

# Modeling the Spread of Infectious Diseases using an Agent-based Spatial Network

George B. Bezerra\* and ThanhVu H. Nguyen

*Department of Computer Science  
Adaptive Computation Lab  
University of New Mexico*

## I. INTRODUCTION

Network models for the spread of infectious diseases in urban areas can be divided into two major classes. The first and more traditional class of models comprises the contact network models [10], [1]. Such networks consist of an undirected graph where nodes represent agents (e.g., people) and edges correspond to the potential interactions among these agents, by means of which the transmission of the disease can occur. Weights associated to the edges define the probability of transmission in case one of the agents is infected. In those models, different network properties are implemented so that their influences to the dynamics of the outbreak can be investigated. A better understanding can then be used to develop new means to deal efficiently with disease outbreaks in the real world. For example, one might be interested in understanding how the removal of the hubs in a power law network would affect the spread of the disease; analyzing the influence of the clustering coefficient or connectivity degree of the network to the velocity in which the disease spreads; simulating the effect of a certain vaccination strategy in different network topologies, etc. [11], [12], [13].

Although these models have served well to increase our knowledge about the dynamics of outbreaks in idealized situations, they are still very limited and do not represent important aspects determinant to these phenomena. First of all, the static structure of the network is usually biased. It is not possible to know in advance which and how people interact, and thus a static network in which interactions are pre-defined is not adequate. Moreover, making the edges correspond to interactions with the acquaintances of a certain individual, for example, is unrealistic, since people often acquire the infection from people they don't know.

Another problem is that there is no spatial information in such models. For example, the disease tends to spread to the nearby regions first before reaching farther places [2]; regions of the space may have increased density than others and thus have different transmission rates. These aspects have strong influence to the dynamics of the infection and cannot be represented in contact network models. Besides, measures to deal with infections that consist of isolating certain regions in a city or vaccinating specific neighborhoods cannot be implemented.

The locations where people meet each other is another important information that is usually not represented in the network models. Infections occur at specific places, like schools, churches, theaters, etc. Whether the place is crowded or not or whether it attracts people from different and farther places, for example, will shape the spatial and temporal patterns of the transmission.

Finally, nodes have no internal dynamics and have no behavior. In real life people are constantly moving about and their daily routine is going to define dynamically with whom they will most frequently interact. Acquiring a disease might change the behavior of the agent, and that will modify the places it usually visits, consequently altering the structure of the interaction network of that agent. Thus, the behavior of the agents is crucial to the dynamics of the infection, and potential change in this behavior confers plasticity to the network.

While the contact network model is over simplistic, the second class of models, the so-called realistic or detailed models [14], represents the other extreme in the complexity scale. Realistic models like Episims [4] intend to simulate an entire city in its minimal details. Synthetic households have gender, role in a family, age, and well-defined daily activities and schedules. Episims estimates the coworkers of a given household, the type of transport it uses, and the time of travel to different locations. The model creates an almost living city where the spread of an infection can be simulated.

Although a myriad of details are included in these models in order to increase their realism, the tremendous complexity that results from this approach has disadvantages. First, such models are extremely computationally intensive. Simulating an entire city may take days even in a supercomputer, and researchers often avoid having to perform these expensive experiments. Many of the works on Episims, for example, use the detailed estimated data to generate a static contact network that can be more easily analyzed and does not require simulations [4], [3].

Second, more “realism” in a scientific model does not imply more accuracy. The realistic models have dozens of parameters to be adjusted [5], which consist of a great source of noise. A slight change in a single parameter may potentially affect the entire output of model — it becomes hard to distinguish between what is really fundamental to the phenomenon being modeled and what is a superfluous detail. It is well known from Model Selection theory that a model should be made as simple as possible and in order to be robust and reliable [8], and the addition of any extra parameters generally requires an exponential increase in the amount of data to validate this model, an effect called curse of dimensionality [9]. For the spread of infectious diseases in urban areas for which the data is often noisy, sparse, and outdated, a model with increased complexity is not justifiable, and may potentially result in decreased reliability [5].

In this paper we propose a network model for the spread of infectious diseases that overcomes the main limitations of the contact networks by considering spatial information, explicit locations, and agent behavior in its representation. The model is also very simple and has a small number parameters

— it does not require a supercomputer to run and avoids the complexity issues of realistic approaches.

In order to demonstrate the applicability of the model, we perform experiments that contrast three different measures to control outbreaks in a city. The results obtained provide important insights to the spread dynamics of an infection in urban areas as well as to how this spread can potentially be controlled.

We acknowledge that other models exist that do not exactly fit into the two categories described above [15], [16], [17]. Since it is not possible to address all of them, we focused on the ones we consider most relevant to this work.

## II. THE MODEL

The model proposed consists of a bipartite graph where the agents (people) are connected to locations (schools, offices, restaurants, etc.), and there are no connections between locations or among people. Each agent has a certain number of connections to locations, representing the places it usually goes to. During the simulation, each agent randomly traverses its own set of known locations, where they will potentially meet other people that also visit those places, and a transmission can take place if one or more of them are infected. Each location has a capacity, that is, the number of people that usually go there, and the connectivity of the location is proportional to its capacity. These capacities are distributed according to a power-law with a negative exponent  $k$ , so that there is a large number of locations with very small capacity, and a small number of locations with very large capacity.

Using the data provided by [6] about the city of Portland in Oregon, we can generate graphs that resemble a real city in terms of the properties being modeled here. According to this data, a person goes on average to approximately 3.5 different places a day, and the size of the population is about 8 times the number of locations. The exponent  $k$  of the power-law that defines the capacity and connectivity of a location is reported to be  $-2.8$ . In order to generate our bipartite graph, the position of the locations in a plane is first defined at random and their sizes follow the distribution defined by the power-law. People are also assigned to random initial positions in this plane, and each person will visit (i.e., be connected to) 3 different locations.

Our last detail concern is defining the locations a given person visits — this is when the spatial information of the model is introduced. Following the idea of the gravity model used in transportation theory [6], a person tends to go more often to places that are close to where they live, thus the probability of having a connection to a location is inversely proportional to the distance to that location. Furthermore, people can travel longer when the capacity of a location is greater, since there are fewer of them. Thus, in general one would travel a long distance to go to the Walmart or to the city’s airport (which have great capacity and are in relatively small number) if they have to, but they would probably not have to go far in order to find a fast-food restaurant or a grocery store, which are in much greater number. Therefore, the probability of making a

connection to a location should also be directly proportional to the capacity of the location.

In order to implement this idea a hill curve will be used as indicated in figure 1, where the  $x$ -axis corresponds to the distance of a location relative to the initial position of the agent divided by the capacity of the location, and the  $y$ -axis represents the probability of the agent in establishing a connection to that location. Thus, according to the figure, the shorter the distance and the greater the capacity the greater the probability of establishing a connection with that location. Gravity model was also used in [2] to model infections diseases, though the implementation presented here is slightly different.

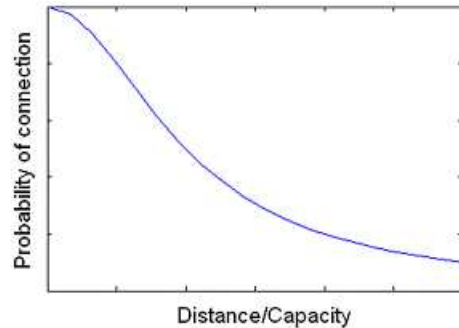


FIG. 1: Hill curve modeling the probability of connection to a location.

By defining random positions for the locations and for the agents, using a power-law to define the capacity of the locations, and using the gravity model described above to set the connections of the graph, a network as the one shown in figure 2 can be generated.

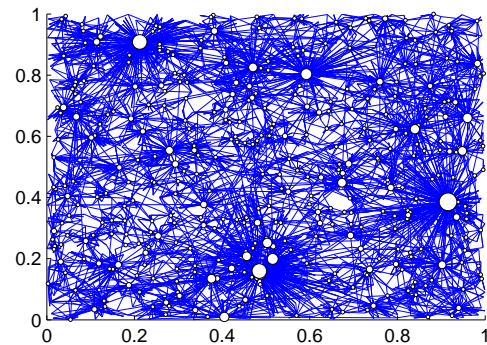


FIG. 2: Sample city generated using the rules described in the text.

The figure does not show the agents explicitly, only their connections to the locations. The radius of the circle representing a location is proportional to its capacity. Note that the larger the circle, the larger its connectivity in the graph and the longer these connections tend to be.

The city in figure 2 has 3000 agents and 375 locations. It is small instance of an urban area with the same ratio between

population size and number of locations as Portland. Larger samples could be obtained with a linear increase in the computational cost of the simulation, but for this paper we restrict our virtual cities to small samples.

As said before, at each iteration every agent moves to a random location to which it is connected. In a given location the agent is paired to another randomly chosen agent that is also in that location, and if one of them is infected there is a probability  $p$  that this infection will be transmitted. We start the city with only one infected seed and watch the spread of the disease over the city after a number of iterations. Figure 3 depicts this process with  $p$  set to 0.1. Healthy people are represented as black dots, and infected ones as red dots. Figures 3a, b, c, and d show the respective states of the network after 10, 20, 30, and 40 iterations. Notice the network is now different from the one in figure 2 — the statistical properties that generated the graph are the same, but the sample is a different one.

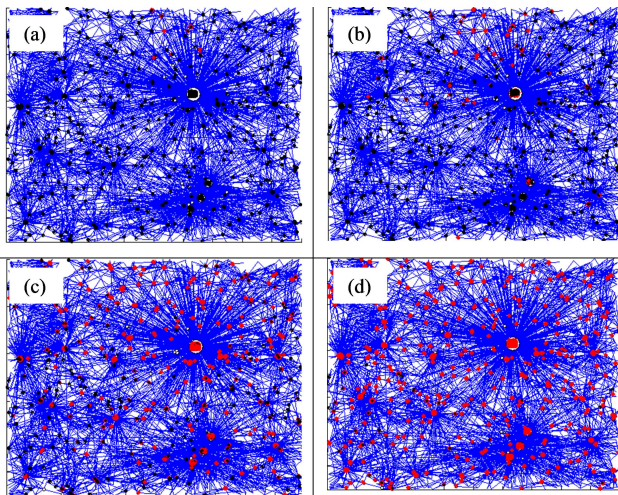


FIG. 3: Evolution of the infection in the virtual city.

Observe that the infection starts spreading at the north region of the city, nearby regions are infected first, and soon the entire city is infected. As expected, the spread of the infection follows a logistic curve, with an exponential growth in the beginning and a later saturation when almost all inhabitants are infected. Figure 4 shows the evolution curve for this outbreak.

### III. EXPERIMENTS AND RESULTS

In this section we use the proposed model to implement three different control measures for disease outbreaks. The experiments show how the model can be used to analyze these measures and provide a better understanding of the dynamics of the phenomenon being modeled.

During the epidemics of SARS (Severe Acute Respiratory Syndrome), the Chinese government implemented an interesting measure to control the disease [7]. Whenever a person shows up with the symptoms of SARS, the place where that

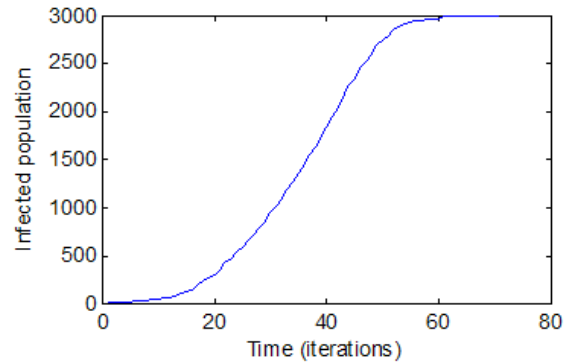


FIG. 4: Logistic growth of the infected population.

person was found sick (e.g., a school) is isolated and nobody can get out or in that place. The government of course provides food and other resources, but people have to stay there in quarantine for a secure period of time.

We use our model to test the effectiveness of this measure. First, it is necessary to consider some delays between the time when the person acquires the infection and then becomes a transmission agent to the time it starts to show the symptoms. If this delay is very small or zero, isolating places where people are found sick will work perfectly, for all the transmission agents will be caught in this process. Unfortunately, there is a delay in the infection, and the larger the delay the farther away the transmission agents may be from the place where they were infected.

We test the model on different delay values and the performance is then evaluated in terms of the number of locations that have been isolated as well as the number of people in the city that have been infected. The results obtained are shown in figures 5 and 6 (solid line). Note that this measure actually works very well if the delay is not too large. But when the delay achieves about 12 iterations  $\frac{1}{4}$  of the city is infected and this number grows even more as the delay increases. In the end, pretty much the entire city has to be stopped, and almost the whole population gets infected (recall that the city has a total of 3000 habitants and 375 locations). This implies that the measure was unable to control the spread of the disease for large delays.

We now modify this measure used by the Chinese government for something slightly different. When a person is found sick, isolate not only the place where it was found at, but also all the locations the person usually visits. This of course would be very hard to implement in practice, but a government could try to isolate instead the most recent places the person has visited. In our model since each agent visits only three places, this is easy to implement. One could expect a great improvement when this measure is applied, since we are backtracking all the places the infected agent has visited, thus other infected people may be found in some of these locations. The results are shown in figures 5 and 6 (dashed line).

Note that although there was improvement for some values of the delay, the performance of this method is still very sim-

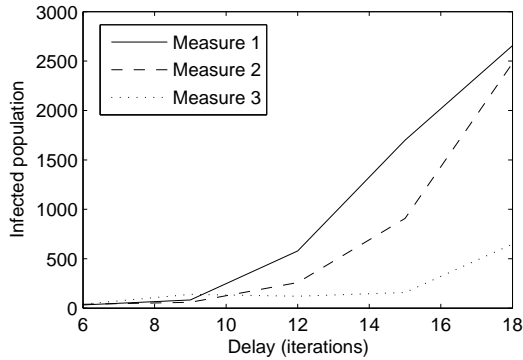


FIG. 5: Performance of the three measures over different delay values with respect to infected population.

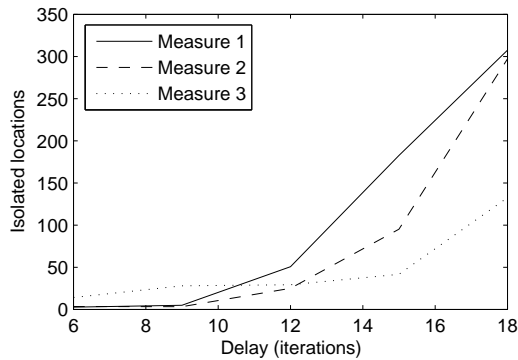


FIG. 6: Performance of the three measures over different delay values with respect to isolated locations.

ilar to that of the first measure applied. Particularly for very large delays, it appears from the graph that the performance of the second measure decreases fast enough so that it will be surpassed at some point by the performance of the first measure. It seems, then, that backtracking the agent’s path does not give a good estimate for where the infected agents are located — we need something else.

For the third and last measure, a different strategy is used. Instead of isolating only the location where the disease was found (as was done in the first measure), all the nearby locations within a radius  $r$  are also isolated. Notice we are taking advantage of the spatial properties of the infection, which tends to spread to nearby regions first. Hence, the spatial information is being used to guess where the transmission agents might be next, thus anticipating the infection. For our virtual city located in a one-by-one square a radius of 0.1 was

used.

The results are shown in figures 5 and 6 (dotted line). Observe that this measure works much better than the previous ones for large delays, leading to a smaller infected population. Also the number of isolated locations is smaller, thus is appealing to the government due to lower cost. For very small delays, though, the first and second measures are more efficient, since fewer locations have to be isolated, but the performance in terms of infected population is approximately the same. It seems from the comparison of these methods that space representation, often ignored in computer models, carries more valuable information from an epidemics control point of view than the structure of the graph.

#### IV. CONCLUSION

The experiments performed demonstrate the potential of the proposed model to represent relevant aspects of the dynamics of an infection, and show how it can be used to test epidemics control measures that can be implemented in the real world. This is just one potential application of our method — the model is extensible, different behavior patterns of the agents can be explored as well as different topologies and infection transmission characteristics.

The results reveal how the spatial information is an important factor to the spread of an infection, and how this information can be used to efficiently deal with an outbreak. Since people tend to go to nearby places more often, the infection spreads locally first, and this can be used as good a guess to where other infected agents might be located. And exception to this rule is made by the highly connected/great capacity locations, the hubs of the bipartite network. Locations with great capacity are able to attract people from farther regions, and this considerably distorts the spatial behavior of the spread. Fortunately, the number of such locations is small and the assumption of the locality of spread still works well enough when trying to control the infection. Probably, even better control policies could be achieved by first isolating the great capacity locations in order to increase the locality of the spread, followed by the application of local measures to control the spread of the disease.

#### V. ACKNOWLEDGMENTS

This work was supported by a grant to UNM from the Howard Hughes Medical Institute and by the Adaptive Computation Lab.

- 
- [1] Meyers, L.A., Pourbohloul B, Newman ME, Skowronski DM, Brunham RC, “Network theory and SARS: predicting outbreak diversity.”, *Journal of Theoretical Biology* 232 (2005), 71—81.
  - [2] Viboud, C. et al. (2006), “Synchrony, Waves, and Spatial Hier-

- archies in the Spread of Influenza”, *Science* 312, 447.
- [3] Eubank, S. et al. (2004), “Modeling disease outbreak in realistic urban social networks”, *Nature*, Vol. 429, 180.
- [4] Eubank, S., (2005), “Network based models of infectious dis-

- ease spread". *Jpn. J. Infect. Dis.* 58.
- [5] Ferguson, NM. et al. (2007), "Planning for smallpox outbreaks", *Nature*, Vol. 425, 681
  - [6] Cliff, A.D., Haggett, P., Smallman-Raynor, M. (1994), "Spatial Processes", Blackwell, New York, 1994.
  - [7] Gumel, A.B. et al. (2004), "Modeling strategies for controlling SARS outbreaks", *Proc Biol Sci.* 2004 November 7; 271(1554), 2223—2232.
  - [8] Burnham, K. P., and D. R. Anderson, (2002), "Model Selection and Multimodel Inference: A Practical-Theoretic Approach", 2nd ed. Springer-Verlag. ISBN 0-387-95364-7.
  - [9] Bellman, R.E. (1961), "Adaptive Control Processes", Princeton University Press, Princeton, NJ.
  - [10] Newman, MEJ. (2002), "Spread of epidemic disease on networks", *Phys Rev E Stat Nonlin Soft Matter Phys* 66:016128.
  - [11] Andersson, H. "Epidemic models and social networks", *The Mathematical Scientist* 1999;24:128.
  - [12] Romualdo Pastor-Satorras and Alessandro Vespignani, (2001), "Epidemic spreading in scale-free networks", *Physical Review Letters*, 86(14), pp. 3200–3203.
  - [13] Boguna M, et al., (2003), "Absence of Epidemic Threshold in Scale-Free Networks with Degree Correlations", *Phys. Rev. Lett.* 90, 028701
  - [14] Models of Infectious Disease Agent Study (MIDAS), <http://www.nigms.nih.gov/Initiatives/MIDAS/>, last access date 02/27/2008.
  - [15] Grais, et al., (2004), "Modeling the spread of annual influenza epidemics in the U.S.: The potential role of air travel", *Health Care Manage. Sci.* 7, pp. 127–134.
  - [16] Colizza V, et al., (2007), "Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions", *A PLoS Medicine* Vol. 4, No. 1.
  - [17] Pourbohloul, et al., (2005), "Modeling control strategies of respiratory pathogens", *Emerg Infect Dis.*