different basic reproduction numbers on the number of infected. Basic scenario.

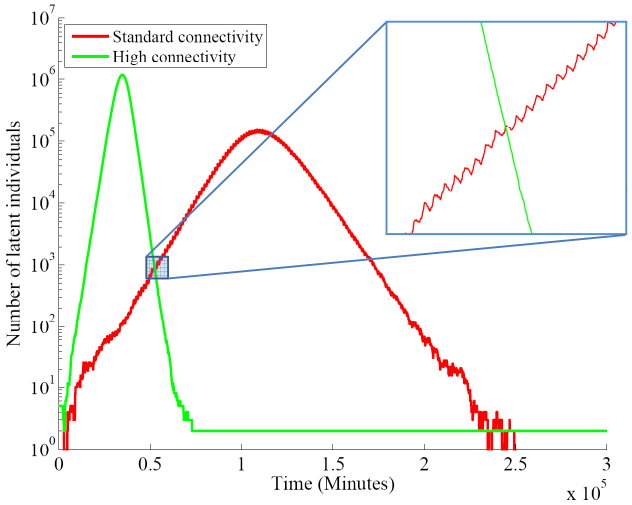


Fig. 4: Effect of different graph configurations on the latent cases. 200-day simulation.

The time step determines the frequency of computations for each individual. By default we use a 10 minute step, which means that we apply the propagation model and update the system state six times per hour. A smaller time step implies a more detailed simulation at the expense of a longer execution time. We execute the basic scenario using the following time steps: 1, 5, 10, 30, 60 and 120 minutes. We observe that the loss of accuracy when using larger steps is not important. More specifically, the peak of infected individuals for all of these executions reaches a mean value of 205,168 with a standard deviation (in percentage of the mean value) of 1.17% and confidence interval of 1.21%. This peak is reached at the simulation time of 118,770 minutes, with a standard deviation of 2.83% and confidence interval of 2.97%.

To evaluate the variability of EpiGraph we run ten times the same scenario with the same initial conditions, including the same set of individuals that are initially infected. After repeatedly simulating the epidemics for 200-day intervals, results show a variability in the number of immune individuals of 0.28%. Similar results are obtained for susceptible and dead individuals. Based on these results we conclude that EpiGraph is able to precisely model the epidemic with a small variability in the results.

## 4.2 Exploiting the features of EpiGraph

EpiGraph employs a highly detailed social model which allows customizing the interactions of each individual as well as the effect of time on the individual relationships. These features allow the simulation of infection and transmission process for individual cases.

We have performed experiments aimed at evaluating the effect of different basic reproduction numbers and different graphs structures on the epidemic propagation. Figure 3 evaluates the effect of different reproduction numbers. We can see that the epidemic propagation is faster and the number of infected individuals is larger when the basic reproduction number grows. For instance, for values of $R_0$ of 1.373, 2, and 4 the overall numbers of infected individuals are 1,933,901 and 2,783,435 and 3,597,751, respectively.

We evaluated two different graph structures called standard connectivity and high connectivity. Standard connectivity corresponds to the basic scenario; high connectivity corresponds to a scenario where the graph is flattened. Specifically we are considering only the graph connections corresponding to workers and we assume that the working hours are from 9am to 9am of the next day. That is, in this case we are considering a global graph that contains only one group type which is active during the whole day. Figure 4 illustrates the evolution of the latent cases for the scenarios of standard connectivity and high connectivity; infected cases exhibit a similar behavior. The figure shows that differentiating between social groups has a significant impact on the evolution of the epidemics. We can observe that when we assume standard connectivity there exists a periodic variation of the latent cases. This is related to the existence of different daily time slices that exhibit different propagation patterns. In the case of high connectivity this pattern doesn't appear due to the unique time interval, that of working hours.

We have evaluated the effect of different vaccination policies on the basic scenario. Figure 5 shows the evolution of the infected cases for five different strategies: no vaccination (reference), vaccination at the beginning of the outbreak, before reaching the peak of the outbreak (day 52), at the peak of the outbreak (day 82) and after the peak (day 97). For each of these cases 28% of the population is vaccinated and the reproduction number for vaccinated people is $R_v = 0.047$ [25]. We can observe the following behavior: vaccinating at day 0 is the most efficient approach in terms of minimizing the number of infected individuals. When vaccinating at day 52 there is a large number of

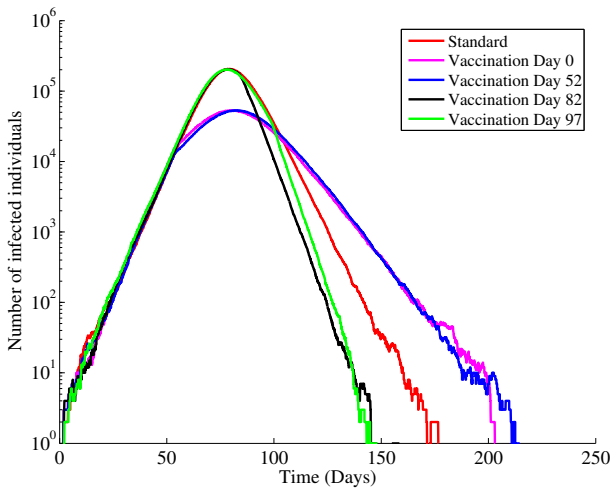


Fig. 5: Impact of different vaccination strategies. Basic scenario, 200-day simulation.

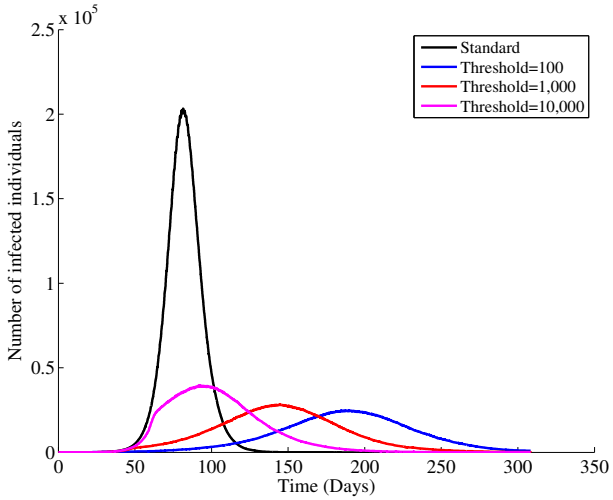


Fig. 6: Impact of different quarantine policies. Basic scenario, 300-day simulation.

individuals in infected and latent stages; the vaccine reduces the number of infected cases but also delays its propagation, thus increasing the duration of the outbreak. This effect is also manifested when vaccinating at day 0. In contrast, for the vaccination campaigns at days 82 and 97 the peak of infected cases has already been reached; vaccination thus contributes to an early end of the outbreak.

Lastly, we evaluated different quarantine policies. For the basic scenario we specify a given threshold in number of infected cases. When this threshold is reached all the school and work activities are cancelled, keeping only two leisure hours per day; during the rest of the day all the individuals stay at home with their family. Figure 6 shows the simulation results when quarantine is applied based on different threshold values. We observe that there is a decrease in the number of infected at the expense of a larger propagation time.

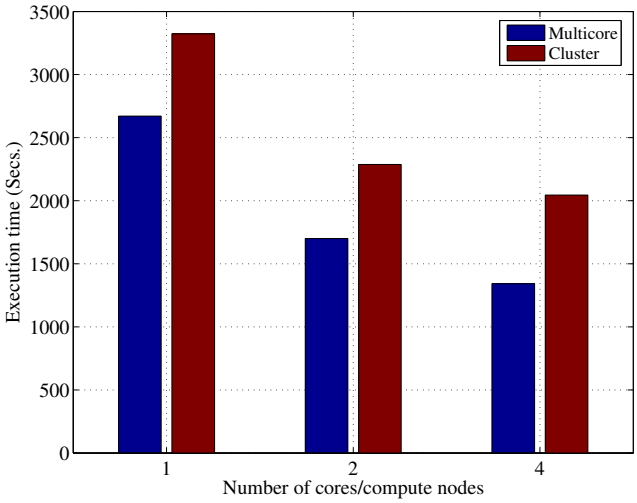


Fig. 7: EpiGraph execution time on a multicore processor and a cluster. Basic scenario, 200-day simulation.

### 4.3 Performance evaluation

We measured the execution time of EpiGraph on two different parallel architectures: a multicore processor and a cluster. The multicore is an Intel Xeon X7350 quadcore processor with a frequency of 2.93 GHz, 3 MB of cache and 16GB of RAM. The cluster consists of 4 computers connected with a GigaBit network, each of them with a single Intel Xeon E5405 at 2GHz with 6MB of cache and 4GB of RAM. Figure 7 shows the EpiGraph execution time for the basic scenario when simulating 200 days of epidemic outbreak. Given the faster interconnection system of the multicore architecture, this achieves better performance than the cluster system. We can observe that in both cases EpiGraph reduces its execution time when more processors are used.

## 5. Conclusions

This paper presents a novel approach to modeling the propagation of the flu virus via a realistic interconnection network based on actual individual interactions extracted from social networks. We have implemented a scalable, fully distributed simulator and we present an extensive analysis of the effects of the new features of our approach on the results of the simulations. Work in progress and future work involve studying the effects of introducing new states in the epidemic model and making use of the individual values such as age and gender in implementing different social and medical propagation characteristics. We are also interested in investigating the characteristics of our social models—such as clustering, node distance, and so on—and estimate to what degree disease propagation occurs differently for different types of real social networks.

# Acknowledgements

The work has been performed under the HPC-EUROPA2 project (project number: 228398) with the support of the European Commission-Capacities Area-Research Infrastructure and the Spanish Ministry of Science and Education under the MEC 2011/00003/001 contract.


# References


[1] M. Eichner, M. Schwehm, H. P. Duerr, and S. Brockmann, "The influenza pandemic preparedness planning tool influsim," *BMC Infectious Diseases*, vol. 7, no. 17, pp. e–pub, 2007.

[2] M. J. Keeling and K. T. D. Eames, "Networks and epidemic models," *Journal of The Royal Society Interface*, vol. 2, no. 4, pp. 295–307, Sept. 2005. [Online]. Available: http://dx.doi.org/10.1098/rsif.2005.0051

[3] I. Doherty, N. Padian, C. Marlow, and S. Aral, "Determinants and consequences of sexual networks as they affect the spread of sexually transmitted infections," *The Journal of Infectious Diseases*, vol. 191, no. S1, p. 42–54, 2005.

[4] K. T. D. Eames and M. J. Keeling, "Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases." *Proc Natl Acad Sci U S A*, vol. 99, no. 20, pp. 13 330–13 335, Oct. 2002. [Online]. Available: http://dx.doi.org/10.1073/pnas.202244299

[5] S. Bansal, B. T. Grenfell, and L. A. Meyers, "When individual behaviour matters: homogeneous and network models in epidemiology," *Journal of The Royal Society Interface*, vol. 4, no. 16, pp. 879–891, Oct. 2007. [Online]. Available: http://dx.doi.org/10.1098/rsif.2007.1100

[6] R. M. Christley, G. L. Pinchbeck, R. G. Bowers, D. Clancy, N. P. French, R. Bennett, and J. Turner, "Infection in social networks: Using network analysis to identify High-Risk individuals," *American J. of Epidemiology*, vol. 162, no. 10, pp. 1024–1031, 2005.

[7] L. A. Meyers, B. Pourbohloul, M. E. Newman, D. M. Skowronski, and R. C. Brunham, "Network theory and SARS: predicting outbreak diversity." *Journal of theoretical biology*, vol. 232, no. 1, pp. 71–81, Jan. 2005. [Online]. Available: http://dx.doi.org/10.1016/j.jtbi.2004.07.026

[8] M. E. J. Newman, *The spread of epidemic disease on networks*, Apr. 2002. [Online]. Available: http://arxiv.org/abs/cond-mat/0205009

[9] J. M. Read, K. T. Eames, and W. J. Edmunds, "Dynamic social networks and the implications for the spread of infectious disease." *Journal of the Royal Society, Interface / the Royal Society*, vol. 5, no. 26, pp. 1001–1007, Sept. 2008. [Online]. Available: http://dx.doi.org/10.1098/rsif.2008.0013

[10] A. Vazquez, "Spreading dynamics on heterogeneous populations: Multitype network approach," *Phys. Rev. E*, vol. 74, no. 6, p. 066114, Dec 2006.

[11] F. Harary, *GRAPH THEORY*. Addison Wesley Longman Publishing Co, 1969. [Online]. Available: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/B000OLF0P0

[12] D. B. West, *Introduction to Graph Theory*, 2nd ed. Prentice Hall, Sept. 2000.

[13] T. Zhang, S. H. Soh, X. Fu, K. K. Lee, L. Wong, S. Ma, G. Xiao, and C. K. Kwoh, "Hpcgen a fast generator of contact networks of large urban cities for epidemiological studies," in *International Conference on Computational Intelligence, Modelling and Simulation*, 2009, pp. 198 –203.

[14] F. C. Coelho, O. G. Cruz, and C. T. Codeco, "Epigrass: a tool to study disease spread in complex networks." *Source code for biology and medicine*, vol. 3, no. 1, Feb. 2008. [Online]. Available: http://dx.doi.org/10.1186/1751-0473-3-3

[15] S. Eubank, A. V. S. Kumar, M. V. Marathe, A. Srinivasan, and N. Wang, "Structural and algorithmic aspects of massive social networks," in *SODA '04: Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2004, pp. 718–727. [Online]. Available: http://portal.acm.org/citation.cfm?id=982792.982902

[16] F. Chung and L. Lu, "Connected components in random graphs with given expected degree sequences," *Annals of Combinatorics*, vol. 6, pp. 125–145, 2002, 10.1007/PL00012580. [Online]. Available: http://dx.doi.org/10.1007/PL00012580

[17] E. Volz and L. A. Meyers, "Susceptible-infected-recovered epidemics in dynamic contact networks," *Proc Biol Sci.*, vol. 274, no. 1628, 2007.

[18] T. C. Germann, K. Kadau, I. M. Longini, and C. A. Macken, "Mitigation strategies for pandemic influenza in the united states," *Proceedings of the National Academy of Sciences*, vol. 103, no. 15, pp. 5935–5940, Apr. 2006. [Online]. Available: http://dx.doi.org/10.1073/pnas.0601266103

[19] K. Carley, D. Fridsma, E. Casman, A. Yahja, N. Altman, L.-C. Chen, B. Kaminsky, and D. Nave, "Biowar: scalable agent-based model of bioattacks," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 36, no. 2, pp. 252 – 265, 2006.

[20] R. M. Anderson, R. M. May, and B. Anderson, *Infectious Diseases of Humans: Dynamics and Control*, new ed ed. Oxford University Press, USA, Sept. 1992. [Online]. Available: http://www.worldcat.org/isbn/019854040X

[21] F. Brauer, P. v. d. Driessche, and J. Wu, Eds., *Mathematical Epidemiology*, 1st ed. Springer, June 2008. [Online]. Available: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/3540789103

[22] M. E. Alexander, C. S. Bowman, Z. Feng, M. Gardam, S. M. Moghadas, G. RÃűst, J. Wu, and P. Yan, "Emergence of drug resistance: implications for antiviral control of pandemic influenza," *Proceedings of the Royal Society B: Biological Sciences*, vol. 274, no. 1619, pp. 1675–1684, 2007.

[23] *U. S. Census Bureau*, http://www.census.gov/.

[24] I. M. Longini, E. M. Halloran, A. Nizam, and Y. Yang, "Containing pandemic influenza with antiviral agents," *Am. J. Epidemiol.*, vol. 159, no. 7, pp. 623–633, Apr. 2004. [Online]. Available: http://dx.doi.org/10.1093/aje/kwh092

[25] L. R. Elveback, J. P. Fox, E. Ackerman, A. Langworthy, M. Boyd, and L. Gatewood, "An influenza simulation model for immunization studies." *American Journal of Epidemiology*, vol. 103, no. 2, pp. 152–165, 1976. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/814808

[26] *MPI: A Message-Passing Interface Standard*, Message Passing Interface Forum, 1995.